Structured Metadata for the Digital Corpora of Ancient Epigraphic Monuments from Bulgaria

Dimitar Iliev
Sofia University St. Kliment Ohridski

Abstract: *The paper presents the main features of the EpiDoc-compliant XML template according to which ancient inscriptions in Greek and Latin from the lands of today's Bulgaria are encoded and then indexed and displayed in the browsable and searchable databases TELAMON and TITVLI*

Keywords: Digital Epigraphy, EpiDoc, metadata, schema, encoding
Ключови думи: дигитална епиграфика, EpiDoc, метаданни, схема, кодиране

Dimitar Iliev is an Assistant Professor at the Department of Classics to the University of Sofia. He holds a PhD in Greek Linguistics and MA in Computational Linguistics. His scholarly interests are in the fields of Greek poetry, Greek and Latin linguistics, Late Antiquity, Digital Humanities.
Email: diiliev@uni-sofia.bg
ORCID: 0000-0002-5231-818X

1. INTRODUCTION: DIGITAL EPIGRAPHY AND EPIDOC

The "digital turn" in the Humanities in the last several decades, with all its achievements and peculiar challenges[1] could not exclude the study of the past, together with its rich and multifaceted documentary heritage. This heritage encompasses different physical objects which often also bear symbolic content in the form of images and/or text: coins, seals, inscriptions, papyri, ostraca, manuscripts, stamps, etc. Such items, known by the common term of "text-bearing objects", provide valuable information by the means of their materiality as well as of their textual content, and both these aspects of them need to be equally described and studied, together with the archaeological context in which they were discovered and their fate after the discovery[2]. Among the prominent text-bearing objects whose research has always been a significant part of *Altertumswissenschaft* are the epigraphic monuments. The past two centuries have seen the development of epigraphy as a scholarly discipline, the publication of thousands of Greek and Latin inscriptions and

[1] See *Viola* 2023: 1-37.
[2] *Tsouparopoulou* 2016.

the appearance of large corpora such as *CIL* and *IG*, new volumes of which continue to appear to this day. With the dawn of the new millennium, the methodology of the digital description and publication of epigraphic monuments was developed. It is based on the principles and standards of application of mark-up languages (SMGL at first and later XML) established by the Text Encoding Initiative, or TEI[3]. On the basis of TEI XML, with its vast variety of elements, attributes and values, a subset was elaborated which was specifically designed for the purposes of the electronic publication of inscriptional corpora previously published in the traditional analogue way[4]. It was later applied to born-digital collections of other historical documents such as wooden tablets[5] and papyri[6]. The EpiDoc guidelines[7], together with additional documentation and tools (e.g. transformation stylesheets and a regularly updated RNG schema)[8], not only allow the detailed edition of a monument's text and the description of its metadata in a simple, yet powerful and interoperable way. They also enable the exportation of the encoded data into various publication formats, online or offline, electronic or printed, aimed at different audiences[9].

In the 2010's, many digital collections of Greek and Latin inscriptions applying the EpiDoc schema went online: the *Inscriptions of the Northern Black Sea*[10], the *Inscriptions of Roman Cyrenaica*[11], the *Dodona Online* collection of oracle lamellae[12], etc. Some of these corpora are still regularly updated with new content. Gradually, EpiDoc came to be applied to epigraphic material written in other languages and belonging to cultures other than the ancient Graeco-Roman world[13]. Being a TEI subset, EpiDoc is also undergoing its own customizations for the digital description and publication of other historical text-bearing objects. An international team of Byzantine scholars created SigiDoc[14] for Byzantine lead seals[15]. Currently, the data from the various EpiDoc projects related to Classical antiquity need to be mapped against each other so that meta-queries can be made through as many of them as possible. Similar issues are addressed by projects such as FAIR Epigraphy[16].

