Informative description of a project implementing a CIDOC-CRM based native graph database for representing megalithic information

Isto Huvila based on peer reviews by 2 anonymous reviewers

Submitted: 01 June 2023, Recommended: 05 January 2024

Cite this recommendation as:
Huvila, I. (2023) Informative description of a project implementing a CIDOC-CRM based native graph database for representing megalithic information. *Peer Community in Archaeology*, 100338. 10.24072/pci.archaeo.100338

Published: 05 January 2024

The paper "Transforming the CIDOC-CRM model into a megalithic monument property graph" describes an interesting endeavour of developing and implementing a CIDOC-CRM based knowledge graph using a native graph database (Neo4J) to represent megalithic information (Câmara et al. 2023). While there are earlier examples of using native graph databases and CIDOC-CRM in diverse heritage contexts, the present paper is useful addition to the literature as a detailed description of an implementation in the context of megalithic heritage. The paper provides a demonstration of a working implementation, and guidance for future projects. The described project is also documented to an extent that the paper will open up interesting opportunities to compare the approach to previous and forthcoming implementations. The same applies to the knowledge graph and use of CIDOC-CRM in the project.

Readers interested in comparing available technologies and those who are developing their own knowledge graphs might have benefited of a more detailed description of the work in relation to the current state-of-the-art and what the use of a native graph database in the built-heritage contexts implies in practice for heritage documentation beyond that it is possible and it has potentially meaningful performance-related advantages. While also the reasons to rely on using plain CIDOC-CRM instead of extensions could have been discussed in more detail, the approach demonstrates how the plain CIDOC-CRM provides a good starting point to satisfy many heritage documentation needs.

As a whole, the shortcomings relating to positioning the work to the state-of-the-art and reflecting and discussing design choices do not reduce the value of the paper as a valuable case description for those interested in the use of native graph databases and CIDOC-CRM in heritage documentation in general and the documentation of megalithic heritage in particular.

References:

Câmara, A., de Almeida, A. and Oliveira, J. (2023). Transforming the CIDOC-CRM model into a megalithic monument property graph, Zenodo, 7981230, ver. 4 peer-reviewed and recommended by Peer Community in Archaeology. https://doi.org/10.5281/zenodo.7981230

Reviews

Evaluation round #1

DOI or URL of the preprint: https://doi.org/10.5281/zenodo.7981230
Version of the preprint: 1

Authors' reply, 08 December 2023

Download author's reply
Download tracked changes file

Decision by Isto Huvila, posted 11 October 2023, validated 11 October 2023

Revision

The reviewers find the text interesting but also point out major questions and issues that would need to be thoroughly addressed before an eventual recommendation.

Reviewed by anonymous reviewer 1, 05 July 2023

In this article, a motivation for organizing archeological data in knowledge graphs is presented. The motivations and justification are well presented. But it is not a novelty per se.
It is based on an interesting case study, the presentation/explanation of the model could be improved for a publication, some elements are missing, (such as Fig.2), or are not clear (Fig. 1). Session titles, "Requirements" and "Methodology" are not quite appropriate for their session´s contents.
Fig 2. is referred to in the text, but not found in the paper, if that was referring to Table 2, it does not seem to show or exemplifies the terminology but presents references to vocabulary source.
A brief overview of the literature is given pointing out the differences in their proposal of KGs. However, they pose a too strong claim that no work was found for the representation of buildings and architectural remains in archaeology, especially aiming at the extraction, reusability, and interpretation of the information by machines, while Santos 2022 and Gergatsoulis et al. 2022 are intended for that.
It could be helpful if the "Requirements" section includes also examples of representing and querying architectural monument data and the trade-offs involved in choosing Neo4j (or any NGDBs) for the study.
The authors claim: In this paper, we only use CIDOC-CRM definitions and it is not discussed its extensions for Archaeology. Here a justification of this decision, and/or a comparison with such extensions would be required.
The CSV is available in Github, which is nice, but it requires seeing the CSV to have a better idea of the model, a

description would be important.

It would be interesting to quantify the elements represented in the graphs, corresponding to the 94 dolmens analysed, including missing data, and also present some query examples, or discuss future applications in more concrete scenarios.

"Megalistimo Alentejano" and other non-English words should be in italic, for example, and the definition could be more clear for non-Portuguese speakers.

"Archaeologica Letter" has a typo and it's not the correct translation for Carta Arqueológica.

