Describing Inscriptions of Ancient Italy. The ItAnt Project and Its Information Encoding Process

Describing Inscriptions of Ancient Italy.
The ItAnt Project and Its Information Encoding Process

Francesca Murano, Dipartimento di Lettere e Filosofia, Università di Firenze, Italy
Valeria Quochi, Istituto di Linguistica Computazionale "A. Zampolli", CNR, Italy
Angelo Mario Del Grosso, Istituto di Linguistica Computazionale "A. Zampolli", CNR, Italy
Luca Rigobianco, Dipartimento di Studi Umanistici, Università di Venezia, Italy
Mariarosaria Zinzi, Dipartimento di Lettere e Filosofia, Università di Firenze, Italy

Abstract
This article discusses the challenges addressed in the digital scholarly encoding of the fragmentary texts of the languages of Ancient Italy according to the TEI/EpiDoc Guidelines in XML format. It describes the solutions and customisations that have been adopted for dealing with the peculiarities of our epigraphical documentation and with the formalisation of epigraphical information deemed interesting for data retrieval in a historical linguistic perspective. The making of a digital corpus consisting of new critical editions of selected inscriptions is a work carried out in the context of the project "Languages and Cultures of Ancient Italy. Historical Linguistics and Digital Models", which aims to investigate the languages of Ancient Italy by combining the traditional methods, proper to historical linguistics, with methods and technology proper to the digital humanities and computational lexicography. More specifically, the purpose of the project is to create a set of interrelated digital language resources which comprise: (1) a digital corpus of texts editions; (2) a computational lexicon compliant with the Web Semantic requirements; (3) a relevant bibliographic reference dataset encoded according to the FRBRoo/LRMoo specifications. Additionally, selected textual data and scientific interpretations will be encoded using CIDOC CRM and its extensions, namely CRMtex and CRMinf. The present contribution thus tackles one of the main aspects of the project, and proposes significant innovations in the encoding of critical editions for epigraphic texts of fragmentary languages, which will hopefully foster future interoperability and integration with other external datasets, a paramount concern of the project.

1 INTRODUCTION

Over the last few years, the digital resources available for the study of ancient languages have grown considerably, both in the field of linguistic resources (e.g., treebanks and dictionaries) and in the field of digital libraries [54, 59]. In the perspective of contributing to the effort of creating a Knowledge Base of language resources for the study of ancient languages, this article discusses the challenges addressed in the process of encoding the fragmentary texts of the languages of Ancient Italy according to the TEI/EpiDoc Guidelines and format, and describes the solutions and customisations that have been adopted for dealing with the peculiarities of our epigraphic documentation and with the formalisation of epigraphic information deemed interesting for data retrieval in a historical linguistic perspective.

The task of compiling a digital corpus of new critical editions of selected inscriptions is a work carried out in the context of the project *Languages and Cultures of Ancient Italy. Historical Linguistics and Digital Models* (ItAnt hereafter), which aims to investigate the languages of Ancient Italy by combining the traditional methods, proper to historical linguistics, with methods and technology proper to the digital humanities and computational lexicography. The present contribution thus tackles one of the main aspects of the project, and proposes significant innovations in the encoding of critical editions for epigraphic texts of fragmentary languages, which is believed to foster future interoperability and integration with other external datasets, a paramount concern of the project.

The article is structured as follows: a short introduction on TEI/EpiDoc standard will be given in Section **Errore. L'origine riferimento non è stata trovata.**, followed by state-of-the-art concerning the application of the TEI/EpiDoc vocabulary to Digital Epigraphy projects (Section 3.1). The issues related to the digital scholarly encoding of the epigraphical data of Ancient Italy will make the object of Section 4, by focusing respectively on the description of the support and layout (Section 4.1), some linguistic elements (Section 4.2), the description of language(s) and script(s) (Section 4.3) and, lastly, vocabularies and gazetteers (Section 4.4). In Section 5, the

contribution of CIDOC CRM in integrating data will be introduced. Final remarks and further steps of the research will be elicited in the conclusions.

2 THE ITANT PROJECT

The project *Languages and Cultures of Ancient Italy. Historical Linguistics and Digital Models* [29] is an initiative funded by the Italian Ministry of University and Research and involves a consortium composed by the University Ca' Foscari of Venice, the University of Florence, and the Institute for Computational Linguistics "A. Zampolli" of the National Research Council of Italy. Its goal is to investigate the languages of Ancient Italy, combining the traditional methods, proper to historical linguistics, with the setting up of digital and computational technologies specifically designed to create a set of interrelated and interoperable digital resources.

Ancient Italy, during the period dating from the appearance of writing (8th c. BCE) until complete linguistic Romanisation (1st c. BCE to 1st c. CE), is characterised by the presence of numerous linguistic varieties that are both Indo-European and non-Indo-European (Figure 1).

Fig.1. Map of the languages of ancient Italy. Image by ItAnt Project.
This file is licensed under the CCBY-NC-SA4.0 license, via https://www.prin-italia-antica.unifi.it.

The attestations of these languages only consist in epigraphic evidence with a very restricted text typology, the evidence being not homogeneous as for quantity, quality, geographical distribution, and chronology (*Restsprachen*). Among these languages, the project focuses on four representative ones, i.e., Oscan, Faliscan, Venetic, and Cisalpine Celtic, varieties belonging to linguistic (sub)groups that are diverse as regards their genetic classification [56]. Because of their status as *Restsprachen*, the languages of Ancient Italy often present specific philological and linguistic issues concerning the reading of the inscriptions, their segmentation into words, the linguistic analysis, and their interpretation. Therefore, one of the main challenges of the ItAnt proposal is to adapt the digital tools, practices, and methodologies of digital epigraphy and computational lexicography to the highly fragmentary nature of such epigraphic documentation. More specifically, the purpose of the project is threefold:

1. to create a digital archive of texts from Ancient Italy which contains their formal representation leveraging the latest recommendations defined within the TEI/EpiDoc XML encoding schema [46]. Furthermore, the digital scholarly edition of the texts is enriched with shared and standard metadata, thus allowing for an accurate description of each of them both as a linguistic and a material object. A facsimile of the

inscriptions is also provided. The encoded corpus pertaining to Ancient Italy following the TEI/EpiDoc guidelines represents a novelty in digital epigraphy[1];

2. to create a multilingual computational lexicon, specifically adapted to languages of fragmentary attestation, which contains lexical information about the items/word-forms occurring in the texts. The corpus and the lexicon are linked to each other and share the bibliographical apparatus, realised according to digital bibliographical standards;

3. to encode additional relevant information—such as bibliography, citations, and scientific interpretations—using FRBRoo/LRMoo [18, 61], and the CIDOC CRM extensions CRMtex [7, 47] and CRMinf [6, 45], in order to propose a description of the texts by means of conceptual classes defined in standard formal ontologies. This will lead to the publication of interlinked datasets well-endowed in information and will foster semantic interoperability of both textual entities and their scientific interpretations.

