Check for updates

RESEARCH ARTICLE

REVISED Vale do Côa rock art as linked open data: Application profile, schemas, and data

[version 3; peer review: 2 approved, 1 approved with reservations]

Ana Baptista[1], Inês Miguel[1], Juliana Galvão[1], Natália Botica[2]

[1]Information Systems Department / ALGORITMI Research Center, University of Minho, Guimarães, 4804 - 533, Portugal
[2]Archaeology Unit, University of Minho, Braga, 4710-229, Portugal

v3 **First published:** 03 Mar 2026, **6**:63
https://doi.org/10.12688/openreseurope.21901.1
Second version: 09 May 2026, **6**:63
https://doi.org/10.12688/openreseurope.21901.2
Latest published: 28 Jul 2026, **6**:63
https://doi.org/10.12688/openreseurope.21901.3

Abstract

Background

The Côa Valley Archaeological Park, a UNESCO World Heritage, preserves more than 1,200 engraved rocks spanning a long chronological sequence from Prehistory to the Contemporary period. Among these, Iron Age engravings form a significant secondary corpus with no known parallel in the Iberian Peninsula. The archaeological documentation of these engravings is stored and managed in the University of Minho's 2ArchIS database. This article presents the process of converting these previously restricted archaeological records into an openly accessible dataset, with the aim of improving data sharing, reproducibility, and future research on Iron Age rock art.

Methods

The work involved transforming the 2ArchIS data into Linked Open Data (LOD). This process began with analyzing the existing data model and content. Subsequently, a series of standards-based artifacts were developed: a Dublin Core Tabular Application Profile (DCTAP), a Resource Description Framework (RDF) Schema (RDFS) vocabulary, a Simple Knowledge Organization System (SKOS) controlled vocabulary, and a Shape Expressions (ShEx) schema for data validation. The data exported from 2ArchIS was transformed into RDF using OpenRefine, validated against the ShEx schema, and finally hosted on a triplestore.

Open Peer Review

Approval Status ? ✓ ✓

	1	2	3
version 3 (revision) 28 Jul 2026			✓ view
version 2 (revision) 09 May 2026		✓ view	? view
version 1 03 Mar 2026	? view	? view	

1. **Philip Riris**, Bournemouth University, Poole, UK
2. **Rich Potter**, University of Gothenburg, Gothenburg, Sweden
3. **Diego Garate**, Universidad de Cantabria, Santander, Spain

Any reports and responses or comments on the article can be found at the end of the article.

Results

The project successfully generated and published a Linked Open Data dataset based on the validated Iron Age rock art records from the Vermelhosa area, together with the semantic artifacts required for future integration of additional datasets from the Côa Valley. All resulting developmental artifacts, the DCTAP, RDFS and SKOS vocabularies, and the ShEx schema, are openly available for the scientific community's use and reuse.

Conclusions

This initiative has significantly broadened access to Iron Age rock art data by publishing it as Linked Open Data, thereby enhancing its potential for discovery, interoperability, and reuse. The semantic framework developed in this work also provides a foundation for the future integration of additional rock art datasets from the Côa Valley. Future efforts will include creating a user-friendly interface to help non-technical users query the triplestore.

Keywords
Archeology, Rock Art, Côa Valley, Linked Open Data (LOD), Resource Description Framework (RDF), Application Profile

This article is included in the COST Actions gateway.

This article is included in the Digital Humanities collection.

Corresponding authors: Ana Baptista (analice@dsi.uminho.pt), Natália Botica (nb@uaum.uminho.pt)

Author roles: Baptista A: Conceptualization, Methodology, Project Administration, Software, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Miguel I**: Data Curation, Investigation, Software, Writing – Original Draft Preparation; **Galvão J**: Data Curation, Investigation, Software; **Botica N**: Conceptualization, Data Curation, Funding Acquisition, Project Administration, Resources, Supervision, Validation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Framework Programme for Research & Innovation as part of the COST Action under grant agreement No [CA23141] as supported by the COST Association (European Cooperation in Science and Technology). [RARAA project, Foundation for Science and Technology (FCT) "COA / OVD / 0097 / 2019- Rock Art Open Access Repository]
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Baptista A *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Baptista A, Miguel I, Galvão J and Botica N. **Vale do Côa rock art as linked open data: Application profile, schemas, and data [version 3; peer review: 2 approved, 1 approved with reservations]** Open Research Europe 2026, **6**:63 https://doi.org/10.12688/openreseurope.21901.3

First published: 03 Mar 2026, **6**:63 https://doi.org/10.12688/openreseurope.21901.1

REVISED **Amendments from Version 2**

We revised the wording in the abstract's background section to clarify which data we are addressing (Vermelhosa) and to which archaeological period they refer (Iron Age). We also revised the opening of the Introduction to specify the three Côa Valley areas the RARAA project worked on—Vermelhosa (10 engraved rocks), Vale de José Esteves (4), and Meijapão (3)—and to clarify that only the Vermelhosa dataset was published as Linked Data. Finally, we added the term "Triple" to the glossary, specifying that it refers to an RDF triple.

Any further responses from the reviewers can be found at the end of the article

Glossary

The integration of archaeological data into the Semantic Web relies on a sophisticated intersection of domain-specific knowledge and computational frameworks. Due to the technical complexity of this interdisciplinary field, establishing a precise terminological consensus is a prerequisite for meaningful analysis. Consequently, this article opens with a foundational clarification of core concepts. By defining these essential instruments at the outset, we aim to provide a clear conceptual roadmap that will guide the reader through the subsequent discussion on data modeling and interoperability.

- Classes and Properties vocabulary – These consist of terms defined as RDF classes or properties. Although primarily structured as simple schemas that establish internal and external relationships, they are often characterized as ontologies in the literature.[1] They are usually expressed using RDFS or the Web Ontology Language (OWL). Notable examples include the FOAF Vocabulary [1] or the AO-Cat Ontology [2].

- Dublin Core Tabular Application Profile (DCTAP) – A model to define application profiles. It uses a table format that requires no specialized technical knowledge beyond the metadata use case.[2]

- Linked Data – "Data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets".[3] Instead of isolating data in data silos, it uses standard web technologies to provide context and link it to related data from different sources. For example, a rock art motif might be linked to a geographical location, which in turn might be linked to a historical period, creating a knowledge network. As Tim Berners-Lee (as cited by Ref. 4) puts it, "Linked Data is the Semantic Web done right, and the Web done right". One of the most relevant technologies for Linked Data is RDF.

- Linked Open Data (LOD) - Linked Data that is open for use and reuse, as it requires that data be openly licensed, allowing anyone to freely access, use, and share it.

- Metadata Application Profile – "defines metadata usage for a specific application".[2] It "is a metadata design specification that uses a selection of terms from multiple metadata vocabularies, with added constraints, to meet application-specific requirements".[1]

- Open data – Data that «anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)»[5]

- Resource Description Framework (RDF) – A standard framework designed for "representing information on the Web".[6] It employs a graph-based data model where information is structured as triples, consisting of a subject, a predicate, and an object.

- RDF Graph – Sets of subject-predicate-object triples.[6]

- RDF Schema (RDFS) – It "provides a data-modelling vocabulary for RDF data"[7]

- Shape Expressions (ShEx) – A "language for describing RDF graph structures"[8] that supports validation of RDF data against a schema. ShEx schemas are useful for any entity that wishes to analyze and/or validate data against it. ShEx schemas may be derived from application profiles.

[1]https://xmlns.com/foaf/spec/
[2]https://doi.org/10.5281/zenodo.7818375

- Simple Knowledge Organization System (SKOS) - A "common data model for knowledge organization systems such as thesauri, classification schemes, subject heading systems and taxonomies".[9] It allows the representation of such knowledge organization systems as machine-readable data that can be published or reused across computer applications.

- SPARQL Protocol and RDF Query Language (SPARQL) – Query language for RDF.[10]

- Triple - In this context, 'triple' refers to an RDF triple (or RDF statement), which consists of three elements: subject, predicate, and object—or, in alternative terminology: resource, property, and value. A triple "expresses a relationship between two resources. The subject and the object represent the two resources being related; the predicate represents the nature of their relationship".[11] Whenever possible, persistent IRIs [3] are used to represent and uniquely identify the subject, predicate, and object. For example, the subject can be represented by [https://doi.org/10.34622/datarepositorium/NY5JQS], the predicate by [https://ariadne-infrastructure.eu/aocat/1.1/has_description], and the object can be the literal string "Cavaleiro em combate", which translates to 'Knight in combat'. This means that the description of the resource identified by https://doi.org/10.34622/datarepositorium/NY5JQS is "Cavaleiro em combate".

- Triplestore – These are "tools that can be installed to serve as RDF databases".[12] It is a database of RDF triples.

- Values vocabulary – Sets of concepts that can be related as values with specific RDF properties. These structures range from simple lists to hierarchical trees or complex networks of associations. They are usually expressed using SKOS or OWL. Notable examples include the UNESCO Thesaurus [4] or the EuroSciVoc [5].

Introduction

The Côa Valley (Côa Valley Archaeological Park), located in the Guarda district of Portugal, covers a vast area with more than 1,200 engraved rocks. It is internationally recognised as the largest open-air ensemble of Palaeolithic rock art in the world and has been inscribed on the UNESCO World Heritage List since 1998. While Palaeolithic engravings constitute the dominant component of the rock art record, the valley also preserves later phases of occupation, including Iron Age engravings, which represent the second most frequent chronological phase.[13,14]

Within this broader archaeological context, the Open Access Rock Art Repository project (RARAA - Repositório de Arte Rupestre de Acesso Aberto), funded by the Portuguese Foundation for Science and Technology (FCT), focused specifically on Iron Age rock art, aiming to improve knowledge of both the engravings and the communities that produced them. Of the approximately 1,200 engraved rocks currently documented in the park, Iron Age engravings form a relatively small but archaeologically significant subset of the record. Within this subset, the project concentrated on three areas: Vermelhosa (10 engraved rocks), Vale de José Esteves (4 engraved rocks), and Meijapão (3 engraved rocks). Consequently, the dataset worked in RARAA corresponds to 17 engraved rocks from three Iron Age rock art areas, representing a small but well-defined subset of the broader rock art record of the Côa Valley.

All data gathered by the RARAA project is stored in a relational database called 2ArchIS. While this database serves the project and the Archaeology Unit of the University of Minho well, the team sought to connect their data with those of other research groups and, conversely, to make it available for use and reuse by researchers and the wider public. In short, the goal was to make the data open and semantically interoperable—that is, to publish it as Linked Open Data (LOD). This led to the pilot initiative presented here, in which the Vermelhosa dataset was converted and published as LOD, establishing a workflow that can be extended to the data of the remaining areas.

This work aligns with broader trends in the digital transformation of cultural heritage. In recent years, global initiatives promoting open, structured, and reusable data have significantly influenced the digital transformation of cultural heritage. The digital transformation of cultural collections has strengthened preservation practices and opened new opportunities for research, education, and public engagement.[15,16] This paradigm shift has been accompanied by international initiatives that promote open, structured data, ensuring that heritage knowledge remains accessible and reusable sustainably. Movements such as Open GLAM and Collections as Data advocate for the responsible sharing and ethical reuse of cultural data. This approach is exemplified by the principles articulated in documents such as the Vancouver Statement on Collections as Data.[17] These initiatives align closely with the FAIR principles (Findable, Accessible,

[3]Internationalized Resource Identifier.
[4]http://vocabularies.unesco.org/thesaurus
[5]https://cordis.europa.eu/about/euroscivoc

Interoperable, and Reusable),[18] which enhance machine-actionable data discovery and usage while supporting human reuse, and with the CARE principles (Collective Benefit, Authority to Control, Responsibility, Ethics), which provide a foundational approach to Indigenous Data Governance.[19]

In this context, the main goal of this work was therefore to transform data in a relational database into LOD and make it available for use and reuse, adhering to principles and practices that promote structured, reusable, and interconnected data, including the FAIR principles and the World Wide Web Consortium (W3C) Data on the Web Best Practices.[20,21] The expected outcome is to enable integration of RARAA's data with datasets from other research teams and to support broader reuse by the academic community and the general public. By enabling such integration and comparison, we aim to generate new insights into Iron Age rock art that would not be possible when data remains confined to isolated systems, often called data silos. To achieve this, we designed an application profile and related artifacts and then used them to expose 2ArchIS data as LOD.

Our end-to-end pipeline starts from existing data in a closed environment and results in open, queryable Linked Data. In summary, we: (1) analyzed the 2ArchIS data model; (2) designed an application profile[2]; (3) developed and published a controlled vocabulary for values; (4) created a project namespace schema with new classes and properties where necessary; (5) transformed exported 2ArchIS data into Resource Description Framework (RDF) data; (6) validated the resulting RDF data against a Shape Expressions (ShEx) schema; and (7) published the validated RDF data in a triplestore (an RDF database) and made the files available for download in a data repository.

Despite many advancements in LOD solutions for cultural heritage, a significant gap persists in the domain of rock art data, where no comprehensive LOD solution exists to address the unique challenges of describing and linking data on motifs, scenes, and their archaeological context. In many cases, LOD is created from scratch, but there are also numerous cases where the objective is to open and link data that already exists in closed environments. This work, therefore, describes a complete methodological pipeline to open existing data in a closed environment and proposes a set of artifacts specifically designed for rock art documentation.

As contributions for other researchers and professionals, we make explicit the process we followed to generate LOD from closed data and the main decisions taken across the process. This process follows and aligns with current best practices for publishing data on the Web.[18,20,21] Furthermore, we reflect on the overall process and on the use of the Dublin Core Tabular Application Profile (DCTAP), a recent construct designed to support the development and representation of application profiles.[2] Also, we present several artifacts and the resulting Linked Data openly available for use and reuse.

We expect this work to be helpful to researchers and professionals in the fields of Linked Data and metadata, particularly in rock art or archaeology, as well as to students seeking guidance on an effective workflow for converting relational data to Linked Data.

Beyond this introduction, the article is organized into the following sections: Related work (Section 2); Information about the 2ArchIs database (Section 3); Methodological procedures (Section 4); Results (Section 5); Conclusions and future work (Section 6).

Related work

In this work, we analyzed the schemas and vocabularies produced across various projects and initiatives, some specific to cultural heritage and others cross-domain, and reviewed development processes used to publish cultural heritage data as Linked Data. The following subsections summarize the most relevant outcomes of these analyses, focusing on resources and approaches that directly informed our modeling decisions and publication workflow.

Ontologies and vocabularies for cultural heritage and archaeology

To contextualize our approach, we first identify several key initiatives that have served as guiding references in the vast landscape of cultural heritage LOD. Among these are the Advanced Research Infrastructure for Archaeological Dataset Networking in Europe (ARIADNE) and its successor, ARIADNEplus; the International Committee for Documentation of the International Council of Museums Conceptual Reference Model (CIDOC CRM); Archaeological Finds on the Semantic Web (FindSampo); and the Europeana Data Model.

