



The banner features a row of six icons: a globe, a book, a network of nodes, a money bag with a Euro symbol, a scale of justice, and a bicycle. Below the icons, the text reads: "AIUCD 2021", "DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale", and "10° congresso annuale PISA 19-22 gennaio". On the right side, a list of topics is displayed: "DIGITAL PUBLIC HUMANITIES", "OPEN CULTURE", "RETI SOCIALI", "TECH ECONOMY", "E-PARTICIPATION", and "TECNOLOGIE ASSISTIVE". The background includes binary code and a classical building.

AIUCD 2021

DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale

10° congresso annuale **PISA** 19-22 gennaio

DIGITAL PUBLIC HUMANITIES
OPEN CULTURE
RETI SOCIALI
TECH ECONOMY
E-PARTICIPATION
TECNOLOGIE ASSISTIVE

Versione PROVVISORIA del contributo presentato al Convegno Annuale

DISCLAIMER

Questa versione dell'abstract non è da considerarsi definitiva e viene pubblicata esclusivamente per facilitare la partecipazione del pubblico al convegno AIUCD 2021

Il Book of Abstract contenente le versioni definitive e dotato di ISBN sarà disponibile liberamente a partire dal 19 gennaio sul sito del convegno sotto licenza creative commons.

CRMtex. An ontological model for ancient textual entities

Francesca Murano¹, Achille Felicetti²

¹DILEF, Università degli Studi di Firenze, Italy – francesca.murano@unifi.it

²PIN, Università degli Studi di Firenze, Italy – achille.felicetti@pin.unifi.it

ABSTRACT

This poster presents the CRMtex, an ontological model based on CIDOC CRM developed since 2015 to support the study of ancient documents by identifying relevant textual entities and by modelling the scientific process related to the investigation of ancient texts and their features in order to foster integration with other cultural heritage research fields. CRMtex is intended to identify and define in a clear and unambiguous way the main entities involved in the study and edition of ancient handwritten texts and to describe them by means of appropriate ontological instruments in a multidisciplinary perspective.

The CRMtex extension provides tools for managing this kind of complexity by defining classes and properties for describing a handwritten text in all its aspects, from its creation (and/or destruction) in the past, down to its present conservation, investigation and study by scholars, including its transcription, translation, interpretation and publication.

The full compatibility of CRMtex with the CIDOC CRM ontology and its extensions ensures persistent interoperability of data encoded by means of its entities with other semantic information produced in cultural heritage and digital humanities.

PAROLE CHIAVE

Ancient texts, Ontologies, Linguistics, CIDOC CRM, Epigraphy.

1. INTRODUCTION AND BACKGROUND

Although from the semiotic point of view there is a single mechanism of production of written texts (i.e. supports, techniques, etc. are not relevant), traditionally, their study falls within different disciplines, generally grown around the specific characteristics of each class of documents (e.g., papyrology for the study of papyri and epigraphy for inscriptions). Nevertheless, an interdisciplinary approach is essential, and the identification of common elements is paramount in order to confer uniformity and interoperability to all these disciplines, as well as to exploit complementary skills from different approaches.

In comparison to modern digital or printed texts, ancient ones are characterised by their uniqueness, being they the result of manual work rather than a mechanised process, as occurs with modern printing. Therefore, a special relationship exists between the texts and its supports. This and other characteristics make the study and digitisation of this type of documentation particularly arduous: the close relationship between the text and its support requires careful analysis since they are inextricably linked to form a unique object of study.

The gigantic integration effort established by Papyri.info [1] and Trismegistos [2] and the various attempts made by projects such as EAGLE [3] to develop a semantic model in the field of epigraphy, testify to a constant and growing interest in the use of advanced and efficient conceptual tools for the generation of standardised, integrated and interoperable information in these disciplines. In the same perspective, many European and international initiatives are also focussing their attention on information concerning ancient texts and on the interoperability challenges in which they are involved [4].

2. OBJECTIVES

The first aim of CRMtex is to identify and define in a clear and unambiguous way the main entities involved in the study and edition of ancient handwritten texts and then to describe them by means of appropriate ontological instruments in a multidisciplinary perspective [5][6].