In order to ensure a long-term preservation and easy accessibility of all project outcomes, software and data will be made available through relevant European Research Infrastructures such as CLARIN [5]. In particular, they will be catalogued and archived according to the FAIR principles [17, 48] in the CLARIN-IT trusted repository (ILC4CLARIN [23]), which ensures their findability and accessibility by providing rich metadata descriptions using standard CLARIN metadata profiles. Each deposited resource will be assigned with a Persistent identifier (PID), and given high visibility through the Virtual Language Observatory (VLO), the CLARIN ERIC European discovery service. Accessibility will also be facilitated by an online full-fledged platform that allows for cross-searching in the interrelated datasets and visualising the connected results[2]; the platform is also to become a CLARIN service, which ensures its sustainability [50]. Interoperability is guaranteed by the adoption of widely recognised and used standards models and formats such as Ontolex-lemon for lexical resources, TEI Epidoc for textual data, and FRBRoo for bibliographies. Finally, all digital resources and software are distributed according to open licences that will allow for reuse (at least) for research purposes. Software is released as open source and data under Creative Commons, non-commercial, share-alike open licences [27]. Reusability is facilitated by the high degree of interoperability and by the PIDs mentioned above, which also provide easy citations for reference. Since this article describes the activities concerning the digital scholarly encoding process of the project, we will specifically focus on the TEI/EpiDoc digital encoding issues. The lexicon and the CIDOC CRM mapping will make the object of forthcoming papers.

3 TEI/EPIDOC AND DIGITAL EPIGRAPHY

TEI/EpiDoc is an international consortium which establishes guidelines and implements tools for digital encoding scholarly editions of ancient documents according to the Leiden conventions [49, 66]. In particular, it specifies a subset of the Text Encoding Initiative's standard vocabulary specifically defined for text criticism and transcription in order to represent epigraphical texts in digital form using the XML format, which is optimised for compatibility, interchange, and durability [42].
TEI/EpiDoc is the de facto standard schema adopted by digital epigraphers, providing mark-up elements for the transcription and edition of the texts (edition, translation, apparatus, commentary, bibliography) and for the description of the history (provenance, location, date, repository) and materiality (physical description) of the objects on which the text appears [13]. The digital description of inscriptions and the semantic mark-up of the text are given in a flexible and exchangeable model [40, 42]. This fosters compatibility and interoperability with other text projects created according to the XML/TEI format and, thus, their availability for the research community, according to the chosen licence policies. Moreover, XML can be processed and queried by mostly any modern search engine.Since XML is an extensible technology, TEI/EpiDoc is a versatile and customisable tool which can be easily adapted to specialised needs and thus enhanced by creating new solutions, including custom-new-values for attributes. ItAnt proposed some ad hoc solutions to manage the peculiarities presented by a language of

---

[1] Lately, nevertheless, the interest in the digital treatment of the languages of Ancient Italy has been growing, as shown, e.g., by the recently published linguistic database *Corpus of the Epigraphy of the Italian Peninsula in the 1st Millennium BCE* (CEIPoM), see Reference [55] . CEIPoM is a full-text database which provides manually annotated linguistic information on all levels of linguistic structure. The corpus is not encoded leveraging the TEI/EpiDoc schema and does not provide specific epigraphical metadata, except for chronological and geographical information; the texts, however, are linked to Trismegistos [38] and epigraphic corpora such as EDR [15]. No user interface is provided for queries and consultation of the database.

[2] The platform is itself an outcome of the project and, at the time of the writing of this article, it is a work-in-progress [27, 59].

fragmentary attestation by describing more carefully its linguistic issues. Furthermore, as pointed out in [42], since XML allows for structured and semantic markup, the files encoded according to the TEI/EpiDoc schema can also be changed in format and structure, as well as transformed into another encoding schema or translated into database systems, and their long-term preservation is expected to be guaranteed.

For its wide use in digital epigraphy, its versatility, and the easy-to-render visualisation of texts, TEI/EpiDoc has been the ideal candidate for the digital scholarly encoding of the documents selected by the project.[3] Nonetheless, as already pointed out in the literature, "a TEI/EpiDoc document consists of a monolithic, self-descriptive and self-standing information unit which hardly exposes an easy way for cross-linking different documents" [43] within the same or between different digital archives. For example, heterogeneous material encoding, semantic-related description of data and metadata, linked open data queries, and effective support for data integration are poorly served matters through the EpiDoc technology ecosystem, based entirely on the Extensible Mark-up Language (XML) format. To overcome this issue, our project aims at taking a further step, namely a description of the texts and their features, such as their production and investigation, by adopting the formal CIDOC CRM ontological ecosystem. This allows us to encode the documents in a more expressive way, establishing a deeper and wider semantic interoperability.

3.1 State-of-the-art of the Application of TEI/EpiDoc to Digital Epigraphy

In the last decades, digital epigraphy has grown and several projects of digitisation of inscriptions have been launched. Many of them, mostly completed, deal with the TEI/EpiDoc encoding schema.[4] The TEI/EpiDoc guidelines have been first and foremost applied to Greek and Latin epigraphical texts coming from all over the Mediterranean area. In this regard, the most relevant project is EAGLE, the Europeana network of Ancient Greek and Latin Epigraphy co-funded by the European Commission [15]. Among the others, TEI/EpiDoc corpora are provided by the *Ancient Inscriptions of the Northern Black Sea* project (IOSPE) [2], the *Inscriptions of Roman Tripolitania* project [25], the *Cretan Institutional Inscriptions* project [64, 65], and the *I.Sicily* project [26, 58]. The project *Crossreads. Text, materiality and multiculturalism at the crossroads of the ancient Mediterranean* [8] aims, among other objectives, at completing the corpus of *i.Sicily*, collecting texts from ancient Sicily dating from the late 7th c. BCE to the Byzantine period (7th c. CE), written in several languages and alphabets (e.g., Greek, Latin, Elymian, Sikel, Punic). A particular mention deserves the *Iscrizioni Latine Arcaiche* (ILA) project [62], dealing with the archaic Latin inscriptions; these documents, dating to the 7th-5th c. BCE, present peculiar characteristics which distinguish them from later "classic" Latin material (e.g., particular directions of writing), for which a customisation of the TEI/EpiDoc schema was necessary. The TEI/EpiDoc encoding schema has been also leveraged for encoding the ancient papyrological documents in the Papyri.info initiative [31]. The portal aggregates three major databases of documentary papyri, namely the *Duke Databank of Documentary Papyri* (DDbDP), the *Heidelberger Gesamtverzeichnis der griechischen Papyruskunden Ägyptens* (HGV), and the *University of Michigan Advanced Papyrological Information System* (APIS).

Starting from the Latin and Greek epigraphic culture, the TEI/EpiDoc mark-up schema has been leveraged for epigraphical texts in different languages. It has been used, e.g., for the following projects: 1. the *Inscriptions of Israel/Palestine* (IIP) project [24], which collects documents of this region from the Persian period through the Islamic conquest (ca. 500 BCE–640 CE) written primarily in Hebrew, Aramaic, Greek, and Latin; 2. the ERC project *Digital Archive for the Study of pre-Islamic Arabian Inscriptions* (DASI) [9, 39], gathering the pre-Islamic Arabian epigraphic material into a comprehensive online database; 3. the collection of Celtic inscriptions written using the ogham alphabet [30].

The adoption of the TEI/EpiDoc XML vocabulary has spread far beyond the ancient Mediterranean world and currently counts projects focusing on texts produced up to Early Modern times. Specifically, it is being leveraged within the project *TITULUS—Corpus des inscriptions de la France médiévale* [37], aiming at creating a digital edition of the newly published volumes of the collection. The TEI/EpiDoc guidelines are also being used for creating *A database of Middle Turkic* [1], which aim at providing a comprehensive digital edition, including transcription and

---

[3] Although the TEI/EpiDoc guidelines have been adopted to manually encode the primary sources of the project, a proof-of-concept alternative has been developed within the project to ease the text encoding process. In particular, a Domain Specific Language (DSL) has been defined in order to produce the digital edition. The use of the DSL preserves the expressivity of the TEI/EpiDoc schema and automatically converts the DSL-encoded text into XML documents. This approach enhances readability, compactness, and familiarity with respect to the traditional editorial practices. A publication on this activity is in progress.

[4] For the sake of brevity, only a few projects leveraging the TEI/EpiDoc schema are listed in this article.

translation, of Middle Turkic sources, spanning from the 13th c. up to the Early Modern times. The project *Beta maṣāḥaft: Manuscripts of Ethiopia and Eritrea* [4, 41] at the University of Hamburg also leverages the TEI/EpiDoc schema as part of digitising work meant to create a virtual research environment managing the information related to the (predominantly Christian) manuscript traditions of the Ethiopian and Eritrean Highlands. The schema is also being used within the running ERC project *DHARMA. The Domestication of Hindu Asceticism and the Religious Making of South and Southeast Asia* [10], to digitally encode South and Southeast Asian manuscripts, which preserve a rich textual archive relevant to the history of Hinduism. The digital editions of the texts, which are still unpublished, come with translations.

4 ENCODING THE EPIGRAPHIC DATA OF ANCIENT ITALY: NEEDS, PROPOSALS, AND CUSTOMISATION

4.1 Describing Support and Layout

Although TEI/EpiDoc provides the necessary structure for an in-depth description of data, the research needs of the current project, emerging in the phase of analysis, required customisation of the TEI/EpiDoc schema. On the one hand, the proposed customisation deals with the fragmentary nature of the languages of Ancient Italy and the peculiarities of their epigraphical documentation and, on the other hand, with the formalisation of epigraphic information interesting for some specific needs in data retrieval. Since ItAnt is part of a historical linguistic study of materials from Italy, the description of specific epigraphic characteristics intertwined with linguistic features is particularly relevant for the project purposes.

The customisation we operated consists of a series of additions to the standard set of TEI/EpiDoc tags, in order to have a more flexible set of tools to deeply describe the different conditions and situations occurring in our studies. Another intervention consists in the differentiation of the information concerning languages and scripts. From a scientific point of view, in fact, within Linguistics, language and writing belong to different conceptual levels and, therefore, their distinction represents a necessary theoretical requirement. According to the bottom-up approach characterising the TEI/EpiDoc community, the solutions we adopted might be useful for other projects which deal with structurally similar epigraphic materials. A schematic description of the solutions adopted is provided below. The inscriptions of Ancient Italy present a variety of typologies of word parsing, ranging from a division of the words using marks, such as simple points or double points, to a complete lack of spaces in writing (the so-called "scriptio continua"). In order to describe the disposition of the words, we customised the TEI/EpiDoc schema adding a `<tei:rs>` element with a `@type="wordDivision"` attribute within the `<tei:scriptNote>` element. This type of encoding is used in TEI/EpiDoc for similar descriptions, e.g., for the description of the form of execution of the text (`<tei:layout><tei:rs type="execution">inscribed</tei:rs></tei:layout>`). The formalisation of this information is important to establish, for instance, the provenance or the dating of a text. Possible values for `<tei:rs type="wordDivision">` are "scriptio continua", "punctuation", "blank spaces", and "mixed". Furthermore, different types of syllabic punctuation are used in the Venetic inscriptions, for which a more specific description is provided using the same method [53]:

```
<tei:scriptNote>
    <tei:rs type="wordDivision">syllabic punctuation</tei:rs>
    <tei:rs type="syllabicPuntuactionSimplification">(un)simplified</tei:rs>
    <tei:rs type="syllabicPuntuactionApplication">(un)properly applied</tei:rs>
</tei:scriptNote>
```

The addition of a `<tei:rs>` element with a @type attribute used to specify various types of information has been used also within the `<tei:support>` element to clarify the object shape (`<tei:rs type="objectShape">`) and its possible reuse (e.g., `<tei:rs type="reuse">`). The values of the "support" element come from an open vocabulary (see below, Section 4.4), the ones of the "reuse" are represented using binary yes/no options. The description of the inscription carrier is complemented by information about its conservation state: according to the TEI guidelines, we use the `<tei:condition>` element within the `<tei:supportDesc>` element, whose values are defined by a controlled list of terms, as "well-preserved", "fragmentary", "very fragmentary", or "lost inscription". An example:

```
<tei:supportDesc>
    <tei:support>
        <tei:objectType>cippus</tei:objectType>
        <tei:rs type="objectShape">cylindrical</tei:rs>
        <tei:rs type="reuse">no</tei:rs>
    </tei:support>
    <tei:condition>fragmentary</tei:condition>
</tei:supportDesc>
```

