

---

# The nature of Humanities data in a case study: a digital collection of Greek epigraphic heritage from Bulgaria

Dimitar Iliev\*<sup>1</sup>

<sup>1</sup>University of Sofia – Bulgaria

## Abstract

The present paper's point of departure is the premise that, in order to better understand the nature of data in the Humanities, the bottom-up approach is the most fruitful one. This implies case studies made for each project or initiative working on a particular set of objects (texts, images, monuments, artefacts, etc.) and applying particular DH methods and tools. Such case studies can successfully describe the specificities of processing the given objects as well as the issues and results of the given approach. When many such case studies are compared, trends and patterns can be detected and least common denominators can be drawn that enable a better understanding of what Humanities data at large are like and what is special about dealing with them in comparison with other types of datasets (e.g. scientific, statistical, etc.).

The topic of this paper will be the Telamon Project which aims at creating an online collection of the Ancient Greek inscriptions from the territory of today's Bulgaria. It was started some years ago by a team of scholars at the Department of Classics to the University of Sofia, it has received technical and financial aid from the University's IT Centre and the Centre for Excellence in the Humanities and it currently undergoes the finishing stages of its preparation in the framework of the National CLaDA-BG Consortium (conducting activities and implementing national contributions for CLARIN-EU and DARIAH-EU in Bulgaria). Its web site with a small initial collection of inscriptions will be launched for the public in 2019.

Ancient Greek epigraphic heritage in Bulgaria encompasses monuments from the period between 6. c. BCE and 4. c. CE covers the territory of the whole country and counts more than 4,500 inscriptions (the number is growing as new monuments are constantly found and published). For the time being, the project team concentrates on presenting the inscriptional heritage of only two major Greek cities in Roman Thrace, Philippopolis (today's Plovdiv) and Augusta Traiana (today's Stara Zagora). The perspectives are to gradually expand the territory covered. But even the fullest and exhaustive collection of monuments would be relatively small from the point of view of data science. The whole of ancient inscriptions from Bulgaria would still be quite far from big data. In Digital Epigraphy and the related DH fields, even large collections of texts, images, and objects from a single given project initiative are mostly amenable to processing and analysis and their data management does not necessarily require heuristic approaches.

But, in order to fully cover even the first several hundred or so monuments from Roman

---

\*Speaker

Thrace, a significant amount of time, effort and specialized skills need to be invested. The nature and content of the monuments can be diverse and complex. For the description of all the features thereof, TEI XML according to the standards and schemas of the EpiDoc subset for epigraphic and papyrological documents is applied. This makes possible the thorough and detailed description of all the characteristics of the text and the material object that carries it, together with its history, its editors in digital and analogue corpora and other similar metadata. The visualisation, cataloguing and indexing of such data which is done with the EFES front-end service and editing platform depends very much on this accurate description. All this implies a great stress on the adequate encoding of single monuments and its important place in the workflow as compared to the subsequent processing which is much easier if the good quality of the encoded data is ensured.

This leads to the question of the human resources needed to conduct such work. The most valuable contributions come from trained epigraphers who have additionally acquired XML encoding skills. That is why regular trainings are essential for the sustainability of such a project. On the other hand, the cross-section of scholars skilled both in the complex discipline of Greek epigraphy and in the technicalities of mark-up will always be relatively small and few people at the same time could be recruited for such initiatives resulting in the gradual pace of the enlargement of the digital epigraphic corpora.

Crucial information conveyed by historical documents like ancient inscriptions are the names, events, people and places mentioned in them: in other words, named entities. The encoding of NE's, together with dates and periods, and their proper organising in authority lists are among the most important tasks of an EpiDoc encoder. Its importance is increased by the linking of the respective NE's with external authorities such as prosopographies and gazetteers such as Pleiades, together with the content linked to it from Pelagios, which enables the integration of monuments from a particular collection to a larger set of linked open data.

These are some of the important characteristics of a digital epigraphic collection that became evident in the work on the Telamon Project: great importance of adequate description of the (meta)data by specialists in the field; relatively small volume of the encoded and processed data in the scope of a single project; the priority of named entities, and the importance of linked open data. These features will be demonstrated with examples from the team's practice and the inscriptional material we have at our disposal in search for connections and comparisons with other initiatives in order to draw more general conclusions for the nature of data in the Humanities.

**Keywords:** digital epigraphy, TEI, XML, EpiDoc, CLaDA, BG, Greek inscriptions, Bulgaria