In our case, the most relevant are ARIADNE and ARIADNEplus, which focus on developing semantically interoperable archaeological data repositories to make data more accessible and easily searchable for researchers.[22] By enabling partners to integrate their datasets within a common infrastructure, ARIADNE successfully connected archaeological data from multiple sources,[23] making them available as LOD for use and reuse. ARIADNE ran between 2012 and 2016.[24]

The backbone of the ARIADNE Reference Model is CIDOC CRM, described as "a global extensible schema in the form of a formal ontology that allows for integration without loss of meaning".[25] The ARIADNE Catalogue Data Model was developed to describe the various standards, formats, and services already used by content providers. This model was one of the pillars for developing the AO-CAT ontology within the ARIADNEplus project.[26] This ontology comprises 66 properties and 22 classes with minimal constraints to avoid making their application too restrictive. ARIADNEplus was an extension of ARIADNE that ran from 2018 to 2022. Both projects use CIDOC CRM as their reference model.[22,25] The ARIADNE Research Infrastructure AISBL (Association Internationale Sans But Lucratif; International Non-Profit Association) was established by the end of 2022, and a LOD triplestore was established, providing access to 3.5 million resources encompassing the archaeology and cultural heritage of over 40 countries across four continents.[24,27]

CIDOC CRM is widely recognized as both a theoretical and practical framework for integrating information in the field of cultural heritage.[28] It provides a "shared understanding" of cultural heritage information, acting as a "semantic glue" to mediate cultural heritage data originating from a variety of sources.[29,30] CIDOC CRM is maintained as a "living standard" consisting of a base ontology (CRMbase) complemented by several modular extensions, two of which are for archaeological data: CRMba for archaeological buildings and CRMarchaeo for excavations. The latest stable release of CIDOC CRM (version 7.1.3) defines 81 classes and 160 properties, while CRMarchaeo ontology consists of 10 classes and 32 properties.[30,31]

Another relevant ontology is MAO/TAO [6], developed within the FindSampo initiative, a portal and LOD service that provides access to information about Finnish archaeological finds, particularly those discovered and reported by the public.[32] The MAO/TAO ontology, maintained by the Finnish Heritage Agency, is particularly well-suited for indexing and describing materials related to museums and art.[33] It results from the integration of two domain ontologies: MAO, focused on museum collections, and TAO, related to applied arts.[34] This ontology was valuable to our project, especially during the early stages of controlled vocabulary design, as it provided concrete examples of terminology alignment, hierarchical structure, and naming conventions in the museum and archaeological domains.

Finally, the Europeana Data Model aims to be "an integration medium for collecting, connecting and enriching the descriptions provided by Europeana's content providers".[35] Europeana provides access to Europe's digital cultural heritage,[36] including libraries, museums, archives, and galleries. It uses 18 classes and 77 properties from several namespaces, mainly the Europeana Data Model, DC Terms, and Object Reuse and Exchange. The Europeana Data Model is key in the Common European Data Spaces [7], particularly for cultural heritage [8].[37] Common European Data Spaces are a European Union initiative that aims to make more data available for use and reuse through a "trustworthy and secure environment for the benefit of European businesses and citizens".[38]

General-purpose metadata Ontologies and properties' vocabularies

Using general-purpose vocabulary schemes can enhance semantic interoperability across domain and application boundaries, contributing to a more direct relationship between datasets from different knowledge domains. In our case, we have, whenever possible, tried to adopt terms from vocabularies specific to archaeology or cultural heritage, such as those identified in the previous section. However, when this was not feasible, we relied on general-purpose vocabularies, either using their terms directly or employing them as properties or classes to relate to our own. Some schemas, due to their broad scope and relevance to our project, must be mentioned explicitly.

The Dublin Core Metadata Initiative (DCMI) Terms (DC Terms) specification is a DCMI recommendation. Among its properties are the well-known 15 elements that constitute the Dublin Core Metadata Element Set, often just called Dublin Core. DC Terms has many more properties; some are even sub-properties of other DC Terms properties (e.g., issued is a sub-property of date). These properties are used in various domains and applications, including digital repository platforms (e.g., DSpace [9]) and the Data Catalogue Vocabulary [10] (DCAT), whose application profile (DCAT-AP [11]) is key for describing datasets in the Common European Data Spaces.[39] Several DC Terms properties and classes have been published as the ISO standard 15836-2:2019.[40]

[6]https:/finto.fi/maotao/en/
[7]https://digital-strategy.ec.europa.eu/en/policies/data-spaces
[8]https://www.dataspace-culturalheritage.eu
[9]https://dspace.org/
[10]https://www.w3.org/TR/vocab-dcat-3/
[11]https://op.europa.eu/pt/web/eu-vocabularies/dcat-ap

DCAT is "an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web".[41] It enables the description of datasets and data services, improving their discovery and access. A good description of datasets also enhances their interoperability and, therefore, their use and reuse, as advocated by the FAIR principles. DCAT-AP, an application profile of DCAT, is used to describe data from European public organizations. It is used in the European Data Portal [12] and the Common European Data Spaces.

DBpedia [13] is a "crowd-sourced community effort to extract structured content from the information created in various Wikimedia projects".[42,43] It provides structured data "for more than 100 Wikipedia language editions as well as Wikimedia Commons, has a mature ontology and a stable and thorough Linked Data publishing lifecycle".[44] At its core is the DBpedia ontology (DBO).[45] DBO is a crowd-sourcing effort resulting in a cross-domain ontology that currently contains 768 classes and 3000 properties,[45] many of which are used in other projects for interoperability and reusability. Another DBpedia namespace, with the prefix dbp, represents "properties extracted from the raw infobox extraction," whose data is considered lower quality than the DBO's.[46] There are other namespaces in DBpedia that have various purposes. Examples are dbr [14], for entities or resources, and dbt, for properties from infobox templates [15].

Like DBpedia, Wikidata [16] is a collaborative project associated with the Wikimedia Foundation [17]. Wikidata and DBpedia both overlap and complement each other. All DBpedia data is extracted from Wikimedia projects,[44] while Wikidata's data is created collaboratively and feeds projects such as Wikipedia and others.[47] Its open knowledge base acts as a central storage of structured data for Wikimedia projects.[48] The data is available in several formats under a Creative Commons Universal License [18].

Schema.org [19] is "a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond".[49] It is an initiative founded by Google, Microsoft, Yahoo, and Yandex that provides a schema whose terms are used extensively on the Web. It currently consists of 817 types and 1518 properties, among other kinds of terms.[50] Because it may be of interest to data providers and catalog publishers, version 3 of the DCAT specification proposes an alignment with Schema.org.[41]

There are other schemas that, although not general-purpose, are applicable across various domains, such as those for geographic data or units of measure. Examples of this type of schema include GeoNames [20] or the Getty Thesaurus of Geographic Names (TGN) Online [21]. GeoNames is a service based on a database containing over 25 million geographic names available for download free of charge under a Creative Commons Attribution license [22].[51] Each geographic name in GeoNames has a unique identifier[51] that can be used in other Linked Data projects. The names are also related to each other. The same is true of the TGN. The TGN includes more than 5 million geographic names available in various formats, including several LOD formats. The Open Data Commons Attribution License (ODC-By) 1.0 [23] is the license used. On its website, TGN describes other geographic initiatives for Linked Data, including GeoNames, and compares them in some respects, including geographical coverage, temporal coverage (including historical data), and data quality.[52] Many other namespace schemas exist, which may be found at services such as the Linked Open Vocabularies [24]. These resources are particularly relevant for future enrichment of rock art datasets with external geographic identifiers to improve linking and comparative analysis across sites and regions.

Linked data development processes in cultural heritage

The transition from legacy cultural heritage data to the Semantic Web has been addressed through various methodological lenses. While some approaches prioritize the preservation of metadata richness through multi-step ingestion processes, others focus on the scalability of database migrations or the agility of ontology design. The following projects illustrate the diversity of frameworks developed to bridge the gap between heterogeneous data sources and ontological structures.

[12]https://data.europa.eu
[13]https://www.dbpedia.org/
[14]http://dbpedia.org/resource/
[15]http://dbpedia.org/resource/Template
[16]https://www.wikidata.org
[17]https://wikimediafoundation.org/
[18]https://creativecommons.org/publicdomain/zero/1.0/
[19]https://schema.org/
[20]https://www.geonames.org/
[21]https://www.getty.edu/research/tools/vocabularies/tgn/
[22]https://creativecommons.org/licenses/by/4.0/
[23]https://opendatacommons.org/licenses/by/1-0/
[24]https://lov.linkeddata.es/dataset/lov/

De Boer and colleagues[53] present a process "for ingesting, converting, and linking cultural heritage metadata into Linked Data" that, according to the authors, was designed to maintain the richness of the original metadata, as opposed to practices of mapping to specific schemas where they claim part of this richness to be lost. Several tools support this process, which had the following steps: 1) XML ingestion; 2) Direct transformation to 'crude' RDF; 3) RDF restructuring; 4) Creation of a metadata mapping schema; 5) Alignment of vocabularies with external sources; and 6) Publishing as Linked Data. Hong and colleagues[54] developed a transformation framework to migrate large volumes of data from relational databases to ontology-based datasets. They created a scripting language "which allows the specification of complex transformation rules from data objects to ontologies". This language was put into practice using an extended version of CIDOC CRM. Daquino and colleagues[55] describe the work carried out to make the photo archive Zeri available as Linked Data through new ontology extensions to existing models. They then mapped the archive's descriptive elements into RDF using the ontologies they created and CIDOC CRM with examples of usage. These developments enabled them to generate an RDF dataset that includes descriptions of approximately 50,000 entries from the Zeri Archive. Nishanbaev and colleagues[56] developed a feature comparison of tools, including OpenRefine, for generating geospatial RDF data. They also propose a methodology for producing linked geospatial data that consists of five steps: 1) Data preparation; 2) Mapping the data into the ontology; 3) Interlinking RDF datasets; 4) Storing RDF data; and 5) Geospatial Semantic Web app & LOD cloud. Carriero and colleagues[57] used and extended eXtreme Design to develop and validate ArCo, the knowledge graph of Italian cultural heritage, represented as Linked Data. eXtreme Design is an ontology design methodology inspired by eXtreme Programming, an agile software development methodology. The authors extended eXtreme Design to deal with Cultural Heritage ontology projects. Binding and colleagues[58] report on the development of Linked Data departing from exported data, where they mapped existing fields to CIDOC CRM and converted the data to RDF using a "template-based data conversion method". They also encoded and published specific controlled vocabularies as Linked Data. In our work, we adopt a similar template-based approach but extended it by using ShEx schemas to validate the resulting RDF graphs against the structure defined in our application profile.

Our development process is generally aligned with some of the presented processes, but in some parts, we have taken different approaches. Firstly, we chose to use a generic open data repository, in this case the University of Minho's data repository, to store motifs' data and images, as well as project outputs, and to generate persistent identifiers, in this case Digital Object Identifiers (DOIs), for each item. We created a record for each motif in the data repository, thereby assigning a DOI to each. In this same registry, we include other information, such as images, as well as tabular data. This way, each motif has a DOI, and its associated information is preserved in a stable, long-term repository. We also stored all developed artifacts in the open data repository, generating a DOI for each, thereby supporting their long-term digital preservation. Secondly, we chose to use DCTAP, aligning the development of the application profile with the most recent developments in DCMI, a leading international initiative in metadata. Detailed information about the entire process is presented in Section 4.

To the best of our knowledge, our project is the first to dedicate itself exclusively and in depth to making LOD for rock art available, developing a range of artifacts and the data itself. Our project aims to make a first contribution in this direction. Drawing on multiple sources, building on them, and expecting to contribute to the Linked Data community in archaeology and cultural heritage, this work presents an application profile, a namespace schema, a controlled vocabulary, a ShEx schema, and actual rock art data. In addition, it presents in detail the process that led to making this data openly available from an existing relational system, with the goal of supporting reuse, extension, and interlinking by the wider community.

The 2ArchIS relational database

The rock art of the Côa Valley is an in situ heritage whose study contributes to a better understanding of the societies that created it. Analyzing artistic representations on rocks provides insights into ancient communities' beliefs, religious practices, social structures, rituals, and the environments in which they lived. The animal representations in rock engravings can offer clues about the environment and ecosystem in which these communities lived, providing information about extinct or changing fauna, climatic patterns, and ancient landscapes. Additionally, hunting scenes depicted in the art may indicate survival and subsistence techniques and strategies. Therefore, the study of rock art representations is a cross-disciplinary endeavor where data sharing is essential for a better understanding of the cultural landscape in which the engravings are situated. Consequently, their archaeological record, often unique and irreplaceable, represents a valuable heritage whose preservation is crucial, allowing for new interpretations and studies. Rock engravings, particularly susceptible to natural factors such as erosion, wind, rain, freeze-thaw cycles, and vandalism, underscore the need to safeguard this cultural legacy to study and preserve its memories.

However, safeguarding archaeological data without FAIR data requirements is not sufficient. It is imperative to build trust in the data and facilitate its reuse in subsequent studies, integrating data and developing consistent results. Rock art

documentation should be structured to ensure interoperability and reusability, following the FAIR principles. This endeavor involves standardizing records, establishing controlled vocabularies, and attaching metadata to ensure the data's quality and reliability.[59]

In this context, 2ArchIS[60] was created by the Archaeology Unit of the University of Minho as an operational information system to support archaeological work from field recording to post-fieldwork analysis and reporting. Its primary intent is to ensure consistent data capture in the field and in the office, to centralize and normalize records produced by different team members, and to maintain long-term institutional memory of projects. Beyond data entry, the system supports day-to-day research workflows by enabling browsing and filtering of records, production of inventories and summary tables for reports, and systematic association of archaeological entities with documentation (bibliography, photographs, drawings, cartography, and 3D models).

In the specific case of rock art, such structured recording is particularly relevant given that these datasets constitute a primary source for interpreting past social practices, symbolic systems, and the construction of cultural landscapes, as well as human–environment interactions reflected in the representation and spatial distribution of motifs.[61,62]

For the RARAA project, 2ArchIS has served as the authoritative source for recording and relating archaeological sites, rocky outcrops, motifs, represented parts, scenes, and associated documentation, supporting both ongoing research and project outputs. For dissemination and preservation purposes, exported records have also been deposited in an open repository with persistent identifiers, while 2ArchIS remained the primary environment for data creation and curation.

Data registration for the RARAA project was conducted using the 2ArchIS information system (see Figure 1 for a portion of its data model and the complementary data[63] for information on the tables' attributes), developed at the University of

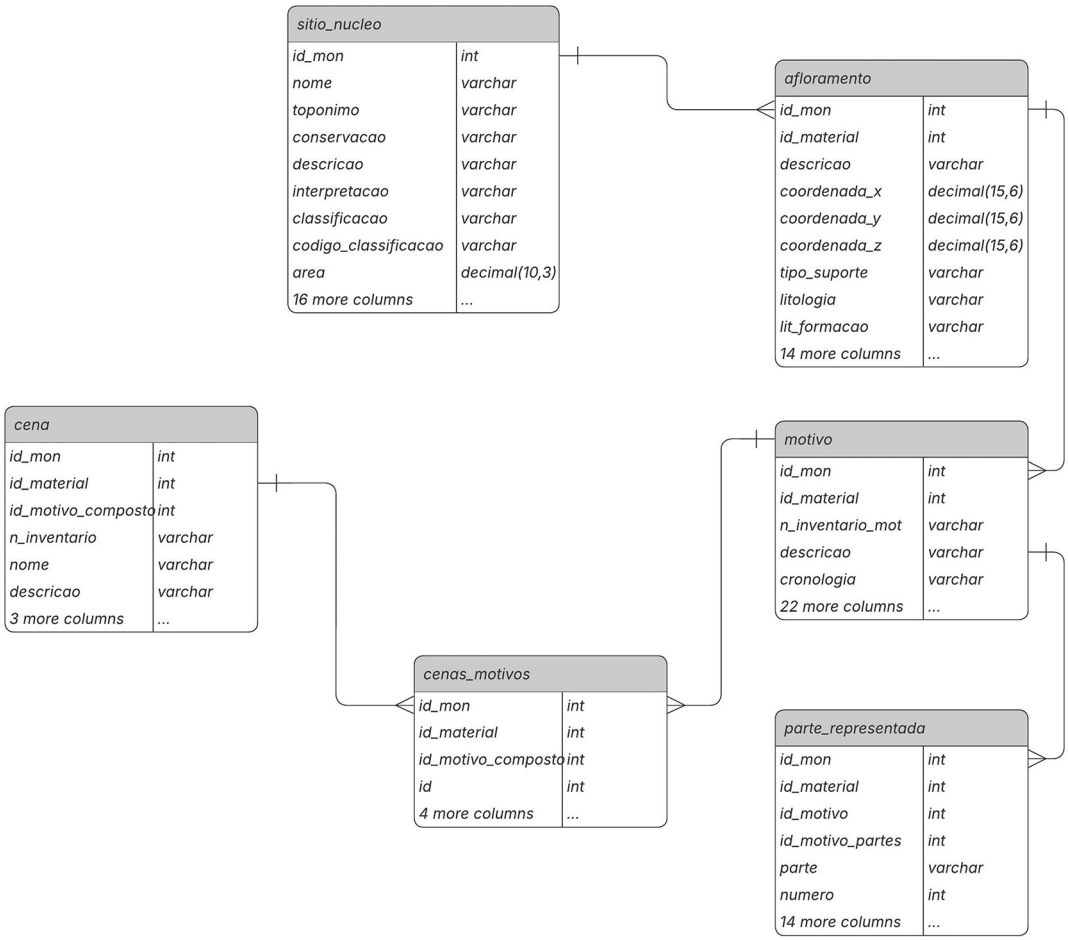

Figure 1. A view of the 2ArchIS entity-relationship model.

