EpiSearch.
Epigraphic Manuscripts and Digital Humanities

EPIGRAPHY info

The EpiSearch Project

EpiSearch is a collaborative and interdisciplinary pilot project that explores the possibilities offered by Digital Humanities technologies to retrieve the data contained in epigraphic manuscripts and to make them more accessible.

While developed on a single manuscript, EpiSearch will provide a method and a corresponding set of reusable tools, whose value extends beyond the immediate results of the project.

Method

The work has been divided into three main steps:

1. The first one, performed by Federico Boschetti, examines how Handwritten Text Recognition (HTR) technologies can be used to analyse epigraphic manuscripts.

2. The second, developed by Daniele Fusi, implies designing an integrated system, created collecting data from four epigraphic databases available online: EDR, EDH, EDCS and PHI. This will be the basis for the creation of a new tool that will provide semi-automatic identification of the inscriptions.

3. The last step will produce a visually annotated version of the manuscript chosen as a case study, with hyperlinks to the current digital editions of the inscriptions transcribed.
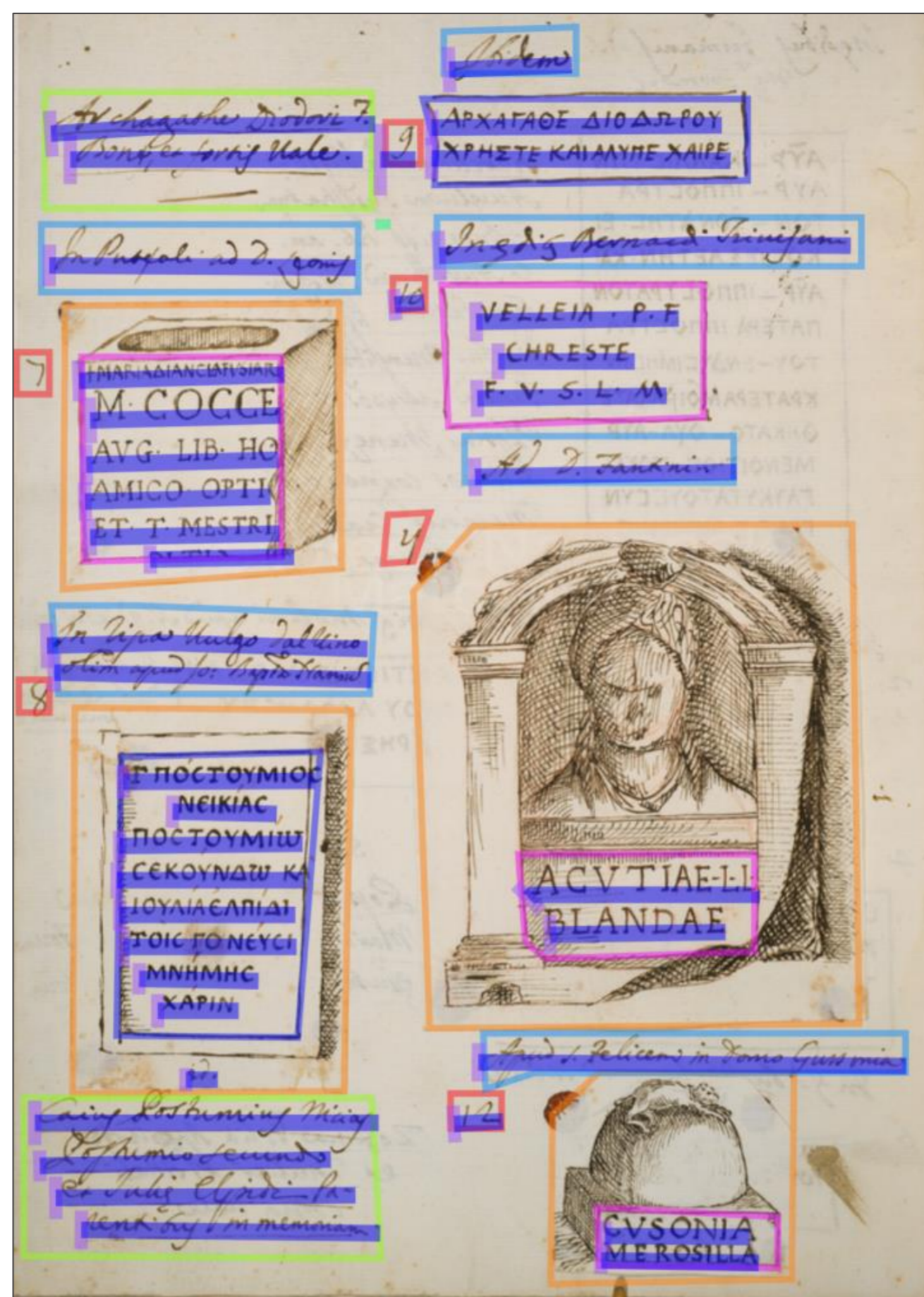

Figure 1. Baseline recognition and regions of interest coloured by type. Venice, BNM, ms. Marc. Lat. XIV, 200 (4336), f. 1v.

