

# Linked Open Data and the Enrichment of Digital Editions: The Potential Contributions of CLARIN to the Digital Classics



Monica Monachini, Fahad Khan<sup>1</sup> Francesca Frontini<sup>2</sup> Anika Nicolosi<sup>3</sup>

<sup>1</sup>Istituto di Linguistica Computazionale “A. Zampolli”, CNR Pisa, Italy

<sup>2</sup>Université Paul-Valéry Montpellier 3, Praxiling UMR 5267 CNRS - UPVM3 France

<sup>3</sup>Department of Discipline Umanistiche, Sociali e delle Imprese Culturali, Parma University, Italy  
name.surname@ilc.cnr.it francesca.frontini@univ-montp3.fr anika.nicolosi@unipr.it



## Motivation

- The success of the Semantic Web in many fields has paved the way towards using Linked Data (LD) technologies to enable **open access to cultural data**.
- LD has become a way of sharing structured information in a flexible and extensible manner across the Web, this is especially true in the **Humanities**.
- The existence of large repositories of freely available digitised texts and datasets is encouraging a greater **uptake of LD technologies in the Classics**.
- Several ongoing projects are contributing towards the provision of the necessary support, especially for use in **enriching digital editions**.

## Main Challenges and Opportunities

- How can we build a platform which responds to the **desiderata of scholars** themselves, facilitating the **enrichment and exploration of digital editions using Semantic Web technologies**?
- Not only produce an **enriched digital edition** of a corpus but also **demonstrate the potential of the LD approach** in the creation of **openly available, scientifically reliable, linked, digital texts** and its usefulness to a range of users including **scholars, students and the wider public**.
- The heavy demands that these approaches make wrt computational capacity and long term support call for a strong **involvement on the part of research infrastructures**.

## Digital Edition of Archilochus (DEA)

- Concentrates on a **particularly complex type of text**, namely **texts of fragmentary authors**
- Develops and builds on **previous work in this area**: this new edition will both integrate already available digital resources as well as incorporating new datasets and resources
- Creates an **integrated and standardised platform for the consultation, updating and searching of textual materials and structured information**.

## Project Goals

- The digitisation, with critical and philological control, of the **whole corpus of fragments of an ancient Greek author Archilochus of Paros** (VII BC), an important figure in archaic Greek Poetry, closely linked to Homer.
- Publication of a **complete, scientifically reliable, born digital, TEI/XML, edition** of the corpus of Archilochus' texts which will pay particular **attention to scholarly users' needs and requirements, as well as to usability and portability**
- Utilisation and (where needed) development of digital methods that allow for the **enrichment of the texts using different levels of linguistic and textual annotation** (morpho-syntactic, semantic, etc.)
- An Investigation into the **interaction between the text and already existing terminologies, ontologies, and lexicons available as Linked Open Data**

## The role and impact of CLARIN

- CLARIN can assist in the promotion of significant **developments in the field of Digital Classics** by offering opportunities to store, develop, share and access datasets of enriched digital editions.
- This newly **created resource will be deposited in CLARIN and integrated into the Tundra exploration system**.
- In order to allow for the deposit of the resource, **all texts will be provided with adequate metadata so as to become findable via the CLARIN metacatalogue (VLO)**
- From the CLARIN perspective, this project provides an example of **how it is possible to integrate and support the proof-reading, encoding and enrichment of the texts in a CLARIN repository**
- Transform the corpus into a Treebank, thereby making an **important contribution to the already existing Ancient Greek Treebank**, this will enable the **improvement of currently existing parsers for ancient Greek** with regards to their performance on fragmentary texts
- The **textual analysis**, with the use of a Treebank, and **translation** of each fragment will be a **important development for the study and teaching of Ancient Greek**

## Example

Archil. fr. 3 Nicolosi (= fr. 5 West) shows how the **Pleiades dataset provides a place reference for the Greek ethnonym Saiwn** (line 1) which is an ancient clan of Thrace. Other datasets such as Geonames or DBpedia may be used for the same purpose, adding useful information on the ancient region and its inhabitants.

## Crucial issues

- The identification of standards and best practices** for the enrichment of our digital edition **with structured knowledge from the Semantic Web**
- In order to cater for the needs of classicists, both the enrichment and the exploration phases should **be supported by easy to use interfaces, requiring little knowledge of the LD technicalities**
- The investigation of such technological aspects should hopefully **serve as a important test bed for future projects in the field of classical studies**

