

# *Scripta Manent.* A CIDOC CRM Semiotic Reading of Ancient Texts

Achille Felicetti · Francesca Murano

Received: date / Accepted: date

**Abstract** This paper tries to identify the most important concepts involved in the study of ancient texts and proposes the use of CIDOC CRM to encode them and to model the scientific process of investigation related to the study of ancient texts in order to foster integration with other cultural heritage research fields. After identifying the key concepts, assessing the available technologies and analysing the entities provided by CIDOC CRM and by its extensions, we introduce more specific classes to be used as the basis for creating a new extension, CRMtex, which is more responsive to the specific needs of the various disciplines involved (including papyrology, palaeography, codicology and epigraphy).

**Keywords** CIDOC CRM extensions · Ancient Manuscripts · EpiDoc · CRMtex

## 1 Introduction

“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 Friedrich 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 surely occupy a prominent place. We cannot say whether technology could worthily incarnate the ‘architecture of the future’ that Nietzsche was thinking about but, nevertheless, it currently represents one of

the most important keys to interpret the ‘ancient oracle’ in order to give us a better understanding of our ancestors. The profitable application of IT to the study of ancient sources for expanding our knowledge of the past is the inspiring principle of the work presented in this paper.

The first written documents date back to the IV millennium BC. With the evolution of this technology, humans began to write texts on different supports using different techniques: inscriptions, papyri, manuscripts and other similar documents. Traditionally, the study of this heterogeneous documentation falls within different disciplines, generally grown around the specific physical characteristics of each class of documents (*e.g.*, papyrology for the study of papyri and epigraphy for epigraphs). 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.

The first and most obvious feature that catches the eye when examining these documents is the fact that all of them bear a text. The second thing that we can observe, specifically in ancient textual sources, is the special relationship between the text and its support. In comparison to modern texts, ancient ones are characterised by their uniqueness because they are the result of manual work rather than a mechanised processes, as occurs with modern printing.

This and other characteristics render particularly arduous the study and digitisation of this type of documentation: the close relationship between the text and its support requires careful analysis since they are inextricably linked to form a unique object of study. In fact, even in the case of texts written by the same person on identical media and with an identical technique, such as the codices produced by scribes in European monas-

---

Achille Felicetti  
PIN, VAST-LAB, Prato, Italy  
E-mail: achille.felicetti@pin.unifi.it

Francesca Murano  
Università degli Studi di Firenze, Italy  
E-mail: francesca.murano@unifi.it

teries during the Middle Ages, the resulting copies are never identical since, as with any human activity, writing also happens *hic et nunc*, which is why our handwriting is never completely identical with itself; by contrast, modern printed copies of books and documents are totally indistinguishable from one specimen to another, since the characters are etched from an identical matrix.

In the ancient world, however, there are certain inscriptions that were created through mechanised processes, such as the legends of coins, medals stamps and seals. Nevertheless, even for these classes of objects it is fundamental to investigate the close relation linking the text with the archaeological object that carries it. The uniqueness of the written text remains unchanged in this case also, since it is characterised by the peculiar history of the support. The first aim of this work is therefore 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. As a guide, we shall use the CIDOC CRM intellectual model [1] with its extensions (mainly CRMsci [2] and CRMarchaeo [3]) to provide the identified entities with shape and consistency, and to try to sketch a new extension (CRMtex) suitable for the study of ancient texts.

## 2 Semiotics, writings and text

### 2.1 The manifold natures of a text

The first and most important concept that needs to be clarified is what is meant by ‘text’, the definition of which remains somewhat debated in the literature and is at present largely dependent on the discipline within which it is analysed.

It is absolutely 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 concepts represented by these same physical features.

Writing is a particularly sophisticated human technology inasmuch as it allows the encoding of a text through a series of semiographic or glottographic signs that have been specifically selected for this purpose. 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 in order

to be properly understood, thus creating, to paraphrase F. de Saussure, a *circuit de la parole (écrite)* that will remain unbroken as long as the code remains shared by the sender and the receiver or is otherwise recovered by the latter. Thus the reading of any text implies decoding and, therefore, interpretation. Specifically in the study of ancient texts, this operation is extremely delicate: each reading is necessarily interpretative and dependent on the decoding, which in turn can be more or less complex, being operated by an editor, distant in time and space from the writer, who has acquired understanding of the code by means of indirect knowledge. In this light, each variant of the reading has to be considered as a different semiotic decoding of a text.

In any case, it must be stated that our investigation must obviously exclude all unintentional and consequently meaningless physical features that were not created with a will to communicate a message. In general terms, on the basis of a semiotic analysis of written signs [4], a physical feature occurring on a given support and intentionally created by humans to convey a message can appear in the following forms:

- a non-glottographic feature, which is a figurative decoration even when it has the value of an icon or symbol (*e.g.*, the sign of the cross in Christian inscriptions), but also a sign of pure semiographic, ‘language-independent’ systems of writing, used to codify the meaning only;
- a glottographic feature, not necessary codifying a linguistic expression, since a sign can be used with other purposes, *e.g.*, decoratively as with the use of  $A$  and  $\Omega$  signs in the Christian tradition symbolising the beginning and end.

Consequently, a text (in the meaning used in this paper) consists of a number of signs that are physically traced on a support and intended to encode a linguistic expression. Only in this case is it possible to speak of written communication.

### 2.2 Scripts and ‘hands’. A stylistic view

Since the texts we are dealing with in this article are, as we have stated, unique and unrepeatable, it is necessary to provide a better definition of the relationship between the linguistic expression that is intended to be conveyed and its concrete realisation through writing, putting aside for the moment the consideration of glottographic writing as a semiotic secondary code and the relationship between linguistic expression and ‘mental image’.

Referring to the Structuralist paradigm [5], we define glyphs as the concrete expression or the physical features the writer has traced (codifying the corresponding linguistic expression) and the reader understands by decoding. Since every person has his own unique manner of writing, there exists the possibility of recognising a specific ‘hand’ in the text. This is of particular interest for manuscripts and papyri, in which, for instance, recognition of different hands can be helpful in order to date the text or at least to understand its history better.

Glyphs that are physically traced on the support are the material manifestation of graphemes, *i.e.*, the abstract units with distinctive value in a given writing system. In an alphabetic system these units are essentially the letters of the alphabet in question.

### 3 Standards for representation of ancient texts

#### 3.1 Traditional standards

The edition of ancient texts boasts one of the earliest and most consistent systems of standardisation in the field of Humanities: the Leiden Conventions, which arose from the need to publish texts using a shared notation to describe the various observable phenomena they present. As of today, many of the well-established and growing database-based corpora, including those of the *Trismegistos* [6] and *Papyri.info* [7] initiatives for papyrology, the Epigraphic Database of Rome [8] and the inscriptions section of the Deutsches Archäologisches Institut [9] for epigraphy, also provide an extensive text field containing the text in Leiden format, besides the typical descriptive fields used for metadata, such as find location, date, dimension and so on. The Leiden Conventions specify how features of the text should be represented in print by using a set of standard symbols and text decorations to reproduce the state of the original document and to report the editors’ interpretations. With the advent of the digital era, the need arose for an electronic format that could allow digital publishing, storage and exchange of epigraphic information in a consistent and shared format. EpiDoc was a response to this need. It is a collaborative format designed to transcode Leiden-encoded printed editions in digital format [10]. EpiDoc at present provides features for recording the materiality and history of text-bearing objects, as well as features for scholarly editions of the text such as commentaries, illustrations, bibliographies, and publication data.

Despite the undoubted merits of the EpiDoc system, it still presents some problems, especially with respect to the inline text encoding features, arising from the

fact that there are no native tools fully able to support the EpiDoc format for sessions of text editing and thus to simplify the encoding operations. EpiDoc is also unable to guarantee the typical ‘relational’ features offered by a database since it lacks all the paraphernalia necessary to describe the complex web of relations between the text, the support it is carried by and the real world entities they refer to. Only ontologies and similar semantic tools would seem to be able to combine the advantages and flexibility typical of XML with the characteristic ‘relationality’ of databases.

#### 3.2 CIDOC CRM investigations

In past years very few attempts to define the ontological layer of textual entities have been made. As an example, a first reply for epigraphy was provided by the VBI-ERAT-LVPA project [11] that in 2004 tried to use CIDOC CRM for the integration of epigraphic digital archives using conceptual tools. However, it did not provide any definitive conclusions on the subject.

The EAGLE project [12], which is still on-going, is investigating the same field and trying to combine EpiDoc and CIDOC CRM entities in order to harmonise the features they provide. The EAGLE metadata model specification provided in one of its deliverables [13] admittedly makes use of the core CIDOC CRM model only “to describe the physical aspects of an epigraph, together with events related to its creation, finding and conservation”, but in no way deals with textual entities and their ambivalent, multifaceted nature in the context of epigraphy, an exercise we already attempted in CRMepi and that we are going to further discuss and focus in the present paper. TEI/EpiDoc remains the preferred way to describe inscriptions within EAGLE and, even if a preliminary mapping exercise to CIDOC CRM is provided, only the general classes of both models are taken into account: no further investigation is carried out to describe the relationships existing between the physical and the conceptual aspects of textual entities, a very important element to focus on, especially when scientific or methodological questions related with the analysis of “semiotic features” arise.

Another interesting research activity is the Menota project [14], which aims to preserve and publish mediaeval texts in digital form by adapting, developing and maintaining encoding standards necessary for this work. The standards Menota proposes are to date limited to a TEI-based model, but attempts to involve CIDOC CRM entities in order to provide richer semantic descriptions of manuscripts and their content is currently under way within this initiative.

Recently, the ARIADNE project [15], which is mainly focused on the integration of archaeological archives, has extended its area of interest to epigraphy by developing a series of tools based on CIDOC CRM in order to describe epigraphic entities in conceptual terms.

Within this project, a tentative epigraphic extension (CRMepi [16]) has been defined as an embryonic attempt to define epigraphic entities in the CIDOC CRM fashion. The CRMepi extension has served as the foundation for the research topics discussed in this paper. The previous definition and conceptualisation work performed for CRMepi are now fully incorporated into the new CRMtex extension; in future works, the epigraphic classes of CRMepi will probably be defined as CRMtex specialised subclasses, as the new extension is designed to represent a more general conceptual level, *i.e.*, ancient texts in general, not only inscriptions.

## 4 A tentative CIDOC CRM representation

### 4.1 The physical support

If we focus on the support, which is a key element to be taken into account, we note that CIDOC CRM offers plenty of concepts with which we could describe it. The physical support, being often an archaeological object, constitutes one of the main points of contact with archaeology; in terms of integration and interoperability, its related specific archaeological aspects (discovery, provenance, context, etc.) can easily be documented using the CRMarchaeo extension. It should be noted that very often the physical support was designed and built specifically to accommodate the text (e.g. a papyrus or a codex); in this case the CIDOC CRM *E84 Information Carrier* entity could be used. However, this condition does not always happen, especially in epigraphy, where inscriptions can occur on objects created at a different time or for different purposes; in this case the use of a more generic class, such as *E22 Man-Made Object*, seems more appropriate. There are also cases in which the text is inscribed on natural surfaces not created by human activities; the use of the superclass *E19 Physical Object* is preferable in this case. Each of these classes can still be linked with the physical features they bear, via the *P56 bears feature* property, having the *E19* class as domain and thus being inherited by all its subclasses. The EpiDoc elements used to mark archaeological information concerning physical objects or monuments (such as the *supportDesc*, *material*, *objectType* and *dimension* tags) can easily be mapped using these CIDOC CRM entities.

### 4.2 The written text: a semiotic feature

In CIDOC CRM, textual entities are conceived as immaterial, and essentially conceptual, entities. Both the classes *E33 Linguistic Object* and *E34 Inscription* belong to the domain of conceptual objects, defined as “non-material products of our minds and other human produced data”, something that renders only in part the essence of what a text is, not taking into account its ‘materiality’ which is a fundamental component of its identity. Other classes describing features and other less conceptual elements, such as *E36 Visual Item* and its subclasses, still fall within the area of conceptual entities and “do not intend to describe the idiosyncratic characteristics of an individual physical embodiment of a visual item, but the underlying prototype”. Even more ambiguous is the *E37 Mark* class, that “comprises symbols, signs, signatures or short texts applied to instances of *E24 Physical Man-Made Thing* by arbitrary techniques”, which could suggest some ‘physicality’ implicit in its nature; however, since it is also said that it “specifically excludes features that have no semantic significance, such as scratches or tool marks”, it seems to be totally useless for our purposes. FRBRoo [17], the object-oriented version of the FRBR ontology, harmonised with CIDOC CRM, also deals only marginally with the concept of manuscript, and mainly in a modern sense, referring mostly as autographs created by authors, often as an *avant-texte* for possible future publication. Nor does it seem to consider the ‘materiality’ of the text as a set of physical signs produced by certain materials and techniques on a given support. It will be sufficient here to observe that class *F2 Expression*, the closest one to the entities we are considering, is a subclass of the *E73 Information Object* CIDOC CRM class and “comprises the intellectual or artistic realisations of works”, occurring “in the form of identifiable immaterial objects, such as texts, poems, jokes, musical or choreographic notations”, leaving out entirely the physical/conceptual duality of a sign that is of considerable importance when dealing with semiotic aspects of human communication.

On the contrary, the study of ancient textual documents typically starts from the analysis of the physical characteristics of the text itself before moving to their archaeological, palaeographic, linguistic and historical characteristics. The ink traces in a manuscript, the scratching or engravings on an epigraph are fundamental elements for the study not only of the text but also of the (archaeological) objects on which the text appears. The ink, with which an ancient document is written, for example, may be scraped or washed off in order to reuse the support of papyrus or parchment to

create a new document (palimpsest). The text of an inscription can be cancelled by chiselling or otherwise obliterated and destroyed to make way for a new text or simply to represent a form of *damnatio memoriae*.

In this sense, a text understood only as a conceptual object does not seem to capture its true nature fully. A ‘written text’ in this sense seems to present a much closer resemblance to the CIDOC CRM classes created for the description of physical features, and more specifically the *E25 Man-Made Feature*. We have managed to create some new and more appropriate classes to be used in documenting textual concepts, and in particular:

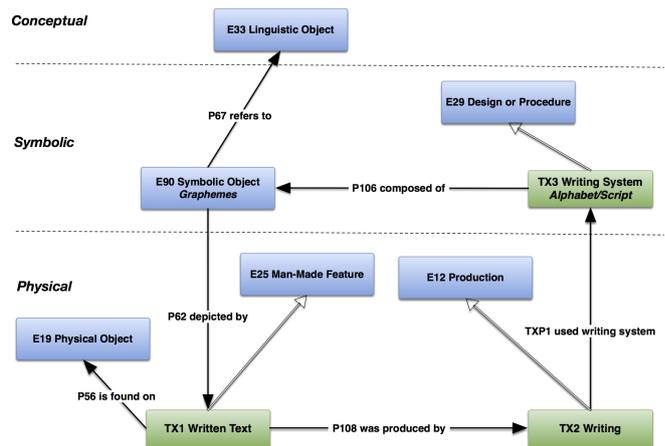
- **TX1 Written Text.** Subclass of *E25 Man-Made Feature* intended to describe a particular feature (*i.e.*, set of glyphs) created (*i.e.*, written) on various kinds of support, having semiotic significance and the declared purpose of conveying a specific message towards a given recipient or group of recipients;
- **TX2 Writing.** Subclass of *E12 Production* indicating the activity of creating textual entities using various techniques (painting, sculpture, etc.) and by means of specific tools on a given physical carrier in a non-mechanical way.

Defining this activity allows us to make a better distinction between the physical creation of the written texts and the production of the physical carriers that host them, two activities that are not necessary contemporary. This distinction is of particular importance in epigraphy, where very often the existence of the carrier precedes the creation of the text, and in the study of palimpsests, *i.e.*, scrolls or books from which a preceding text has been erased so that the pages could be reused to host another text. The relation between *TX1* and *TX2* can be expressed by the *P108 was produced by* property. The relation between *TX1* and the physical support it is carried by (*E19 Physical Object*) can be rendered through the *P56 is found on* property.

Another class has been instantiated to describe another important element in the study of ancient texts:

- **TX4 Writing Field.** Subclass of *E25 Man-Made Feature*, usually understood as the surface or portion of the physical carrier reserved, delimited and arranged for the purpose of accommodating a written text, to highlight and isolate it from the other parts of the object to which it belongs, to enhance and guarantee its readability.

This element becomes important in the very frequent case where more than a single text is found on



**Fig. 1** CRMtex classes and properties. Representation of the Writing event

a specific support, and is paramount in epigraphy, in which a specific element called epigraphic field has been defined by the discipline itself. Its importance is also evident in papyrology and codicology, where a clear distinction between area(s) containing the written text and empty parts of the support (margins, *intercolumnia*, etc.) is significant for the definition of styles and periods of the document.

From a CIDOC CRM perspective, the writing field is a feature designed to accommodate another feature (the text). EpiDoc also provides specific entities for the description of these elements (*e.g.*, the tag *layoutDesc*) that can be easily mapped on the *TX4* class. To define the relation occurring between *TX4* and *TX1*, the new property *TXP2 is included within* has been proposed as a sub property of *P56 bears feature*.

### 4.3 The text as a linguistic production

Since CIDOC CRM already contains classes for describing the abstract aspects of text, we have used the *E33 Linguistic Object* class to represent the abstract linguistic expression, *i.e.*, the message to be conveyed. The production of the physical manifestations of the text is inextricably linked to the intellectual activity of text encoding by means of graphemes, *i.e.*, the conceptual level of encoding a linguistic expression through the activity of writing (*TX2 Writing*), of which the *TX1 Written Text* class is the concrete graphical manifestation. The graphemes have been represented using the *E90 Symbolic Object* class. For the writing system we have created a new class:

- **TX3 Writing System.** Subclass of *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 a natural language. A writing system can be used to notate different natural languages, by means of specific rules in the combination and phonological value assignment of the chosen graphemes. It is used to produce a *TX1 Written Text* during a *TX2 Writing* event.

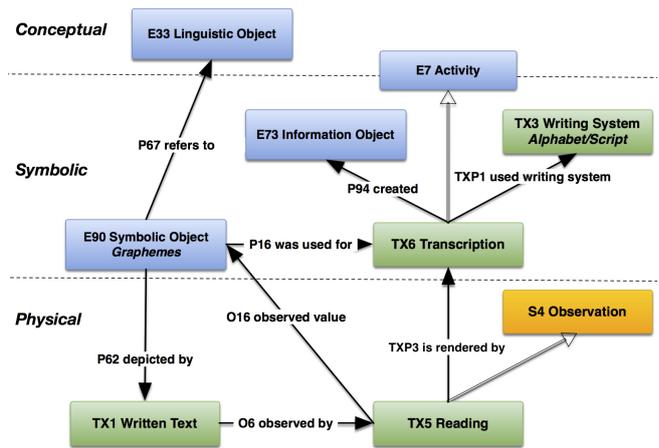
The relation between *TX2* and *TX3* is expressed through the new *TXP1 used writing system*. The relation between the feature (*TX1*) and the writing system graphemes (*E90*) could be further clarified by means of *P62 is depicted by*. The relation between *TX3* and graphemes can be codified through *P106 is composed of*. Eventually, on the most abstract level, the graphemes can be linked to the linguistic expression (*E33*) through *P67 refers to*.

#### 4.4 The text as an object of study. Scientific observation

Study and preservation of ancient documents requires scientific editions of unpublished texts and improved re-editions of already published texts. The various disciplines involved in this kind of study provide various methodological approaches, mostly depending on the various natures of the physical carriers and the number of available copies of the documents. It is impossible to describe all these approaches in detail within the limits of this paper. However, the method of textual analysis presents common traits allowing us to establish general classes in order to describe the text. One of the most important activities carried out in the scientific investigation of texts is that of reading, which we have defined as:

- **TX5 Reading.** Subclass of the CRMsci *S4 Observation* class, referring to the scientific autoptic examination of the document and constituting the first action required in preparation for its study. It consists of an accurate analysis of the surface and the signs and prescribes the use of specific tools and procedures, to establishing as faithfully as possible the exact value of each sign drawn on the physical feature.

The glyphs observed by the scholar (*TX1*) are linked to the reading activity (*TX5*) through the *O6 observed by* property. The result of this activity is the recognition of the graphemes (*E90*) represented by the glyphs; the notion of recognition in this case is expressed through the CRMsci *O16 observed value* property. The relation between the observed glyphs (*TX1*) and the represented graphemes is again rendered using the *P62 is depicted by* property.



**Fig. 2** CRMtex extension: Reading and Transcription events.

The activity of reading is the specular counterpart of the event of writing (*TX2*): where the latter represents the physical encoding of a conceptual object (*E33* through *E90*), the former constitutes the level of the intellectual decoding and understanding of the signs. It is also the basis for the subsequent operations of transcription, for which we have created the class:

- **TX6 Transcription.** Subclass of *E7 Activity*, referring to the activity of re-writing the text conducted by an editor. This operation, in some cases, involves a writing system (*TX3*) different from that of the original text (e.g., Latin characters to render a Coptic text); this results in a re-encoding of the text itself and, from a linguistic point of view, it is indicated more properly as a ‘transliteration’, because it implies a 1 : 1 relation between the signs of the two writing systems. The *P16 was used for* property can be used to specify the role of the original graphemes during the commuting operations.

Moving on from observation to the editorial work of scholars, the two events of reading (*TX5*) and transcription (*TX6*) are linked through the new *TXP3 is rendered by* property, to emphasize the close connection between these two activities. The transcription generally results in the creation (*P94 created*) of a document containing the outcome of the transcription operations in textual form (*E73 Information Object*), e.g., the edition of the text. FRBRoo entities could be used here to link the transcriptions operated by scholars with the related bibliographic information.

Since the activity of transcription (*TX6*) itself is a re-encoding of the text, the use of a writing system can also be specified by means of the *TX3* class and the related *TXP1* property.

Additionally, since the *E33 Linguistic Object* is an expression of the original language (*P72 has language*), it can be provided with a translation into any other language (*P73 has translation*), for example, into English, in order to make the content more understandable.

## 5 An example

In this section we shall present an example to demonstrate the potentialities of our model. We chose a famous artefact, the Derveni Papyrus, the oldest surviving manuscript of Europe according with UNESCO, currently stored at the Archaeological Museum of Thessaloniki. The papyrus was found in 1962, in a tomb of the necropolis of the ancient city of Lete (modern Derveni, Greece). The physical support is a scroll of papyrus written in Greek. Both the scroll and the text have been dated to around the 4th century BC, during the reign of Philip II of Macedon. The papyrus is of great interest and is still being studied by many scholars but the text remained unpublished for a long time. Its first publication dates to 2006, edited by Kouremenos et al. [18]. A very interesting episode in its history is the reading made in 1982 [19] by an anonymous scholar who, unsatisfied with the delay before publication, made an unauthorised reading of the text, which provided a valuable starting point for subsequent studies. We have based the example presented here on this unauthorised account.

If we focus on the physical carrier of this object, we could assume that the scroll was created specifically to host the text. In this case, it seems more appropriate to use the *E84 Information Carrier* class rather than its *E22 Man-Made Object* superclass for the encoding of this element.

The detailed archaeological story of the artefact could easily be traced by using the entities provided by CRMsci and CRMarchaeo, and specifically by means of the *S19 Encounter Event* (and the related *O19 was found by* property to connect *E84* encoded objects) and the *A1 Excavation Process Unit* classes. These descriptions could then be immediately connected with the modelling of the textual entities provided by the CRMttx extension. The use of *TX1 Written Text* and the related production event (*TX2 Writing*) allows us to distinguish the event of creation of the text from that of the archaeological object, although in this case the two events happened simultaneously, since the text was in fact written contextually to the production of the papyrus. The definition of instances of the same *E52 Time Span* class for both the events can be useful to indicate in an unambiguous way their contemporaneity.