2. BULGARIA'S EPIGRAPHIC HERITAGE

The land of today's Bulgaria has a rich and diverse epigraphic heritage. The ancient Greek inscriptions date from a time span of approximately 12 centuries, from 6th c. BCE to 6th c. CE. More than 5,000 in number, the largest part of them is published in two big corpora, edited by Georgi Mihailov[17] and by Veselin Beševliev[18]. Many monuments, however, are either left outside the big collections, often scattered in inaccessible publications, or in need of serious revisions even when they are known and quoted. The current state of the study of Latin inscriptions from Bulgaria is even more complicated:

[3] See *The TEI Consortiu* 2024.
[4] Such as *Reynolds, Roueché, Bodard* 2007.
[5] *The Vindolanda Tablets Online* at http://vindolanda.csad.ox.ac.uk/ (accessed 23.07.2024).
[6] See the largest and the most exhaustive meta-collection of documentary and literary papyri currently available online at https://papyri.info/ (accessed 23.07.2024).
[7] *Eliott, Bodard, Mylonas, Stoyanova, Tupman, Vanderbilt* 2007-2022.
[8] See https://epidoc.stoa.org/.
[9] *Flanders, Roueché* 2006.
[10] https://iospe.kcl.ac.uk/index.html.
[11] https://ircyr2020.inslib.kcl.ac.uk/en/.
[12] https://dodonaonline.com/.
[13] See, e.g., *Forschungsplattform für jüdische Grabsteinepigraphik* (http://www.steinheim-institut.de/cgi-bin/epidat), *Die Inschriften des deutschen Sprachraumes in Mittelalter und Früher Neuzeit* (https://www.inschriften.net/), the *Corpus of Pyu Inscriptions* (http://hisoma.huma-num.fr/exist/apps/pyu/index2.html) and many others.
[14] *Sopracasa, Filosa*, 2020.
[15] The SigDoc v.1.1. Guidelines are published here: http://sigidoc.huma-num.fr/ . A sample corpus can be seen here: https://sigidoc.raketadesign.com/en/ .
[16] More about the project and its demo meta-platform can be seen here: https://inscriptiones.org/. See also *Heřmánková, Horster, Prag* 2024.
[17] *Mihailov* 1956-1995.
[18] *Beševliev* 1964.

there is no comprehensive publication meeting modern scholarly standards. The number of the Latin inscriptions is comparable to that of the Greek inscriptions, although they belong to a much shorter period, from the beginning of the 1st to the end of the 6th c. CE. Only a small fraction of them was edited by Boris Gerov and published posthumously[19] as an intended first part of a large corpus which never materialized. Some inscriptions are included only in rather old editions with multiple errors and incomplete data, while many others still remain dispersed, obscure, or unknown to the scholarly public.

3. THE BULGARIAN EPIDOC COLLECTIONS

The *Telamon* project[20] initiated by the Department of Classics to St. Kliment Ohridski University of Sofia aims at creating a digital library of the ancient Greek inscriptions found in Bulgaria using an EpiDoc-compliant template according to which inscriptions from the rich ancient Greek epigraphic heritage in Bulgaria would be encoded. The .xml documents of the separate monuments contain the text of the inscription itself (diplomatic as well as editorial), together with additional commentaries, bibliography, and metadata concerning both the content of the text and the features of its physical carrier. Most of the inscriptions included in the collection are also contained in one of the two large corpora mentioned above, or are presented in other publications which are not part of *IGBulg* or *ILBulg*. One of the aims of the project is to revise the previous editions of the inscriptions, check once again the monuments *ex autopsia*, be it in the museum repositories or elsewhere, and to examine and present them in the light of the newest discoveries and publications. Sometimes, the inaccessibility of either some of the publications[21] or the monuments themselves could be a challenge. For the first time, these monuments are now being collected into a bilingual digital corpus with translations added. Their revisions, corrections, the publications of new inscriptions and also the commentaries are mainly the result of the research activities of Dr. Nicolay Sharankov. Some of his research on the topic has already seen the light of the day as articles and monographs[22], some of it is born-digital and appears, for the first time, in his notes and comments on the inscriptions published as a part of the *Telamon* collection.

