Publication, Testing and Visualization with EFES: A tool for all stages of the EpiDoc XML editing process

Gabriel Bodard and Polina Yordanova
University of London, University of Helsinki

Abstract: EpiDoc is a set of recommendations, schema and other tools for the encoding of ancient texts, especially inscriptions and papyri, in TEI XML, that is now used by upwards of a hundred projects around the world, and large numbers of scholars seek training in EpiDoc encoding every year. The EpiDoc Front-End Services tool (EFES) was designed to fill the important need for a publication solution for researchers and editors who have produced EpiDoc encoded texts but do not have access to digital humanities support or a well-funded IT service to produce a publication for them.

This paper will discuss the use of EFES not only for final publication, but as a tool in the editing and publication workflow, by editors of inscriptions, papyri and similar texts including those on coins and seals. The edition visualisations, indexes and search interface produced by EFES are able to serve as part of the validation, correction and research apparatus for the author of an epigraphic corpus, iteratively improving the editions long before final publication. As we will argue, this research process is a key component of epigraphic and papyrological editing practice, and studying these needs will help us to further enhance the effectiveness of EFES as a tool.

To this end we also plan to add three major functionalities to the EFES toolbox: (1) date visualisation and filter—building on the existing "date slider," and inspired by partner projects such as Pelagios and Godot; (2) geographic visualization features, again building on Pelagios code, allowing the display of locations within a corpus or from a specific set of search results in a map; (3) export of information and metadata from the corpus as Linked Open Data, following the recommendations of projects such as the Linked Places format, SNAP, Chronontology and Epigraphy.info, to enable the semantic sharing of data within and beyond the field of classical and historical editions.

Finally, we will discuss the kinds of collaboration that will be required to bring about desired enhancements to the EFES toolset, especially in this age of research-focussed, short-term funding. Embedding essential infrastructure work of this kind in research applications for specific research and publication projects will almost certainly need to be part of the solution.

Keywords: Text Encoding, Ancient Texts, Epigraphy, Papyrology, Digital Publication, Linked Open Data, Extensible Stylesheet Language Transformations

1. Background

The benefits of online publication are well understood and sought after by epigraphists.[1] EpiDoc is a tool-set and a community of practice for the digital encoding of edited ancient texts, including inscriptions, papyri, seals, coins, and related objects, in TEI XML. EpiDoc is used by dozens of projects relating to the classical world and beyond, and hundreds of people worldwide have been trained in the use of EpiDoc practices and tools.[2]

However, after having invested a substantial amount of effort and time in encoding their materials in a digital format, researchers are often met with another obstacle before publishing. The transformation of EpiDoc XML documents and their visualization online require a different set of advanced technical skills, or the support of a dedicated IT unit or a development team, which might ultimately render publication without substantial institutional support or project funding a daunting challenge.

The increasing demand from the epigraphic community for a tool that is free, specifically designed to reflect the research and encoding practices of the discipline, but also customizable to fit the particular requirements of the individual projects, and accessible to users without advanced technical knowledge, was the prompt behind the creation of the EpiDoc Front End Services (EFES) from 2017.[3]

EFES was aimed at facilitating the creation of a rich output similar to those of some of the prominent epigraphic projects such as the Ancient Inscriptions of the Northern Black Sea (IOSPE), Inscriptions of Roman Tripolitania (IRT), Inscriptions of Aphrodisias (IAph), and Roman Inscriptions of Britain (RIB), which are all large-scale projects with extensive technical support or in-house expertise.[4] The platform was designed to provide the components that are theoretically a common interest for most publications of epigraphic material—display of individual texts in interpretive and diplomatic editions along with their historical and descriptive data, multiple indices, browse and search interface, and possible export of information through Linked Open Data, and make them accessible to the non-technical user via an "out-of-the-box" publishing platform.

---

[1] The authors would like to thank Elina Boeva, Martina Filosa, Tamara Kalkhitashvili, Jamie Norrish, Charlotte Roueché, Alessio Sopracasa, Simona Stoyanova, Charlotte Tupman, Irene Vagionakis, Valeria Vitale and the journal's anonymous reviewer for invaluable comments on a draft of this essay.
[2] EpiDoc: Epigraphic Documents: epidoc.stoa.org.
[3] EFES: EpiDoc Front-End Services (University of London, 2018–20): github.com/EpiDoc/EFES.
[4] These corpora and other digital resources are listed at the end of the paper, alongside the bibliography.

As base for building a publication platform to match the desiderata of epigraphers, the developers chose the Kiln application, developed and used at the King's Digital Lab, King's College London.[5] Since 2010, Kiln—and its predecessor under another name—has been used in the creation of over fifty digital humanities projects, among which the above-mentioned IOSPE.[6] The platform is designed to work with content primarily in the form of XML and is itself largely written in XML, allowing users who are familiar with editing XML to perform customization, templating and project-specific code adjustments.

Kiln is built on Apache Cocoon and integrates several independent components—Apache Solr as a searching platform, Sesame 2 as an RDF store, built-in Jetty web server, XSLT-based templating system with inheritance infrastructure, and Foundation, a front-end framework based on HTML, CSS and JavaScript. Thanks to its specific file structure, which separates different components into distinct files and directories, Kiln provides a 'batteries included' experience, allowing the user to see an organized display of the XML content by simply placing the documents in the content directory, and easily to index and harvest the data through the online administrative panel. The general framework is put in place for more project-specific customizations, such as Schematron validation of XML files and infrastructure for multilingual sites. In its EFES incarnation, Kiln has been supplemented with specific display templates for EpiDoc XML, further search facets, and indexing features. Kiln comes with documentation on each of its main functional components, and a tutorial that guides the user through the process of setting it up and customising its various parts.[7] A user guide has been created for EFES, covering its specific EpiDoc-related enhancements and additions (Yordanova).