Minho's Archaeology Unit. This data concerns the excavated sites, the outcrops with engraved rock art, the motifs discovered, their parts, and the scenes represented. Data access was facilitated by creating a back-office application in PHP and HTML, running on a web platform and usable both in the field and in the office. All data recorded in 2ArchIS was also deposited in a repository with persistent identifiers, dataRepositóriUM, ensuring their preservation and dissemination and making them locatable and accessible.

The modular structure of the 2ArchIS information system integrates and interconnects modules for characterizing archaeological sites, stratigraphy, excavation materials, geological and geographical contexts, and rock art. In what follows, we summarize the modules most relevant to rock art documentation and to the Linked Data transformation described in this article.

Archaeological Site: The archaeological site characterization module (sitio_nucleo) records the site's location, toponymy, and typology data. However, additional information is essential to associate with the site, such as its geographical, geological, hydrographic context, or land use, as these contextual data are crucial for better understanding the site and the criteria for choosing its location. Thus, we have specific tables for recording this data, namely sitio_hidrografia, sitio_contexto_geog, sitio_contexto_geol, sitio_uso_sol. Sites may also have occupation over time, meaning they are associated with various chronologies (table sitio_rela_percultural) and have multiple configurations and uses during these periods of occupation (sitio_rela_tipo). This information results from studies derived from archaeological excavations, whose material records are stored in the sitio_rela_materiais table. This data, integrated into the 2ArchIS information system, provides a more comprehensive framework for rock art.

Rocky Outcrops: This module characterizes the existing rocky outcrops at the site that support rock art. Detailed information includes the type of outcrop, such as location, size, lithology (shale, granite, others), surface morphology (flat, concave, or convex), surface appearance (fractures, nodules, polishing, etc.), and conservation status.

Motifs: On the surface of the outcrop, motifs can be identified, which can be of various types: figurative, geometric, and alphanumeric. Details recorded for each motif cover inventory number, chronology, engraving techniques, style, typology, and other characteristics. The module highlights the diversity of motifs, ranging from isolated figures to narrative scenes.

Represented Parts in motifs: The motifs characterized in the Motif module are not always complete, and even when they are, they may present unique characteristics of the design of each component, such as the shape of the head, the number and shape of the limbs represented, in this case of anthropomorphs or zoomorphs, or the shape and type of each weapon component. Therefore, in this module, we characterize each represented part of the motif, including its typology, shape, position, engraving technique, and associated measurements.

Scenes: Not all motifs stand alone; some form scenes, interacting with each other in a narrative context. The Scene module is designed to associate motifs that create these narrative compositions.

Documentation: In this module, we emphasize the system's ability to associate all entities in the 2ArchIS system with bibliographic references, photographs, drawings, cartography, and 3D models. This comprehensive documentation enriches the archaeological data and provides a more holistic understanding of the findings.

Portions of the attributes of the 2ArchIS tables are available as complementary data.[63] Specifically, these attributes relate to entities directly relevant to this work, such as Archaeological Site, Outcrop, Motif, Represented Part, and Scene. We highlight the case of motifs and represented parts, which share some common attributes and descriptors. A motif characterizes the representation of a figure, whether geometric, letters and numbers, or figures of people or animals, through a set of lines engraved on the rock surface. However, this engraving, done by incision using a tool, which can be a fragment of stone or metal, is not always uniform throughout the entire motif, depending on the force and/or inclination applied by the tool used for engraving. Thus, a motif may exhibit varying line depths and thicknesses.[64] For example, we have motifs in which the head of the anthropomorph is drawn with a deeper incision and even reinforced by more than one line, whereas the same does not occur in the representation of the torso and limbs. Therefore, while the motifs record the general characteristics of the figure, namely in the attributes of technique and variant technique, for the represented parts, we specify the same attributes to characterize the features of the rock carvings in that specific part.

The transformation of this specialized data structure into LOD presents unique challenges that require a tailored approach, which we address through our methodological framework.

Methodological procedures

This work uses archaeological data previously collected in a relational database, developed within the RARAA project and implemented in the 2ArchIS information system. The contribution of this article lies in the semantic modeling, transformation workflow, and publication of this data as LOD, along with the development of reusable semantic artifacts and software tools that enable FAIR access, interoperability, and reuse of rock art documentation.

To support the development of the application profile, we used the DCTAP, which provides a tabular structure for defining application profiles in a clear and understandable format.[2] It defines how the data should be used and what to expect from it. The current vocabulary version comprises 12 terms[65] used to name the DCTAP columns. The number of columns can be increased or decreased based on users' needs. Adding columns means adding terms, which should be defined for clarity. Clear definitions are essential to ensure that data can be reused effectively in open environments.

Using the Entity-Relationship (E-R) model (Figure 1) as a basis, we first identified the shapes to include in our DCTAP. It is worth highlighting that a shape can be related to one or more entities, given that in Linked Data the properties (attributes in the E-R model) can be repeated (cardinality >1), i.e., there are no restrictions on m:n relationships. This means that entities in the E-R model may be partially or fully absorbed by others, resulting in a single shape. Another situation may arise in which an entity gives rise not to a shape but to a controlled vocabulary.

Still on the DCTAP, we replaced the propertyLabel column with an attribute column to indicate each attribute present in the E-R model. Then we listed the attributes associated with each shape. Next, for each attribute, we looked for a property with the appropriate semantics that could represent it without restrictions that would prevent its use, especially regarding domain and range. We started by looking at the ARIADNE and CIDOC CRM ontologies. Finding the property, we moved on to the following attribute. Not finding it, we turned to other schemes such as DC Terms, DBpedia, or Schema.org. Sometimes, searching in Linked Open Vocabularies was necessary. When DBpedia dbp properties had limited information, to ensure correct semantics, we created specific properties and linked them to the corresponding dbp terms via `rdfs:subPropertyOf`. To relate the shapes, we also sought specific properties that would allow us to express the semantics of those relationships.

We sometimes could not find properties with the appropriate semantics, domains, and ranges that suited our needs. In these cases, we had to create new properties. The same happened with some classes. The lists of values associated with the attributes in the 2ArchIS database were collected, yielding a controlled vocabulary that we encoded using the Simple Knowledge Organization System (SKOS). New properties and classes have been encoded in RDF Schema (RDFS). We related these properties and classes to existing ones (through `rdfs:subPropertyOf` or `rdfs:subClassOf`) to get their semantics whenever possible.

The classes and properties vocabulary, encoded in RDFS, and the controlled vocabulary for values, encoded in SKOS, were placed in the dataRepositóriUM to obtain a DOI for each. These DOIs persistently identify the respective vocabularies. The RDFS and SKOS files were then modified to include these identifiers, generating new versions uploaded to the same record in the dataRepositóriUM. The DCTAP was also placed in the dataRepositóriUM.

We also used dataRepositóriUM to store data for each motif and obtain a persistent identifier (a DOI) for each motif. Only motifs have a DOI, so other resources have internal identifiers and, when applicable, are described in relation to the motifs. For example, scenes and outcrops relate to motifs using the `ao:has_part` property, and represented parts relate to motifs using the `ao:is_part_of` property. This option is duly represented in the DCTAP.

After completing the DCTAP, we exported the data from the 2ArchIs database into XML files[66] and then transformed them to RDF/XML using OpenRefine, following the property mappings defined in the application profile. The data transformation was an iterative process that required some adjustments until the data was correctly transformed. The data was then validated using the ShEx schema. At first, there were some validation issues that, after analysis, revealed minor typos in the shapes but not in the data, which were promptly corrected. To fully comply with the FAIR principles, we also included CC0 or CC-BY licenses in RDF files.

To house the data and create the RARAA repository, we used GraphDB [25]. We also put the RDFS and SKOS vocabularies there to facilitate the queries. The triplestore can be interrogated via a SPARQL Protocol and RDF Query Language (SPARQL) endpoint. The data was also made available as RDF/XML and Turtle files at the dataRepositóriUM.

[25]https://www.ontotext.com/products/graphdb/

Table 1. Used namespace schemas.

Prefix	URL
ao	https://ariadne-infrastructure.eu/aocat/
cidoc	http://www.cidoc-crm.org/cidoc-crm/
dbo	https://dbpedia.org/ontology/
dbp	http://dbpedia.org/property/
raraa	https://doi.org/10.34622/datarepositorium/VVCF9M#

Results

Selected namespace schemas

The used namespace schemas are shown in Table 1. The "raraa" prefix refers to the namespace we created for this project. The corresponding RDFS file is stored in the dataRepositóriUM so a DOI identifies it.

RARAA properties and classes

Adhering to the best practices of the Semantic Web community, to maximize semantic interoperability, we tried to reuse existing properties and classes as much as possible, prioritizing standardized or widely used namespace schemas. The semantics, ranges, and domains of existing properties were verified for compatibility. However, despite our best efforts, we were unable to do so in some cases, leading us to create 29 new properties (Table 2) [26]. These properties were related to existing properties to get their semantics whenever possible (Table 3). Also, in this case, the semantics, domain, and range of existing properties were considered for compatibility. The RDFS code for these new properties can be found in the RARAA namespace URI,[67] as shown in Table 1.

The `rdfs:subPropertyOf` relationships between RARAA properties and existing properties may be found in Table 3. The third column in this table shows the expected types of values. We did not constrain the domain or range of the properties we created so that they can be used in other situations (provided the semantics is adequate).

The RARAA namespace also contains three classes: `raraa: Motif`, `raraa: RepresentedPart`, and `raraa: Scene`. All three are subclasses of `cidoc: E25_Human-Made_Feature` (via `rdfs:subClassOf` property). Creating these classes was necessary to distinguish the resources by their class using the `rdf:type` property.

Controlled Vocabularies

A SKOS-encoded controlled vocabulary was created based on the 2ArchIS lists of values as presented by Botica, Luís, and Silva[59] and available as complementary data, which provides an English translation of the original values.[63] Each term has an identifier and at least one prefLabel. We encoded and made available one SKOS vocabulary containing 297 concepts with English, Portuguese, and Spanish labels. The vocabulary is organized as a tree structure with six levels. The vocabulary structure reflects part of the shapes and properties with which the values/concepts are used. Leaf concepts were reused as much as possible within the vocabulary.

This vocabulary is available online.[68] Figure 2 shows the first three levels of the values vocabulary.

Dublin core tabular application profile

We considered five shapes in the DCTAP: Motif, Scene, Represented Part, Site, and Outcrop. The shapes are related via `ao:is_part_of` and `ao:has_part` properties. As mentioned earlier, only motifs have a DOI, so scenes, represented parts, and outcrops point to motifs, and motifs have no outgoing links to other shapes' resources. This choice preserves the integrity of the relationships since the motifs' identifiers are persistent.

The instructions in the DCTAP Primer were strictly followed. Since it was not mandatory, we replaced the propertyLabel column with the attribute column for better reference when designing the DCTAP. In properties whose values should come from a controlled vocabulary, we use the value constraint type IRIstem. We used the `rdf:type` property with a value constraint for the corresponding class in each shape. For example, we used the value constraint `raraa: Motif` for the Motif shape. Figure 3 shows the DCTAP for the shape Motif. The complete DCTAP for the RARAA project is available online.[69]

[26]See the complementary data, for a complete list of 2ArchIS tables and respective attributes (terms translated to English) used in this project.

Table 2. 2ArchIS attributes x RARAA namespace properties.

E-R Diagram		RARAA namespace		Definition
E-R Entity	**Attribute**	**DCTAP shape**	**Property**	
sitio_nucleo		Site		
	acessos		raraa:accessDescription	A site's access description, e.g., access to the site via a steep slope, sometimes covered with loose stones and occasional vegetation.
	tipo_acesso		raraa: accessType	A site's access type, e.g., public, private, or restricted.
	interpretacao		raraa:interpretation	Site interpretation.
	classificacao		raraa: siteClassification	The classification assigned to the site, such as national monument or a special protection area.
	nome		raraa:siteName	The site name.
	cod_classificacao		raraa:siteNationalCode	The site's national code issued by a national authority.
motivo		Motif		
	animacao		raraa:animationTechnique	The animation technique used to create a dynamic visual effect that suggests movement.
	grupo		raraa:graphicalGroupType	The group classifies the motifs into three categories based on their visual and symbolic characteristics: figurative, geometric, and alphanumeric.
	sub_tipo		raraa:graphicalSubType	The subtype provides a more detailed classification within the broader categories defined in group. For example, within the figurative group, subtypes include anthropomorphic, zoomorphic, and weapons.
	tipo		raraa:graphicalType	The type offers a more detailed classification within each subtype; for example, we can identify types such as equine, deer, or aurochs within the zoomorphic subtype.
	motivo_figura		raraa: completeFigure	Indicates whether the figure is complete or not.
	invertida		raraa:invertedFigure	Indicates whether the figure is inverted or not.
	orientacao		raraa:partOrientation	Orientation of motifs in degrees, e.g., 20° or 45°.
	patine		raraa:patina	Indicates whether the motif's trace has patina or not.
	perfil		raraa:perspective	Indicates the motif's perspective, which can be either frontal or absolute profile.
	relacao_suporte		raraa:relationWithRock	Indicates how the motif utilizes the support, such as fissures or inclusions.
	tecnica_variante		raraa:variantTechnique	Indicates the type of incision used in engraving the motifs. Examples include simple incision, repeated incision, abrasion, and others.

