Challenges and Perspectives in Creating a Digital Epigraphy Collection Using TEI XML and Front-end Services

Dimitar Iliev[1], Elina Boeva[1], Polina Yordanova[2]

[1] St. Kliment Ohridski University of Sofia, Sofia, Bulgaria
[2] University of Helsinki, Helsinki, Finland
diiliev@uni-sofia.bg, eboeva@uni-sofia.bg, po-
lina.yordanova@helsinki.fi

Abstract. The article presents the main tools and methods applied in the creation of the Telamon database of the ancient Greek inscriptions from Bulgaria encoded in TEI XML. The work so far on the project is reported, the modifications to the existing services are enumerated and some future perspectives are discussed.

Keywords: Digital Epigraphy, TEI XML, EpiDoc, Front-end Services.

1   The Telamon Project: a Short History

The Greek inscriptions from the territory of today's Bulgaria are a rich part of our local cultural heritage that is not very accessible even to specialists today and remains virtually unknown to the general public. Currently, their number is above 4,500 for the period of Antiquity alone (approx. 6. C. BCE - mid-3. C. CE). A large part of these monuments has already been published by Georgi Mihailov in his corpus *Inscriptiones Graecae in Bulgaria repertae* (Mihailov, 1956-1995), further referred to as IGBulg, as well as in other articles and monographs[1]. However, many of the previously published inscriptions need update and revision, not to mention the monuments which remain unpublished to this day.

The idea for a digital collection of the ancient Greek inscriptions from our lands was prompted by Prof. Charlotte Roueché of King's College, London, who visited Bulgaria recently after the advent of the online edition of her corpus of Late Antique Inscriptions from Aphrodisias in Caria (Roueché, Reynolds, & Bodard, 2007). The project was initiated by a team at the Department of Classics to the University of Sofia consisting of Prof. Mirena Slavova, Nicolay Sharankov and Dimitar Iliev. Various specialists were also involved in our work at different times. In the current phase of the project which will finally see the online publication of a part of our huge and ever-growing corpus,

---

[1] Among the most important recent monographs much useful for the work on the project is (Sharankov, 2015).

the team consists of Dimitar Iliev, Elina Boeva, Polina Yordanova, seven BA and MA students from the Department, a designer, a developer and a web-hosting assistant. For the last 18 months (February 2018 – August 2019), the project activities aimed at producing a publicly accessible online corpus of Greek inscriptions from the period of Antiquity found in Bulgaria can be summarized as follows:

- Creating a version a TEI XML template valid to the EpiDoc schema (see below, 2.1.) that would suit best the requirements of the particular source material and also the needs of the potential target audience.
- Recruiting and training assistant XML encoders, mainly BA and MA students (see 2.2.).
- Adapting and developing EFES, the existing front-end services for the visualization and indexing of ancient documents written on stone and other durable materials (see below, 2.3.).
- Improving the design of the front-end services.
- Creating and maintaining a GitHub repository of epigraphic content and the customized services for its processing.

These tasks and activities will be the subject of the present paper. Currently, in August 2019, the customized Telamon front-end service tool is being automated and deployed and its functionalities are to be tested on a server. The website in its final form will be publicly launched in September 2019 as a part of the research activities of the National CLaDA-BG Consortium and as a Bulgarian contribution to the DARIAH-EU ERIC. Its web address will be at http://telamon.uni-sofia.bg. During the preparation of the current article, the Telamon collection has not yet been made available as a public website. However, it can already be downloaded and launched locally as a pilot platform for the project's ever-growing GitHub repository (see 2.3).

2 Creating a Digital Corpus of the Ancient Greek Inscriptions in Bulgaria

2.1 The Epidoc Subset of TEI and the Creation of the Telamon Template

EpiDoc XML is a subset of TEI XML specifically aimed at the needs of epigraphers and papyrologists, with its own encoding guidelines and recommendations supported by other resources such as schemas, templates, stylesheets, etc. (Bodard & Stoyanova, 2016, pp. 52-53). Many EpiDoc training events are held every year in different locations around the world, including, since 2013, regular workshops in Bulgaria involving epigraphers and other specialists from the country and the region (Bodard & Stoyanova, 2016, p. 53). The educational materials for such events, a great part of which is openly available online[2] have been translated and localized by the Telamon project team for the purpose of trainings at the University of Sofia such as the ones in May 2018 and in

---

[2] See https://wiki.digitalclassicist.org/EpiDoc_Workshops.