With Kiln chosen as the foundation of the publication platform, we employed Jamie Norrish, one of the specialists involved in its creation and development, to build upon its core components a specification suited for the particular needs of epigraphists and papyrologists. In defining the parts shared by philological editions that needed to be provided with the EFES package as the base of the platform, Norrish worked in close collaboration with Yordanova, who brought in the perspective of the epigraphists regarding the *sine qua non* of an online publication. Yordanova had had previous experience in epigraphic projects applying the EpiDoc standards of encoding and conducted multiple consultations with EpiDoc developers and scholars who had expressed interest in working with the platform, in order to determine the most essential common features for projects with different backgrounds.

We used as the starting point several existing projects in the field, and discussed with a selected group of stakeholders what features of these projects they thought would be applicable to their material, what they would need to handle differently in accordance with the specifics of their data, and also what the crucial elements in terms of representation were for their respective projects. These features

---

[5] Kiln (King's College London, 2012–2019): github.com/kcl-ddh/kiln. Kiln Documentation (2012): kiln.readthedocs.io/en/latest/.
[6] Kiln Documentation (2012), "Projects": kiln.readthedocs.io/en/latest/projects.html.
[7] Kiln Documentation (2012), "Tutorial": kiln.readthedocs.io/en/latest/tutorial.html.

were then added in iterations, using the responses we received from stakeholders, learning from feedback we received during the first training workshop in London in 2017 at which EFES was introduced to epigraphists, and building on user experience in order to improve the platform. This cycle was repeated with a second training event in Sofia, Bulgaria, later in the year.

Thus EFES was first equipped with discipline-specific customizations in the form of templates for some of the most common edition layouts, represented by the EpiDoc Reference Stylesheets (Elliott *et al.*, *Stylesheets*), code infrastructure for indexing with pre-existing matrices for some indices that could be populated with project data, potential for RDF export, and search facets based on categories shared by most EpiDoc projects. Further refinement with project-specific customizations was expected to be easily achievable by users who are well-versed in XML, but do not necessarily have other advanced technical experience, on the basis of training and documentation made available.

1.1 Status quo

The vast majority of the specified technical objectives of the EFES tool were therefore met during the first phase of the project: the platform can be easily downloaded and installed on any operating system, allowing users to add their own EpiDoc files into a specified folder and view a basic transformation of their texts in the web browser. Several customized templates for viewing the texts can be selected, via existing parameters in the EpiDoc Reference Stylesheets. If EpiDoc recommendations for tagging have been followed, several core features of the editions will be indexed, and search results will be filtered according to the selected facets. Instructions for selecting pre-existing templates, indices and facets for the core display, and for creating new ones, are well-documented and have proved easy to follow. A few active projects (Inscriptions of Roman Cyrenaica, Telamon: Greek Inscriptions from Bulgaria, SigiDoc: text encoding for Byzantine sigillography, Epigraphic Corpus of Georgia, Epigraphic Database Vernacular) are now using EFES as part of their pre-publication editorial workflow. They have in particular reported that one benefit of the system is its use to make evident and diagnose encoding errors in their electronic texts as they go along.

```xml
<div type="edition" xml:lang="grc" xml:space="preserve">
  <ab>
    <lb n="1"/><placeName ref="#lato" type="ethnic"><supplied reason="lost">Λάτιο</supplied>ι</placeName>
      <persName type="divine" key="Aphrodite">Ἀφροδίται</persName>
    <lb n="2"/><w lemma="νικάω">νικάσ<g type="𐅀"/>αντες</w>
    <lb n="3"/><rs type="institution" subtype="tribe" role="eponym" key="Aischeis" ref="#lato">ἐπὶ τῶν
      <w lemma="Αἰσχεῖς">Αἰσχέων</w></rs>
    <lb n="4"/><placeName ref="#olous" type="ethnic">Βολοντίος</placeName>.
  </ab>
</div>
```

Figure 1: Example of encoded text from an EpiDoc file (© 2020 Vagionakis)

Divinità, ninfe ed eroi

Le parentesi quadre indicano che il nome è o parzialmente o totalmente integrato.

Nome ▲	Occorrenze ⇕
Acheloios	ic1_17_7.1 [ic1_17_7.3]
Achilles	ic3_4_37.4
Amphitrite	chaniotis_61b.180 [ic1_16_5.74] [ic1_17_1.6] [ic1_17_1.18] ic1_30_2.5 [ic2_16_2.4] [seg_26_1049.83] [seg_33_638.3]
Aphrodite	chaniotis_54-56a.12 chaniotis_54-56d.1 chaniotis_61b.162 chaniotis_61b.182 ic1_14_2.3 ic1_16_18.7 ic1_16_25.10 [ic1_16_27.8] [ic1_16_5.70] [ic1_16_5.75] [ic1_18_9.C.6] [ic1_22_2.4] ic1_8_4.b.14 ic1_9_1.a.27 ic3_3_3ab.B.14 ic3_3_5.14 [ic4_171.15] [ic4_174.59] ic4_174.75 [ic4_183.21] [seg_23_547.B-C.52] [seg_26_1049.63] [seg_26_1049.84] [seg_33_638.4] [seg_41_743.13] seg_54_841.6 seg_60_996.A.3 [seg_61_722.B.9]
Apollo	chaniotis_55b-56b.11 [ic1_16_3.20] ic1_16_4.A.15 [ic1_19_2.8] ic1_23_1.17 ic1_5_20a.II.19 ic1_8_12.48 ic1_8_6.8 ic2_1_2.B.23 ic3_3_3ab.B.11 [ic3_3_3ab.B.12] [ic3_4_43.5] [ic4_3.2] ic4_51.2 [id_1602.4] ig_iv2_1_96.11 ig_ix2_3_783.2 seg_28_746a.4 [seg_45_1528.2] syll3_560.33 syll3_737.22

Figure 2: Example of index in EFES based on encoding in fig. 1 (© 2020 Vagionakis)

Where our aims were less well met, however, was in the area of the ease of customization and deployment by a non-technical user. Of the approximately forty people to whom we offered training in the processes of creating and customizing indices and search facets for EFES in that first year, three individuals (all exceptionally motivated and having access to personal support), plus the five project teams mentioned above, have gone on to deploy the platform and were able to master the skills required to create their own display, index, search and similar features. They all report finding EFES extremely useful as an editing and testing environment—viewing the results of the encoding in real-time—even if not all are quite ready to deploy it for publication. While this is a pleasing result in its own right, feedback from the vast majority of users, some of whom left the training workshops before completing the week of practice, was that the coding and technical skill required for creating new indices was too steep a learning curve for most editors, even those who have acquired facility in editing EpiDoc XML. On aggregate this was a disappointing outcome, and we came to the conclusion that the target user of the EFES platform (a non-technical editor who had created an EpiDoc corpus without institutional support) could not easily on her own install and customize the tool to create an online publication, even with the help of the very thorough documentation and user guide.