Case study

As a case study we identified an epigraphic manuscript written in Venice in the early 1700s by a local antiquarian, Giovanni Antonio Astori, and currently kept in Venice, Marciana National Library (BNM): Marc. Lat. XIV, 200 (4336).

Step 1: HTR applied to Astori's manuscript

HTR technologies can be used successfully to map legacy manual transcriptions on the manuscript facsimile and to improve its layout analysis, even with few pages (Figure 1).

We also tried to face the problem of the small amount of written text, using different techniques according to the languages and the scripts.

The most relevant results have been obtained from the analysis of the Latin texts. For the Latin inscriptions, in upper-case letters, we used the *Modèle imprimé 16-18e Fra+Lat*, which gave quite good outputs (Figure 2).

For the cursive parts we searched for other documents written by the same author (Figure 3) and used them for fine-tuning an existing model created from data available on HTR-United; this gave quite satisfactory results once applied to Astori's manuscript (Figure 4).

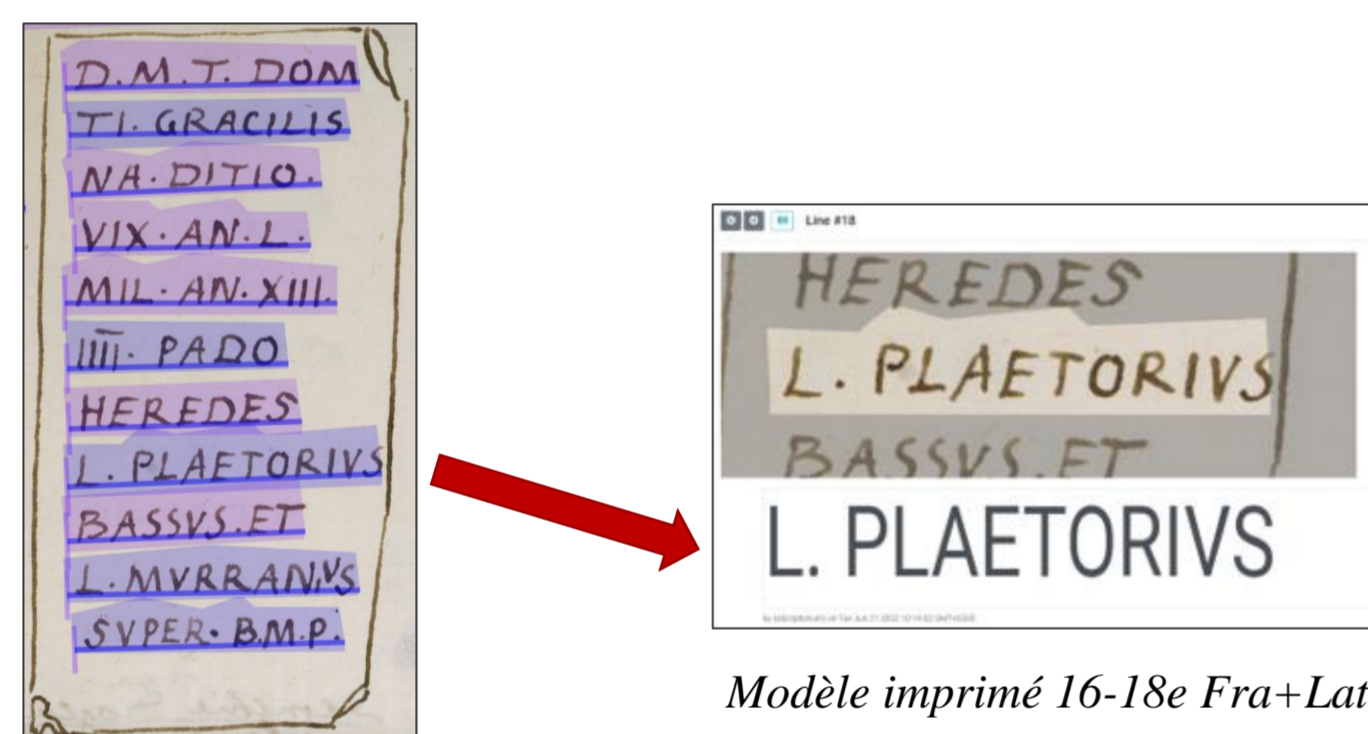*Modèle imprimé 16-18e Fra+Lat*

Figure 2. HTR applied to a Latin inscription. Venice, BNM, ms. Marc. Lat. XIV, 200 (4336), f. 4r, nr. 21.
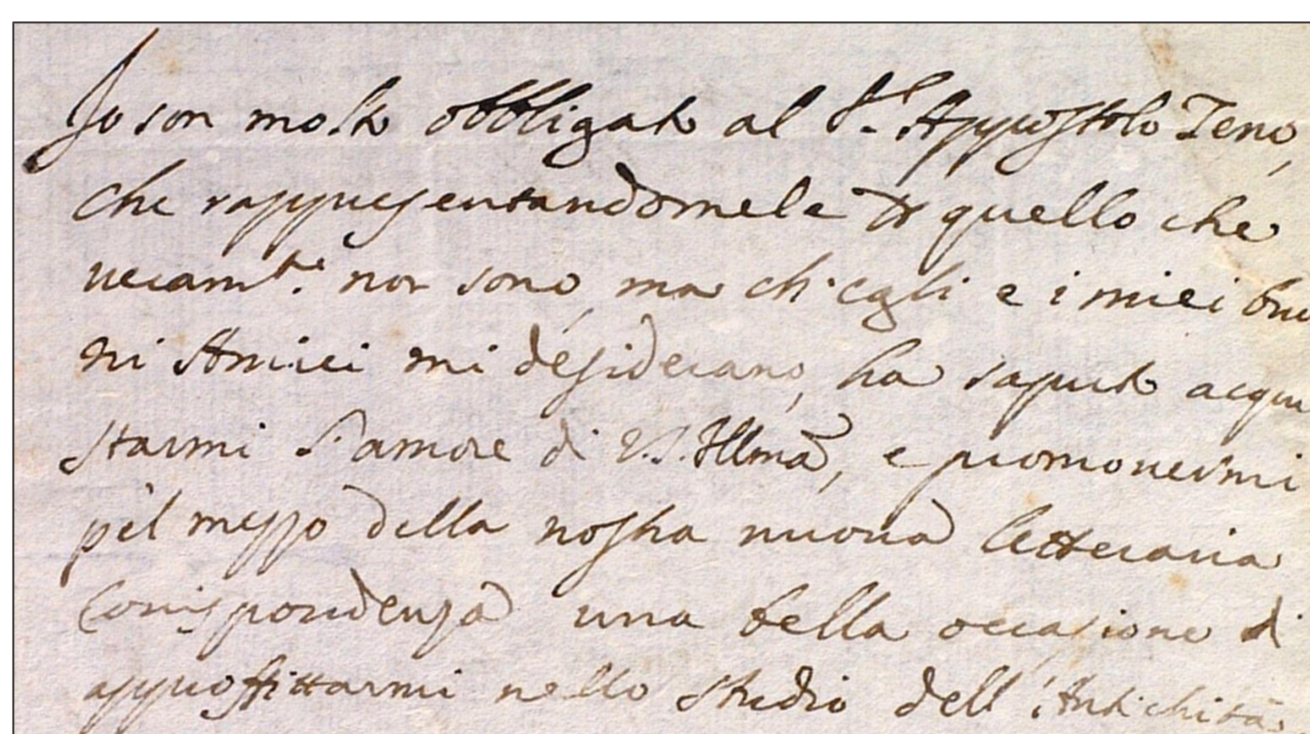