Table 2. *Continued*

E-R Diagram		RARAA namespace		Definition
E-R Entity	**Attribute**	**DCTAP shape**	**Property**	
afloramento		Outcrop		
	direcao		raraa:direction	General orientation of the rock outcrop measured in degrees, for example, 30°N or 25°E.
	inclinacao		raraa:inclination	The inclination of the outcrop panel measured in degrees, for example, 87°NW.
	estratigrafia_inclinacao		raraa:inclinationStratigraphy	The stratigraphic inclination.
	lit_formacao		raraa:lithologyFormation	Name of the lithological formation, such as Desejosa or Pinhão.
	litologia		raraa:lithology	Lithology type, e.g., shale or granite.
	altura		raraa:maxHeight	Maximum height of the outcrop.
	comprimento		raraa:maxLength	Maximum length of the outcrop.
	termo		raraa:modificationTerm	Indicates whether the panel has the term “alteration” or not.
	asp'sup		raraa: surfaceAppearance	Characterization of the surface appearance, which can be, for example, smooth or heavily fractured.
	morfologia_sup		raraa:surfaceMorphology	The surface morphology, which can be flat or concave, for example.
parte_representada		Represented Part		
	parte		raraa:representedPart	Part of the motif represented, such as the head or limbs.
cena		Scene		
	nome		raraa:sceneName	The name given to the scene.

Table 3. RARAA namespace properties, their `rdfs:subPropertyOf` relations with existing properties, and their expected types of values.

RARAA property	rdfs:subPropertyOf	Expected types of values
raraa:accessDescription	ao:has_description	xsd:string
raraa:accessType	dct:type	xsd:string
raraa:animationTechnique	dbo:technique	xsd:string
raraa:direction	dbp:direction	xsd:string
raraa:graphicalGroupType	ao:has_type	IRI* (term from a controlled vocabulary)
raraa:graphicalSubType	ao:has_type	IRI (term from a controlled vocabulary)
raraa:graphicalType	ao:has_type	IRI (term from a controlled vocabulary)
raraa:completeFigure	ao:has_description	xsd:string
raraa:invertedFigure	ao:has_description	xsd:string
raraa:inclination	dbp:inclination	xsd:string
raraa:inclinationStratigraphy	dbp:inclination	xsd:string
raraa:interpretation	ao:has_description	xsd:string
raraa:lithologyFormation	N/A	IRI (term from a controlled vocabulary)
raraa:lithology	dbo:material	IRI (term from a controlled vocabulary)
raraa:maxHeight	dbo:height	xsd:decimal (centimeters)
raraa:maxLength	dbo:length	xsd:decimal (centimeters)
raraa:modificationTerm	ao:has_description	xsd:string
raraa:partOrientation	dbp:direction	xsd:string
raraa:patina	N/A	IRI (term from a controlled vocabulary)
raraa:perspective	N/A	xsd:string
raraa:relationWithRock	ao:has_description	xsd:string
raraa:representedPart	dbo:pictureDescription	IRI (term from a controlled vocabulary)
raraa:sceneName	ao:has_title	xsd:string
raraa:siteClassification	dbo:classification	xsd:string
raraa:siteName	ao:has_place_name	xsd:string
raraa:siteNationalCode	ao:has_identifier	xsd:string
raraa:surfaceAppearance	N/A	IRI (term from a controlled vocabulary)
raraa:surfaceMorphology	N/A	IRI (term from a controlled vocabulary)
raraa:variantTechnique	N/A	IRI (term from a controlled vocabulary)

*Internationalized Resource Identifier - https://www.ietf.org/rfc/rfc3987.txt.

RDF files

We automatically generated RDF files using OpenRefine. Having, on the one hand, the data in XML format as exported from 2ArchIS and, on the other, the DCTAP and the links of the namespace schemas and controlled vocabularies, we created RDF skeletons. In these RDF skeletons, we defined the prefix table. Then, we chose the properties using the prefix of the respective namespace, followed by information about the corresponding database column. This process specifies the transformations from entities' attributes to RDF properties, ensuring the data conforms to the project's requirements. All files regarding each table had their RDF skeleton defined. Using these RDF skeletons allowed OpenRefine to export RDF files with the defined structure so that we could later, after validation, upload the data to the GraphDB triplestore. See Figure 4 for part of the RDF skeleton for the table Motivo (Motif).

The RDF files are available online.[70] Figure 5 illustrates the RDF graph for scene 'VRM003-E,' titled 'Cena de combate' ('Combat scene'). It should be noted that while the illustrative drawings of the motifs are included in the figure to facilitate interpretation, they are not formal components of the RDF graph itself.

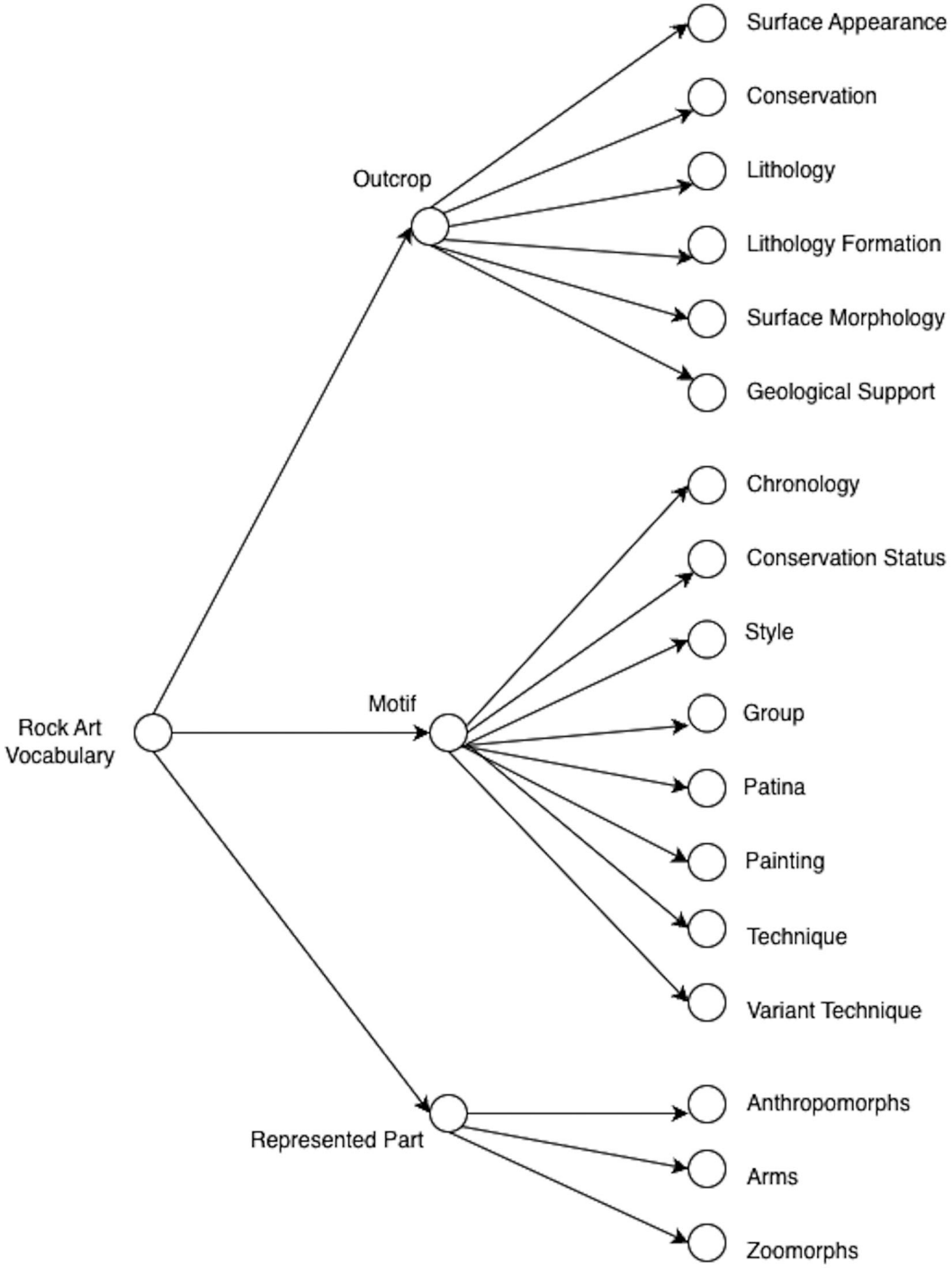

Figure 2. The first three levels of the RARAA values vocabulary for rock art.

Shape expressions

A ShEx schema is "a collection of shape expressions that describe an RDF graph".[8] Our ShEx schema has expressions for five shapes, one for each DCTAP shape, with two purposes: 1) to validate the data generated by OpenRefine; and 2) to provide the rock art community with shapes that allow them to create and validate data using our application profile. In this way, we are creating conditions for others to provide data that can be easily added and linked to ours to advance research into rock art.

DCTAP for the RARAA project

shapeID	shapeLabel	propertyID	2ArchIs attribute	mandatory	repeatable	valueNodeType	valueDataType	valueConstraint	valueConstraint
motif	Motif	rdf:type		TRUE	FALSE	IRI		https://doi.org/10.34622/datarepositorium/VVCF9M#Motif	
		ao:has_original_id	Número de Inventário	TRUE	TRUE	Literal	xsd:string		
		ao:has_description	Descrição	FALSE	TRUE	Literal	xsd:string		
		cidoc:P44_has_condition	Conservação	FALSE	FALSE	IRI		https://doi.org/10.34622/datarepositorium/2SGCAG	IRIstem
		ao:has_native_period	Cronologia	FALSE	FALSE	IRI		https://doi.org/10.34622/datarepositorium/2SGCAG	IRIstem
		dbo:style	Estilo	FALSE	FALSE	IRI		https://doi.org/10.34622/datarepositorium/2SGCAG	IRIstem
		cidoc:P32_used_general_technique	Técnica	FALSE	FALSE	IRI		https://doi.org/10.34622/datarepositorium/2SGCAG	IRIstem
		raraa:variantTechnique	Técnica Variante	FALSE	FALSE	IRI		https://doi.org/10.34622/datarepositorium/2SGCAG	IRIstem
		raraa:patina	Patine	FALSE	FALSE	IRI		https://doi.org/10.34622/datarepositorium/2SGCAG	IRIstem
		raraa:maxLength	Largura	FALSE	FALSE	Literal	xsd:decimal		
		raraa:maxheight	Altura	FALSE	FALSE	Literal	xsd:decimal		
		raraa:graphicalGroupType	Grupo	TRUE	FALSE	IRI		https://doi.org/10.34622/datarepositorium/2SGCAG	IRIstem
		raraa:graphicalType	Tipo	FALSE	FALSE	IRI		https://doi.org/10.34622/datarepositorium/2SGCAG	IRIstem
		raraa:graphicalSubType	Sub-Tipo	FALSE	FALSE	IRI		https://doi.org/10.34622/datarepositorium/2SGCAG	IRIstem
		raraa:completeFigure	Figura	FALSE	TRUE	Literal	xsd:string		
		raraa:invertedFigure	Invertida (Figura)	FALSE	TRUE	Literal	xsd:string		
		raraa:inclination	Orientação	FALSE	TRUE	Literal	xsd:string		
		raraa:perspective	Perspetiva	FALSE	TRUE	Literal	xsd:string		
		raraaa:animationTechnique	Animação	FALSE	TRUE	Literal	xsd:string		
		raraa:relationWithRock	Relação com o Suporte	FALSE	TRUE	Literal	xsd:string		
		ao:has_identifier	Identificador	TRUE	FALSE	IRI			

Figure 3. Part of the DCTAP for the Motif shape.



Figure 4. The first part of the RDF Skeleton for the Motivo (Motif) table.

As for validating the structure of our RDF code, the ShEx shapes ensured that each triple data type was validated and that the properties defined in each shape were appropriately used. In this process, we used an online tool [27]. We manually developed ShEx code and utilized RDF code exported from OpenRefine in Turtle format (Terse RDF Triple Language). Figure 6 shows an example of ShEx validation of a motif's description.

The validation was straightforward once we aligned the ShEx schema with the DCTAP. Since the shapes specified that most properties are optional, the presence of objects lacking specific properties, often due to null values, did not pose an issue. Another contributing factor to the data's validity is that the data types were strictly aligned with those defined in the shapes, as documented in the ShEx schema available online.[71]

The triplestore

The RARAA triplestore was created using GraphDB. The new namespace schema, controlled vocabulary, and RDF files were uploaded to the triplestore, resulting in 6.085 explicit RDF statements. The data is available through a SPARQL endpoint[72] and in dataRepositóriUM.[67,68,70]

Figure 7 shows a screenshot of the results of a SPARQL query to list all motifs' descriptions. As with other triplestores, its practical use depends on understanding the application profile, its shapes, and the relationships between them. To help maximize the potential for reusing this data, an upcoming article will detail and provide more examples of potentially

[27]https://shex.io/webapps/shex.js/doc/shex-simple.html

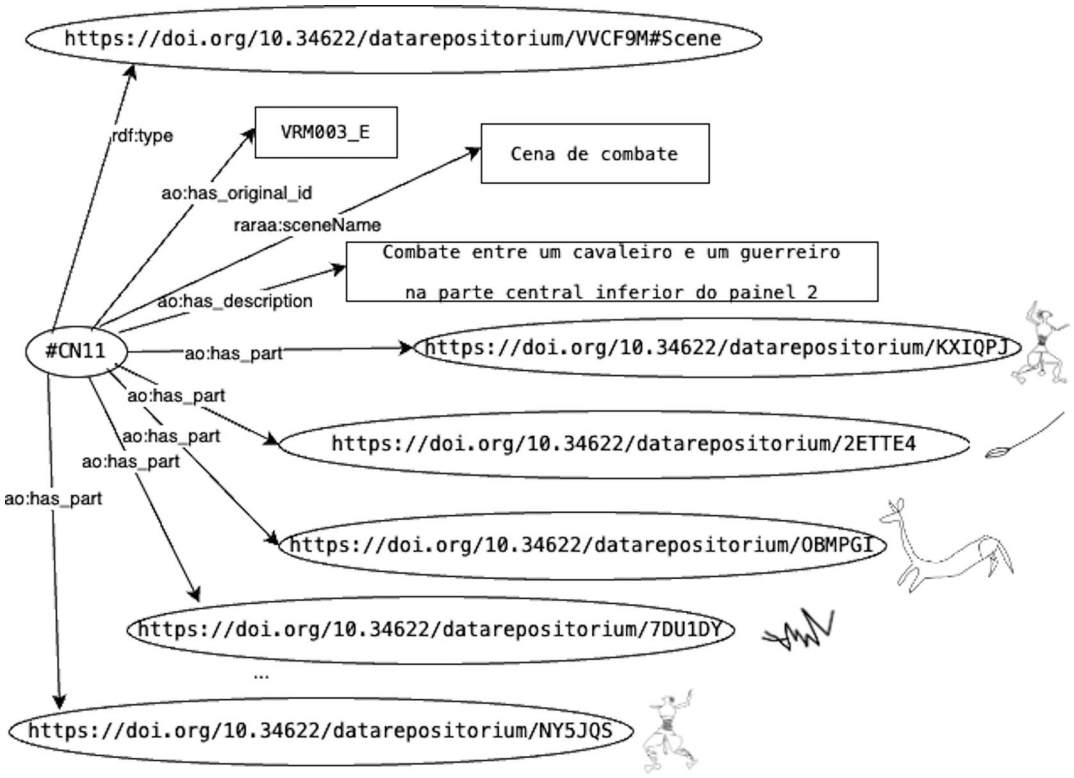

Figure 5. **RDF graph for the "VRM003-E" scene named as "Cena de Combate".**

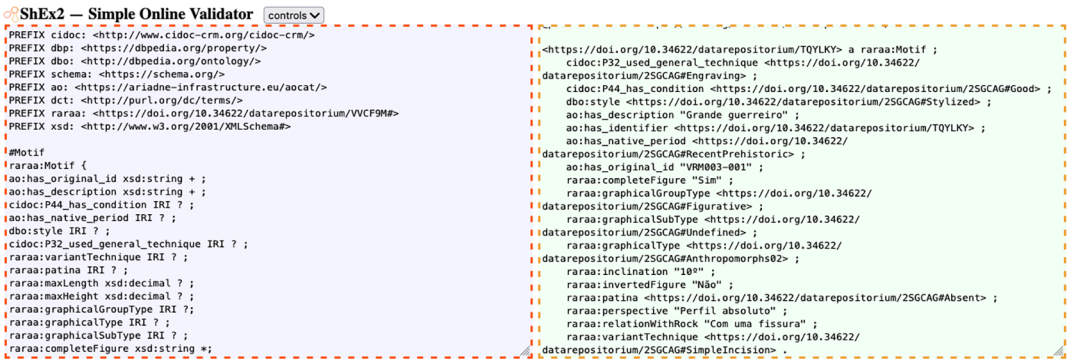

Figure 6. **Screenshot of ShEx validation for a motif (tool: ShEx2) [27].**

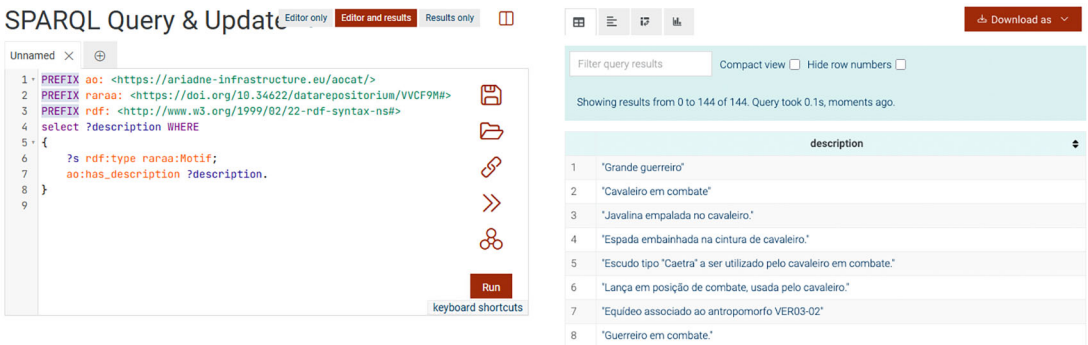

Figure 7. **SPARQL query to list all the descriptions of all motifs and part of its results.**

