

Response to Reviewers

We would like to thank the reviewers for their comments and suggestions, which have helped us to improve our manuscript. Below, we provide our responses to each reviewer's comments.

Response to Reviewer A

"Heritrace introduces an innovative approach to traceability and data management in cultural heritage digitization projects. Unlike other LOD platforms, it stands out for its emphasis on easy of use and provenance tracking, ensuring transparency in the evolution and modification of data. However, a structured comparison with existing applications, such as Europeana or national LOD initiatives, and semantic editors (such as Omega, Wikipedia, ec..) is missing. This would better highlight Heritrace's added value. The study is significant for the field, but it should better articulate its practical contribution for users, providing concrete examples of how the platform simplifies operations compared to other tools."

Thank you for recognizing the innovative aspects of HERITRACE, particularly regarding usability and provenance tracking! We'd like to point out that we've actually included quite an extensive comparison with relevant semantic editors in Section 2.2, "Comparative Analysis of Semantic Editors in GLAM." This section specifically compares HERITRACE with OmekaS, Semantic MediaWiki, Research Space, and CLEF, all semantic data editors designed for cultural heritage data management.

Regarding the comparison with Europeana or national LOD initiatives, there's a key distinction we should clarify: HERITRACE isn't a LOD repository or database, but rather a tool for managing and editing RDF data. Comparing HERITRACE (a data editor) with Europeana (a data repository) would be a bit like comparing a word processor with a library catalog: they're designed for fundamentally different purposes. We do mention Europeana in Section 2.1 as an example of a prominent LOD initiative in the cultural heritage domain that highlights the need for tools like HERITRACE. In addition, we've have expanded Section 2.1 to include ARCo (the Italian Cultural Heritage Knowledge Graph).

As for Wikipedia, it's not actually a semantic editor but a wiki platform. The semantic variant mentioned in our comparison is Semantic MediaWiki, which we've included in our analysis.

On the practical front, we've taken your suggestion and added a concrete example to Section 4 that shows how a common task – correcting an author's name in a bibliographic record and tracking its provenance – would be handled in different systems compared to HERITRACE. This example highlights how HERITRACE eliminates the fragmentation and technical complexity found in other systems, making data curation tasks in RDF much more accessible.

"The problem is well defined, but the specific objectives of Heritrace should be stated more clearly. Clearly articulate the practical goals of the platform."

We admit the objectives were a bit implicit in the Introduction. We've now added a concise paragraph that explicitly lays out HERITRACE's five primary objectives: (1) providing a user-friendly interface for domain experts who don't have technical knowledge; (2) implementing comprehensive provenance management to document metadata changes; (3) delivering robust change-tracking for reconstruction of previous data states; (4) offering flexible customization through standard languages rather than proprietary solutions; and (5) enabling seamless integration with existing RDF data collections.

"Describe practical experiences of users with Heritrace. Integrate more references to usability studies and user experiences."

Just to clarify, all the examples, screenshots, and workflows in the paper come from the real-world deployment of HERITRACE in the ParaText project, as mentioned in Section 3. This project has essentially served as what developers call "guerrilla testing" – a practical, real-world deployment that's helped us identify bugs, usability issues, and workflow optimizations.

Regarding formal usability studies, we completely agree they're important! As outlined in Section 4, comprehensive usability testing is planned as a key next step: "While the system has been designed with user accessibility in mind, real-world testing with GLAM professionals is essential to identify potential barriers and optimize the interface."

Since we haven't conducted these formal studies yet, we can't include their results in the current paper. But we're definitely committed to conducting these studies and publishing their results in future work.

Response to Reviewer B

"Reduce somehow Section 2.1. It contains only examples from National libraries and European. It is well known that they are rich and can afford projects in LOD"

We've thought carefully about this suggestion, but Section 2.1 serves several important purposes that justify keeping it:

1. It establishes that Semantic Web technologies are widely used in GLAM institutions, providing context for why tools like HERITRACE are needed.
2. While these examples may be familiar to experts, they might not be as well-known to all readers, especially those new to the intersection of Semantic Web and cultural heritage.
3. The section addresses a common question: "Why not use simpler models for cultural heritage data?" By showing that major institutions choose semantic technologies despite having resources for any

model, we demonstrate that this choice isn't just about budget but about the inherent benefits for cultural heritage data.

4. The examples show how semantic technologies work in various contexts, from medieval manuscripts to contemporary digital collections.

Interestingly, Reviewer A suggested the opposite: expanding our comparison with Europeana and national LOD initiatives! This difference in feedback suggests this content is valuable to some readers.

That said, we've tightened the language and removed some repetition while keeping the core examples that show why HERITRACE's semantic approach makes sense.

"Reduce somehow Section 2.2, possibly referring to the previous paper, which has almost the same content of this section."

There seems to be a misunderstanding here. This paper is actually meant to be an extended version of our AIUCD 2024 conference contribution. The citation to "Di Silvestro and Spampinato 2024" refers to the proceedings where our shorter contribution appears, not a separate previous paper.