Figure 3. A sample of Astori's letter. Modena, Biblioteca Estense Universitaria, Archivio Muratori, 52.1. From *Internet Culturale* (CC BY-NC-SA 3.0).
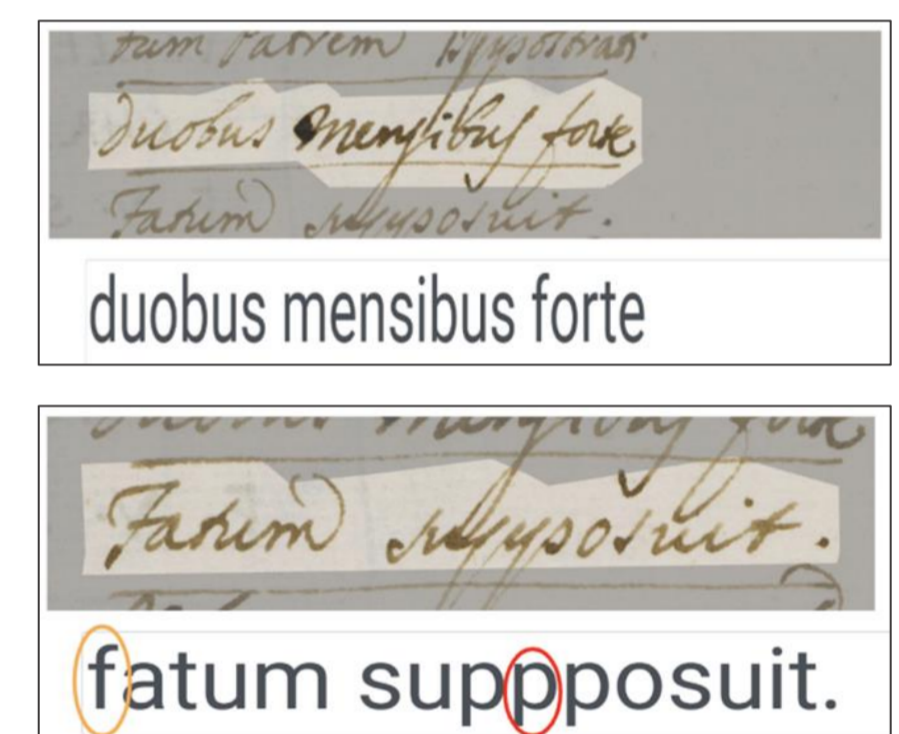

Figure 4. HTR applied to a part of Astori's manuscript in cursive script. Venice, BNM, ms. Marc. Lat. XIV, 200 (4336), f. 1r, nr. 1.
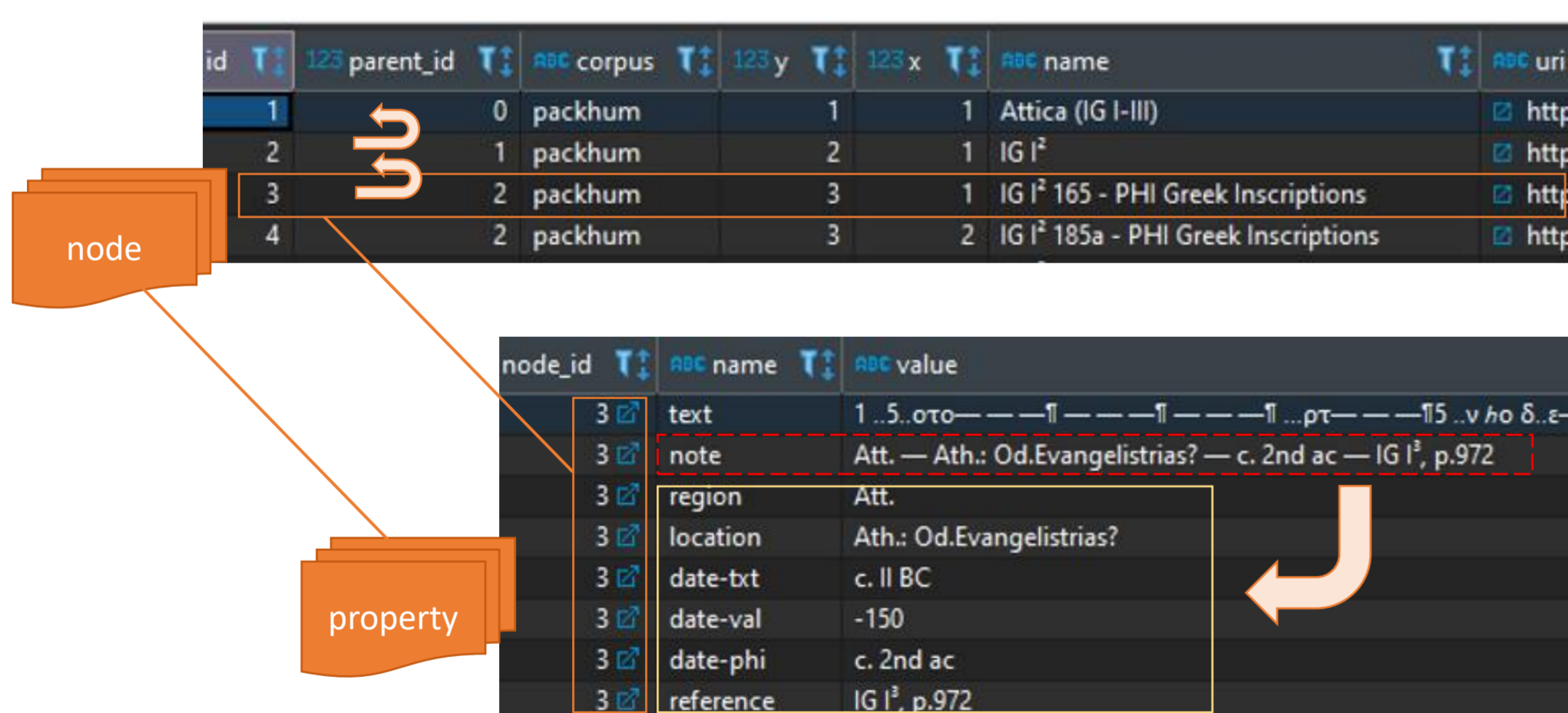