Our intention is now to mitigate this shortcoming in the usability of the EFES platform in two ways. Firstly we need to rethink the approach whereby EFES provides a bare-bones handful of example indices and search facets by way of demonstration, and users are expected to customize the platform themselves to add any further indexing functionality desired for their project. The core version of EFES included indices for: symbols, abbreviations, fragments, numerals, words (base form or lemmatised) and personal names. The *Cretan Institutional Inscriptions* (CII) project added indices for prosopography, divinities, toponyms and institutions. The hope is that other projects will further equip the distribution version of EFES with twenty or so

further indices, to include: bibliographical concordance; names, people and places all subdivided into multiple criteria and subgroups; as well as categories of text, object, material, offices, languages, calendars and the like. Between them, these indices should cover the majority of basic needs of most classical and other ancient corpora, as identified in the outreach phase of the initial EFES project.

Further customization will always be necessary for less common indexing and searching features, but this is no longer likely to be a blocker to the start of a project, or even in some cases to publication itself. The indices covered will include some features that would be challenging for even a confident developer to implement independently, such as lemmatized word searching, sorting of search results, and the handling of multiple search terms in a single XML tag (such as a word with two simultaneously valid lemmata). There are already (as of 2020) concrete proposals to include elements of this future development in project applications for SigiDoc (Sopracasa-Filosa) and the Epigraphic Corpus of Georgia (Doborjginidze). We are also working with a newly identified community of scholars working on Hittite and other Cuneiform inscriptions in EpiDoc XML to develop new desiderata and needs in this related subdiscipline;[8] this initiative should additionally help to track the usability of the new indices and documentation going forward.

1.2. A workflow tool

One of the barriers to the adoption of digital editions, including but by no means exclusively epigraphic corpora, as the standard for philological publication has been the lack of suitable software that supports the workflow developed by editors who were trained in the compilation and publication of printed editions (Rosselli Del Turco 227; Burghart §1.1). EFES aims not only at providing an easy-to-use solution for publishing the final digital corpus, but also at addressing some of these steps in the process of preparing the edition.

The steps involved in the traditional editorial workflow include, not necessarily in this order and certainly with multiple iterations of some aspects: fieldwork for assembling materials; outlining the questionnaire required to complete individual records; critical treatment and addition of information; data management; discovering patterns and preparing historical commentary; checking for consistency and completeness; and finally publication. The production of a digital edition in EFES is able to mirror several of these steps, while the separation of labour made possible by the specific file structure of the platform allows for these processes to be running not only as series, but also in parallel.

Data management in EFES is available directly out-of-the-box by placing the EpiDoc XML files into the 'content' directory in the files system and indexing them through the admin panel. This gives immediate access to the list of documents in the project and reports any obvious technical issues, since the platform will return errors for ill-formed documents and other unexpected features. The platform can also be

---

[8] HittiDoc (ed. Gabriel Bodard, İlgi Gerçek, Katherine Shields). 2019-20. "Hittite and Cuneiform in EpiDoc." github.com/EpiDoc/HittiDoc.

customized to validate the files against a specialist schema to ensure consistency of the annotation throughout the project. Since XML content, controlled vocabularies, EpiDoc transformation stylesheets, and CSS files and libraries are all kept separately, the EpiDoc encoding of the documents, creation and maintenance of controlled vocabularies, editorial decisions for organizing and displaying the data, and graphic design of the site can all be performed simultaneously or in parallel, perhaps by different team members, and will complement one another.

Although EFES is therefore highly customizable, allowing users to shape and specify it locally according to the particular needs of each project, the platform provides a default structure based on EpiDoc encoding practice. Thus, projects making their first steps into digital epigraphy can have a solid frame around which to create their content. Many of the pre-made indices in EFES are generated on the basis of authority lists, controlled lists of terms or names, which are XML documents that give all items unique identifiers and may include other information, variant spellings or localisation. These lists may be defined *ad initio*, but can also be built retrospectively or iteratively throughout the preparation of the edition, helping maintain the consistency of vocabulary and terminology. The information stored in the authority lists can be exported to RDF and shared with other projects. This has the potential to encourage the development and maintenance of community standards, as well as working towards data compatibility, sharing and reuse.

The possibility to visualize each document and the indices generated from the markup and through references to the controlled vocabularies, proves useful in keeping an overview of the corpus as new texts are added. It is also a tool for validation and "sanity checking" while assembling the individual edition. The process of annotating and indexing may for example aid the editor in finding patterns and structuring the content, say into chapters or sections, by grouping and visualising relevant criteria.

Most editors will appreciate that access to both philological and data querying and visualisation tools are an essential part of the editing process, not just an added feature for the final publication. As an editor of an early EpiDoc project (who shall remain anonymous) complained, "Our readers can now go and find all the errors in my work, thanks to the tools we've created for them, to which I didn't have access while writing it!" Using a tool like EFES as part of the workflow of editing a corpus both enables the editors to verify their own analysis and output at every stage in the process, and potentially offers the opportunity to pre-publish and make provisional results available for peer review at an early stage, so that the final publication both benefits from community contributions and more closely suits the needs of scholarly readers and users.