March 2019 conducted in the framework of the CLaDA-BG Consor-tium. An especially important place among the supporting materials for the EpiDoc initiative is occupied by the schema for validation as well as by the template for encoding EpiDoc XML documents (Elliott & Bodard, 2018). Following these, a number of projects in the field of current digital epigraphy are tailored, more important among which are *IOSPE: Ancient Inscriptions of the Northern Black Sea[3]*, *Inscriptions of Roman Tripolitania[4]*, *Inscriptions of Greek Cyrenaica and Greek Verse Inscriptions of Cyrenaica[5]* and others.

Telamon also adapted and developed the EpiDoc TEI template in accordance with the particular nature of the source material and the target audience[6]. In the template, the basic structure of a TEI document is preserved: all the metadata describing the monument, as well as the previous printed and the current digital editions of its text, are nested under the `<teiHeader>` element. Next comes the `<facsimile>` element referencing to an image of the monument in .jpeg format when available (some of the monuments still lack proper accompanying visualization). Then, we have chosen to include under the principal `<text>` element the text of the monument itself (`<div type="edition">`), the relevant editorial comments to it derived from IGBulg, from other printed sources or especially added for the purposes of the digital edition (`<div type="apparatus">`) and also historical or cultural commentary when needed (`<div type="commentary">`). The only piece of metadata not nested under `<teiHeader>` but included as a `<div @part>` under `<text>` is the bibliography, almost always referring to an *IGBulg* volume and sometimes also to (Sharankov, 2015) and other *addenda et corrigenda* to Mihailov's large corpus. For the purposes of creating of bilingual content almost all of the elements are duplicated in English and Bulgarian via either an `@xml:id` attribute with the respective language value or, where the general EpiDoc schema does not allow it, via the insertion of the universal `<seg>` element in the respective place in the code which does allow for multilingual treatment. A special attention is accorded to the classification of the source material by type of object and text category. Both are described in separate authority files - `material.xml` and `document-type.xml` – references to which occur in several places in the template. Proper and unambiguous references to geographical places of origin or discovery of the monuments (which can coincide or differ) are also of great importance for the collection. Such place names are referenced both internally, by their ID's in the `findspot.xml` authority file, and externally, by their URI's in existing gazetteers such as Geonames[7] or Pleiades[8]. The same principle of double reference is applied to the encoding to personal names and historical persons – a task in which onomastic and prosopographical annotation have to be carefully distinguished. The template is also

rich in instructions for the encoders in commentaries accompanying nearly all the elements that require specific attributes or content.

2.2 Learning Epigraphy by Doing XML

Telamon project makes use of the assistance of students enrolled in the BA and MA Programmes in Classics to the University of Sofia. All of them have received some formal training in EpiDoc XML, be it in the optional EpiDoc XML course in the framework of their curriculum (last held in the winter semester of 2018-2019) or through participations in some of the EpiDoc training events. However, none of them has had much previous formal training in Greek epigraphy. One of the expected benefits of the Telamon project is that XML encoding introduces students to the complex discipline of ancient epigraphy step by step, through "learning by doing". So far, this approach is justified and the Telamon investigators are witnessing a gradual advancement on the part of the students in their competences in the field of epigraphy proper. Some of the encoders have the intention of further developing their epigraphic skills and one is writing his MA thesis on the same monuments which he is encoding for the purposes of the project.

Each encoder receives a given number of inscriptions taken from the printed *IGBulg* corpus. The corpus was originally published in Latin which renders it rather inaccessible to many potential readers nowadays and also presents a challenge to the encoders. In the process of dealing with the monuments, the students get acquainted with the type and structure of metadata typically used for the description of cultural heritage objects written on stone. They learn how to isolate separate elements of metadata description in the Latin text of *IGBulg* and identify them with particular lines of XML code in the Telamon template. The formalized description of an item from a printed epigraphic corpus according to an EpiDoc XML template often requires a very thorough study of the available information about the monument. Points which require special attention are giving the document a distinct title in the collection, classifying it as monument type and text type, properly encoding its dating criteria and other peculiarities, as well as providing it with translations in Bulgarian and English usually not available in the original editions.

