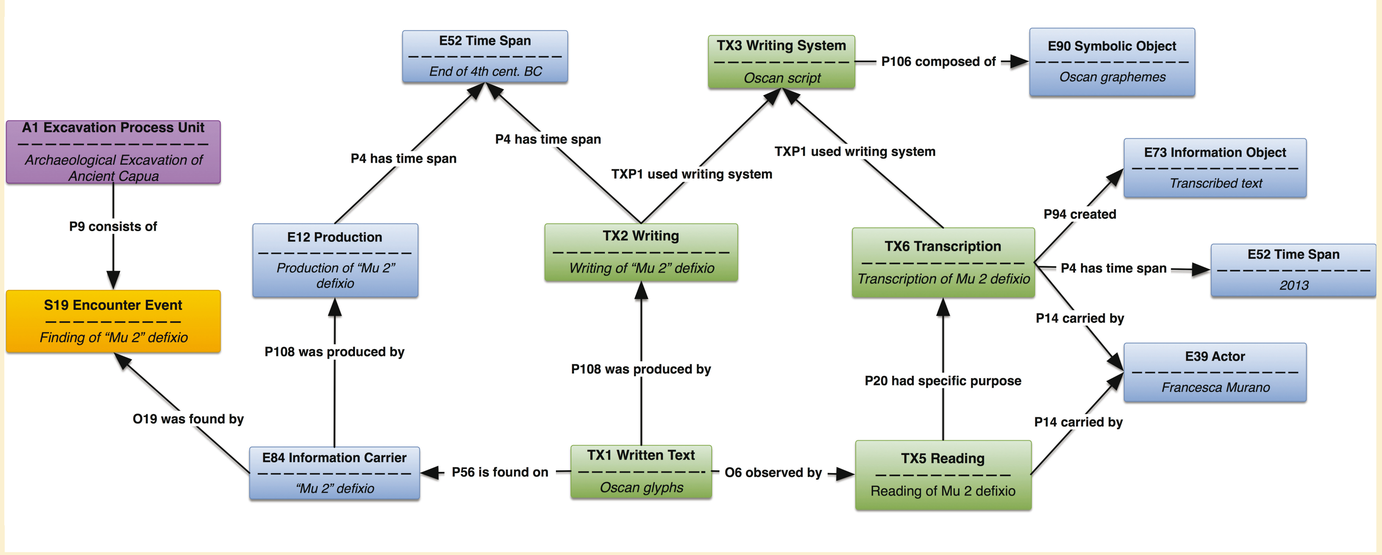
XV CONGRESSUS INTERNATIONALIS EPIGRAPHIAE GRAECAE ET LATINAE

VINDOBONAE MMXVII

WIEN 2017
XV. INTERNATIONALER KONGRESS FÜR GRIECHISCHE UND LATEINISCHE EPIGRAPHIK

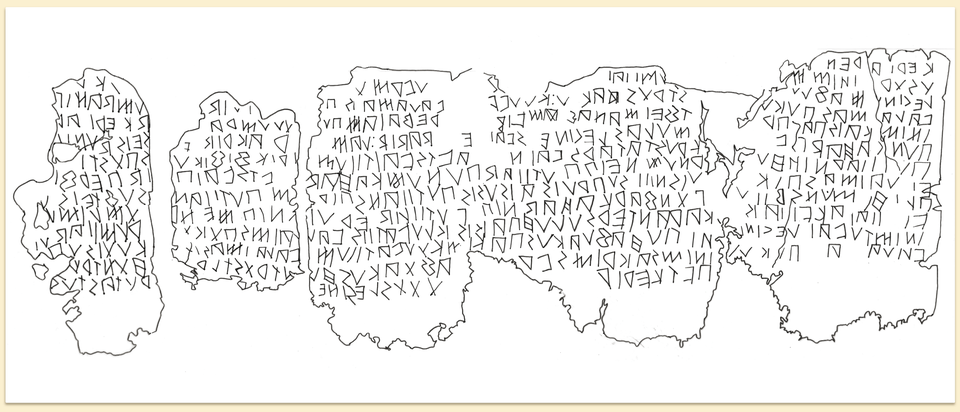
CRMtex: Models and Strategies for the Encoding of Epigraphic Documents