2. Visualization possibilities

One of the core priorities for EFES is to produce user-friendly display of indexed information using tools that are both generic and robust enough to be deployed by a non-technical editor. This includes not only static indices that reproduce the display from a conventional print book, and faceted search filters, but more novel and dynamic

visualizations, such as maps, sliders and customisable tabular data. For those indices that are generic enough to be deployed "out of the box" for a majority of corpora, the editor would ideally be able to select, activate and deactivate options from a graphical user interface, without needing to edit stylesheets, XML templates or Cocoon sitemaps in most cases.

The main visualisation tools that we would like to see added to the EFES platform, which are likely to be of value to the largest number of text corpora, ancient or otherwise, include:

1. a date slider that both displays the distribution of search results as a histogram, and allows the selection of a data range as a search modifier;
2. geographical visualisation to display search results on a map, with links back to individual text editions, and identifiers from a standard gazetteer;
3. the ability to export search results and other data in tabular or graph form, in as many open standards as possible, so that external tools that are better suited to the handling of data or particular user needs may be employed to analyse or display results as needed.

Even with all of the enhancements suggested above, from a larger menu of indices and search facets, to visualisation tools and export formats, EFES would still be a relatively bare-bones platform. Inevitably, any range of features implemented will be but examples of the possibilities enabled by digital publication of textual corpora, and its greatest strength remains the extensibility of the platform so that editors can create new features to serve their unique needs.

2.1 Date slider

One of the most widespread and useful means to organize, sort and filter an epigraphic corpus, or any other large collection of historical texts, is by date of creation. To a lesser degree, dates of discovery, accession, publication, and for that matter other dates in the history of the text or its support may be used to filter or navigate a corpus. Conventional epigraphic or papyrological editions are very often sorted by date, perhaps within other criteria such as locations or categories of text, and almost every historical corpus is explicitly delimited by chronological extent in some way, whether by editorial choice or as a result of the occupation or epigraphic habit of a site.

Dates as an item of metadata for a textual edition may be expressed in several fundamental ways (and arguably one edition might choose several of all of these forms of data expression in parallel, for different purposes), including:

- Textual expression of the date: "Late first century," "presumably within *circa* fifty years of the death of Hadrian," "third century, but not before the citizenship edict of Diocletian in 212," etc. For human-readability by scholars, such an expression contains information, including argument, that cannot be replaced by any numerical or computational formula.
- Dates gathered into named periods, defined by more or less granular start and end dates: "Hellenistic," "Byzantine," "Julio-Claudian," "Middle Bronze Age" and so forth. There are authority lists and ontologies of such named

    periods, which reflect the contingency of naming, geography, scholarly tradition, and other features on the dating of said periods. See for instance the PeriodO project and ontology (Rabinowitz).

- The original date expression in the text: "the 23rd day of Hathur, year 3," "in the archonship of Phainippos the second," "the 14th year of the reign of Antoninus," "after the death of Sekhemre Shedtawy." These may or may not be resolvable to Gregorian dates by modern scholars. For instance GODOT project supplies identifiers for ancient date and calendar expressions (Grieshaber 2017).

Arguably the most powerful format for capturing dates for searching, sorting or filtering historical documents, is to express the dates (where possible, of course) as numerical data, capable of having mathematical functions performed upon it. A date expressed numerically can be sorted, can be used to calculate periods ("how long did this reign last?"), to find objects within a date range ("-0100 – 0100") or within a certain number of years of a fixed point ("0212±25 years"), and so forth. Once all of these dates are in a single format—Gregorian dates, including proleptic Gregorian years for events in the ancient world, are the conventional rendering—such operations and calculations are computationally trivial. One of the most elegant ways both to express the parameters and represent the results of such a date search or browse is via a visual slider with start- and end-points that can be moved along the range of available dates in the corpus. The standard search interface of projects built using the Kiln tool, upon which EFES is based, includes a basic date slider, coded with CSS and Javascript. Although this date slider does not recalibrate and scale when the time-range of the search results is limited by the application of other facets in the search interface, making it misleading as a visualisation of date, it is still highly intuitive to use and easier than entering numerical dates in text fields.[9]

The industry standard for the date slider view and input field, however, is exemplified by the dynamic search results visualisation in the Pelagios Project's geographical search and visualization tool, Peripleo.[10] In this view, the results of a search are shown, in addition to the weighted geographical distribution on a map, by a simple histogram in the date slider; that is, a series of bars showing the relative proportions of results by chronological period. When the endpoints of this date slider are manipulated by the user, the search results and geographical display dynamically update, showing only those results within the new date criteria.[11] The advantages of this implementation include the fast-updating dynamic display, and the visually pleasing and intuitive histogram. One could criticise the lack of a horizontal scale on

---

[9] See for example the Ancient Inscriptions of the Northern Black Sea (IOSPE) project search page iospe.kcl.ac.uk/search/en/-500/1800/ and the People of Mediaeval Scotland project search page poms.ac.uk/search/.

[10] *Peripleo: The Pelagios Exploration Engine.* (Rainer Simon *et al*. Austrian Institute of Technology, Exeter University, The Open University, University of London, 2016–2019). peripleo.pelagios.org. The "Timerange Selector" is a free-standing open source tool published in 2017, with code at github.com/pelagios/timerange-selector.