Apart from the up-to-date and accessible content, another contribution of the *Telamon* initiative for the development of Digital Epigraphy in Bulgaria and beyond is the indexing and visualization tool for the .xml files developed especially for the aims of the project. AIAX[23] is a CMS and a front-end tool used to produce the output of the separate inscriptions as well as the indices based on the internal XML authority files of the collection[24]. Initially designed to work with the custom EpiDoc-compliant template used for the purposes of *Telamon*, the tool can be adjusted for other purposes. Recently, it was applied to Latin monuments in the framework of a pilot collection of Latin inscriptions from Bulgaria created by the National Archaeological Institute with Museum to the Bulgarian Academy of Sciences under the name of *Tituli*[25]. Both *Telamon* and *Tituli* use the same basic template which applies the same description structure for the metadata, the text of the monument, its translations in Bulgarian and English, the apparatus and the commentaries to it. We

[19] *Gerov* 1989.

[20] https://telamon.uni-sofia.bg/.

[21] Examples of such works are *Dimitrov* 1931, used as a secondary source for the electronic publication of *IGBulg* 727 (https://telamon.uni-sofia.bg/en/epi/view_ins/IGBulg_0727) and *Botusharova* 1959, used as a secondary source for the electronic publication of *IGBulg* 1460 (https://telamon.uni-sofia.bg/en/epi/view_ins/IGBulg_1460). Not only are such works not very accessible in terms of availability, but some of them require knowledge of Bulgarian which makes them virtually unknown to the international scholarly public. The bilingual output of the *Telamon* platform (Bulgarian-English) allows for the dissemination of Bulgarian research hitherto unknown to international scholars.

[22] See e.g. *Sharankov* 2016; *Sharankov* 2024.

[23] Available for download and installation as a desktop and server package at: https://telamon.uni-sofia.bg/en/page/project.

[24] See more details about AIAX and how it stores and processes data in *Iliev, Boeva* 2023.

[25] The demo version of the web platform can be seen at: https://tituli.epistone.net/.

shall now proceed to examine the basic structure of the EpiDoc .xml file used across these digital projects and, in particular, the part of it containing the metadata.

4. THE MAIN PARTS OF THE DOCUMENT

XML is a mark-up language derived from SGML and similar to HTML[26]. It basically consists of strings of metatext enclosed in triangular brackets giving additional information about the strings of text they encompass in the following way:

<tag> text </tag>

The opening and the closing tags should be identical except for the / symbol at the beginning of the latter. The unity of an opening and a closing tag together with the content enclosed between them forms an XML element[27]:

<title>Two stories by Edgar Allen Poe: electronic version</title>[28]

Most of the elements can have daughter elements, or sub-elements, wholly nested within them and forming with them a hierarchical tree-like structure:

```
<persName>
    <forename>Franklin</forename>
    <forename>Delano</forename>
    <surname>Roosevelt</surname>
</persName>
```
[29]

Within the opening tag of an element, different attributes with their values can be added for supplementary information:

<name type="city">Glasgow</name>[30]

Thus, TEI XML (and its subsets such as EpiDoc) provides a detailed and flexible mechanism to formally describe all the peculiarities of a text or an object for the purposes of scholarly research and publication. This description can be stored in the lightweight .xml format and then exported for user-end visualization as a web page, a PDF file, etc.

Like all the TEI-based documents, the EpiDoc-compliant .xml template used for the Bulgarian epigraphic collections uses the *<TEI>* root element within which three sibling sub-elements are nested for the three main parts of the document:

```
<TEI>
   <teiHeader>...</teiHeader>
     ...
   <facsimile>...</facsimile>
     ...
   <text>...</text>
</TEI>
```

Of these, the *<teiHeader>* element contains all the metadata describing the text and the text-bearing object. The *<facsimile>* element has the shortest content of the three and serves to link to the image of the monument which is usually stored locally on the web site's server:

```
<facsimile>
    <graphic url="1069.jpg"/>
</facsimile>
```
[31]

The *<text>* element contains not only the text of the inscription itself, but also paratexts such as translation(s), commentary, critical apparatus, bibliography:

```
<text>
    <div type="edition">...</div>
    <div type="apparatus">...</div>
    <div type="commentary">...</div>
    <div type="bibliography">...</div>
</text>
```

It is the first of these main components, the *<teiHeader>*, which will be the object of detailed presentation in the next pages.

---

[26] Concerning the history and the basic features of XML, see *Yott* 2015.

[27] With the exception of the so-called "empty elements" which we will not discuss in detail.