useful queries. Because these data were designed to maximize semantic interoperability, they can be easily integrated with other datasets from the same or different domains, enabling cross-domain analyses and comparative studies.

Files made available for download

The following files were made openly available for use and reuse:

- DCTAP; Distribution in a.XLSX Excel file and corresponding distributions in three.CSV files: the first with the metadata, the second with the DCTAP, and the third with information about the terms of the RARAA namespace - https://doi.org/10.34622/datarepositorium/IWOZHJ[69]
- RDF schema; Distribution in RDF/XML and Turtle (TTL) of the RARAA namespace RDF schema - https://doi.org/10.34622/datarepositorium/VVCF9M[67]
- Controlled vocabulary; Distribution in RDF/XML and TTL of the SKOS file with the controlled vocabulary for values - https://doi.org/10.34622/datarepositorium/2SGCAG[68]
- ShEx schema; Distribution in Shape Expressions Compact format (ShExC) for the five shapes: Site, Outcrop, Scene, Motif, and Represented Part - https://doi.org/10.34622/datarepositorium/TVJMZY[71]
- RDF files; Distributions in RDF/XML and TTL of the descriptions of 144 motifs, seven scenes, 181 represented parts, six outcrops, and one site (Vale da Vermelhosa) - https://doi.org/10.34622/datarepositorium/CQ42YJ[70]

These openly available resources are structured and standardized to facilitate reuse, not only within rock art studies but also in other domains that require semantic interoperability, such as cultural heritage documentation, archaeological data integration, or linked data research in environmental and spatial sciences. For instance, the RDF files and controlled vocabularies could support comparative studies of symbolic systems, integration of archaeological datasets for predictive modeling, or enrichment of museum and heritage digital platforms. At the same time, the SPARQL endpoint, combined with visualization tools, enables automated querying, visualization, and analysis of spatial, temporal, or thematic patterns across diverse datasets.

Conclusions and future work

Rock art represents a unique cultural heritage that intertwines artistic expression with its natural landscape. Through the RARAA project, we have prioritized documenting Iron Age rock art in the Côa Valley, adhering to FAIR principles to ensure our data are openly accessible and reusable for new interdisciplinary studies and interpretations.

The Semantic Web provides powerful tools for archaeology and rock art studies, enabling researchers to access, integrate, and interpret data more effectively while fostering broader collaboration. By working with semantic standards, associating detailed metadata with archaeological resources, and creating controlled vocabularies, and classes and properties vocabularies, we facilitate deeper discovery and understanding of this precious heritage.

Data interoperability enables archaeologists to connect and integrate data from different sources and formats across various disciplinary areas, such as geography and anthropology, which is fundamental to archaeological study and interpretation. Efficient and comprehensive access to and sharing of data are urgent needs in archaeology, promoting the work of researchers and fostering collaboration among researchers, projects, and institutions.

To make data from rock art studies in the Côa Valley conducted by the RARAA project available to other researchers, we created an open data repository aligned with the FAIR principles and emblematic projects in the area, including ARIADNE and ARIADNEplus. This repository is implemented in a triplestore, which can be accessed via a SPARQL endpoint.[72] Its creation involved several steps, from the study of the 2ArchIS database, the definition of the application profile using DCTAP, the definition of a new namespace schema (RDFS) and one controlled vocabulary schema (SKOS), the definition of a ShEx Schema, the transformation of data using OpenRefine, the validation of the data against the ShEx Schema, and, finally, uploading the data to the triplestore. All these outputs are available in dataRepositóriUM, the University of Minho's data repository.[67–71] Both the data made available via triplestore and those made available via dataRepositóriUM are helpful for other archaeology or Linked Data researchers, students, data managers, and interested citizens.

Contributions to theory and practice

This work provides three distinct contributions to the field of cultural heritage data management—methodological, technical, and theoretical.

From a methodological perspective, we explicitly describe a complete and reproducible workflow for converting specialized rock art data from relational databases into LOD, using DCTAP as the core model for developing application profiles. One of the purposes of adopting DCTAP was to assess its ability to express the various requirements that emerge during the design of application profiles. In our case, it successfully addressed almost all our needs throughout the process. Nevertheless, we could not express the intention of using part of a vocabulary (a branch, for example) as a value constraint. The possibility of using an IRIstem, combined with an adequate design of SKOS vocabularies, could overcome this expression difficulty. However, it would be more interesting to use a more expeditious and transparent way without resorting to subterfuges like that. Therefore, as a theoretical contribution, we can state that DCTAP effectively fulfilled its purpose, while also identifying a potential improvement: the ability to indicate one or more branches of a vocabulary as value constraints.

The University of Minho's open data repository was used as a key component to assign persistent identifiers to both motifs and developed artifacts, while simultaneously ensuring the long-term digital preservation of these resources. Documenting these specific steps within the overall project workflow is crucial, particularly for low-budget initiatives that lack the internal capacity to ensure identifier persistence and long-term preservation autonomously.

As technical contributions, we provide a set of reusable artifacts that other researchers and developers can adopt in their own projects involving either open or closed data: an application profile for the description of rock art expressed using DCTAP; a new schema of properties encoded in RDFS; a controlled vocabulary encoded in SKOS; a ShEx schema corresponding to the application profile shapes; and RDF files relating to rock art data stored in the 2ArchIS database. All these resources are openly available for use and reuse in dataRepositóriUM, alongside a triplestore accessible via its SPARQL endpoint. The developed vocabularies, being semantically interoperable, can be reused in other rock art or archaeology projects. The published LOD can also be integrated with data from different domains, fostering new perspectives and insights.

Our work also demonstrates how to balance domain-specific requirements through customized properties while maintaining cross-domain interoperability by aligning with major semantic frameworks, including CIDOC CRM, Ariadne, DBpedia, and schema.org. The online availability of the property, class, and value vocabularies supports their discussion, refinement, and potential integration with existing or future vocabularies, thereby contributing to both practice and theory in rock art knowledge organization. This process enhances the reusability of rock art data in new projects and interpretations, facilitating broader analyses at national and international scales.

Overall, this approach provides a sustainable model for other specialized archaeological subdisciplines seeking to publish their data as LOD, combining methodological rigor, semantic alignment, and practical reusability.

Reuse and access guide

All outputs produced in this work are openly available for access and reuse via dataRepositóriUM (DOI-identified deposits) and via the project triplestore (SPARQL endpoint). To make reuse easier—especially for users unfamiliar with RDF or SPARQL, we provide a README document (included in the main deposit record) that explains what each artifact contains and how they relate (DCTAP application profile, ShEx validation shapes, RDF instance data; plus the RDFS namespace schema and the SKOS controlled vocabulary).

The materials can be reused in at least three ways. First, as a research dataset, users may download the RDF files (RDF/XML or Turtle) and query them locally, or query the online triplestore via SPARQL, for example to list motifs and their descriptions, retrieve motifs by typology/chronology/technique, explore represented parts, or analyze motif participation in scenes. Second, as reusable semantic specifications, other teams may adopt the DCTAP and the SKOS vocabulary to describe compatible rock art resources and validate their RDF against the ShEx schema before publication or integration. Third, as interoperability building blocks, data managers and developers may reuse the RARAA namespace schema (RDFS) and align local terms and identifiers to facilitate interlinking and enrichment with other datasets.

The README provides a "start here" checklist, indicates recommended download formats, and includes a small set of example SPARQL queries and validation instructions to support immediate adoption.

Limitations and future work

Like all research projects, this one has limitations, summarized below. The RARAA project brought together specialists from different fields, archaeologists responsible for identifying and characterizing the motifs, and information technology experts in charge of digital surveys and LOD modeling. With the end of the project's funding, continuing this interdisciplinary work requires the rock art research team to create new funding opportunities or to rely on the voluntary contribution of data specialists to maintain and expand the LOD component.

The first limitation is that the source data (the 2ArchIS database) was not originally designed for data sharing and primarily reflects the perspective of Archaeology Unit of the University of Minho researchers. The structure of the source data, such as the attributes used, the controlled vocabularies and their organization, has a substantial impact on the design of the data to be published as Linked Data. Aligning these data with existing namespace schemas and controlled vocabularies is challenging, especially if those schemas and vocabularies already have predefined structures or explicit constraints that limit the reuse of their terms. For example, some of the properties we use must be able to be related to terms from our controlled vocabulary. This implied that certain properties from relevant schemas could not be used due to their range constraints. Our alignment efforts focused mainly on the ARIADNE and ARIADNEplus projects. However, these projects have not yet considered rock art-specific data, focusing more on archaeological excavation datasets, such as stratigraphy and materials. Consequently, several of the properties and classes we required were absent. To address this, we used more general vocabularies and ontologies, such as DBpedia, Schema.org, and Dublin Core, and even created new properties. Our application profile can therefore serve as a starting point for other projects that include or will include data on rock art, including ARIADNEplus.

In this project, we did not establish any connection between the 2ArchIS database and the triplestore. We first exported the data from 2ArchIS to XML and then imported it into OpenRefine, where we transformed it into RDF. As the 2ArchIS database is expected to continue operating and remain the primary source of data for the triplestore, it is crucial that we implement, in the future, a mechanism that performs these operations automatically or, at least, semi-automatically, based on the ShEx expressions we developed and made available.

Currently, we only make the data available via the SPARQL endpoint, which can be difficult for users unfamiliar with SPARQL. This limitation may jeopardize timely access to data for some researchers, underscoring the need to develop a user-friendly interface that enables non-technical users to access the data. This interface, expected to be available soon, will replicate some of 2ArchIS's functionalities, allowing external researchers to query the data like Archaeology Unit of the University of Minho archaeologists do. Exporting and downloading data in open formats, such as CSV, should also be possible in the future. Providing such features may encourage other archaeological research units to adopt similar practices, mainly Portuguese, Ibero-speaking, and European. Other actions may include developing tutorials or training materials and sessions, and making them available online.

Currently, our triplestore provides access only to the data, the schema, and the controlled vocabulary we created. It would be interesting in the future to enrich these data with data from other sources, including other institutions with rock art or related themes. It may, for instance, include geographical information in the form of GeoNames or TGN IRIs. It could also be enriched with specific ARIADNE, DBpedia, or Wikidata data. In this case, the interface for non-technical users should allow querying all data to discover new relationships and insights. Our data can also contribute to other data stores, such as Wikidata – efforts in this direction are needed. Future work may also include defining mappings between our controlled vocabulary and other vocabularies or ontologies, such as MAO/TAO, to enhance semantic interoperability.

Finally, new opportunities will emerge from integrating our data into the Social Sciences and Humanities Open marketplace and the Common European Data Space for Cultural Heritage. With our Linked Data we have already met several technical requirements. Still, others will need to be met, such as the use of the Data Catalogue Vocabulary (DCAT), and other operational and institutional requirements will need to be addressed in future work.

Ethics and consent

Ethical approval and consent were not required.

Data availability

The data used in this study originates from the 2ArchIS relational database developed at the University of Minho within the scope of the RARAA project. No new primary archaeological data were generated for this work.

All datasets and semantic artifacts produced and used in the context of this study are openly available through the University of Minho's institutional data repository, dataRepositóriUM. These include the complementary original data, the LOD representations of rock art motifs (in RDF/XML and Turtle formats), the application profile defined using DCTAP, the namespace schema encoded in RDFS, the controlled vocabularies encoded in SKOS, and the ShEx schema used for data validation.

Each artifact is individually identified with a persistent Digital Object Identifier (DOI), ensuring long-term preservation, citability, and unambiguous reference. The available data last versions are the following:

1. dataRepositóriUM. RARAA DCTAP. https://doi.org/10.34622/datarepositorium/IWOZHJ[69]
   - DCTAP_RARAA_FinalV3_1.tab -.XLSX file with the RARAA Application Profile (DCTAP).
   - DCTAP_Metadata_V3_1.tab-.CSV file with the Application Profile catalog metadata.
   - DCTAP_RARAA_FinalV3_1CSV.tab-.CSV file with the RARAA Application Profile (DCTAP).
   - Terms_RARAA_NamespaceV3_1.tab-.CSV file with information about the terms of the RARAA namespace.

Data is available under the terms of the CC0 - "Public Domain Dedication" - https://creativecommons.org/publicdomain/cc0/.

2. dataRepositóriUM. RARAA Properties and Classes Vocabulary for Rock Art. https://doi.org/10.34622/datarepositorium/VVCF9M[67]
   - RARAA_RockArtPropertiesVocabularyV2_1.rdf - RDF/XML file with RARAA namespace RDF schema.
   - RARAA_RockArtPropertiesVocabularyV2_1.ttl - Turtle file RARAA namespace RDF schema.

Data is available under the terms of the CC0 - "Public Domain Dedication" - https://creativecommons.org/publicdomain/cc0/.

3. dataRepositóriUM. RARAA Values Vocabulary for Rock Art. https://doi.org/10.34622/datarepositorium/2SGCAG[68]
   - RARAA_RockArtValuesVocabularyV2_1skos.rdf - RDF/XML file with SKOS RARAA Values Vocabulary for Rock Art.
   - RARAA_RockArtValuesVocabularyV2_1skos.ttl - Turtle file SKOS RARAA Values Vocabulary for Rock Art.

Data is available under the terms of the CC-BY-4.0 - "Creative Commons Attribution" - https://creativecommons.org/licenses/by/4.0/.