Additionally, there are the typical editorial conventions used in the publication of inscriptions. Such conventions are usually applied in a highly regularized and unified fashion in the printed and digital epigraphic corpora (Krummrey & Panciera, 1980). Most of them have quite straightforward encoding equivalents. For example,

[a text included in square brackets]

denotes a part of the inscription lost with the time and is to be uniformly encoded as:

```
<supplied reason="lost"> </supplied>,
```

while round brackets signify an abbreviation, etc. Such highly standardized equivalents between analogue and digital publishing in epigraphy allow for the relatively easy generation of XSLT transformation scenarios for the EpiDoc community some of which are then embedded into the EFES front-end service tool (see below, 2.3.).

The peculiarities of the monuments and their encoding are discussed by team and assistants almost in real time via shared Q&A documents. These will serve as the basis of the project's documentation which is to be enriched and developed in English as a future contribution of the Telamon project to the EpiDoc community.

By June 2019, about 45 inscriptions were encoded and their proper visualization and indexing under EFES (see 2.3.) was tested locally by the project team. As of the beginning of August 2019, there are 130 inscriptions in the database and their number is expected to approach 200 as the Telamon collection goes online.

2.3 Customisation and Development of the EpiDoc Front-End Services (EFES)

EFES (EpiDoc Front-End Services)[9] is a publishing and indexing platform for XML files valid according to the EpiDoc schema. It provides several built-in XSLT transformation scenarios (see above, 2.2.) that enable the visualization of the XML files according to the existing conventions for publication aimed at users with little technical knowledge.

EFES was created by the developer Jamie Norrish as a modification of the Kiln platform[10]. It is comprised of several free self-standing components supported by a lightweight server that comes as an inbuilt function with the installation of the platform. The search engine and the data storage are provided by the Apache Solr searching platform, which can be used for lemma search, grouped search, and Boolean queries. A templating system creates the general framework for the public URLs by organizing an inherited website structure which can be maintained and curated even by a non-technical person. The main transformations needed for the visualization of the encoded inscription are carried out by the EpiDoc stylesheets that the Telamon team has customized to better serve the project's goals. These components are held together and accessed through Apache Cocoon operating on a number of sitemaps listing instructions for generating and transforming the site's content. The tool documentation (Norrish & Yordanova, 2018) contains a technical part with information about the implementation of the application and a user section aimed at epigraphers trained to encode their material in XML.

Since EFES was launched only recently, the testing provided by the Telamon team proved useful for eradicating existing bugs and improving certain functionalities. The improvements which resulted from this testing have been fed back into the main EFES code so that they can be used by other scholars working with the platform. The Telamon team has customized and developed so far the EFES tool in the following ways:
- Modifications in the number and type of indices for the collection of inscriptions.
- Combining three transformation scenarios for the visualization of the text of the inscriptions.

---

[9] Available at https://github.com/EpiDoc/EFES/.
[10] https://kiln.readthedocs.io/en/latest/.

- Changing the design of the stylesheets for the visualization of the main page, the inscription list and the template for the representation of individual inscriptions (to be merged with the EFES customization).
- Fixing minor visualization bugs.

Our current EFES customization, together with all the XML content it processes, can be viewed and downloaded at https://github.com/DHLabUniSofia/Telamon-EFES/. The collection can be explored locally after the initiation of the `build` file that comes with it and launches a Jetty server. The Telamon EFES can then be opened in a browser as a local host page[11]. The platform is currently being automated and deployed on a web server.

EFES comes with the basic indices that might be desired in an epigraphic project. Out of those that come with the main installation bundle, the indices for word forms and fragments, not relevant to the scope of the project, were removed. New indices were created listing personal names mentioned in the inscriptions, as well as emperors, attested Roman and local officials, and divinities.