[28] Example taken from *TEI P5 Guidelines* "2.2.1. The Title Statement" [Last modified 2024-07-08] https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html (accessed 20.07.2024).

[29] Example taken from *TEI P5 Guidelines* "14.1.2. Personal Names" [Last modified 2024-07-08] https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html (accessed 20.07.2024).

[30] Example taken from *TEI P5 Guidelines* "14.2.2. Organizational Names" [Last modified 2024-07-08] https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html (accessed 20.07.2024).

[31] This is an example of an "empty" element which has no textual content and thus lacks a closing tag.

5. THE STRUCTURE OF THE METADATA IN THE BULGARIAN EPIDOC-COMPLIANT TEMPLATE

The *<teiHeader>* element containing all the metadata about the digital publication and its analogue source consists of three main sibling sub-elements:

<fileDesc>**...**</fileDesc>
<encodingDesc>**...**</encodingDesc>
<revisionDesc>**...**</revisionDesc>

Of these, *<fileDesc>* is where almost all of the principal metadata of the encoded file (and its source) are stored. The *<encodingDesc>* element may contain additional information about the methodological particularities of the encoding, if such information is relevant and needed. Its content may not be displayed in the user-end view of the encoded monument. But, if the raw .xml file is made available somewhere (which is a good practice generally recommended in Digital Humanities[32]), users, researchers, encoders, etc. may draw valuable workflow examples from such data. The same goes for the *<revisionDesc>* which contains information about the team members having worked on the file. It may look as follows:

<revisionDesc>
  ***<change** when="2022-07-20" who="E.B.">*
  encoded, added links to authority
  </change>
  ***<change** when="2022-07-23" who="N.Sh.">*
  corrected, added description
  and Bulgarian translation
  </change>
</revisionDesc>[33]

This is an information that the project team might or might not choose to display as a part of the front-end publication of the inscription. However, if contained at least in the raw .xml file, it gives its creators (usually the work on one .xml publication is collaborative) authority and responsibility akin to those of the traditional editors of ancient inscriptions in the printed publications. And such contributions to the digital publishing of epigraphic monuments should receive acknowledgement equal to the authorship of paper editions: a step which is still a *desideratum* in the framework of national and international research assessment policies[34].

Of these first-level sub-elements, *<fileDesc>* is the only one that contains daughter elements in an at least two-level-deep further hierarchy. The first level consists of the following sub-elements:

<titleStmt>**...**</titleStmt>
<publicationStmt>**...**</publicatiionStmt>
<sourceDesc>**...**</sourceDesc>

The title statement contains the title of the online publication displayed at the head of the web page of the inscription. It also contains the names of the main scholarly editor of the online publication, as well as those of all the previous scholars whose editorial decisions and observations have been taken into account in the digital publication (also notice the bilingual rendition via the duplicated elements with the *@xml:lang* attribute):

<titleStmt>
  ***<title** xml:lang="bg">*Почетен декрет
  <title** xml:lang="en">*Honorary decree for Akornion</title>***
  <editor>
    <persName** xml: lang="bg">*Василий Латышев</persName>***
    <persName** xml:lang="en">*Vasiliy Latyshev</persName>***
    <persName** xml:lang="bg">*Ернст Калинка</persName>***
    <persName** xml:lang="en">*Ernst Kalinka</persName>***
    <persName** xml:lang="bg">*Георги Михаилов</persName>***
    <persName** xml:lang="en">*Georgi Mihailov</persName>***
    <persName** xml:lang="bg">*Николай Шаранков</persName>***

[32] The *Telamon* collection provides the downloadable .xml files with the respective raw data within the display page where each inscription is published. Otherwise, whole project datasets may be published in Github repositories, see e.g. the repository of the EpiDoc initiative here: https://github.com/epidoc.
[33] Source code taken from *Tlmn3_0001.xml*, front-end publication at: https://telamon.uni-sofia.bg/en/epi/view_ins/Tlmn3_0001.
[34] On the subject of digital research and its assessment see *Tasovac, Romary, Tyth-Czifra, Ackermann, Alves, et al.* 2023, particularly p. 5-6.

```
        <persName xml:lang="en">Nicolay
        Sharankov</persName>
    </editor>
</titleStmt>
```
[35]

The <publicationStmt> element is also related to the electronic publication of the inscription rather than to the printed publication(s) thereof which are described in the bibliography to the text (*TEI/text/div type="bibliography"*). In it, the project responsible for the digital edition of the inscription is indicated, as well as the ID number of the inscription in the digital collection.