The concept of “written text”, provided by CRMtex, is intended to identify a common entity consisting in a particular feature (i.e., set of glyphs) created (i.e., written) on various kinds of support, having semiotic significance and the declared purpose of conveying a specific message towards a given recipient or group of recipients. The modelling of the

scientific autoptic examination of the document, consisting of an accurate analysis of the surface and the signs and prescribing the use of specific tools and procedures, is another key aspect taken into account by the CRMtex extension. In addition to dealing with text as an object, our model also focuses on the aspects of the research and provides classes and relationships to describe the typical operations that scholars from different disciplines put in place to gain knowledge about texts. It is evident, in this perspective, that the study of ancient texts typically starts from the analysis of the physical characteristics of the text itself before moving to the investigation of their archaeological, palaeographic, linguistic and historical features.

The model we propose allows to describe the textual entities in their entirety, capturing the differences between the various writing systems. We intend to enrich this model through a more detailed semantic coding of the relationship between grapheme and codified linguistic unit, taking into account not only the phonographic scripts, but the glottographic scripts as a whole. A future development of the model will allow to extend the description also to writing systems based on logographic and ideographic models [7].

3. METHODOLOGIES AND EXPECTED OUTCOMES

The need to create a new ontology for ancient texts started from the assumption that, unlike printed texts, non-mechanised written texts (including inscriptions, papyri and manuscripts) have specific features that must be taken into account for their study.

Since writing is an intellectual process aimed at the encoding of a language, CRMtex has its foundation in the semiotic aspects of language and text. The core concept of our model is therefore the notion of “text” as the product of a semiotic process involving an encoding (“writing”) and a decoding (“reading”) process. Writing is in turn a particularly sophisticated human technology allowing the encoding of a linguistic message through a series of signs specifically selected for this purpose [8][9].

In writing, as in any semiotic system, every component (sign) possesses a dual nature, one physical and another conceptual. Writing, therefore, appears as a code requiring an encoding process by the creator or writer and a decoding one by the receiver or reader to be properly understood. It is necessary to distinguish between the physical manifestation of the text, understood as a set of physical features shown on a given support through the use of a specific technique (e.g. scribbled with ink, painted, engraved, etc.), from its abstract dimension, i.e. from the set of mental images represented by these same physical features.

We have based our model on the solid foundations of CIDOC CRM [10] because it constitutes one of the most widely used ontologies in the field of Cultural Heritage. In its core version, CIDOC CRM already provides most of the entities necessary to model common elements such as actors, objects, places, events and their mutual interrelations on a chronological basis.

Because of the native ability of CRMtex to describe relationships between text and artefacts by efficiently placing the text in the context of the life and history of ancient objects, our model has been chosen as a candidate for the encoding and integration of inscription and *graffiti* data in the semantic infrastructure that the ARIADNEplus project [11] is building. CRMtex has also been selected among the basic models for the ontology in the process of definition by the community of epigraphists in the framework of the epigraphy.info initiative [12] for the interoperability of epigraphic data.

BIBLIOGRAPHY

- [1] Papyri.info Initiative, <http://papyri.info>.
- [2] Trismegistos Portal, <http://www.trismegistos.org>.
- [3] EAGLE Project, <https://www.eagle-network.eu>.
- [4] PARTHENOS Project, <http://www.parthenos-project.eu>.
- [5] Felicetti, Achille, and Francesca Murano. “Scripta manent: a CIDOC CRM semiotic reading of ancient texts.” *International Journal on Digital Libraries* 18 (2017): 263–270.
- [6] Felicetti, Achille, Francesca Murano, Paola Ronzino and Franco Niccolucci. “CIDOC CRM and Epigraphy: a Hermeneutic Challenge.” In *Extending, Mapping and Focusing the CIDOC CRM. CRMEX 2015 Workshop, 19th International Conference on Theory and Practice of Digital Libraries (Poznan, September 14-18, 2015)*, edited by Paola Ronzino and Franco Niccolucci, 55–68. <http://ceur-ws.org/Vol-1656/paper5.pdf>.
- [7] Pulgram, Ernst. “The Typologies of Writing-Systems.” In *Writing Without Letters*, edited by William Haas, 1–28. Manchester: Manchester University Press, 1976.
- [8] Harris, Roy. *La Sémiologie de l’écriture*. Paris: CNRS, 1993.
- [9] Harris, Roy. *Signs, Language and Communication*. London: Routledge, 1996.
- [10] CIDOC CRM v.6.2.3, <http://www.cidoc-crm.org/Version/version-6.2.3>.
- [11] ARIADNEplus Infrastructure, <https://ariadne-infrastructure.eu>.
- [12] Epigraphy.info Initiative, <http://epigraphy.info>.