Achille Felicetti, PIN - Università degli Studi di Firenze
Francesca Murano, DILEF - Università degli Studi di Firenze



Exemple of CRMtex encoding

The Oscan curse tablet Murano 2013, nr. 2



Oscan curse tablet

From Capua, end of 4th cent. BC (Murano 2013, nr. 2)

Naples, Museo Archeologico Nazionale

Encoding Ancient Texts: the CRMtex semantic and semiotic approaches

«*The voice of the past is always the voice of an oracle; only if you are architects of the future and connoisseurs of the present will you truly understand it* », as Nietzsche once said. Knowledge of the past is entrusted to the direct and indirect sources the past itself has bequeathed to us and, among these, written documents occupy a prominent place. The profitable application of IT to the study of ancient written sources for expanding our knowledge of the past is the inspiring principle of CRMtex, a CIDOC CRM extension particularly responsive to the specific needs of the various disciplines involved in the study of ancient texts, including epigraphy, papyrology, palaeography and codicology.

CIDOC CRM is a standard ontology designed for the modelling of Cultural Heritage entities in semantic format, meant to be a shared conceptual framework for the construction and integration of complex datasets. CIDOC CRM uses a conceptual approach for the unambiguous description of the elements and objects involved in Humanities. Its ontological model is now enriched through various extensions: CRMtex is the one devoted to the semantic and semiotic encoding of textual entities.

Ancient texts are heterogeneous and require an interdisciplinary approach. The identification of common elements is paramount to confer interoperability to the disciplines involved in their study. In ancient written documents a special relationship between text and its support exists: ancient texts are unique, being the result of manual work rather than mechanised processes, as occurs with modern printing.

CRMtex was designed with the purpose of defining unambiguously the main entities involved in the study and the edition of ancient (hand)written texts, describing them by means of appropriate ontological instruments in a multidisciplinary perspective.

Semiotic Perspective of Text and CRMtex

Following Structuralist theories, both material manifestation (i.e. physical features shown on a support, created using specific techniques and tools) and abstract dimension (message to communicate) need to be considered when dealing with a text. CRMtex was developed to take into account these aspects and the events related with them: the writing and the reading processes, considered in their semiotic dimension of encoding and decoding signs. Following Peirce's semiotic theory, we can categorise these signs in two broad classes: glottographic and not-glottographic features. CRMtex is capable of dealing with both of them.

CRMtex classes and properties

TX1 Written Text. Subclass of the CIDOC CRM *E25 Man-Made Feature* intended to describe a particular feature (i.e., 'set of glyphs') created (i.e., written) on a support, having semiotic significance and the purpose of conveying a message. This class focuses on the analysis of the physical characteristics of the text itself. For instance, engravings on an inscription are fundamental elements for the study not only of the text but also of the (archaeological) object on which the text appears.

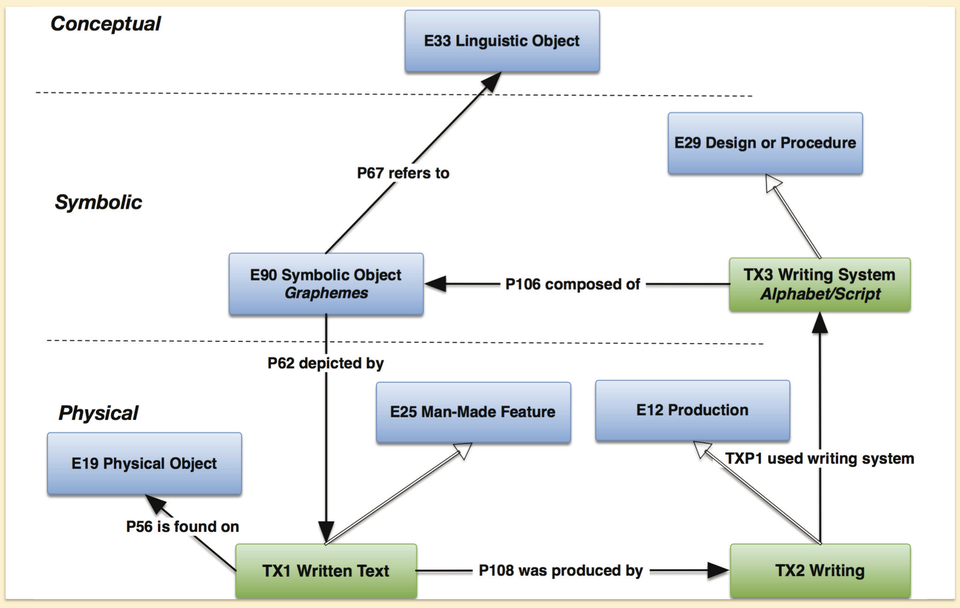
TX2 Writing. Subclass of CIDOC CRM *E12 Production* indicating the activity of creating textual entities using various techniques (painting, sculpture, etc.) and by means of specific tools on a physical carrier in a non-mechanical way. The class allows the distinction between the creation of the text and the production of the physical carrier, an important distinction in the study of inscriptions.