<sourceDesc> contains the description of the physical source of the inscriptional text, its material medium. As such, it has all the further levels of hierarchy embedded within it. Its main daughter element is <msDesc> which stands for "manuscript description". Although the physical bearer of an epigraphic text is different than a parchment codex or another type of manuscript, the name of the element is inherited in EpiDoc from the superset of TEI which was initially designed mainly with the digital edition of Mediaeval codices in mind. <msDesc>, in turn, contains the following daughter elements:

```
<msIdentifier>...<msIdentifier>
<msContents>...</msContents>
<physDesc>...</physDesc>
<history>...</history>
```

<msIdentifier> contains the inventory number of monument in a museum repository, if the inscription is listed or stored in such. In case a museum is indicated, a link is also given to its official website in order to give it credit as a collaborating institution (since usually providing an up-to-date inventory number of an item requires the cooperation of the museum's employees):

```
<repository>
    <ref target="http://naim.bg/">
    Национален археологически
    институт с музей към БАН
    </ref>
</repository>
```
[36]

<msContents> classifies the inscription by topic linking it to a *document-type.xml* internal authority list where 23 different categories of inscriptions are described. The <physDesc> and the <history> elements contain information about the monument's physical characteristics and history. Of these, within the former the following daughter elements are embedded:

```
<objectDesc>...</objectDesc>
<handDesc>...</handDesc>
<decoDesc>...</decoDesc>
```

The <handDesc> element is where the palaeographic description of the script and the hand(s) is indicated. <decoDesc> is dedicated to the representation of the monument's decorative elements, reliefs, etc., if such are available. The rest of the physical features of the monument, however, are described by sub-elements embedded withing the <objectDesc> element. They are as follows:

```
<supportDesc>...</supportDesc>
<layoutDesc>...</layoutDesc>
```

The inscription's layout, i.e. the exact position of the text on the surface of the monument in relation to its other material constituents is handled by the <layoutDesc> element. All else is covered by the <supportDesc> element containing the following sub-elements:

```
<support>...</support>
<material>...</material>
```

<material> describes the material of which the monument is made, most often different types of stone such as marble, limestone, etc. that are also linked to an internal *material.xml* authority list. As for the <support> element, it contains information about the monument's physical type (<objectType> daughter element linked to an *object-type.xml* authority list describing more than 15 different types of monuments such as altar, slab, statue base, etc.) as well as about its dimensions:

[35] Source code taken from *IGBulg_0013.xml*, front-end publication at: https://telamon.uni-sofia.bg/en/epi/view_ins/IGBulg_0013.

[36] Source code taken from *TTL2_003.xml*, front-end publication at: https://tituli.epistone.net/epi/view_ins/TTL2_0003.

<dimensions>
<height quantity="66">66</height>
<width quantity="51">51</width>
<depth quantity="10">10</depth>
</dimensions>[37]

The <history> element is related to the physical monument's history and has three main daughter elements embedded within it:

<origin>...</origin>
<provenance type="found">...</provenance>
<provenance type="observed">...</provenance>

Similar as they may be in their names, they refer to three different episodes in the monument's history. <origin> indicates the ancient place where the monument was produced and first put on display. It may be linked to a gazetteer or another list of ancient places. <provenance type="found"> indicates the modern findspot of a monument and may refer to a gazetteer or another list of modern places. <provenance type="observed"> can be used in two cases. The first one is if an important previous editor or observer has noted or recorded the inscription in a historical moment between its creation and its current acquisition. The second one is if the monument is currently to be observed in a place outside a museum (in which case it is an alternative to the repository number we discussed above and both can't be indicated at the same time).

In this way, all the metadata connected with a certain epigraphic monument can be described in an exhaustive and flexible way. A database of various inscriptions encoded following this template allows for a rich, dynamic, an informative user experience of the digital collection from the point of view of scholars and the general public alike.