The *E90 Symbolic Object* is represented by a set of characters (graphemes) belonging (*P106*) to the Greek alphabet (*TX3 Writing System*) used during the writing activity (*TX2 Writing* -> *TXP1 used writing system*). Through analysis of the papyrus (*TX5 Reading*, subclass of *S4 Observation*), in 1982 the anonymous scholar provided a reading and a transcription of the text (*TX6 Transcription*), using the Greek alphabet (*TX3*), published in a scientific international journal (*P94 created* -> *E73 Information Object*). This same set of classes and properties can be instantiated several times for any new or different readings, transcriptions and interpretations of the same text by other scholars, in order to create a chain of events able to represent the history of the object.

According with the general CRMttx schema, the transcription refers to the *E33 Linguistic Object*, i.e., the text intended as a linguistic expression, encoded by means of a given writing system (in this case, Greek alphabet) in the papyrus. The *E33 Linguistic Object* is therefore linked both to the *E90 Symbolic Object* and to the *TX1 Written Text* (the concrete manifestation of such Greek units as physical features), through the *P67 refers to* and the *P62 is depicted by* properties respectively. Since the *E33 Linguistic Object* is an expression of the Greek language (*P72 has language*), it can be provided with a translation into any other language (*P73 has translation*), for example, into English, in order to make the content available to a wider audience.

## 6 Conclusions and further work

The activities described in this paper are only the tip of the iceberg of the complex work that would require the definition of all entities involved in the study of ancient texts. The styles, the techniques, the materials used, the historical events and the concepts involved are so numerous that this work would never be completed.

From a palaeographic point of view, for instance, the study of the stylistic variations of glyphs (e.g. an 'A' uppercase, lowercase, italics, round, printed or written by hand, or in different font families) has great importance in the description of ancient texts, using different styles for different purposes or in different times and places. We can for instance describe all the entities to a given epoch and place, e.g. the *Ptolemaic cursive* of the Hellenistic Egypt, the (majuscule) *uncial script* (3rd-8th cent. AD), used both for Greek and Latin alphabets, or the more recent *Carolingian minuscule*, used from the beginning of the 8th cent. AD. Therefore, in palaeography the concepts of stylistic class, style and canon are

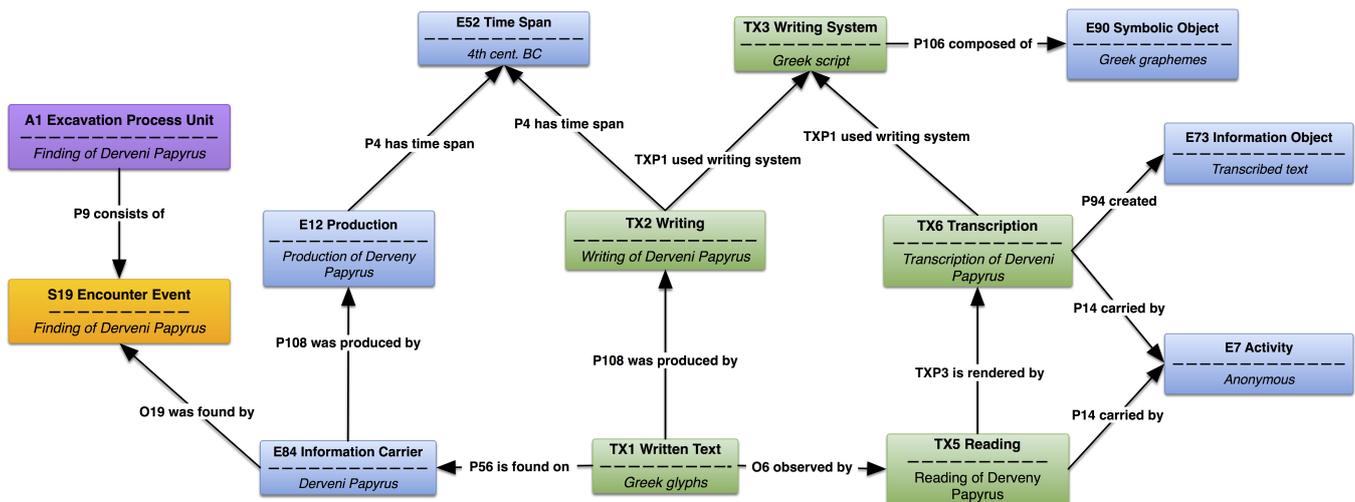


Fig. 3 CRMtex encoding of the Derveni Papyrus example

fundamental to underline different meaningful observable aspects. The specific study of these stylistic variations needs to be properly addressed. The *P15 was influenced by* and the *E83 Type Creation* classes could constitute an optimal starting point for this activity. A thesaurus containing specifications of the various writing systems and styles would tremendously improve our work.