TX3 Writing System. Subclass of CIDOC CRM *E29 Design or Procedure*, refers to a conventional system (e.g., the Greek alphabet) consisting of a set of characters (graphemes, *E90*) used to codify one or more natural languages. It is used to produce a *TX1 Written Text* during a *TX2 Writing event*. The relation between *TX2* and *TX3* is expressed through the CRMtex *TXP1 used writing system* property.

TX4 Writing Field. Subclass of CIDOC CRM *E25 Man-Made Feature*, is a portion of the physical carrier arranged for accommodate and highlight a text, enhancing its readability.

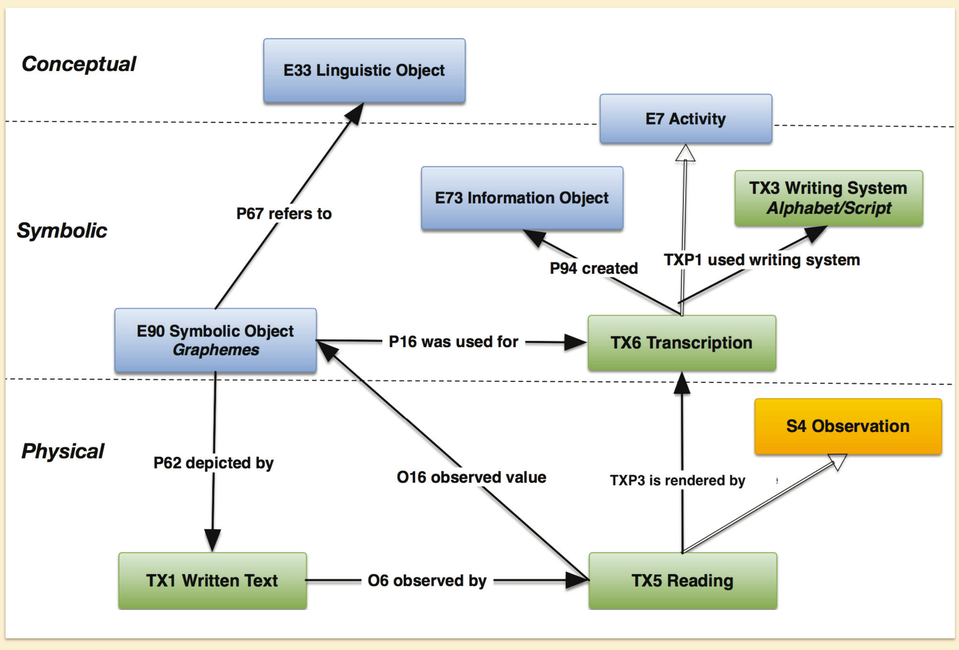
TX5 Reading. Subclass of CRMsci scientific extension *S4 Observation* class, refers to the scientific autoptic examination of the document to establish as faithfully as possible the exact value of each sign drawn on the support.

TX6 Transcription. Subclass of *E7 Activity*, referring to the editor's activity of re-writing the text. This activity can involve a writing system (*TX3*) different from that of the original text (e.g., Latin characters to render an Oscan text), resulting in a re-encoding of the text (transliteration).



CRMtex classes and properties

Representation of the writing event



CRMtex classes and properties

Representation of the reading event

Contact details:
achille.felicetti@pin.unifi.it
francesca.murano@unifi.it

universität wien
Historisch-Kulturwissenschaftliche Fakultät

HOLZHAUSEN DER VERLAG

ÖAW ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN