

**Gregor Pobežin**

Univerza na Primorskem,  
Fakulteta za humanistične študije  
ZRC SAZU,  
Inštitut za kulturno zgodovino  
gregor.pobezin@fhs.upr.si

**Zrinka Mileusnić**

Univerza na Primorskem,  
Fakulteta za humanistične študije

© 2025 Authors(s)

## Big Data or Many Small Data? XML and CIDOC CRM in a Project on Medieval and Early Modern Epigraphic Heritage

Digital epigraphic projects are by no means a novelty. The Epigraphy.info platform, for instance, currently lists dozens of active initiatives. Since the emergence of the first digital epigraphic, palaeographic, and manuscript projects, scholars have grappled with a persistent challenge: how to translate more than a century and a half of established scholarly conventions – such as the Leiden standard in epigraphy – into a coherent digital format?

The research community has largely addressed this issue through the adoption and refinement of guidelines for textual encoding in XML, particularly the TEI standard (Text Encoding Initiative), which has become the foundation for the digital transcription of epigraphic texts. Within this framework, the EpiDoc subset of TEI provides a robust and well-tested schema for encoding inscriptions, and it remains the most widely accepted format for the digital transcription and interpretation of epigraphic sources.

However, a persistent limitation of the EpiDoc model is its insufficient capacity to represent event-based phenomena – those connected to the archaeological context of inscriptions and their intra- and extra-textual relationships. These relationships are essential not only for the interpretation of epigraphic material itself but also for its meaningful integration into broader historiographical or archaeological data environments. To address this challenge, several initiatives, such as the EAGLE Project, have proposed methodological frameworks for harmonising EpiDoc with the CIDOC Conceptual Reference Model (CIDOC CRM).

For medieval and early modern inscriptions, this harmonisation is not merely desirable but essential. Such inscriptions – simultaneously material and immaterial artefacts – often refer to entities that also appear in palaeographic, archival, and other historical sources. The integration of the TEI/EpiDoc and CIDOC CRM data models thus provides a coherent response to the central question of how to interlink diverse yet related datasets.

By analysing a sample epigraphic record encoded in both the EpiDoc and CIDOC CRM frameworks, this paper demonstrates how an inscription can be described, concluding with a discussion of why both perspectives are indispensable for a comprehensive understanding of epigraphic heritage.

### Acknowledgements

The research paper was elaborated within the research project no. J7-60128: AID HCH – Breakthrough in humanities and cultural heritage with artificial intelligence, funded by the Slovenian Research and Innovation Agency (ARIS) and the Research Programme Core Funding No. P6-0460: Heritage as an Object and Reflection of Social Processes.

3rd Annual Conference  
of the Slovenian Node  
of the European Research  
Infrastructure  
for Heritage Science

E-RIHS Slovenia  
Koper, 18–19 November 2025