To clear this up, we've corrected the citation from:

"Di Silvestro, Antonio, and Daria Spampinato. 2024. Me.Te. Digitali. Mediterraneo in Rete Tra Testi e Contesti, Proceedings Del XIII Convegno Annuale AIUCD2024. AIUCD.
<https://doi.org/10.6092/UNIBO/AMSACTA/7927>."

to:

"Massari, Arcangelo, and Silvio Peroni. 2024. "HERITRACE: A User-Friendly Semantic Data Editor with Change Tracking and Provenance Management for Cultural Heritage Institutions." In Quaderni Di Umanistica Digitale. Catania: AMSActa. <https://doi.org/10.48550/arXiv.2402.00477>."

Since this is being submitted for a special issue specifically for extended versions of AIUCD2024 contributions, the expanded content in Section 2.2 is exactly what's expected. This comparative analysis is actually one of the core contributions of this extended version, providing important context for understanding HERITRACE's value compared to existing solutions.

"In Section 3 clarify the relationship between Time Machine and Time Vault. Also, clarify how to import existing graphs. Is it enough to import them in the triple store? Is it necessary to update the SHACL descriptions?"

We have addressed clarifications to both points.

Regarding the Time Machine and Time Vault, we have added:

"While existing entities are accessible through the main catalog interface, when resources are deleted, they disappear from this catalog but are never permanently removed from the system. This is where Time Vault complements the Time Machine functionality. Time Vault serves as a dedicated "recycle bin" or catalog specifically for deleted entities, making them easily discoverable through a specialized interface. Users can review essential information about deleted resources, such as the deletion timestamp, the responsible agent, and the associated modifications. Additionally, the Time Vault enables quick access to the most recent version of a deleted resource, allowing users to restore it if necessary."

Regarding the process for importing existing graphs and the necessity of SHACL descriptions, we have included an additional explanation in Section 3.1:

"HERITRACE works with existing RDF data without requiring SHACL or YAML configurations, though these optional customizations enhance the system's functionality. Connecting to a triple store is sufficient to import RDF graphs—the system automatically discovers entities and applies provenance and change-tracking to subsequent modifications. SHACL descriptions, while not mandatory, improve the user experience by enabling validation and customized editing forms. This flexibility allows HERITRACE to adapt to various deployment scenarios, from quick implementations with existing data to fully customized installations."

"At the end of Section 3.1 provide some more examples of YAML"

We've added a comprehensive YAML configuration example (Listing 1) that shows how administrators can configure how a Journal Article entity appears in HERITRACE. The example demonstrates different property types, including simple display properties and properties that use SPARQL queries to transform values. We've also included a detailed explanation of each component and how it affects the interface.

As mentioned in the paper, the complete SHACL and YAML files from the ParaText project are available on Zenodo (<https://doi.org/10.5281/ZENODO.14741864>) for anyone who wants to see more examples.

"Section 4 repeats many of the concepts expressed in Section 2.2. They might be reduced in favor of describing possible future extensions of HERITRACE."

While Section 4 necessarily compares HERITRACE with other systems, we agree that there's no reason to repeat the same detailed analysis already provided in Section 2.2. We have therefore streamlined Section 4 by directly referencing the comparative analysis in Section 2.2 rather than repeating it: "While Section 2.2 provided a detailed comparative analysis of existing solutions, this section focuses on HERITRACE's approach to solving these challenges..."

Instead of redundantly listing each system's limitations again, we've incorporated mentions of these alternative systems within a concrete example that addresses another reviewer's request for practical comparisons. This example shows how a common task—correcting an author's name in metadata—would be handled in OmekaS, Semantic MediaWiki, Research Space, CLEF, and HERITRACE. Through this approach, we maintain the paper's structural balance where Section 2.2 analyzes other systems in detail,

while Section 4 demonstrates how HERITRACE specifically addresses their limitations using a real-world scenario.

As suggested, we've also expanded the future directions part of Section 4, particularly the discussion of usability evaluation. We have replaced the general statement about usability testing with a detailed description of our planned evaluation framework, which will employ task-based testing with 15-20 GLAM professionals, and collect both quantitative measurements (task completion rates, time-on-task analysis, error frequencies, and System Usability Scale scoring) and qualitative feedback through post-task interviews.

"There are a few typos here and there. Second paragraph of the Introduction: "Hannemann and Kett 2010)2.1Errore. L'origine riferimento non è stata trovata". In the last paragraph of the Introduction there are wrong references to Section numbers In Section 4, in the paragraph beginning with "This streamlined usability is particularly critical in..." there is a wrong reference to a Section number. The text in most of the figures is hardly readable"

We've addressed all the reference errors in the manuscript. For the readability of figures: we've kept the navigation bar only in Figure 1 to show the interface's complete structure, while removing it from all subsequent images. This allows the actual content to use the full page width in all other figures, significantly improving their readability and clarity.

Response to Reviewer C

"Please check the bibliographic references in the text: there seems to be a missing reference in the first paragraph of the introduction, and at several spots the references are inconsistent (S. Peroni et al > Peroni et al; M.Daquino et al > Daquino et al)"

Yes, there were inconsistencies. We've gone through the entire manuscript and standardized all citations following APA style guidelines. In particular, we've made sure that all citations use the same format (Peroni et al. instead of S. Peroni et al.).

In the bibliography section, we've also ensured consistency by using full names for all authors (not just initials or surnames).