The EFES platform allows the displaying of the text of the inscriptions in different ways. Inbuilt transformation scenarios visualize in parallel the same source code as both diplomatic and interpretative editions. The interpretive edition presents the text in normalized form together with all the editorial changes and emendations to it. The diplomatic edition presents the text as it appears on the stone: in capital letters, without spaces between words, with lacunas unrestored, abbreviations not expanded, etc. With these two, the Telamon team also combined the transformation scenario for displaying verse which treats differently the line breaks in the inscription (encoded as `<lb/>`) in cases where versification is encoded with the elements `<lg>` and `<l>`.

With the aid of a web designer the visualizations of the main page, of the inscription list and of the individual inscription template have been modified. The CSS modifications conducted by the designer are currently being merged with the existing Telamon EFES customization.

In adapting the original EFES tool to the Telamon collection, the following minor bugs were fixed: the visualization of the inventory numbers of the inscriptions in museum repositories, where applicable, and also the correct visualization of dimension measure units in the physical description of the monuments.

3   Conclusion

The Telamon project is currently in the stage of merging the improved design elements into the customized services as well as automation and deploy for web. A beta version of the online collection is expected to be launched in September 2019 at https://telamon.uni-sofia.bg. The documentation of our workflow and the choices made by our team in the editing and encoding of the inscriptions is also to be expanded and published in the wiki to our GitHub repository page for the benefit of similar future initiatives. A

---

[11] More on how to run EFES locally see at https://github.com/EpiDoc/EFES/wiki/Installation.

particular contribution to the Bulgarian localization of various digital epigraphy resources will also be the creation of controlled vocabularies of terms describing in a unified way features such as the material of the monuments, object types, epigraphic formulas, etc. Such controlled vocabularies will be based on the EAGLE Europeana vocabularies of the same type[12] and will ideally include English, Bulgarian, Latin and, when applicable, the original Ancient Greek.

As a further perspective, the Telamon collection will provide a significant dataset of structured and annotated data to be processed and included in a future knowledge graph for the representation of the cultural heritage of Bulgaria: one of the primary tasks before the research infrastructure under development in the CLaDA-BG consortium of which The University of Sofia, and the project team in particular, are an active part.

References

Bodard, G., & Stoyanova, S. (2016). Epigraphers and Encoders: Strategies for Teaching and Learning Digital Epigraphy. In G. Bodard, & M. Romanello (Ed.), *Digital Classics Outside the Echo Chamber. Teaching, Knowledge Exchange and Public Engagement*, (pp. 51-68). London.

Elliott, T., & Bodard, G. (2018). *Examples. In: EpiDoc: Epigraphic Documents in TEI XML*. (T. Elliott, G. Bodard, & H. C. all., Editors) Retrieved June 04, 2019, from Examples. In: EpiDoc: Epigraphic Documents in TEI XML: https://sourceforge.net/p/epidoc/wiki/Examples/

Krummrey, H., & Panciera, S. (1980). Criteri di edizione e segni diacritici. *Tituli 2*, 205–215.

Mihailov, G. (1956-1995). *Inscriptiones Graecae in Bulgaria Repertae. Vol. I-V.* Serdicae: ALB.

Norrish, J., & Yordanova, P. (2018). *EpiDoc Front-End Services (EFES) User Guide and Technical Documentation*. Retrieved June 02, 2019, from EpiDoc Front-End Services (EFES) User Guide and Technical Documentation: https://github.com/EpiDoc/EFES/wiki

Roueché, C., Reynolds, J., & Bodard, G. &. (2007). *Inscriptions of Aphrodisias*. Retrieved June 01, 2019, from Inscriptions of Aphrodisias: http://insaph.kcl.ac.uk/iaph2007/

Sharankov, N. (2015). Notes on Greek inscriptions from Bulgaria. In M. Slavova, & N. S. M., *Studia classica Serdicensia V. Monuments and Texts in Antiquity and Beyond* (pp. 305-361). Sofia: Sofia University Press.

---

[12] https://www.eagle-network.eu/resources/vocabularies/.