Additionally, if we consider that EpiDoc not only provides entities for the description of the text and its structural characteristics but also a series of tags for the identification of actor and place names it contains, we shall notice that the text itself may contain semantically relevant elements that need to be captured in some way. Actor appellations can for instance relate to the commissioners of a given monument or to the people to whom a certain epigraph was dedicated, and place appellations could refer to places where the inscription was located in the past, or to which the text refers in various ways.

Interpretation of texts as inferences is a very hot topic in IT research, not only in Cultural Heritage, and the integration and interoperability with bibliographic information would constitute a natural complement and a privileged transversal activity for the future development of our study. Integrating this huge variety of data will allow us to enrich the international knowledge network on ancient studies and also to expand our historical knowledge beyond the scope of our field of research.

**Acknowledgements** The present work has been partially supported by the ARIADNE project, funded by the European Commission (grant 313193) under the FP7 INFRA-2012-1.1.3 call.

## References

1. CIDOC CRM v. 6.2.1, [http://cidoc-crm.org/official\\_release\\_cidoc.html](http://cidoc-crm.org/official_release_cidoc.html)
2. CRMsci: The Scientific Observation Model, [http://www.ics.forth.gr/isl/index\\_main.php?l=e&c=663](http://www.ics.forth.gr/isl/index_main.php?l=e&c=663)
3. CRMarchaeo: The Excavation Model, [http://www.ics.forth.gr/isl/index\\_main.php?l=e&c=711](http://www.ics.forth.gr/isl/index_main.php?l=e&c=711)
4. Pierce C S (1931-1958), Collected Papers of Charles Sanders Peirce, Burks A W (ed.), Harvard University Press, Cambridge, MA
5. Sanders C (ed.) (2004), The Cambridge Companion to Saussure, Cambridge University Press
6. Trismegistos, An interdisciplinary portal of papyrological and epigraphical resources, <http://www.trismegistos.org/>
7. Papyri.info, papyrological navigator and editor, <http://papyri.info/>
8. Epigraphic Database Roma, <http://www.edr-edr-it>
9. Deutsches Archaeologisches Institut, Arachne iDAI Objects, <http://arachne.dainst.org/>
10. EpiDoc: Epigraphic Documents in TEI XML, <http://sourceforge.net/p/epidoc/wiki/Home/>
11. Doerr M, Schaller K, Theodoridou M (2010) Integration of Complementary Archaeological Sources. In Nicolucci F, Hermon S (eds.) Beyond the Artifact. Digital Interpretation of the Past. Proceedings of CAA 2004, Archaeologia, Budapest, p6469
12. EAGLE Project, <http://www.eagle-network.eu/>
13. EAGLE Deliverable D3.1, <http://www.eagle-network.eu/about/documents-deliverables/>
14. Menota Project, Medieval Nordic Text Archive, <http://www.menota.org/>
15. ARIADNE Project, <http://www.ariadne-infrastructure.eu/>
16. Felicetti A, Murano F, Ronzino P, Nicolucci F (2016) CIDOC CRM and Epigraphy: a Hermeneutic Challenge. In Ronzino P, Nicolucci F (eds.) Extending, Mapping and Focusing the CIDOC CRM. Proceedings of CRMEX 2015. Under Publication
17. FRBRoo, [http://www.cidoc-crm.org/frbr\\_intro.html](http://www.cidoc-crm.org/frbr_intro.html)
18. Kouremenos T, Parssoglou GM, Tsantsanoglou K (2006) The Derveni Papyrus. Firenze, Olschki
19. (1982) Der Orphische Papyrus von Derveni. ZPE 47:1-12