4. dataRepositóriUM. RARAA Shape Expressions. https://doi.org/10.34622/datarepositorium/TVJMZY[71]
   - RARAAShexFileV2.shex – Shape Expressions Compact format (ShExC) file for the five shapes: Site, Outcrop, Scene, Motif, and Represented Part.
   - ShEx_ValidationV2.jpg - Screenshot with a validation example.

Data is available under the terms of the CC0 - "Public Domain Dedication" - https://creativecommons.org/publicdomain/cc0/.

5. dataRepositóriUM. RARAA Rock Art Data. https://doi.org/10.34622/datarepositorium/CQ42YJ[70]
   - motifV2_1.rdf - RDF/XML data for motifs.
   - motifV2_1.ttl - Turtle data for motifs.
   - outcropV2_1.rdf - RDF/XML data for outcrops.
   - outcropV2_1.ttl - Turtle data for outcrops.
   - representedPartV2_1.rdf - RDF/XML data for represented parts.
   - representedPartV2_1.ttl - Turtle data for represented parts.

- sceneV2_1.rdf - RDF/XML data for scenes.
- sceneV2_1.ttl - Turtle data for scenes.
- siteV1_1.rdf - RDF/XML data for sites.
- siteV1_1.ttl - Turtle data for sites.

Data is available under the terms of the CC-BY-4.0 - "Creative Commons Attribution" - https://creativecommons.org/licenses/by/4.0/.

6. dataRepositóriUM. 2ArchIS tables attributes values. https://doi.org/10.34622/datarepositorium/XPSP7L[63]
   - 00_ReadMe_file.txt – ReadMe texto file with the description of the dataset.
   - 01_2ArchIS tables and attributes values.tab-.XLSX file with information about 2ArchIs tables, attributes and values used within the RARAA project.
   - 02_2ArchIS tables and attributes values.csv - 2ArchIS enumerated values and respective tables and attributes, excluding Group, Type and Sub-type (terms translated to English).
   - 03_2ArchIS tables and attributes values.csv - 2ArchIS enumerated values and respective tables and attributes for Group, Type and Sub-Type (terms translated to English).
   - 2ArchIS tables attributes values.csv - 2ArchIS enumerated values and respective tables and attributes, excluding Group, Type and Sub-type (terms translated to English).

Data is available under the terms of the CC-BY-4.0 - "Creative Commons Attribution" - https://creativecommons.org/licenses/by/4.0/.

7. dataRepositóriUM. 2ArchIS XML Data Export (RARAA Project). https://doi.org/10.34622/datarepositorium/KZEINV70
   - afloramento.xml - Records describing rock outcrops associated with archaeological and rock art contexts, preserving information on natural geological surfaces where engravings or other archaeological features occur.
   - cena.xml - Records describing rock art scenes, understood as groups of motifs that appear to be conceptually or visually connected, forming structured narrative or compositional units.
   - motivo.xml - Records describing rock art motifs, representing individual engraved figures identified within rock art panels.
   - parterepresentada.xml - Records describing the represented parts within each rock art motif, detailing the specific anatomical, symbolic, or compositional elements depicted.
   - sitionucleo.xml - Records describing archaeological sitio_nucleo, representing spatially defined archaeological site units within the 2ArchIS database.
   - README.txt - Readme file about the dataset with the following contents: Overview, Content, Data Structure, Project Context, Purpose, Citation, License and Contact.

Data is available under the terms of the CC-BY-4.0 - "Creative Commons Attribution" - https://creativecommons.org/licenses/by/4.0/.

Acknowledgments

We thank Professor Julian D. Richards and Maria Theodoridou for their availability and assistance regarding the ARIADNE and ARIADNE Plus projects.

We also thank the reviewers of this article, whose commitment and generosity contributed to clarity and more valuable content.

References

1. Dublin Core Metadata Initiative: **Dublin Core [Internet].** [cited 2026 Apr 7]. Glossary.
Reference Source
2. Coyle K: **DC Tabular Application Profiles (DC TAP) - Primer.** *Dublin Core Metadata Initiative.* Sep. 26, 2023. Accessed: Sep. 30, 2023.
Reference Source
3. Bizer C, Heath T, Berners-Lee T: **Linked Data - The story so sar.** *Int. J. Semant. Web Inf. Syst.* 2009; **5**(3): 1–22.
Publisher Full Text
4. Heath T: **Humans and the Web of Data.** Sep. 05, 2008. Accessed: Feb. 09, 2026.
Reference Source
5. Open Knowledge International: **Open Definition - Defining Open in Open Data, Open Content and Open Knowledge.** Accessed: May 23, 2017.
Reference Source
6. Kellogg G, Hartig O, Champin PA, *et al.*: **RDF 1.2 Concepts and Abstract Data Model - W3C Working Draft 03 April 2026 [Internet].** *World Wide Web Consortium*; 2026 [cited 2026 Apr 6].
Reference Source
7. Tomaszuk D: **RDF 1.2 Schema - W3C Working Draft 28 March 2026 [Internet].** *World Wide Web Consortium*; 2026 [cited 2026 Apr 8].
Reference Source
8. Baker T, Prud'hommeaux E: *Shape Expressions (ShEx) 2.1 Primer - Final Community Group Report 09 October 2019.* W3C; Oct. 09, 2019. Accessed: Jan. 16, 2025.
Reference Source
9. Miles A, Bechhofer S: **SKOS Simple Knowledge Organization System Reference [Internet].** 2009 [cited 2009 May 26].
Publisher Full Text
10. Hartig O, Seaborne A, Taelman R, *et al.*: **SPARQL 1.2 Query Language - W3C Working Draft 23 March 2026 [Internet].** *World Wide Web Consortium.* 2026.
Reference Source
11. Champin PA, Lindström N: *RDF 1.2 Primer - W3C Group Note Draft 16 April 2026 [Internet].* World Wide Web Consortium (W3C); 2026. [cited 2026 Jun 2].
Reference Source
12. World Wide Web Consortium: *Category: Triple Store.* W3C Semantic Web Standards; Accessed: Jan. 28, 2025.
Reference Source
13. CoaParque: **A Região. Côa Parque.** *A Região.* Côa Parque; Accessed: Jan. 13, 2022.
Reference Source
14. CoaParque: **Parque. Côa Parque.** *Parque.* Côa Parque; Accessed: Jan. 13, 2022.
Reference Source
15. Wright H, Richards JD: **Reflections on Collaborative Archaeology and Large-Scale Online Research Infrastructures.** *J. Field Archaeol.* Oct. 2018; **43**(sup1): S60–S67.
Publisher Full Text
16. Kansa E, Kansa SW: **Digital Data and Data Literacy in Archaeology Now and in the New Decade.** *Adv. Archaeol. Pract.* Feb. 2021; **9**(1): 81–85.
Publisher Full Text
17. Padilla T, Scates Kettler H, Varner S, *et al.*: **Vancouver Statement on Collections as Data.** *Zenodo.* Sep. 13, 2023.
Publisher Full Text
18. Wilkinson MD, *et al.*: **The FAIR Guiding Principles for scientific data management and stewardship.** *Sci. Data.* Mar. 2016; **3**(1): 160018.
PubMed Abstract | Publisher Full Text | Free Full Text
19. Carroll SR, Herczog E, Hudson M, *et al.*: **Operationalizing the CARE and FAIR Principles for Indigenous data futures.** *Sci. Data.* Apr. 2021; **8**(1): 108.
PubMed Abstract | Publisher Full Text | Free Full Text
20. GO FAIR: **GO FAIR.** Accessed: Jan. 19, 2024.
Reference Source
21. Lóscio BF, Burle C, Calegari N: **Data on the Web Best Practices.** Jan. 2017. Accessed: Jun. 09, 2017.
Reference Source
22. Meghini C, *et al.*: **ARIADNE: A Research Infrastructure for Archaeology.** *J. Comput. Cult. Herit.* Aug. 2017; **10**(3): 1–27.
Publisher Full Text
23. Ronzino P, Felicetti A, Giorgio SD: **Integrating Archaeological Datasets: the ARIADNE Portal.** *Proceedings of the 10th International Workshop on Artificial Intelligence for Cultural Heritage.* Genova, Italy: CEUR Workshop Proceedings; Nov. 2016; pp. 42–49. Accessed: Apr. 18, 2024.
Reference Source
24. ARIADNE Research Infrastructure AISBL: **What is the ARIADNE RI?.** Accessed: Nov. 10, 2025.
Reference Source
25. Aloia N, Binding C, Cuy S, *et al.*: **Enabling European Archaeological Research: The ARIADNE E-Infrastructure.** *Internet Archaeol.* 2017; **43**.
Publisher Full Text
26. Felicetti A, Meghini C, Richards J, *et al.*: **The AO-Cat Ontology.** *Zenodo.* Mar. 2023.
Publisher Full Text
27. Salzburg Research InstituteGeser G: **Achievements of the ARIADNE Initiative for Archaeological Data Sharing and Research.** *Internet Archaeol.* Nov. 2023; **64**.
Publisher Full Text
28. **What is the CIDOC CRM?.** *CIDOC CRM|Conceptual Reference Model.* 2024. Accessed: Apr. 18.
Reference Source
29. CIDOC CRM Special Interest Group: **CIDOC CRM.** *What is the CIDOC CRM?* Accessed: Nov. 24, 2021.
Reference Source
30. CIDOC CRM Special Interest Group: **Volume A: Definition of the CIDOC Conceptual Reference Model.** Apr. 2021. Accessed: Dec. 20, 2021.
Reference Source
31. **Definition of the CRMarchaeo - An Extension of CRMbase to support the archaeological excavation process.** Apr. 2024. Accessed: Jan. 18, 2025.
Reference Source
32. Semantic Computing Research Group (SeCo): **'About the FindSampo Portal', FindSampo - Archaeological Finds on the Semantic Web.**
33. Hyvönen E, *et al.*: **Citizen science archaeological finds on the Semantic Web: the FindSampo framework.** *Antiquity.* Aug. 2021; **95**(382): e24.
Publisher Full Text
34. finto - Finnish Thesaurus and Ontology Service: **MAO/TAO - Ontology for Museum Domain and Applied Arts.** Accessed: Mar. 02, 2022.
Reference Source
35. **Definition of the Europeana Data Model v5.2.8.** Oct. 06, 2017. Accessed: Jan. 18, 2025.
Reference Source
36. Europeana: *About.* Europeana; Accessed: Jan. 22, 2025.
Reference Source
37. European Commission: *Directorate General for Communications Networks, Content and Technology., Common European data space for cultural heritage: annual report 2023/2024.* LU: Publications Office;

2025. Accessed: Jan. 22, 2025.
Publisher Full Text

38. European Commission, DG CONNECT: **Common European Data Spaces.** *Shaping Europe's digital future.* Accessed: Feb. 07, 2025.
Reference Source

39. *Data Spaces Blueprint v1.0.* Data Spaces Support Centre, Blueprint; Mar. 2024. Accessed: Jan. 22, 2025.
Reference Source

40. Dublin Core Metadata Initiative: **DCMI Metadata Terms.** Accessed: May 23, 2017.
Reference Source

41. Albertoni R, Browning D, Cox SJD, *et al.*: *Data Catalog Vocabulary (DCAT) - Version 3.* W3C Recommendation; Aug. 22, 2024.
Reference Source

42. DBpedia Association: *About DBpedia.* DBpedia; Accessed: Jan. 22, 2025.
Reference Source

43. Auer S, Bizer C, Kobilarov G, *et al.*: **DBpedia: A Nucleus for a Web of Open Data.** *The Semantic Web.* Aberer K, Choi K-S, Noy N, *et al.*, editors. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007; vol. **4825**: pp. 722–735. Lecture Notes in Computer Science, vol. 4825.
Publisher Full Text

44. Ismayilov A, Kontokostas D, Auer S, *et al.*: **Wikidata through the eyes of DBpedia.** *SW.* 2018; **9**(4): 493–503.
Publisher Full Text

45. DBpedia Association: **Ontology (DBO).** Accessed: Jan. 22, 2025.
Reference Source

46. Freudenberg M: **Dbo vs Dbp.** 2024. users@jena.apache.org. Accessed: Sep. 02
Reference Source

47. *Wikidata: Introduction.* Wikidata; Accessed: Jan. 23, 2025.
Reference Source

48. *Welcome to Wikidata.* Wikidata; Accessed: Jan. 24, 2025.
Reference Source

49. Schema.org: Accessed: Jan. 17, 2025.
Reference Source | **Reference Source**

50. Schema.org: **Schemas.** Accessed: Jan. 22, 2025.
Reference Source | **Reference Source**

51. **About GeoNames.** Accessed: Jan. 23, 2025.
Reference Source

52. **About the TGN.** *Getty Thesaurus of Geographic Names.* Accessed: Jan. 23, 2025.
Reference Source

53. De Boer V, *et al.*: **Supporting Linked Data Production for Cultural Heritage Institutes: The Amsterdam Museum Case Study.** *The Semantic Web: Research and Applications.* Simperl E, Cimiano P, Polleres A, *et al.*, editors. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012; vol. **7295**: pp. 733–747. Lecture Notes in Computer Science, vol. 7295.
Publisher Full Text

54. Hong Y, Solanki M, Foxhall L, *et al.*: **A Framework for Transforming Archaeological Databases to Linked Ontological Datasets.** *CAA 2010, Fusion of Cultures. Proceedings of the 38th Annual Conference on Computer Applications and Quantitative Methods in Archaeology, Granada, Spain, April 2010.* 2013; pp. 569–574.