Information about the layout is enriched through a further `<tei:rs>` element added within the `<tei:layout>` element to specify whether the inscription is opisthographic (i.e., inscribed on both sides) or not. Values consist of a binary yes/no option. In case the editor of a text deems it appropriate to also take into account the dimension of the lines, in accordance with the TEI/EpiDoc guidelines, the `<tei:dimensions>` element is used, with the addition of a `@type="lineDimension"` attribute. To allow for the description of each line, lines are referred to by the `@corresp` attribute. An example:

```
<tei:layout columns="1" writtenLines="2">
    <tei:rs type="execution" ana="http://vocab.getty.edu/page/aat/300053829">engraving</tei:rs>
    <tei:rs type="opistography">no</tei:rs>
    <tei:dimensions type="lineDimension" unit="cm" precision="high" corresp="#1">
        <tei:height>3</tei:height>
        <tei:width>28</tei:width>
    </tei:dimensions>
    <tei:dimensions type="lineDimension" unit="cm" precision="high" corresp="#2">
        <tei:height>3</tei:height>
        <tei:width>23</tei:width>
    </tei:dimensions>
</tei:layout>
```

4.2 Describing Linguistic Elements

Concerning the edition of the text, the comment of the linguistic elements is carried out by means of `<tei:w>` and `<tei:name>` elements only, since, within ItAnt, the actual linguistic description is entrusted to the electronic lexicon, providing all the morpho-syntactic, semantic, and etymological information [59]. The `<tei:w>` and `<tei:name>` elements are linked to the lexicon via their unique identifiers, consisting of the value of their `@xml:id` attribute. The identifier is built using information about the language, the line number, and the position of the word in the line, so as to be transparent and easily readable also by scholars. For example, the value "Fal_6_l_1_w_2" stands for 'second word of the first line of the sixth Faliscan inscription of the ItAnt collection'. Lines are also provided with an identifier built with the same method.[5] An example:

```
<tei:div type="edition" xml:space="preserve">
    <tei:div type="textpart" n="face_a" style="text-direction:r-to-l">
        <tei:ab>
            <tei:lb n="1" xml:id="Fal_6_l_1"/>
                <tei:w xml:id="Fal_6_l_1_w_1">eko</tei:w>
                <tei:name type="praenomen" xml:id="Fal_6_l_1_w_2" ref="#p1">lartos</tei:name>
        </tei:ab>
    </tei:div>
</tei:div>
```

As seen in the example, according to the guidelines, the `<tei:name>` element is further specified through an `@type` attribute meant to declare its typology with values as "praenomen", "gentilicium", "patronymic", and so on. The formalisation of this information allows the retrieval of onomastic data and their integration with datasets coming from other scientific initiatives within the onomastic studies. Furthermore, we use a `@ref` attribute to link each `<tei:name>` element to its related onomastic formula. Since the onomastics are paramount in the study of the languages and cultures of Ancient Italy, ItAnt "reconstructs" the entire onomastic formulas within a `<tei:listPerson type="onomastics">` element within the `<tei:div type="commentary">` element. This

---

[5] Although we are aware that this would be questionable, the solution of generating semantic identifiers represents for us the best compromise between technical soundness and disciplinary scholarly needs, since: (1) in our scenario all material is encoded and linked manually; (2) the number and length of texts are intrinsically limited; and (3) (part of the) textual information will have to be published as Linked Open Data.

facilitates the description of the people mentioned in a text, also when the onomastic formulas appear in a non-standard sequence. In fact, the texts can show syntactic breaks between the components of the onomastic formulas, as well as the sharing of a component between two or more formulas. An example is provided by the Venetic inscription Bl 1 *Enoni Ontei Appioi sselboisselboi Andeticobos ecupetaris* "funerary monument of Enono-for himself, Onti- and Appio- Andetico-", where *Andeticobos* is a gentilicium referring both to *Onti-* and *Appio-*, and indirectly to *Enono-*. The rigidly nested syntax of the XML makes it impossible to describe this kind of situation while tagging a text. Thus, the solution proposed above seems particularly adapted for this kind of condition. An example from Oscan:

```
<tei:listPerson type="onomastics">
    <tei:person xml:id="p1" corresp="#Osc_2_l_1">
        <tei:persName>
            <tei:name type="praenomen">pakis</tei:name>
            <tei:name type="gentilicium">heleviis</tei:name>
            <tei:name type="patronymic">tre(bieís)</tei:name>
        </tei:persName>
    </tei:person>
    <tei:person xml:id="p2" corresp="#Osc_2_l_1b">
        <tei:persName>
            <tei:name type="praenomen">statis</tei:name>
            <tei:name type="gentilicium">betitis</tei:name>
            <tei:name type="patronymic">[-]</tei:name>
        </tei:persName>
    </tei:person>
</tei:listPerson>
```

4.3 Describing Language(s) and Script(s)

Another solution adopted by ItAnt is meant to express an important issue for linguistics, namely the separation in the encoding of language(s) and script(s) information. The TEI/EpiDoc guidelines suggest to describe the language(s) and script(s) as a combination marked up in `<tei:language>` element(s) within the `<tei:Header>` section. The value of the `<tei:language>` tag is defined by the `@ident` attribute containing a composed string (called "language tags") functioning as unique identifiers for any combination of language and script. The "language tags", in accordance with the *RFC 5646: Tags for Identifying Languages* [34], consist of "subtags" for language and, if necessary, for script, regional, and dialectical variation, which should be recorded in the authoritative IANA Language Subtag Registry [21]. For example, the element `<tei:language ident="grc">Ancient Greek</tei:language>` marks a Greek text written in the Greek alphabet, whereas `<tei:language ident="grc-Latn">` marks a text in "Ancient Greek language written in Latin script". This latter text description, in the example provided by the guidelines, is also recorded as textual value of the `<tei:language>` element [12].

When a subtag is not available in the IANA Language Subtag Registry list, a private-use subtag, separated from the other "standard" subtags by the single-character "x", can be used to indicate distinctions in language considered important in a given context. For example, `<tei:language ident="sa-x-OldCam">` Sanskrit language, written in the Old Cam script `</tei:language>`.