BIBLIOGRAPHY:

Beševliev 1964: *Beševliev*, Veselin. Spätgriechische und spätlateinische Inschriften aus Bulgarien. Berlin.

Botusharova 1959: *Botusharova*, Liliya. Оброчни релефи на Аполон от Тракия [Obrochni relefi na Apolon ot Trakiya]. – Annuaire du Musée national archéologique Plovdiv, 3, 145-153.

Eliott, Bodard, Mylonas, Stoyanova, Tupman, Vanderbilt, 2007-2022: *Elliott*, Tom, Gabriel *Bodard*, Elli *Mylonas*, Simona *Stoyanova*, Charlotte *Tupman*, Scott *Vanderbilt*, et al. (2007-2022). *EpiDoc Guidelines: Ancient documents in TEI XML* (Version 9.6). Available: https://epidoc.stoa.org/gl/latest/ (accessed 23.07.2024).

Dimitrov 1931: *Dimitrov*, Dimitar. Zur Geschichte der Stadt Augusta Traiana. – Bulletin de la Société Historique a Sofia, 11-12, 61-69.

Flanders, Roueché 2006: *Flanders*, Julia, Charlotte *Roueché*. Introduction To Markup For Epigraphers. Online at: https://blog.stoa.org/archives/4084 (accessed 06.04.2024).

Gerov 1989: *Gerov*, Boris. Inscriptiones Latinae in Bulgaria repertae. Serdicae.

Heřmánková, Horster, Prag 2022: *Heřmánková*, Petra, Marietta Horster, Jonathan Prag, Digital Epigraphy in 2022: A Report from the Scoping Survey of the FAIR Epigraphy Project (v1.0.0). https://doi.org/10.5281/zenodo.6610696 (accessed 07.04.2024).

Iliev, Boeva 2023: *Iliev,* Dimitar, Elina *Boeva*. From stone to screen: The Telamon database of ancient inscriptions in Greek from Bulgaria – AIP Conference Proceedings, 2939/1. https://doi.org/10.1063/5.0178753.

Mihailov 1956-1995: Mihailov, Georgius. Inscriptiones Graecae in Bulgaria repertae. Serdicae, Voll. I-V.

Reynolds, Roueché, Bodard 2007: *Reynolds*, Joyce, Charlotte *Roueché*, Gabriel *Bodard*. Inscriptions of Aphrodisias. Online at: http://insaph.kcl.ac.uk/iaph2007 (accessed 06.04.2024).

Sharankov 2016: *Sharankov*, Nicolay. Notes on the Greek Inscriptions from Bulgaria. – *Studia Classica Serdicensia* 5, 305-361.

Sharankov 2024: *Sharankov*, Nicolay. Antichnite nadpisi na Dionisopol. Sofia University Press 2024.

Sopracasa, Filosa 2020: *Sopracasa*, Alessio, Martina *Filosa*. Encoding Byzantine Seals: SigiDoc. In: Atti del IX Convegno Annuale AIUCD. La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica. (eds. Marras, Cristina, Marco Passarotti, Greta Franzini, Eleonora Litta). Milano, 240-245.

Tasovac, Romary, Tyth-Czifra, Ackermann, Alves, et al. 2023: *Tasovac,* Toma, Laurent *Romary,* Erzsébet *Tyth-Czifra,* Rahel C. *Ackermann,* Daniel *Alves, et al.* The Role of Research Infrastructures in the Research Assessment Reform: A DARIAH Position Paper. ffhal-04136772f. Online at: https://hal.science/hal-04136772 (accessed 20.07.2024).

The TEI Consortium 2024: TEI P5: Guidelines for Electronic Text Encoding and Interchange. [Version 4.8.0.]. [Last modified 2024-07-08]. (eds. The TEI Consortium). Online at: http://www.tei-c.org/Guidelines/P5/ (accessed 23.07.2024).

Структурирани метаданни за дигитални корпуси на антични епиграфски паметници от България

Димитър Илиев

Статията представя основните черти на EpiDoc-съвместимия XML шаблон, според който антични гръцки и латински надписи от земите на днешна България се кодират, след което се индексират и показват в онлайн базите данни TELAMON и TITVLI с възможност за разглеждане и търсене.