[11] An example search showing off the features of the Peripleo date slider for results for the query string "defixio" can be seen at: peripleo.pelagios.org/ui#q=defixio&filters=true.

the date slider, other than the endpoints, and therefore the unclear granularity of the bars in the histogram, but it remains a very elegant, usable visualisation widget.
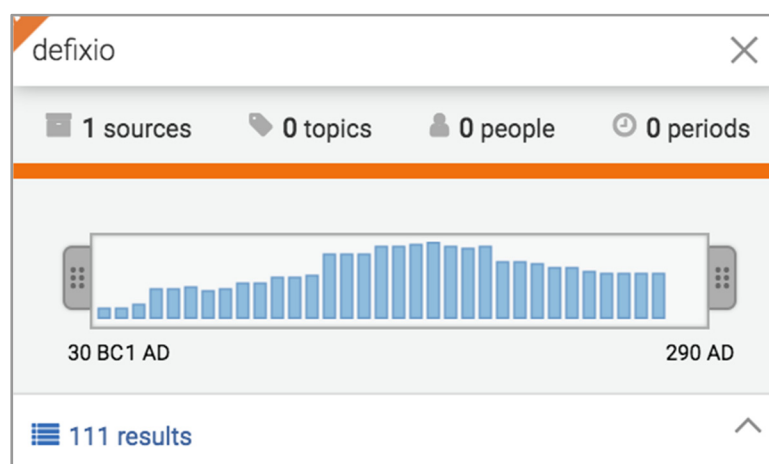

Figure 3: Peripleo time range selector, showing results for "defixio" (© Pelagios Project)

It is a clear priority to explore extending the EFES code-base with the implementation of a Peripleo-like data slider visualisation, especially since the existing code in a partner open source project makes it somewhat low-hanging fruit. We believe that a project intending to customize EFES for their own publication could implement this rather quickly, and ideally contribute the code back to the EFES Git repository for the benefit of the community.

2.2 Geographic organization and visualization

Geographical or spatial distribution of information is arguably the most universal and compatible criterion for organising and dividing bodies of ancient materials, including but not restricted to objects bearing texts, such as epigraphic, papyrological or numismatic editions. It is inevitable that place and space will be among the most important features of any corpus (as arguably for any historical dataset).

It is certainly not the aim of EFES to include full GIS database functionality; many existing tools will perform these functions more effectively, and they may be integrated with the online presentation of a text database or XML corpus in a variety of ways. Given that geographical or spatial data form such an integral part of the organisation of most corpora, and will accordingly be highly desirable as indexing and search features of a digital publication, the visualisation of space and place should form part of the features of the generic platform.

An ancient epigraphic corpus typically may contain several categories of information about the geographic context of the texts and text-bearing objects within it, perhaps including:

- The original or other attested or assumed ancient location of the object, such as a gravestone presumably originally erected in the necropolis found re-used in a late antique fortification wall. A text may also have (one or more) ancient places of origin independent of its object, as in the case of a letter authored in Rome and copied in the provinces, or a copy of a contract placed in an archive.
- The modern place of finding, which although often a proxy for or the best evidence for the original location, is technically a different category of information—and a different category of place, being a modern location usually with coordinates as opposed to a historical settlement or place of known or unknown original name.
- Other modern places and locations in the provenance of the object, including repositories that may hold or have held it, and its current or last known location.
- Places explicitly named or otherwise referenced in the text, whether evidence for the location of the object or further context for people and events mentioned in it.

Some of these places may be principally expressed in a digital edition by means of geographic coordinates in a database, in particular if place of finding or other attested modern locations are captured by a GPS-enabled camera or similar technology. More often, however, a controlled list or gazetteer of named (or otherwise) settlements, monuments and other relevant places is likely to be preferred. Disambiguating places and locations in a text edition or its associated data record to an internal authority list or public gazetteer helps to avoid redundancy and potential errors in the data, and allows common information to be stored in a central or shared resource. Most fundamentally, this would mean that the text records and index files only need to supply the place identifier, while the gazetteer will be the source of coordinates, extent and other geographic information needed to display the items on a map or other visualisation.

In addition to structured text indices and search facets, any or all of these geographical entry-points into the epigraphic data may be sorted, visualised or queried via some form of map interface. The range of texts or objects may be displayed on a map showing distribution, weighting, network of relationships, movement or time. Points or regions on the map may be grouped by chapter, century, type of text, object, or essentially any other facet that is also used to index or search the concept. Ideally the map would be an entry-point into the individual text editions, as well as an overview and visualisation: each item on the map would include a link to the page for the text, as would an entry in an index.

It is our aim to avoid the reinvention of the wheel as far as possible in the development of features in EFES, and of course when it comes to geographical visualisation of ancient and historical data, there is much prior art to draw upon, of which we shall mention only a small selection here.

1. As the gazetteer of record for ancient places, Pleiades (originally founded at Ancient World Mapping Center at UNC but now run by the Institute for the Study of the Ancient World at NYU) captures spatial, historical and philological information, including relationships between places, and provides visualisations of some of these features and relationships.[12] As well as being the key example of a digital gazetteer, which provides the canonical identifiers and key spatial information about ancient places, Pleiades has been at the centre of a community of practice for the querying, visualisation and sharing of such information (Elliott-Gillies).

2. Another project originating at the Ancient World Mapping Centre at UNC is *Antiquity à la Carte*, a custom, web-based map-building tool targeted at scholars wanting to create, customise and publish maps of the ancient world; the platform enables the selection of layers, features and labels to build maps for different delivery media, and is a very useful and user-friendly, entry-level visualisation tool.[13]

3. The Digital Atlas of the Roman Empire (DARE) is a parallel project to, and close collaborator with, Pleiades.[14] As well as providing canonical records and identifiers for ancient locations and places, DARE is the source of a set of widely used historical map tiles showing landscape features such as coastlines, rivers and woodland as they were in the Roman period.[15]

4. The World Historical Gazetteer is a collaborative project to build a database of historical placenames, aligned to existing and new gazetteers, along with a web platform for querying and displaying the records, and an application programming interface for access to shared and interoperable data.[16]

5. The Pelagios Project ran from 2011-19, and built a large community of practice around the representation of historical spatial information as Linked Open Data. Among its key achievements of relevance to this discussion are the Linked Places Format (on which more below), the Recogito geographical annotation and lightweight visualisation tool, and Peripleo (mentioned above), a search and visualisation tool for historical geographical information. As of 2020 the Pelagios Project has been succeeded by a Pelagios Network of independent partners who commit to taking the goals of the community forward (Simon *et al*.).[17]