Concerning the languages considered in the ItAnt project, the IANA provides the subtags "xcg" for Cisalpine Celtic, "xfa" for Faliscan, "osc" for Oscan, and "xve" Venetic. The linguistic situation of Ancient Italy, characterised by a variety of people and cultures which also results in a variety of local writing systems and epigraphies, however, makes it necessary to describe the information about languages and scripts of the analysed inscriptions in a more structured way, separating data concerning the language of the inscription from that concerning the writing system. Limiting to these four languages, in fact, three different scripts are attested in the Oscan inscriptions, namely an epichoric Etruscan-based alphabet, a modified Greek and a modified Latin alphabet [57]. Faliscan uses an epichoric Etruscan-based alphabet, but also other alphabetical forms are attested [60]. The Venetic inscriptions show a script characterised by fundamental traits of unity, albeit with variations due to chronology and local uses [53]. Also, the Cisalpine Celtic shows an articulated situation concerning its scripts [63]. Figure 2 shows, e.g., an Oscan text, written in the Oscan "National" alphabet.
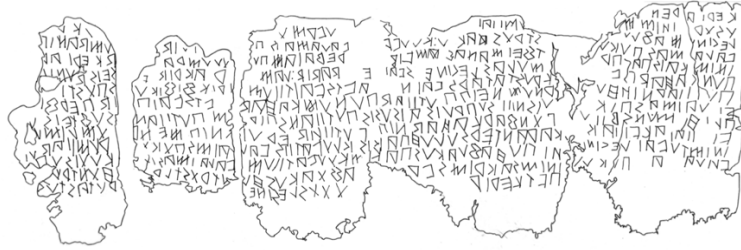
Fig. 2. Example of inscription: the Oscan curse tablet from Capua (4th c. BC). Drawing by Francesca Murano.
This file is licensed under the CC BY-NC-SA 4.0 license.

For these reasons, the mere indication of the language is not sufficient for accurately describing our inscriptions, and it is necessary to also proceed to the indication of the script. However, the ISO 15924 script code provided by the IANA Language Subtag Registry for the epichoric Etruscan-based alphabets of Ancient Italy ("Ital") fails to satisfy all the descriptive needs of the texts ItAnt deals with, since it merges different writing systems under a single label, regardless the elaborate writing situation of Ancient Italy [44].

Although it is recommended to avoid the use of private subtags,[6] for marking up both the difference between script and language and the various scripts attested in the epigraphy of Ancient Italy, we found it necessary to introduce private subtags, such as "osc-Ital-x-oscetr" for Oscan written in the epichoric Etruscan-based alphabet (the so-called Oscan national alphabet).[7] Each private subtag is also linked through a `@source` attribute to the vocabulary of Ancient Italy scripts that the project is creating and will publicly release on its platform. Nevertheless, the combination language(s)-script(s) recommended in the TEI/EpiDoc guidelines creates not only, as seen above, practical issues for the description of our texts, but also theoretical-linguistic issues, being fundamental in language science a clear distinction between a language and the system(s) used for writing it.[8]

From this perspective, specifying the `<tei:language>` element[9] as, for example, "grc-Latn", for "Ancient Greek written in Latin script", in accordance with the TEI/Epidoc guidelines, would be misleading, since the identification of a language has to be independent of the used script. In order to both overcome this issue and ensure interoperability with other digital corpora, we chose to differentiate the information concerning the script(s) and the language(s).

Consequently, in accordance with the TEI/EpiDoc guidelines, we describe the language(s) only with the `<tei:language>` element, and we describe the script(s), regardless of the language, through a `<tei:rs>` element nested in the `<tei:scriptNote>` element. The information is also provided according to the TEI/EpiDoc guidelines (using the private language code) in order to make our corpus compliant and exchangeable with other existing TEI/EpiDoc datasets. An example:[10]

---

[6] See the RFC 5646 guidelines [34] for more details.

[7] The use of private subtags is generally to be avoided (see RFC 5646 [34] for more details).

[8] The ItAnt project is considering sending a proposal for registration of the specific combinations of language-script to IANA. When our proposal will be accepted, the datasets will be updated accordingly.

[9] The `<language>` element characterises a single language or sublanguage used within a text (see Reference [35]).

[10] For the sake of completeness, the example reports the full URI recorded via the attribute `@ref` of the element `<rs>`. The actual corpus will employ a prefix schema, exploiting the "teidata.pointer" data type and the `<prefixDef>` machinery.

```
<tei:scriptDesc>
    <tei:scriptNote>
        <tei:rs type="writingSystem" subtype="alphabet" ref="https://www.prin-italia-antica.unifi.it/p166.html#
        oscan-etruscan">Oscan national alphabet</tei:rs>
    </tei:scriptNote>
</tei:scriptDesc>

<tei:langUsage>
    <tei:language ident="osc" source="https://iso639-3.sil.org/code/osc">Oscan</tei:language>
    <tei:language ident="osc-Ital-x-oscetr" source="https://www.prin-italia antica.unifi.it/p166.html#
    oscan-etruscan" ana="https://unicode.org/iso15924/iso15924-codes.html">Oscan in Oscan national alphabet
    </tei:language>
</tei:langUsage>
```

The same private code is also used as value of the `@xml:lang` attribute in the text edition (see, for example, `<xml:lang="osc-Ital-x-oscetr"` and not `xml:lang="osc">`).

4.4 Vocabularies and Gazetteers

In the perspective of the best possible data integration, ItAnt takes advantage of concepts coming from widely accredited vocabularies and gazetteers.

Concerning the text support description, for the object type ItAnt uses *The Art & Architecture Thesaurus* (AAT) provided by the J. Paul Getty Trust [20] and, as a supplement with regard to natural supports (e.g., cliffs), the *iDAI.thesauri* provided by the Deutsches Archäologisches Institut [22]. These vocabularies are used by European projects such as the ARIADNEplus archaeological platform [3] and the PARTHENOS initiative for digital humanities data integration [32]. In the perspective of integrating our data also with scientific communities outside the field of epigraphy, we used concepts provided by the Getty AAT thesaurus also for the description of materials and for the specification of the engraving technique. For the type of inscriptions, we took advantage of the vocabulary provided by the EAGLE project [11]. The use of the EAGLE vocabularies, specifically designed for epigraphy, is suggested by the TEI/EpiDoc guidelines [15, 52].

For places and geographical names, ItAnt has included entities from both Pleiades [33] and GeoNames [19] for ancient and modern names, respectively.

No vocabularies have been used for dates and periods, since the inscriptions pertaining to our domain are mostly dated using extremely wide ranges of time spans.