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <fileformat>xml</fileformat>
  <encoding>UTF-8</encoding>
  <idno>Archilochus Fr. 3 W. 2. 1</idno>
  <revision number="1"></revision number>
  <url>http://www.tei-c.org/ns/1.0</url>
  <date>2017-11-30</date>
  <author>Archilochus</author>
  <editor>Fahad Khan</editor>
  <title>Archilochus Fr. 3 W. 2. 1</title>
  <text>
    <p>
      ἀσπίδι μὲν Σαίων τις ἀνέλειται, ἥν παρὰ θάμνον,
      ἔντος ἀμώμητον, κάλλιπον οὐκ ἐθέλων·
      ψυχὴν δ' ἔξεσάωσα. τί μοι μέλει ἀσπίς ἐκείνη;
      ἔρρετο· ἐξαῖτις κτήσασαι οὐ κακίῳ.
    </p>
  </text>
  <backmatter>
    <table border="1">
      <tr>
        <th>Line</th>
        <th>Text</th>
        <th>Comment</th>
      </tr>
      <tr>
        <td>1</td>
        <td>ἀσπίδι μὲν Σαίων τις ἀνέλειται</td>
        <td>The shield of the Saiwn is being thrown</td>
      </tr>
      <tr>
        <td>2</td>
        <td>ἔντος ἀμώμητον</td>
        <td>Within the unblemished</td>
      </tr>
      <tr>
        <td>3</td>
        <td>ψυχὴν δ' ἔξεσάωσα</td>
        <td>My soul I have saved</td>
      </tr>
      <tr>
        <td>4</td>
        <td>τί μοι μέλει ἀσπίς ἐκείνη</td>
        <td>What care has it for me, that shield</td>
      </tr>
      <tr>
        <td>5</td>
        <td>ἔρρετο</td>
        <td>It flowed</td>
      </tr>
      <tr>
        <td>6</td>
        <td>ἐξαῖτις κτήσασαι</td>
        <td>In order to buy</td>
      </tr>
      <tr>
        <td>7</td>
        <td>οὐ κακίῳ</td>
        <td>Not for a bad thing</td>
      </tr>
    </table>
  </backmatter>
</TEI>
```

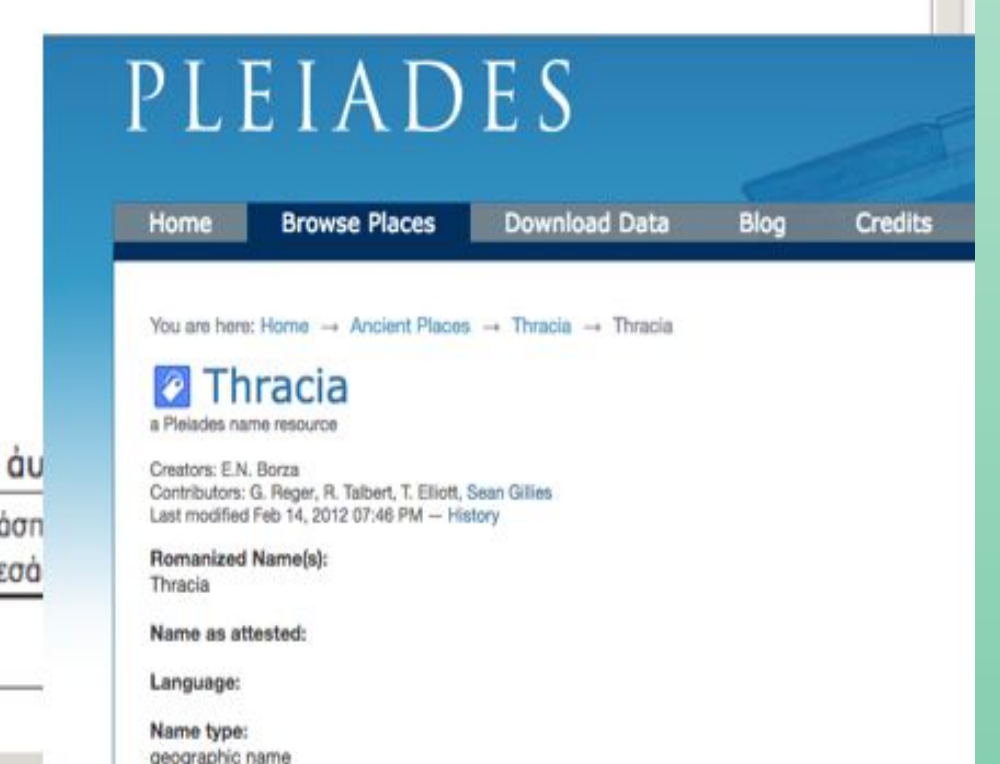
### Archilochus Fr.5 W.2

#### Table of contents

- ἀσπίδι μὲν Σαίων τις ἀνέλειται, ἥν παρὰ θάμνον, ἔντος ἀμώμητον, κάλλιπον οὐκ ἐθέλων· ψυχὴν δ' ἔξεσάωσα. τί μοι μέλει ἀσπίς ἐκείνη; ἔρρετο· ἐξαῖτις κτήσασαι οὐ κακίῳ.

- ἀσπίδα μὲν Σαίων τις ἀνέλειται, ἥν παρὰ θάμνον, ἔντος ἀμώμητον, κάλλιπον οὐκ ἐθέλων· ψυχὴν δ' ἔξεσάωσα. τί μοι μέλει ἀσπίς ἐκείνη; ἔρρετο· ἐξαῖτις κτήσασαι οὐ κακίῳ.

Date: 2017-11-30



- Francesca Frontini, Carmen Brando, and Jean Gabriel Ganascia. 2016. REDEN ONLINE: Disambiguation, Linking and Visualisation of References in TEI Digital Editions. In Digital Humanities 2016: Conference Abstracts, pages 193–197, Krakow. Jagiellonian University & Pedagogical University.
- Fahad Khan, Francesca Frontini, Federico Boschetti, and Monica Monachini. 2016. Converting the Liddell Scott Greek-English Lexicon into Linked Open Data using lemon. In Digital Humanities 2016: Conference Abstracts, pages 593–596, Krakow. Jagiellonian University & Pedagogical University.
- Monica Monachini and Francesca Frontini. 2016. CLARIN, l'infrastruttura europea delle risorse linguistiche per le scienze umane e sociali e il suo network italiano CLARIN-IT. IJCoL - Italian Journal of Computational Linguistics, 2(2):11–30.
- Anika Nicolosi. 2013. Archiloco: elegie. Patron, Bologna. Google-Books-ID: 9uj5oAEACAAJ.