55. Daquino M, Mambelli F, Peroni S, *et al.*: **Enhancing Semantic Expressivity in the Cultural Heritage Domain: Exposing the Zeri Photo Archive as Linked Open Data.** *J. Comput. Cult. Herit.* Oct. 2017; **10**(4): 1–21.
Publisher Full Text

56. Nishanbaev I, Champion E, McMeekin DA: **A Comparative Evaluation of Geospatial Semantic Web Frameworks for Cultural Heritage.** *Heritage.* Aug. 2020; **3**(3): 875–890.
Publisher Full Text

57. Carriero VA, Gangemi A, Mancinelli ML, *et al.*: **Pattern-based design applied to cultural heritage knowledge graphs: ArCo: The knowledge graph of Italian Cultural Heritage.** *SW.* Jan. 2021; **12**(2): 313–357.
Publisher Full Text

58. Binding C, Evans T, Gilham J, *et al.*: **Linked Data for the Historic Environment.** *Internet Archaeol.* Jul. 2022; **59**.
Publisher Full Text

59. Botica N, Luís L, Silva J-P: **Atributos e descritores propostos para a arte rupestre da Idade do Ferro no Vale do Côa.** *Cuadernos de Arqueología.* Jun. 2022; **30**: 61–90.
Publisher Full Text

60. Botica N, Magalhães F, Machado D, *et al.*: **Del sistema de información 2ArchIS al DataRepositóriUM: el estudio de caso de hallazgos arqueológicos.** *Cuad Arqueol. Univ. Navar.* 2021 Apr; **29**(29): 381–396.
Publisher Full Text

61. Bettencourt AMS, Santos-Estevez M, Sampaio HA, *et al.*: **Prehistoric Rock Art and Landscape: the outcrops and its forms were certainly not simple carving surfaces**. In: Ribeiro C, Martins RN, editors. *I International Seminar on Art and Landscape [Internet]* Guimarães, Portugal: Laboratório da Paisagem; 2017. pp. 137–153.
Reference Source

62. Cruz Berrocal M, Vicent GJ: **Rock art as an archaeological and social indicator: The neolithisation of the Iberian Peninsula.** *J. Anthropol. Archaeol.* 2007 Dec; **26**(4): 676–697.
Publisher Full Text

63. Botica N: *2ArchIS tables attributes values.* Repositório de Dados da Universidade do Minho; 2026.
Publisher Full Text

64. Botica N, Luís L, Bernardes P: **Use of photogrammetry to survey Iron Age rock art motifs in the Côa Valley: the Vermelhosa Rock 3 case study (Vila Nova de Foz Côa, Portugal).** *Virtual Archaeol. Rev.* Nov. 2023; **15**(30): 97–109.
Publisher Full Text

65. Coyle K: **DCTAP - Elements for DC Tabular Application Profiles.** *Draft - Request for Comments.* Dec. 16, 2022. Accessed: Jan. 17, 2025.
Reference Source

66. Botica N: **2ArchIS XML Data Export (RARAA Project) [Internet].** *Repositório de Dados da Universidade do Minho*; 2026 [cited 2026 Apr 15].
Publisher Full Text

67. Baptista AA, Miguel I, Galvão J, *et al.*: *RARAA Properties and Classes Vocabulary for Rock Art.* Repositório de Dados da Universidade do Minho; 2024.
Publisher Full Text

68. Botica N, Miguel I, Galvão J, *et al.*: *RARAA Values Vocabulary for Rock Art.* Repositório de Dados da Universidade do Minho; 2024.
Publisher Full Text

69. Miguel I, Galvão JP, Botica N, *et al.*: *RARAA DCTAP.* Repositório de Dados da Universidade do Minho; 2024.
Publisher Full Text

70. Baptista AA, Miguel I, Galvão J, *et al.*: *RARAA Rock Art Data.* Repositório de Dados da Universidade do Minho; 2024.
Publisher Full Text

71. Galvão J, Miguel I, Botica N, *et al.*: *RARAA Shape Expressions.* Repositório de Dados da Universidade do Minho; 2024.
Publisher Full Text

72. Baptista AA, Botica N, Miguel I, *et al.*: *RARAA SPARQL endpoint.* GraphDB; 2024. [RDF].
Reference Source | **Reference Source**

Open Peer Review

Current Peer Review Status: ? ✓ ✓

Version 3

Reviewer Report 04 August 2026

https://doi.org/10.21956/openreseurope.26498.r70885

© **2026 Garate D.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✓ **Diego Garate**
Universidad de Cantabria, Santander, Spain

I approve the changes made by the authors.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 2

Reviewer Report 29 May 2026

https://doi.org/10.21956/openreseurope.25768.r74077

© **2026 Garate D.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

? **Diego Garate**
Universidad de Cantabria, Santander, Spain

This work presents a significant initiative in the field of archaeological data management and openness, addressing the transformation of the information contained in the 2ArchIS database into an accessible and reusable Linked Open Data (LOD) dataset.

It is a technically sound, well-documented contribution, aligned with current trends in the openness, interoperability and digital preservation of archaeological heritage. However, beyond

the quality of the technical development, the manuscript raises several issues that limit its scientific scope, and which should be addressed before considering its final indexing.

The main issue lies in the imbalance between the technological dimension of the work and its archaeological contribution. Although the article is presented within the field of rock art and takes the Côa Valley as its case study, the discussion focuses almost exclusively on aspects related to semantic modelling, the construction of controlled vocabularies and data transformation procedures. The reader is provided with a detailed description of the computational processes developed, but there is little explanation of the implications this new model may have for archaeological research itself.

Throughout the text, it is repeatedly stated that opening up the data will enable new insights into rock art to be generated or facilitate comparative analyses at different scales. However, these claims remain at the level of potential and are not substantiated by concrete examples. It would have been desirable to include a case study illustrating how the proposed structure allows archaeological questions to be formulated that were difficult or impossible to address using the original database. Similarly, the absence of examples of effective integration with other repositories or datasets limits the reader's ability to assess the true scope of the proposal.

In this regard, the manuscript also suffers from a lack of evaluation of the developed model. Formal validation using ShEx ensures the structural consistency of the data, but does not provide information on the scientific utility of the system. It would be advisable to incorporate some form of comparative evaluation to assess the model's advantages over other existing solutions or over the original relational structure itself. Currently, the work demonstrates that the transformation is technically possible, but fails to fully demonstrate its benefits for research.
Another aspect worthy of attention is the actual size of the published dataset. Although the introduction places the work in the context of the more than 1,200 known engraved rocks in the Côa Valley, the model is ultimately applied only to a relatively small set of records corresponding to a small part of the RARAA project. This does not invalidate the work, but it does call for a more explicit reflection on the representativeness of the sample and on the limitations this imposes on the future reuse of the data.

From an archaeological perspective, the literature review is insufficient. The article engages extensively with the literature on ontologies, metadata and the Semantic Web, but makes scant reference to work carried out over recent decades on the digital documentation of rock art or the management of archaeological databases. Particularly striking is the scarcity of references relating to the digital documentation and analysis of rock art, as well as the absence of a more in-depth discussion of international projects that have addressed similar problems from different perspectives.
Certain sections, particularly the introductory glossary and parts of the methodological description, are excessively lengthy in relation to the subsequent scientific discussion. The detailed listing of properties, classes and vocabularies takes up considerable space that could be partially moved to supplementary materials. Such a reorganisation would allow for the reinforcement of aspects that are currently dealt with too briefly, such as the scientific implications of the project, its potential for reuse, or its contribution to the study of rock art.

Consequently, I believe the article is of sufficient merit to be indexed, but it requires revision to strengthen its archaeological dimension and to demonstrate more clearly the scientific

contribution that this new data environment offers to the study of rock art, beyond the technological innovation itself.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Rock Art

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 29 Jun 2026

Ana BAPTISTA

We thank the reviewer for the careful reading of the manuscript and for the constructive comments. We are pleased that the reviewer recognizes the technical soundness of the work and its contribution to archaeological data management, interoperability, and openness. Regarding the balance between the technological and archaeological dimensions of the article, we would like to clarify that the primary objective of this contribution is methodological. The manuscript focuses on the process of transforming archaeological data from a relational database into Linked Open Data and on the development of the semantic artifacts required to support this transformation. For this reason, the emphasis is placed on the design of the application profile, vocabularies, validation mechanisms, and publication workflow rather than on archaeological interpretation. We agree that Linked Open Data environments may ultimately support new archaeological questions and comparative analyses. However, demonstrating such outcomes would require a separate research stage involving the integration of additional datasets and the formulation of specific archaeological case studies. At the current stage, the objective was to establish and

document a reusable semantic framework capable of enabling such future research. We have therefore been careful throughout the manuscript to frame these possibilities as potential benefits rather than as demonstrated outcomes. Concerning the evaluation of the model, our intention was to validate the consistency and completeness of the data representation through the application profile and ShEx validation process. The article does not aim to compare alternative modeling approaches or database technologies, but rather to document a complete and reproducible workflow for publishing archaeological rock art data as Linked Open Data. We acknowledge that comparative assessments of different semantic models and their research outcomes would constitute valuable future work. With respect to the size of the published dataset, we agree that the scope should be made explicit. The manuscript now clarifies that the Linked Open Data workflow was developed and validated using the Vermelhosa dataset, corresponding to a limited number of engraved rocks within the broader RARAA project. We regard this dataset as a pilot implementation that establishes the foundations for the future incorporation of additional records from Vale de José Esteves, Meijapão, and other areas of the Côa Valley. Regarding the literature review, we acknowledge the importance of the broader literature on digital rock art documentation and archaeological information systems. However, given the specific focus of this article on semantic modeling, Linked Data publication, and interoperability, we prioritized references directly related to ontologies, metadata standards, application profiles, and archaeological Linked Open Data initiatives. We recognize that a more extensive review of digital rock art documentation projects could further enrich the archaeological context and represents a valuable avenue for future work. Finally, concerning the length of some methodological sections, we understand the reviewer's concern. However, we intentionally provide a detailed description of the modeling decisions, vocabularies, and workflows because one of the central contributions of the paper is to offer a transparent and reproducible methodological framework that can be adopted and adapted by other researchers and institutions working with archaeological and cultural heritage data. We thank the reviewer once again for these valuable observations, which have helped us clarify both the scope and the intended contribution of the manuscript.

Competing Interests: No competing interests were disclosed.

Reviewer Report 25 May 2026

https://doi.org/10.21956/openreseurope.25768.r73893

© 2026 Potter R. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rich Potter
University of Gothenburg, Gothenburg, Sweden

Much more readable than the first draft and ready for indexing with some very minor

suggestions.

1. The first sentence in the abstract is not quite right. Does the Côa Valley Archaeological Park hold the data or is it the 2ArchIS database? This needs to be reworded.
2. in the background you only talk about Palaeolothic art, but you have worked with only the Iron Age material.
3. It would be good to include a definition of triples in your glossary
4. It might be an idea to be explicit about how many of the 1200 sites you use in your project.

Glossary has otherwise worked well, and new images are an improvement.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Rock Art, Digital Archaeology, 3D Documentation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 29 Jun 2026

Ana BAPTISTA

We are pleased to submit the revised version of our manuscript. We are very grateful to the reviewer for the time and effort he invested in evaluating our manuscript. We highly appreciate his thorough review and positive guidance, which have greatly helped us enhance both the scientific quality and the readability of this work. Below, we address the reviewer's comments and describe the changes made.

1. "The first sentence in the abstract is not quite right. Does the Côa Valley Archaeological Park hold the data or is it the 2ArchIS database? This needs to be reworded."
2. "In the background you only talk about Palaeolothic art, but you have worked with only the Iron Age material."

Response: We agree that the original wording could create ambiguity regarding the repository of the archaeological data. To clarify this point, the abstract has been revised to explicitly identify the University of Minho's 2ArchIS database as the platform where the archaeological documentation of the Iron Age engravings is stored and managed. This revision distinguishes the role of the Côa Valley Archaeological Park as the institution responsible for preserving the engraved rocks from that of the 2ArchIS database as the repository of the associated archaeological records.

revised text The Côa Valley Archaeological Park, a UNESCO World Heritage, preserves more than 1,200 engraved rocks spanning a long chronological sequence from Prehistory to the Contemporary period. Among these, Iron Age engravings form a significant secondary corpus with no known parallel in the Iberian Peninsula. The archaeological documentation of these engravings is stored and managed in the University of Minho's 2ArchIS database. This article presents the process of converting these previously restricted archaeological records into an openly accessible dataset, with the aim of improving data sharing, reproducibility, and future research on Iron Age rock art.

1. "It would be good to include a definition of triples in your glossary"

Response: We included in the glossary the following entry: "Triple - In this context, 'triple'

refers to an RDF triple (or RDF statement), which consists of three elements: subject, predicate, and object—or, in alternative terminology: resource, property, and value. A triple "expresses a relationship between two resources. The **subject** and the **object** represent the two resources being related; the **predicate** represents the nature of their relationship" [72] . Whenever possible, persistent IRIs [26] are used to represent and uniquely identify the subject, predicate, and object. For example, the subject can be represented by [https://doi.org/10.34622/datarepositorium/NY5JQS], the predicate by [https://ariadne-infrastructure.eu/aocat/1.1/has_description], and the object can be the literal string "Cavaleiro em combate", which translates to 'Knight in combat'. This means that the description of the resource identified by https://doi.org/10.34622/datarepositorium/NY5JQS is 'Cavaleiro em combate'."

1. "It might be an idea to be explicit about how many of the 1200 sites you use in your project."

Response: We agree that the scope of the dataset should be stated more explicitly. To address this point, the Introduction has been revised to clarify that, although the Côa Valley Archaeological Park contains more than 1,200 documented engraved rocks, the present dataset focuses on a specific subset of Iron Age rock art. The revised text now specifies that the project concentrated on three areas: Vermelhosa (10 engraved rocks), Vale de José Esteves (4 engraved rocks), and Meijapão (3 engraved rocks), corresponding to a total of 17 engraved rocks. This addition makes the scale and limits of the dataset clearer for readers.
revised text Introduction The Côa Valley (Côa Valley Archaeological Park), located in the Guarda district of Portugal, covers a vast area with more than 1,200 engraved rocks. It is internationally recognised as the largest open-air ensemble of Palaeolithic rock art in the world and has been inscribed on the UNESCO World Heritage List since 1998. While Palaeolithic engravings constitute the dominant component of the rock art record, the valley also preserves later phases of occupation, including Iron Age engravings, which represent the second most frequent chronological phase. [12, 13] Within this broader archaeological context, the Open Access Rock Art Repository project (RARAA - Repositório de Arte Rupestre de Acesso Aberto), funded by the Portuguese Foundation for Science and Technology (FCT), focused specifically on Iron Age rock art, aiming to improve knowledge of both the engravings and the communities that produced them. Of the approximately 1,200 engraved rocks currently documented in the park, Iron Age engravings form a relatively small but archaeologically significant subset of the record. Within this subset, the project concentrated on three areas: Vermelhosa (10 engraved rocks), Vale de José Esteves (4 engraved rocks), and Meijapão (3 engraved rocks). Consequently, the dataset worked in RARAA corresponds to 17 engraved rocks from three Iron Age rock art areas, representing a small but well-defined subset of the broader rock art record of the Côa Valley. All data gathered by the RARAA project is stored in a relational database called 2ArchIS. While this database serves the project and the Archaeology Unit of the University of Minho well, the team sought to connect their data with those of other research groups and, conversely, to make it available for use and reuse by researchers and the wider public. In short, the goal was to make the data open and semantically interoperable—that is, to publish it as Linked Open Data (LOD). This led to the pilot initiative presented here, in which the Vermelhosa dataset was converted and published as LOD, establishing a workflow that can be extended to the data of the remaining areas. Once again, thank you for your review, which helped us improve the article.

Competing Interests: No competing interests were disclosed.

Version 1

Reviewer Report 26 March 2026

https://doi.org/10.21956/openreseurope.23694.r70889

© 2026 Potter R. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Rich Potter
University of Gothenburg, Gothenburg, Sweden

This is an interesting, if at times impenetrable, article relating to the creation of a database for the curation and FAIRification of rock art motifs in Portugal. The article is heavily technical in places, and covers the reasoning behind decisions made during its creation in detail. I am writing this review from the perspective of an end user as someone who works extensively with rock art. It is my opinion that the article should be indexed, but it needs revision first.

The overall content of the article is good, but I would suggest that the introduction and related work section need to be revised. A lot of the concepts that are introduced seem to not be used later on, or in the case of ShEx, reintroduced fully later on. There are sections where a concept is introduced and then never described in full. For example, linked data on page 3 would benefit from a better description than the quote provided. It would be good if these descriptions used simplified terminology or were expanded upon for a wider audience. You also need to introduce and explain terms like triplestore.

The main issue for me was that there are too many acronyms throughout the article, which makes it confusing in places, particularly the introduction and related work. This can be alleviated by perhaps removing some terms, e.g. CH should just be cultural heritage. A glossary, or table of acronyms would help improve readability greatly. Several of the acronyms are never introduced, e.g. TGN, CIDOC-CRM, and even the project name RARAA. CIDOC-CRM is frequently written as CIDOC CRM. Towards the end of the article, some of the acronyms are dropped, e.g. LOD becomes Linked Open Data from the conclusion onwards.

I missed a description of what the intent of the database is. Is it just for data collection/dissemination or are there other reasons behind its creation? Has it already been used for something?

It is admirable that you have made everything open, however, I was unable to find any description of how to reuse any of the material. It might be an idea to include a readme somewhere. How do

you anticipate it will be used?