Trismegistos IDs [38] have been used when available to identify our texts, in order to provide an additional strong integration layer to our data.

Bibliographical records are linked through a specific library built up within the project by using Zotero and the unique identifiers it provides [28].

For words and names, as seen in Section 4.2, ItAnt implements a full integration with the computational lexicon also developed by the project.

5 INTEGRATING DATA

For some time now, the scientific community that deals with Digital Humanities has exploited, for the purposes of data integration and interoperability, the descriptive expressiveness of ontologies. The use of such conceptual models allows to express the entities of a domain in a semantically rich encoding and, consequently, allows for the exchanging of data coming from different research areas. This semantic interoperability fosters the circulation of data in an interdisciplinary perspective.

Within the Digital Humanities, the CIDOC CRM [36] is recognised to be the reference ontology and has become, especially at the European level, the de facto standard for many scientific fields. It is now also adopted for the description of data deriving from hard sciences such as chemistry or physics, with regard, for example, to the scientific analysis of archaeological materials. CIDOC CRM is released by the International Council of Museums (ICOM) and has been certified as an ISO standard since 2014 (ISO21127: 2014).

In this framework, it is worth mentioning projects and initiatives such as Papyri.info [31], Trismegistos [38], or Epigraphy.info [14], aiming to develop semantic models for the generation of standardised, integrated, and interoperable information, also defining and developing interoperable archives based on the renowned FAIR principles (as in PARTHENOS [32], ARIADNEplus [3], or FAIR Epigraphy AHRC-DFG [16] projects).

The ItAnt project fits into this perspective. The possibility of a data exchange arising from the use of shared thesauri and gazetteers and the encoding of the inscription according to the TEI/EpiDoc standard is further fostered by the

use of CIDOC CRM and its extensions, in particular, CRMtex [7], CRMinf [6], and FRBRoo/LRMoo [18]. In fact, although TEI/EpiDoc provides a series of tags for labelling specific elements, introducing some semantic level in the encoding of the text (e.g., tags identifying temporal entities and actors), it does not offer the same descriptive features typical of an ontology [51]. The use of ontologies also facilitates the alignment of both customised and non-standard uses of the TEI/EpiDoc scheme arisen from specific needs of the treated dataset: the implementation of the ontological approach is thus able to overcome the differences typically occurring in data encoded in XML due to the use of customised schemas.

ItAnt specifically uses CRMtex [7], the CIDOC CRM extension designed to investigate the textual entities from antiquity, to encode textual data, by mapping the TEI/EpiDoc entities onto those of the CRMtex ontology, which is being used by ARIADNEplus [3] for the description of the epigraphical data aggregated by this project.

ItAnt is also investigating the use of CRMinf [6] to code scientific interpretations in semantic format, in order to encode and integrate data about the hermeneutical positions of the individual editors. CRMinf also allows the linking between a text and the related bibliography. Bibliographical data, codified in TEI according to the TEI/EpiDoc guidelines, are being encoded by mapping the TEI schema to FRBRoo/LRMoo [18], the conceptual model(s) developed by the International Federation of Library Associations and Institutions (IFLA) and harmonised with CIDOC CRM to provide a semi-formal representation of bibliographical information. An example:

```
<https://www.prin-italia-antica.unifi.it/7622302d-d71b-492e-bb8b-c2ca0e3482c3>
        a                       frbr:F3_Manifestation_Product_Type ;
        rdfs:label              "Product Le tabellae defixionum osche" ;
        frbr:R41i_is_rep_manifestation_product_type_for
<https://www.prin-italia-antica.unifi.it/1088a776-94d7-416a-8587-5dbf69df0541> ,
<https://www.prin-italia-antica.unifi.it/f5853d83-7b8c-4def-b075-9343fadc7672> ,
<https://www.prin-italia-antica.unifi.it/0727db6c-2894-4fc2-a07b-52e3d1693e5f> ,
<https://www.prin-italia-antica.unifi.it/972f58ec-4def-451e-9365-eb1e0c2debe7> ;

crm:P102_has_title    <https://www.prin-italia-antica.unifi.it/fdd49c7c-3f2e-4fc9-ad55-d9e4cd171509> ;
crm:P1_is_identified_by <https://www.prin-italia-antica.unifi.it/2a180b32-6076-4a52-9f82-6a8e207b4ad3> ,
<https://www.prin-italia-antica.unifi.it/1afff7f6-6087-4b27-84d4-4417ae1645fe> ,
<https://www.prin-italia-antica.unifi.it/7c1c1fed-9abb-4623-aff6-ec69334f34f5> ;

crm:P2_has_type      <https://www.prin-italia-antica.unifi.it/287b8a6e-1e3c-473f-b32e-0b4a72db2458> ;
crm:P94i_was_created_by <https://www.prin-italia-antica.unifi.it/83a09c1c-eb36-4d64-bea8-752138f0b075> ,
<https://www.prin-italia-antica.unifi.it/63a005e2-918f-4433-b503-478d1b7529f0> ,
<https://www.prin-italia-antica.unifi.it/e5a3db9f-a4a0-4014-bbe3-969626bc890d> .
```

The encoding of the ItAnt data by means of the ontological models of the CIDOC CRM family has shown an interesting potential. In future works, we will address this issue in more detail and provide concrete examples of encoding and implementation, including mapping proposals.

6 CONCLUSIONS

The ItAnt project aims at producing a series of scientific studies, updated with respect to the available data, and a series of digital tools devoted to facilitating scholarly research and its sharing, thereby also allowing the exploitation of the linguistic and cultural heritage of Ancient Italy. In its general thrust, the project shares the approach of many European projects that are at present under way, the purpose of which is to experiment with digital and semantic technologies for the study and encoding of the information produced in the domain of Cultural Heritage. The use of the TEI/EpiDoc schema and of ontological tools makes the information modelled by ItAnt perfectly compatible and fully integrable with datasets from different research domains and produced in the context of various international initiatives.

At the end of the project, in order to share the results and valorise this segment of the cultural heritage ItAnt is dealing with, we will release the results in LOD format, enabling integration of our data with information coming from other archives and also with those belonging to fields other than linguistics.

The results of ItAnt will also be made available via the CLARIN European research infrastructure for e-Humanities and Cultural Heritage and will be shared with the ARIADNEplus project.