In practice, the EFES implementation of a geographical indexing and visualisation features will need to be relatively light-weight and highly generic and customisable, while working out-of-the-box for the core constituency of classical epigraphy. The backbone of the spatial information will be an index of the Solr component that drives EFES search functions. This index will contain label, type, and

---

[12] Pleiades: pleiades.stoa.org.
[13] Antiquity à la carte application, awmc.unc.edu/wordpress/alacarte/.
[14] Johan Åhlfeldt's DARE: imperium.ahlfeldt.se.
[15] Klokan Technologies' Roman Empire Map Tiles: github.com/klokantech/roman-empire.
[16] World Historical Gazetteer: whgazetteer.org/about/.
[17] Pelagios Network: pelagios.org.

coordinate information for each geographical identifier, along with the individual texts from which the place is referenced. It is expected that as a rule the EpiDoc corpus will store only the references and identifiers, and geographical and spatial information will be retrieved from the record in the gazetteer.[18]

Figure 4: Peripleo place distribution visualisation, showing results for "defixio" (© Pelagios Project)

The default implementation of this index will be based on a current dump of the Pleiades information, retrieved by the user and copied into the EFES file structure. Instructions will be made available for editors from other epigraphic traditions, to transform the gazetteers of relevance for their corpus into the form expect by the EFES indexer, following the practice of the Linked Places Format developed by the Pelagios Project and the World Historical Gazetteer (Grossner 2018).

The resulting index will then be displayed in an embedded map interface such as the Leaflet Javascript library for interactive maps,[19] using the historical map-tiles developed for DARE. This will enable various views of the data, such as:

- all places (mentioned, findspots, etc.) in the whole corpus;
- all places in the current search or filter results from the search page;
- all places mentioned in or associated with a single inscription.

---

[18] The process will need to be aware of place records in the gazetteers that have no associated coordinates, and that can therefore probably not easily be mapped onto a traditional visualisation.
[19] Leaflet: leafletjs.com.

These places may be represented as pins or icons on the map, which may be selected to pop-up an "infobox" containing information, description or images relating to the text or texts that refer to this place, and a link back to the page for the digital edition of each text. The function of the map for users of the corpus will therefore be exactly comparable to an index of place information, but in a more useful and interactive interface.

This mapping functionality will still be generic and limited to features that the designers have considered and that the software available provides. We will not, for example, attempt to implement full faceted search options within the geographic visualisation interface, enabling users to adjust date or other facets in the search page and see the map re-generated or re-centered on the screen in response; or moving or zooming the map interface and having facets updated based on which texts are represented by the area of the world on the current map. One might in the long term envisage incorporating the entire Peripleo tool as a plugin to EFES to achieve this, but it is more likely that an editor wanting this functionality would be advised to use Peripleo or some other GIS library to deliver their corpus, alongside (or perhaps even instead of) EFES.

As noted, we do not intend to reproduce most of the functions and potential of GIS databases and visualisation tools, but all geographic data (including the indexes containing spatial data imported from gazetteers) we be made available for export in common formats such as comma-separated values (CSV), GeoJSON or RDF for use in third-party tools for visualisation or analysis.

2.3 Open data export and connectivity

Ultimately all of the features described here—indexes, date selection tool, geographic display and other visualisations—can only be implemented in a generic (albeit extensible) way, in a platform designed for many users and corpora with diverse needs. For advanced digital and quantitative analysis and visualisation of data, more technical users and editors alike will need access to the underlying data, in a portable format, using open standards, and licensed for unrestricted reuse and republication, to process and study using their own or published tools and algorithms. In addition to the underlying EpiDoc XML and other formats that drive the online publication, EFES will make available for download the processed indexes, vocabularies and other structured data that feed the internal visualisations.

From any given index, search or view of the data, we plan to enable the download, at a single click or via command to an API, the tabular data behind the index in a common format such as CSV, JSON or RDF, with a clear indication of license. Users may then visualise this data, display it on a digital map, process it using social network analysis, or remix it in other ways in combination with related or enhanced data, and redistribute or publish the results. Such re-use is both essential for reviewers wanting to assess and comment on the published work, and the next generation of editors building on and citing the work of earlier authors, as is and has always been scholarly practice.

Open data is a fundamental component of electronic publishing, and especially when fully findable, accessible, interoperable and reusable (FAIR), it enables new forms of collaboration (Wilkinson *et al.*). This interchange and sharing of digital information comes into its own when relationships within and between datasets, vocabularies, entities and classes of data are expressed in semantic and web-accessible formats. In addition to simply being available online, using open licenses and formats, good Linked Open Data should be expressed in semantic web formats such as RDF, and should be linked to other datasets and collections that also use dereferenceable identifiers—*i.e.* HTTP URIs that point to online information about those specific entities or properties (Kim-Hausenblas). In this model, the entities under discussion (inscriptions, places, people, objects, publications, events or scholarly claims), the predicates or relationships that define or comment on them, and the properties or classes used to categorise and describe them, are all represented by HTTP URIs that reference concepts and entities defined in ontologies, taxonomies, vocabularies and related knowledge systems.

The ontologies and LOD communities of particular interest for exposing and sharing data from an epigraphic or other text corpus include:

1. The Linked Places Format, discussed above, developed by the Pelagios and World Historical Gazetteer projects to share and align gazetteers, relevant for exposing geographical data.

2. Guidelines and ontology made available by the Standards for Networking Ancient Prosopographies (SNAP), for the sharing and alignment of person-databases and exposing references to persons and names (Bodard *et al.*; Bodard).

3. Identifiers and online tools relating to ancient dates and calendars that occur in primary texts, made available through the Graph of Dated Objects and Texts (GODOT) project (Grieshaber "Godot").

4. The ChronOntology Project's ontology of dating and types of dates (Schmidle *et al.*), and PeriodO's taxonomy of named historical periods used in scholarship (Rabinowitz), for exposing and sharing various kinds of dating information, as discussed above.