Figure 5. A sample of the Relational Database designed by Daniele Fusi for EpiSearch.
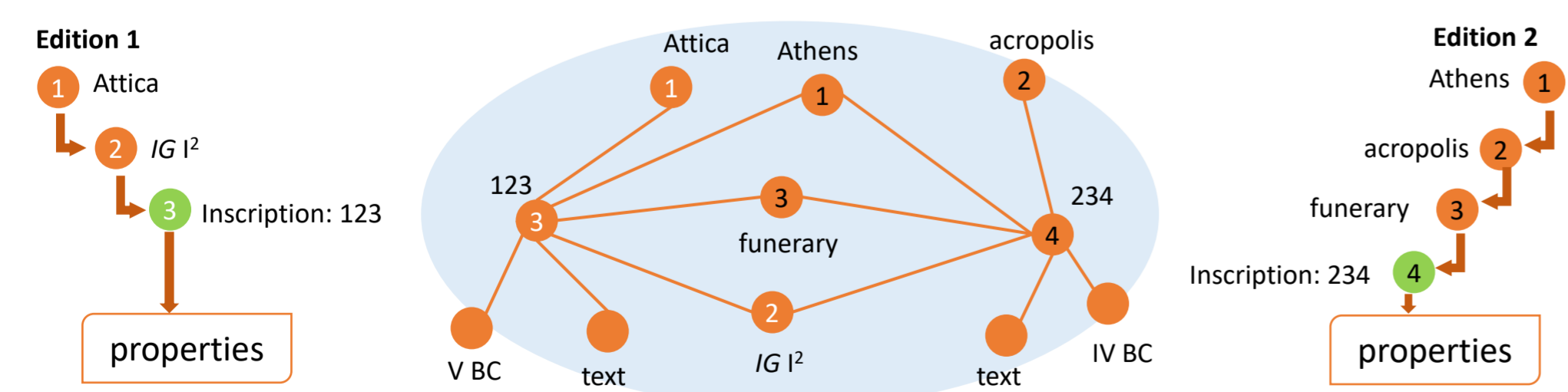

Figure 6. An example of destructuring and merging tree models into a single graph: the nodes are enriched by their connections.

Step 2: Coalescing Models

To collect, remodel and merge data from the epigraphic databases selected (each identifiable as a tree model), Daniele Fusi has designed a Relational Database composed by two tables: one for the nodes of each collection (the hierarchies), the other with all the properties of each inscription (Figure 5). This structure allows to inspect the data extracted and to progressively refine them through metadata injections. At the present, we have scraped PHI and EDCS and we are inspecting and refining the data extracted.

The injection process is fundamental to get uniform data, also with the aid of mapping custom metadata to community-driven datasets like Pleiades and shared ontologies as adopted by more modern epigraphic digital resources, and consequently provides the basis for the creation of a semi-automatic tool for identifying the inscriptions, according to the comparison and matching of metadata. Moreover, only uniform metadata can be linked in a network in which all the nodes are enriched by their connections. This aim is one of the main values of the EpiSearch project (Figure 6).

Conclusions

The evaluation of HTR tools is almost completed. After the refinement of the data extracted from EDCS and PHI we will proceed to the scraping of EDR and EDH.

EpiSearch is conceived as the initial segment of a broader project, of which it constitutes a proof of concept. Therefore, in the future, we will reuse the set of tools created for the EpiSearch project to analyse other epigraphic manuscripts.