REFERENCES

[1]A database of Middle Turkic. Retrieved from https://middleturkic.lingfil.uu.se/

[2]Inscriptions of the Northern Black Sea. Retrieved from https://iospe.kcl.ac.uk/index.html

[3]ARIADNEplus Project. Retrieved from https://ariadne-infrastructure.eu

[4]Beta maṣāḥəft. Manuscripts of Ethiopia and Eritrea. Retrieved from https://www.betamasaheft.uni-amburg.de

[5]CLARIN-ERIC. Retrieved from https://www.clarin.eu/

[6]CRMinf. Argumentation Model. Retrieved from https://www.cidoc-crm.org/crminf/

[7]CRMtex. Model for the Study of Ancient Texts. Retrieved from https://www.cidoc-crm.org/crmtex/home-8

[8]Crossreads. Text, materiality, and multiculturalism at the crossroads of the ancient Mediterranean. Retrieved from https://www.csad.ox.ac.uk/crossreads-textmateriality-and-multiculturalism-at-the-crossroads-of-the-ancient-mediterranean

[9]Digital Archive for the Study of pre-Islamic Arabian Inscriptions. Retrieved from http://dasi.cnr.it/

[10]The Domestication of "Hindu" Asceticism and the Religious Making of South and Southeast Asia. Retrieved from https://dharma.hypotheses.org/project

[11]Eagle Vocabularies. Type of inscription. Retrieved from https://www.eagle-network.eu/resources/vocabularies/typeins/

[12]EpiDoc guidelines. Languages and Scripts. Retrieved from https://epidoc.stoa.org/gl/latest/supp-language.html

[13]EpiDoc guidelines. Provenance. Retrieved from https://epidoc.stoa.org/gl/latest/ref-provenance.html

[14]Epigraphy.info. Retrieved from http://epigraphy.info

[15]Europeana Network of Ancient Greek and Latin Epigraphy. Retrieved from https://www.eagle-network.eu/

[16]FAIR Epigraphy AHRC-DFG. Retrieved from https://www.csad.ox.ac.uk/fair-epigraphy

[17]FAIR Principles. Retrieved from https://www.go-fair.org/fair-principles

[18]FRBRoo. Functional Requirements for Bibliographic Records. Retrieved from https://cidoc-crm.org/frbroo/

[19]GeoNames. Retrieved from https://www.geonames.org

[20]Getty Thesaurus of Geographic Names Online. Retrieved from http://www.getty.edu/research/tools/vocabularies/tgn/index.html

[21]IANA Language Subtag Registry. Retrieved from https://www.iana.org/assignments/language-subtag-registry/.

[22]iDAI.thesauri. Retrieved from http://thesauri.dainst.org/en.html

[23]ILC4CLARIN repository. Retrieved from https://dspace-clarin-it.ilc.cnr.it/repository

[24]Ancient Inscriptions from Israel/Palestine. https://doi.org/10.26300/pz1d-st89

[25]Inscriptions of Roman Tripolitania. Retrieved from https://irt2021.inslib.kcl.ac.uk/en/

[26]I.Sicily. Retrieved from http://sicily.classics.ox.ac.uk

[27]ItAnt GitHub. Retrieved from https://github.com/DigItAnt

[28]ItAnt Zotero Library. Retrieved from https://www.zotero.org/groups/2552746/prin_italia_antica/library

[29]Languages and Cultures of Ancient Italy. Historical Linguistics and Digital Models. Retrieved from https://www.prin-italia-antica.unifi.it

[30]Ogham in 3D. Retrieved from https://ogham.celt.dias.ie/menu.php?lang=en

[31]Papyri.info. Retrieved from https://papyri.info/

[32]Parthenos Project. Retrieved from https://www.parthenos-project.eu

[33]Pleiades. Retrieved from http://pleiades.stoa.org/

[34]RFC 5646: Tags for Identifying Languages. Retrieved from https://www.rfc-editor.org/info/rfc5646

[35]Text Encoding Initiative. Language. Retrieved from https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-language.html

[36]The CIDOC Conceptual Reference Model (CRM). Retrieved from https://cidoc-crm.org

[37]TITULUS - Corpus des inscriptions de la France médiévale. Retrieved from https://titulus.huma-num.fr/exist/apps/titulus/index.html

[38]Trismegistos. Retrieved from http://www.trismegistos.org

[39]Avanzini Alessandra, Santis Annamaria De, and Rossi Irene. 2018. Encoding, interoperability, lexicography: digital epigraphy through the lens of DASI experience. In Crossing Experiences in Digital Epigraphy, Santis Annamaria De and Rossi Irene (Eds.). De Gruyter, Berlin - Boston, 1–18. https://doi.org/10.1515/9783110607208-002

[40]Bauman Syd. 2011. Interchange vs. Interoperability. In Proceedings of Balisage: The Markup Conference 2011 (Montréal, Canada, August 2-5, 2011). Number 7 in Balisage Series on Markup Technologies. Retrieved from http://www.balisage.net/Proceedings/vol7/html/Bauman01/BalisageVol7-Bauman01.html

[41]Bausi Alessandro and Liuzzo Pietro M.. 2018. Inscriptions from Ethiopia. Encoding Inscriptions in Beta Maṣāḥǝft. In Crossing Experiences in Digital Epigraphy. Santis Annamaria De and Rossi Irene (Eds.). De Gruyter, 84–92. https://doi.org/10.1515/9783110607208-007

[42]Bodard Gabriel. 2010. EpiDoc: Epigraphic documents in XML for publication and interchange. In Latin on Stone: Epigraphic Research and Electronic Archives. Feraudi-Gruénais Francisca (Ed.). Lexington Books, Lanham (MD), 101–118.

[43]Casarosa Vittore, Ruiz Paolo Manghi, Andrea Mannocci, Eydel Rivero, and Zoppi Franco. 2014. A conceptual model for inscriptions: harmonizing digital epigraphy data sources. In Information Technologies for Epigraphy and Cultural Heritage. Proceedings of the First EAGLE International Conference. Orlandi Silvia, Santucci Raffaella, Casarosa Vittore, and Liuzzo Pietro Maria (Eds.). Sapienza Università Editrice, Roma, 23–40.

[44]Consortium The Unicode. 2021. Chapter 8. Europe-II. Ancient and Other Scripts. The Unicode Consortium, Mountain View, CA, 341–364. Retrieved from https://www.unicode.org/versions/Unicode14.0.0/ch08.pdf

[45]Doerr Martin, Kritsotaki Athina, and Boutsika Katerina. 2011. Factual argumentation. A core model for assertions making. Journal on Computing and Cultural Heritage 3, 3 (2011), 1–34.