5. A growing community of epigraphic bodies and projects proposing ontologies and vocabularies for sharing data about inscriptions in particular. Early work exemplifying the needs and potential of epigraphic LOD accompanied the release of Epigraphische Datenbank Heidelberg (EDH) content as open data, and has since been taken forward by the Epigraphy.info community, and in particular the Epigraphic Ontology working group; among the specialised vocabularies available, the EAGLE Network published canonical lists of identifiers under headings including

Inscription Type, Object Type and Material (Grieshaber "Epigraphic Database"; Granados García, §19–21; Tupman).[20]

In order for any of these types of information to be exported as LOD, the underlying EpiDoc files would need to contain references to the relevant identifiers in standard elements, of course. A generic publication tool such as EFES should enable export in a few of the most useful ontologies, but an editor will have to decide which information to encode and therefore make available in this form, and further customisation may be required to add connection to new ontologies and data ecosystems to the platform.

3. Future research

In a very real sense, the services and functionalities offered by a visualisation and publication tool like EFES also serve a prescriptive role in the practice of EpiDoc encoding (itself a set of guidelines for narrowing down the application of TEI XML to ancient texts). The implementation of search, index, visualisation and export features that rely on internal authority lists and consistent encoding in the EpiDoc files, demands a more constrained encoding scheme than the flexible options in the EpiDoc Guidelines (Elliott *et al.*, *Guidelines*) themselves. While a certain amount of flexibility is desirable, given the breadth of disciplinary practices and epigraphic cultures EpiDoc and EFES need to support, there is a demonstrable need—and frequent requests— for prescriptive guidance on encoding features (Bodard-Stoyanova).[21] Just as the TEI Guidelines create a narrow practice for existing academic practice, EpiDoc limits TEI to even narrower practices, the expectations of a tool like EFES further delimit EpiDoc, and a single project will no doubt define their own, extremely consistent internal encoding scheme.

It is a truism of the Digital Humanities that open standards are essential for responsible and sustainable online publication of digital editions and other research data. It is equally true however that the tools used to create, prepare, analyse, communicate and disseminate digital data and editions are part of the research process, and so ought to be available, well documented, and reusable by readers and reviewers (Turska; Liuzzo 48-9). Since electronic publication is in effect a form of data visualisation, the open distribution of "source code" should include not only EpiDoc XML files, but also XSLT stylesheets and the customised version of the platform such as EFES that was used to generate indices, search and visualisations. It is recommended that editors who create their publication using EFES do so by "cloning" the Git repository into a new fork of the platform, and make all customisations and

---

[20] See also Epigraphy.info: A Collaborative Environment for Digital Epigraphy: epigraphy.info; Epigraphic Ontology Group: groups.google.com/forum/#!forum/epont; EAGLE Vocabularies: eagle-network.eu/resources/vocabularies/.

[21] See instances of such requests on the Markup list (lsv.uky.edu/archives/markup.html) or the EpiDoc Feature Requests tracker (sourceforge.net/p/epidoc/feature-requests/) at any given time.

further enhancements available to readers; some of these improvements may be offered back to the core EFES code-base by means of "pull requests."[22]

The description of the EFES platform offered in this paper, and the proposed enhancements to its functionality, are by no means intended to suggest a complete, ideal, stable and perfectly sustainable publication tool for the future. No software package, least of all free and open source software, can make that claim. There remain several issues that EFES needs to address in the longer term, in particular complex dependencies on several other open source tools, some of which are no longer regularly updated and supported. There are occasional and unpredictable issues with installation on certain versions of certain operating systems (which of course are not unique to this tool), and a current dependency on a legacy version of the Java Virtual Machine, that will need to be resolved sooner or later. Although EFES is in principle deployment-ready, with a built-in Jetty server instance allowing all the other components to be displayed in a web browser, further support and guidance is currently needed for a user intending to deploy it independently, for instance on a commercial web host without institutional and technical support.[23]

It is our hope and intention that the community of EpiDoc editors and developers will take ownership of EFES as part of the EpiDoc tool set (alongside the Guidelines, Schema and Reference Stylesheets), and that future improvements to the platform will come about as a result of funded project-specific requirements fed back into the code base. Several funding applications in the pipeline have already included EFES development in their budgets, and new scholars request advice or training in the use of the platform on an almost weekly basis. Collaboration, both with individual scholars and with related large publication and infrastructure projects, is essential for the continued development and sustainability of any open source platform such as EFES, and the engagement and generosity of interested scholars and developers in contributing to our collective work has so far been exemplary.

Corpora and Web Resources

CII. *Cretan Institutional Inscriptions*. Irene Vagionakis, Venice Centre for Digital and Public Humanities, 2020. cretaninscriptions.vedph.it.

ECG. *Epigraphic Corpus of Georgia*. Demo version. Nino Doborjginidze *et al.*, Ilia State University, 2019. epigraphy.iliauni.edu.ge.

EDV. *Epigraphic Database Vernacular*. Nadia Cannata, Luna Cacchioli, Alessandra Tiburzi. Sapienza University of Rome, 2020. edv.uniroma1.it.

IAph. *Inscriptions of Aphrodisias*. Joyce Reynolds, Charlotte Roueché, Gabriel Bodard. King's College London, 2007. insaph.kcl.ac.uk/iaph2007/.

IOSPE. *Ancient Inscriptions of the Northern Black Sea*. Irene Polinskaya, Askold Ivantchik, *et al.*, King's College London, 2015–2017. iospe.kcl.ac.uk.

---

[22] An example of a pull request that integrated dozens of improvements by Irene Vagionakis into the EFES code-base can be found at github.com/EpiDoc/EFES/pull/54.

[23] Current web hosting documentation to be found in Vagionakis, "Host server setup" (in Yordanova): github.com/EpiDoc/EFES/wiki/Host-server-setup.

IRT. *Inscriptions of Roman Tripolitania*. Joyce Reynolds and John Bryan Ward-Perkins, enhanced electronic reissue by Gabriel Bodard and Charlotte Roueché. King's College London, 2009. inslib.kcl.ac.uk/irt2009/.