Typos

… its components it's based …

… interest in standartised access …

… it's based at the E22… -> based on?

… that allow describe the monumento …

Reviewed by anonymous reviewer 2, 03 October 2023

Comments by sectionsIntroduction

- The opening sentence appears to be quite a broad statement. In fact, similar information is available in databases from numerous archaeological services throughout Europe. However, the accessibility of these databases presents a separate issue.

Requeriments

- This section provides an update on existing methods for creating graphs. It is too technical for an archaeology article, what it is not a real issue. However, I don't understand how it fits into the structure of the paper, as it is closely related to the methodology

Related Literature

- The cited literature may be sufficient, although many works are missing. A thorough review is certainly not expected, but the selection seems to be very carefully chosen to demonstrate the novelty of the work.

- I find the last sentence of the section especially 'striking' for its imprecision, where it is claimed that none of the works have addressed a representation of architectural elements in Archaeology. As far as I know, it seems that some of these works have indeed addressed it. See for example Table 2, which contradicts your statement. Furthermore, the statement appears to be rather generalistic once again, likely not taking into account a more thorough literature search.

Methodology

- Bueno Ramírez 2013 is mentioned in the text, but is not cited in the bibliography. I find it striking once again that a reference to a Spanish author is used, when in fact the quantifications of megaliths come from Portuguese authors, some of them at the University of Évora, for example, Leonor Rocha.

- "Megalitismo Alentejano" is a Portuguese expression It is not necessary to use Portuguese instead of English in this case, but if that is the choice it should be in italics.

- "Orthostats" is slabs in English.

- The English in this section appears to be ackward in many senses. The description of what a dolmen is supposed to be is unclear. Perhaps someone who is not a specialist in the field would not understand from the description what we are referring to.

- "Archaeologica Letter" is a direct translation from Portuguese and doesn't make any sense in English.

- At the end of the 'Data Model' section, a dolmen is explained again, even though it has already been described earlier. The description doesn't seem to be very 'critical'; it is once more overly general and doesn't take into account the peculiarities of Portugal.

Overview of the Approach

- I understand that the role of this section is purely methodological; it describes the implementation of the model and does not offer tangible results.

Conclusion and Future Work

- I believe the conclusions merely repeat ideas already articulated in previous sections and do not offer anything beyond generalities that are not really useful for assessing the impact of the proposal

General commentsIllustrations

- A map with the study area seems mandatory

Honestly, I understand the good intentions behind the work, but I think it suffers from many issues. It is not a proper case study; it's merely a technical proof of concept where too many elements are being attempted to be tied together, which is ultimately not reflected in either the results or the conclusions. I don't see how this structure can aid semi-automatic remote sensing (of which, by the way, many references are missing), for example. I also don't see how this can translate into a tool for interpreting the past, which is what we archaeologists are looking for. I only see its technical utility with proper development, but not its repercussions.

The structure of the paper appears to be somewhat complex. The text seems to reflect two distinct perspectives: one that is technical and proficient in the use of ontological modeling tools, and another that aims to explore these tools' applicability in the field of Archaeology. Unfortunately, the latter aspect seems quite underdeveloped in comparison to the former. This discrepancy creates a certain level of conceptual ambiguity and imprecision in terminological translation. There are also moments where the paper leans towards generalizations and may benefit from more rigorous bibliographic support. Additionally, there is no discussion comparing this work to others, such as the study by Santos et al. 2022, which focuses on the same area. Does this work represent an improvement or does it complement the previous research?Additionally, I observed that there could be more reflection on the potential utility of the tool in question. In conclusion, while the technical aspects of the paper are well-developed, the overall structure might benefit from clarification. This leads me to think that the article may be more appropriately aimed at an audience specialized in semantic models.

The data table raises several questions, both in terms of its design and the information it contains. Some of its contents appear to be 'constants,' such as the units of measurement. The treatment of chronology also seems to be less than optimal, especially considering the specific characteristics associated with this type of burial sites. The data are limited and often puzzling, such as the unknown status of the funerary chamber for all the sites.

I would suggest reviewing the English in the sections dedicated to Archaeology for clarity and accuracy. It may also be beneficial to delve deeper into the concepts, avoiding broad generalizations. A thorough review of the bibliography and its relevance could also add value to the work. Additionally, I recommend a careful reconsideration of the paper's overall structure for improved coherence.