[46]Elliott Tom, Bodard Gabriel, Mylonas Elli, Stoyanova Simona, Tupman Charlotte, and Vanderbilt Scott. EpiDoc guidelines. Retrieved from http://www.stoa.org/epidoc/gl/latest/

[47]Felicetti Achille and Murano Francesca. 2021. Ce qui est écrit et ce qui est parlé. CRMtex for modelling textual entities on the semantic web. Semantic Web 12, 2 (Jan.2021), 169–180. https://doi.org/10.3233/SW-200418

[48]Jong Franciska de, Maegaard Bente, Fišer Koenraad De Smedt, Darja, and Uytvanck Dieter Van. 2018. CLARIN: Towards FAIR and responsible data science using language resources. In Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018). Calzolari Nicoletta (Ed.). European Language Resources Association, 3259–3264. Retrieved from https://aclanthology.org/L18-1515.pdf

[49]Krummrey Hans and Panciera Silvio. 1980. Criteri di edizione e segni diacritici. Tituli 2 (1980), 205–215.

[50]Lin Dawei, Crabtree Jonathan, Dillo Ingrid, Downs Robert R., Edmunds Rorie, Giaretta David, Giusti Marisa De, L'Hours Hervé, Hugo Wim, Jenkyns Reyna, Khodiyar Varsha, Martone Maryann E., Mokrane Mustapha, Navale Vivek, Petters Jonathan, Sierman Barbara, Sokolova Dina V., Stockhause Martina, and Westbrook John. 2020. The TRUST principles for digital repositories. Scientific Data 7, 1 (May2020), 144.

[51]Liuzzo Pietro M.. 2017. Mapping epigraphic databases to EpiDoc. In Digital and Traditional Epigraphy in Context. Orlandi Silvia, Santucci Raffaella, Mambrini Francesco, and Liuzzo Pietro Maria (Eds.). Sapienza Università Editrice, Roma, 187–199. Retrieved from http://www.editricesapienza.it/node/7610

[52]Liuzzo Pietro Maria and Evangelisti Silvia. 2021. Modeling execution techniques of inscriptions. Semantic Web 12, 2 (Jan.2021), 181–190. https://doi.org/10.3233/SW-200395

[53]Marinetti Anna. 2020. Venetico. Palaeohispanica. Revista sobre Lenguas y Culturas de la Hispania Antigua20 (May2020), 367–401. https://doi.org/10.36707/palaeohispanica.v0i20.374

[54]Passarotti Marco and Mambrini Francesco. 2021. Linking Latin: Interoperable Lexical Resources in the LiLa Project. In Building New Resources for Historical Linguistics. Biagetti Erica, Zanchi Chiara, and Luraghi Silvia (Eds.). Pavia University Press, Pavia, 103–124.

[55]Pitts Reuben J.. 2022. Corpus of the epigraphy of the Italian peninsula in the 1st millennium BCE (CEIPoM). Journal of Open Humanities Data 8 (Jan.2022), 1. https://doi.org/10.5334/johd.65

[56]Poccetti Paolo. 2017. The documentation of Italic. In Handbook of Comparative and Historical Indo-European Linguistics. Klein Jared, Joseph Brian, and Fritz Matthias (Eds.). Handbücher zur Sprach- und Kommunikationswissenschaft = Handbooks of linguistics and communication science, Vol. 2. De Gruyter Mouton, Berlin - Boston, 733–742. https://www.degruyter.com/document/doi/10.1515/9783110523874-001/html

[57]Poccetti Paolo. 2020. Lingue sabelliche. Palaeohispanica. Revista sobre Lenguas y Culturas de la Hispania Antigua20 (May2020), 403–494. https://doi.org/10.36707/palaeohispanica.v0i20.399

[58]Prag Jonathan and Chartrand James. 2018. I. Sicily: Building a Digital Corpus of the Inscriptions of Ancient Sicily. In Crossing Experiences in Digital Epigraphy. Santis Annamaria De and Rossi Irene (Eds.). De Gruyter, 240–252. https://doi.org/10.1515/9783110607208-020

[59]Quochi Valeria, Bellandi Andrea, Khan Fahad, Mallia Michele, Murano Francesca, Piccini Silvia, Rigobianco Luca, Tommasi Alessandro, and Zavattari Cesare. 2022. From inscriptions to lexicon and back. A platform for editing and linking the languages of ancient Italy. In LREC 2022 Workshop Language Resources and Evaluation Conference. Second Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2022). Proceedings. European Language Resources Association, Marseille, 59–67. Retrieved from http://www.lrec-conf.org/proceedings/lrec2022/workshops/LT4HALA/2022.lt4hala2022-1.0.pdf#page=59

[60]Rigobianco Luca. 2020. Falisco. Palaeohispanica. Revista sobre Lenguas y Culturas de la Hispania Antigua 20 (2020), 299–333. https: //doi.org/10.36707/palaeohispanica.v0i20.373

[61]Riva Pat and Žumer Maja. 2018. FRBRoo, the IFLA library reference model, and now LRMoo: A circle of development. In Transform Libraries, Transform Societies. Kuala Lumpur, Malaysia. Retrieved from http://library.ifla.org/id/eprint/2130

[62]Sarullo Giulia. 2016. The encoding challenge of the ILA project. In Off the Beaten Track. Epigraphy at the Borders. Felle Antonio E. and Rocco Anita (Eds.), Archaeopress, 15–27. Retrieved from https://www.archaeopress.com/Archaeopress/download/9781784913229.pdf#page=25

[63]Stifter David. 2020. Cisalpine Celtic. Palaeohispanica. Revista sobre Lenguas y Culturas de la Hispania Antigua20 (May2020), 335–365. https://doi.org/10.36707/palaeohispanica.v0i20.375

[64]Vagionakis Irene. 2021. Cretan institutional inscriptions. A new EpiDoc database. Journal of the Text Encoding Initiative Rolling Issue (Sept. 2021), 1–21. https://doi.org/10.4000/jtei.3570

[65]Vagionakis Irene, Gratta Riccardo Del, Boschetti Federico, Baroni Paola, Grosso Angelo Mario Del, Mancinelli Tiziana, and Monachini Monica. 2022. 'Cretan institutional inscriptions' meets CLARIN-IT. In Selected Papers from the CLARIN Annual Conference 2021, Monica Monachini and Maria Eskevich (Eds.). 139–150. https://doi.org/10.3384/ecp18913

[66]Wilcken Ulrich. 1932. Das Leydener Klammersystem. Archiv für Papyrusforschung 10 (1932), 211–212.