RIB. *Roman Inscriptions of Britain.* Robin G. Collingwood and R. P. Wright, enhanced electronic reissue by Scott Vanderbilt. 2014–2020. romaninscriptionsofbritain.org.

Bibliography

Bodard, Gabriel. "Linked Open Data for Ancient Names and People." *Linked Open Data for the Ancient World: A Practical Introduction,* edited by Paul Dilley, Ryan Horne and Sarah Bond. *ISAW Papers* 20, 2020 (forthcoming). dlib.nyu.edu/awdl/isaw/isaw-papers/20/.

Bodard, Gabriel and Simona Stoyanova. "Epigraphers and Encoders: Strategies for Teaching and Learning Digital Epigraphy." *Digital Classics Outside the Echo-Chamber: Teaching, Knowledge Exchange and Public Engagement,* edited by Gabriel Bodard and Matteo Romanello. London: Ubiquity Press, 2016, 51–68. dx.doi.org/10.5334/bat.d.

Bodard, Gabriel, Hugh Cayless, Mark Depauw, Leif Isaksen, K. Faith Lawrence, Sebastian Rahtz. *SNAP Cookbook*. King's College London, 2014. snapdrgn.net/cookbook.

Burghart, Marjorie. "The TEI Critical Apparatus Toolbox: Empowering Textual Scholars through Display, Control, and Comparison Features." *Journal of the TEI* 10, 2016. doi.org/10.4000/jtei.1520.

Elliott, Tom and Sean Gillies. "Digital Geography and Classics." *Digital Humanities Quarterly* 3.1, 2009. digitalhumanities.org/dhq/vol/3/1/000031.html.

Elliott, Tom, Gabriel Bodard, *et al*. *EpiDoc Guidelines*. Version 9. Stoa Consortium, New York University, 2008–20. epidoc.stoa.org/gl/latest/.

Elliott, Tom, Zenata Au, Gabriel Bodard, Hugh Cayless, Carmen Lanz, Faith Lawrence, Scott Vandebilt, Raffaele Viglianti, *et al*. *EpiDoc Reference Stylesheets*. 2008–20. github.com/EpiDoc/Stylesheets.

Granados García, Paula Loreto. "Hesperia" (review). *RIDE* 7, 2017. ride.i-d-e.de/issues/issue-7/hesperia/.

Grieshaber, Frank. "GODOT. Graph of dated objects and texts, building a chronological gazetteer for antiquity." *Epigraphy Edit-a-thon. Editing chronological and geographic data in ancient inscriptions, April 20–22, 2016*, edited by Monica Berti, University of Leipzig, 2017. nbn-resolving.de/urn:nbn:de:bsz:15-qucosa-221532.

Grieshaber, Frank. "Epigraphic Database Heidelberg – Data Reuse Options." University of Heidelberg, 2019. doi.org/10.11588/heidok.00026599.

Grossner, Karl. "Contributing in Linked Places format." *World-Historical Gazetteer*, 2018. whgazetteer.org/2018/09/11/lp-format/.

Kim, James G. and Michael Hausenblas. "5 ★ OPEN DATA." 2015. 5stardata.info.

Liuzzo, Pietro Maria. *Digital Approaches to Ethiopian and Eritrean Studies*. Aethiopica Supplement 8, Harrassowitz Verlag, 2019. doi.org/10.2307/j.ctvrnfr3q.

Rabinowitz, Adam *et al*. "Making Sense of the Ways We Make Sense of the Past: The PeriodO Project." *Bulletin of the Institute of Classical Studies* 59.2, 2016, 42–55. onlinelibrary.wiley.com/doi/full/10.1111/j.2041-5370.2016.12037.x.

Rosselli Del Turco, Roberto. "The Battle We Forgot to Fight: Should We Make a Case for Digital Editions?" *Digital Scholarly Editing: Theories and Practices*, edited by Matthew Driscoll and Elena Pierazzo, Open Book Publishers, 2016, 219–238. doi.org/10.11647/OBP.0095.

Sahle, Patrick. "What is a Scholarly Digital Edition?" *Digital Scholarly Editing: Theories and Practices*, edited by Matthew Driscoll and Elena Pierazzo, Open Book Publishers, 2016, 19–40. doi.org/10.11647/OBP.0095.

Schmidle, Wolfgang, Nathalie Kallas, Sebastian Cuy and Florian Thiery. "Linking Periods: Modeling and Utilizing Spatio-temporal Concepts in the ChronOntology Project". *Computer Applications and Quantitative Methods in Archaeology* (CAA), 2016. Oslo.

Simon, Rainer, Leif Isaksen, Elton Barker and Pau de Soto Cañamares. "Peripleo: a Tool for Exploring Heterogeneous Data through the Dimensions of Space and Time." *Code4Lib* 31, 2016. journal.code4lib.org/articles/11144.

Sopracasa, Alessio and Martina Filosa. "Encoding Byzantine Seals: SigiDoc." *Atti del IX Convegno Annuale AIUCD. La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica*. Università Cattolica del Sacro Cuore, 2020. aiucd2020.unicatt.it/aiucd-Sopracasa_Filosa.pdf.

Tupman, Charlotte. "Where can our inscriptions take us? Harnessing the potential of Linked Open Data for epigraphy." *Epigraphy in the Digital Age: Opportunities and Challenges in the Recording, Analysis and Dissemination of Epigraphic Texts,* edited by Isabel Velázquez Soriano and David Espinosa Espinosa. Archaeopress, forthcoming 2021.

Turska, Magdalena, James Cummings and Sebastian Rahtz. "Challenging the Myth of Presentation in Digital Editions." *Journal of the TEI* 9, 2016. doi.org/10.4000/jtei.1453.

Wilkinson, Mark D., Michel Dumontier, Barend Mons, *et al*. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3, 2016, 160018. doi.org/10.1038/sdata.2016.18.

Yordanova, Polina *et al*. *EFES User Guide.* University of London via Github, 2017–18. github.com/EpiDoc/EFES/wiki/User-Guide.