While it is obviously not the main theme of the article, it would be good if you expanded a little on the fieldwork on page 3 as the sentence is a little bit standalone. Since this is the basis of where the data in the database comes from, it would be good to know what sort of fieldwork was done.

Images are generally good, though figure 5 could perhaps benefit from being rotated as it is unreadable on paper. Figure 2 could be improved by moving the labels, if possible, though this is a minor issue.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Rock Art, Digital Archaeology, 3D Documentation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 24 Apr 2026

Ana BAPTISTA

We appreciate your valuable input and the thorough review of our work. For clarity, we have addressed your requests by organizing them into the following six points.

1 – Define better “Linked Data”, triplestore and other concepts for a vaster audience. We improved the definitions of the following terms: Linked Data, Triplestore, Application Profile, RDF, RDFS, SKOS, Values Vocabulary, Properties and Classes vocabulary (sometimes called ontology). All these terms were put into the glossary before the Introduction and their

definitions were removed from the body of the text.

2 – Use less acronyms. While there are some acronyms that can be not used (e.g., CH for cultural heritage), others are better recognized as acronyms than by the expanded version (e.g., SKOS, RDF, FAIR, or CARE). We, therefore, alleviated the following acronyms: CH – Cultural Heritage RDB – Relational Database UAUM – Archaeology Unit of the University of Minho LOV – Linked Open Vocabularies EDM – Europeana Data Model DWBP – Data on the Web Best Practices And other acronyms that were expanded once and never again used.

3 - Expand – TGN, CIDOC-CRM (or CIDOC CRM?), RARAA and verify all other acronyms. CIDOC CRM does not have a hyphen in the official documentation, so we have amended the sole occurrence in which a hyphen was used. The acronym was already expanded in the text. TGN was also indeed expanded, but the acronym was not related to the expanded version, which we corrected. We also expanded the RARAA acronym and verified all the acronyms.

4 - It might be an idea to include a readme somewhere. How do you anticipate it [the data] will be used? Under the Conclusions and Future Work, we included a subsection with the following text: “Reuse and access guide All outputs produced in this work are openly available for access and reuse via dataRepositóriUM (DOI-identified deposits) and via the project triplestore (SPARQL endpoint). To make reuse easier—especially for users unfamiliar with RDF or SPARQL, we provide a README document (included in the main deposit record) that explains what each artifact contains and how they relate (DCTAP application profile , ShEx validation shapes, RDF instance data; plus the RDFS namespace schema and the SKOS controlled vocabulary). The materials can be reused in at least three ways. First, as a research dataset, users may download the RDF files (RDF/XML or Turtle) and query them locally, or query the online triplestore via SPARQL, for example to list motifs and their descriptions, retrieve motifs by typology/chronology/technique, explore represented parts, or analyze motif participation in scenes. Second, as reusable semantic specifications, other teams may adopt the DCTAP and the SKOS vocabulary to describe compatible rock art resources and validate their RDF against the ShEx schema before publication or integration. Third, as interoperability building blocks, data managers and developers may reuse the RARAA namespace schema (RDFS) and align local terms and identifiers to facilitate interlinking and enrichment with other datasets. The README provides a “start here” checklist, indicates recommended download formats, and includes a small set of example SPARQL queries and validation instructions to support immediate adoption.”

5 – Improve Figure 5. Figure 5, originally an automated tool output, has been replaced with a schematic representation. This version incorporates illustrative renderings of specific motifs to enhance interpretative clarity and provide better visual context for the underlying RDF structure.

6 – Improve Figure 2 (move the labels). Figure 2, originally an automated tool output, has been replaced with a schematic representation.

Competing Interests: No competing interests were disclosed.

Reviewer Report 20 March 2026

https://doi.org/10.21956/openreseurope.23694.r70881

© **2026 Riris P.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Philip Riris
Bournemouth University, Poole, UK

The authors report on recent work that attempts to transition a database of rock art sites from the Vale do Coa (Portugal), resolved to the motif level, into a Linked Open Data format. As I understand it, the manuscript predominantly focuses on the process and methodological justifications underpinning the end-product, i.e. a complete LOD dataset. My comments chiefly concern: 1) the accessibility of the text itself, and 2) the underlying RARAA dataset. Not being a domain expert in LOD or data interoperability, I am writing from the point of view of a prospective end-user and invite the authors to consider the accessibility of the text to the non-expert. I appreciate that the manuscript concerns a specialist subject, however, as the journal has a general and broad readership, I am emphasising this element more than I perhaps would in a specialist disciplinary journal such as *Journal of Computer Applications in Archaeology* or *Advances in Archaeological Practice*.

In the first instance, the introduction would especially benefit from re-writing in layman's terms. There are 14 acronyms introduced on a single page. Although some are familiar or obvious, e.g. *UNESCO* or *FAIR,* the meaning and overall flow of the text is disrupted by this much information. Additionally, there is liberal use of direct quotes throughout the text (usually definitions, e.g. 'open data', fifth paragraph). **T**he authors may consider making a new table or text box for the manuscript which provides a glossary of terms, and place it in the introduction**.** This would enable the authors to foreground the meaning and significance of their text, rather than losing the reader in a sea of acronyms. The existing text would necessarily be shortened, allowing space for more plain-language explanation of *why* this work is being performed. Even as a proponent and practitioner of Open Science, who understands the fundamental impetus for this sort of work (and I am deeply appreciative of its development in a rock art context), the justification gets lost. In summary: the introduction especially needs strengthening by simplifying it. The authors should consider this general advice for the text as a whole**,** but it's particularly noticeable in the introduction.

Regarding the second point, I find there is a disparity between the reported 1200 engraved outcrops/boulders in the Coa Valley Park, and what the 2ArchIS database on http://www.raraa.pt/ actually contains. I apologise to the authors if I have misunderstood, or if I am attempting to access the wrong dataset. Nevertheless, the fact that I cannot tell where to look or what to look for is, in itself, a justification for a clearer explanation and signposting of the underpinning archaeological evidence (i.e. the actual transcendental value of this research). It only really becomes clear in the final sections that primary rock art imagery (sketches, photos, tracings, rubbings) is *not* included as part of the LOD (to my knowledge). Rather, it is a series of data tables recording the various interpretations *of* the underlying imagery. One of the unspoken principles of

Open Data is "trust, but verify" and without backlinks to the original imagery this is not possible. The imagery apparently *is* available (here) but the Data Availability statement does not contain reference to it. The LOD product is impressive, useful, and worth publishing but end-users will always ultimately want to return to the original data (for example for re-interpretation or comparative work), which currently does not seem to be possible.

Aside from this, the "atlas" linked above only contains sketches (not original images) of Iron Age rock art and I therefore find it misleading to make an implicit link between the entire corpus of Coa Valley rock art (which is extensive, varied, and multi-period) and the rather constrained sample of sites within the U of Minho repository. It must be made clearer in the introduction that the focus is *specifically on* Iron Age rock art here - it is currently mentioned once in passing in the introduction and in the conclusion. How it compares in terms of number of sites, outcrops, motifs, themes, etc. with the more well-known Palaeolithic engravings should be considered, so that readers have a better handle on the content of the LOD.

Other comments

On page 6 (**section Linked data development...**), you must identify which projects you are talking about. "Some of these project" [...] "others" [...] "Many projects" is not specific enough, because multiple examples are provided in the subsequent text, and it is not clear to me which examples relate to, e.g. extensive use of ontologies versus, e.g. value vocabularies. Rewrite this for more clarity.

Page 7, (section **The 2ArchIS relational database**) needs, at a minimum, citations of supporting literature on the aspects of society and environment that rock art can inform us about.

Page 17, (**The triplestore**), it should be "6085" not "6.085".

Page 19, second to last paragraph: "humans" seems to be a false friend from "homens" - replace with "users", "researchers", or "people".

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
No

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Rock art, computational archaeology, simulation & modelling, historical ecology.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 24 Apr 2026

Ana BAPTISTA

We appreciate your valuable input and the thorough review of our work. For clarity, we have addressed your requests by organizing them into the following nine points. 1 – Glossary A glossary was added before the Introduction with the following terms:

- Classes and properties vocabulary
- Linked Data
- Linked Open Data
- Metadata Application Profile
- Resource Description Framework (RDF)
- RDF Graph
- RDF Schema (RDFS)
- Shape Expressions (ShEx)
- Simple Knowledge Organization System (SKOS)
- SPARQL
- Triplestore
- Values vocabulary

2 – Alleviate the acronyms. While there are some acronyms that can be not used (e.g., CH for cultural heritage), others are better recognized as acronyms than by the expanded version (e.g., SKOS, RDF, FAIR, or CARE). We, therefore, alleviated the following acronyms: CH – Cultural Heritage RDB – Relational Database UAUM – Archaeology Unit of the University of Minho LOV – Linked Open Vocabularies EDM – Europeana Data Model DWBP – Data on the Web Best Practices And other acronyms that were expanded once and never again used. 3 – Explain better why this work is being performed. With the glossary and the need to explain better some parts of the work, we reformulated the whole Introduction. The following paragraphs are part of this reformulation: “All the data gathered by the RARAA project is stored in a relational database called 2ArchIS. Although the relational database is beneficial and widely used by the RARAA project and the Archaeology Unit of the University of Minho team, the researchers would like to be able to relate their data to those of other research teams and, conversely, to be able to share their data with other researchers or interested citizens for use and reuse. In other words, the RARAA team wanted its data to be

open and semantically interoperable, that is, to be Linked Open Data (LOD). This need aligns with broader developments in the digital transformation of cultural heritage. In recent years, global initiatives promoting open, structured, and reusable data have significantly influenced the digital transformation of cultural heritage. The digital transformation of cultural collections has strengthened preservation practices and opened new opportunities for research, education, and public engagement. [3, 4] This paradigm shift has been accompanied by international initiatives that promote open, structured data, ensuring that heritage knowledge remains accessible and reusable sustainably. Movements such as Open GLAM and Collections as Data advocate for the responsible sharing and ethical reuse of cultural data. This approach is exemplified by the principles articulated in documents such as the Vancouver Statement on Collections as Data. [5] These initiatives align closely with the FAIR principles (Findable, Accessible, Interoperable, and Reusable), [6] which enhance machine-actionable data discovery and usage while supporting human reuse, and with the CARE principles (Collective Benefit, Authority to Control, Responsibility, Ethics), which provide a foundational approach to Indigenous Data Governance. [7] In this context, the main goal of this work was therefore to transform data in a relational database into LOD and make it available for use and reuse, adhering to principles and practices that promote structured, reusable, and interconnected data, including the FAIR principles and the World Wide Web Consortium (W3C) Data on the Web Best Practices [15,16]. The expected outcome is to enable integration of RARAA's data with datasets from other research teams and to support broader reuse by the academic community and the general public. By enabling such integration and comparison, we aim to generate new insights into Iron Age rock art that would not be possible when data remains confined to isolated systems, often called data silos. To achieve this, we designed an application profile and related artifacts and then used them to expose 2ArchIS data as LOD." In another part of the text, we clarified again why it is important, in our view, to have the data open as Linked Data: "In addition, it presents in detail the process that led to making this data openly available from an existing relational system, with the goal of supporting reuse, extension, and interlinking by the wider community." 4 – Disparity between the 1200 engraved rocks and the number of motifs in the 2ArchIS database and the triplestore. Another part of the reformulation: "Within this broader archaeological context, the Open Access Rock Art Repository project (RARAA - Repositório de Arte Rupestre de Acesso Aberto), funded by the Portuguese Foundation for Science and Technology (FCT), explored the Côa Valley in search of Iron Age rock art, aiming to increase knowledge of the art and the communities that created it. Of the approximately 1200 engraved rocks currently known in the park, a significant proportion date from the Iron Age, the second most represented chronological period in the Côa Valley. Within this subset of Iron Age rock art, the project focused on documenting the engravings from three specific areas: Vermelhosa, Vale de José Esteves, and Meijapão . Consequently, the data presented in this work correspond to a small subset of the extensive rock art record of the park." 5 – Section Linked Data Development – Clarify: "Some of the projects" (...) "others" (...) "Many projects". – We replaced that introductory paragraph with the following: "The transition from legacy cultural heritage data to the Semantic Web has been addressed through various methodological lenses. While some approaches prioritize the preservation of metadata richness through multi-step ingestion processes, others focus on the scalability of database migrations or the agility of ontology design. The following projects illustrate the diversity of frameworks developed to bridge the gap between heterogeneous data sources and ontological structures." 6 - Section "The 2ArchIS Relational

Database" – needs, at a minimum, citations of supporting literature on the aspects of society and environment that rock art can inform us about. We added the following text: "In this context, 2ArchIS[69] was created by the Archaeology Unit of the University of Minho as an operational information system to support archaeological work from field recording to post-fieldwork analysis and reporting. Its primary intent is to ensure consistent data capture in the field and in the office, to centralize and normalize records produced by different team members, and to maintain long-term institutional memory of projects. Beyond data entry, the system supports day-to-day research workflows by enabling browsing and filtering of records, production of inventories and summary tables for reports, and systematic association of archaeological entities with documentation (bibliography, photographs, drawings, cartography, and 3D models). In the specific case of rock art, such structured recording is particularly relevant given that these datasets constitute a primary source for interpreting past social practices, symbolic systems, and the construction of cultural landscapes, as well as human–environment interactions reflected in the representation and spatial distribution of motifs [71,72]. For the RARAA project, 2ArchIS [8] has served as the authoritative source for recording and relating archaeological sites, rocky outcrops, motifs, represented parts, scenes, and associated documentation, supporting both ongoing research and project outputs. For dissemination and preservation purposes, exported records have also been deposited in an open repository with persistent identifiers, while 2ArchIS remained the primary environment for data creation and curation." With references (69,71,72): (69) Botica N, Magalhães F, Machado D, Fontes L. Del sistema de información 2ArchIS al DataRepositóriUM: el estudio de caso de hallazgos arqueológicos. Cuad Arqueol Univ Navar. 2021 Apr 14;29:381–96. doi:10.15581/012.29.005 (71) Bettencourt AMS, Santos-Estevez M, Sampaio HA, Cardoso D. Prehistoric Rock Art and Landscape: the outcrops and its forms were certainly not simple carving surfaces. In: Ribeiro C, Martins RN, editors. I International Seminar on Art and Landscape [Internet]. Guimarães, Portugal: Laboratório da Paisagem; 2017. p. 137–53. Available from: http://www.labpaisagem.pt (72) Cruz Berrocal M, Vicent García J. Rock art as an archaeological and social indicator: The neolithisation of the Iberian Peninsula. J Anthropol Archaeol. 2007 Dec;26(4):676–97. doi:10.1016/j.jaa.2007.02.003 7 – The triplestore – 6.085 -> 6085 – Done. 8 - Page 19, second to last paragraph: "humans" seems to be a false friend from "homens" - replace with "users", "researchers", or "people". We used the term "interface for humans" as opposition of the SPARQL endpoint, which is an interface for machines (and specialized humans). The term "human" in this context is often used, for instance, in Human-Computer Interaction (HCI). 9 – "Are all the source data underlying the results available to ensure full reproducibility? No" We made available the source data in the data repository as an XML file, exported from the **2ArchIS** database at the time of the study. We have also provided the formal citation and the corresponding reference for this file. The reference is: Botica N. 2ArchIS XML Data Export (RARAA Project) [Internet]. Repositório de Dados da Universidade do Minho; 2026 [cited 2026 Apr 15]. Available from: https://doi.org/10.34622/datarepositorium/KZEINV doi:10.34622/DATAREPOSITORIUM/KZEINV

Competing Interests: No competing interests were disclosed.