Greek Literature, the Digital Humanities, and the Shifting Technologies of Reading

Other

How to cite:

For guidance on citations see FAQs.

Version: Accepted Manuscript

Link(s) to article on publisher’s website:
http://dx.doi.org/doi:10.1093/oxfordhb/9780199935390.013.45
http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935390.001.0001/oxfordhb-9780199935390-e-45?rskey=aeAMNi&result=1

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.
Greek Literature, the Digital Humanities, and the Shifting Technologies of Reading
Elton Barker and Melissa Terras

ABSTRACT
Contrary perhaps to expectation, Classical studies is at the vanguard of the latest technological developments for using digital tools and computational techniques in research. This article outlines its pioneering adoption of digital tools and methods, and investigates how the digital medium is helping to transform the study of Greek and Latin literature. It discusses the processes and consequences of digitization, explaining how technologies like multispectral imaging are increasing the textual corpus, while examining how annotation, engagement, and reuse are changing the way we think about “the text”. It also considers how the digital turn is reinvigorating textual analysis, by exploring the broader ecosystem, within which the digital text can now be studied, and which provides enriched contexts for understanding that are constantly shifting and expanding. Classical literature in the digital age has the potential to both challenge dominant modes of thinking about antiquity and disrupt traditional ways of doing research.

KEYWORDS
Classics, Greek, Latin, digitization, digital text, annotation, engagement, reuse, ecosystem, multispectral imaging

1 INTRODUCTION
1.1 Scenes from inside the classics stronghold
In January 1998 a query posted to the Liverpool Classics List asked whether there had ever been an English translation of the Suda. A product of the thriving trade in manuscripts—the information technology of its day—the Suda (“stronghold”, “fortress”) was the 10th-century Byzantine equivalent of Wikipedia, a massive compendium of knowledge covering Greek and Roman antiquity in its entirety, as well as assorted Biblical and Christian material. Fast-forward a millennium and scholars from various backgrounds and with different expertise answered the call-to-arms to help compile the first English Suda—in doing so they transformed not only the text but the process of scholarship itself. Ever since the end of the 14th century, the Suda had been passed down in traditional hard-copy book editions; now, it was to be a digital project. From the beginning, Suda On Line (SOL) was not just about translating the Suda but developing a new paradigm for scholarly publication, from purpose building a computational infrastructure for compiling and working with the submitted material, to making the editing process entirely open and open-ended.
This meant that anyone who could translate ancient Greek was eligible to contribute, regardless of formal credentials; submitted entries would be instantly accessible on the site (though marked as ‘draft’ until vetted); moderation and editing would be done transparently; and “no entry, however well translated and annotated, would ever be considered off-limits for future improvement”.¹

For us the history of the Suda, particularly its rebirth online, raises a number of issues that are important for thinking about ancient Greek literature, the Digital Humanities, and the shifting technologies of reading. At one level it is clear that technologies change the way we both produce and consume information (“the medium is the message”: McLuhan 1964). An encyclopaedia as large and ambitious as the Suda was only possible because of the advance in manuscript production and the concurrent preservation of the knowledge out of which it was compiled (Dickey 2007). In turn the SOL project took advantage of the new online space to bring together scholars in an ongoing collaborative effort that would hardly have been feasible even a decade before (Mahoney 2009, Hutton 2014). At another level the technological shift heralds a change not only in production but also in the consumption of information: the way we read, as much as the way we write, is changing.

This review article attempts to trace some of those developments and reflect on related questions and concerns. We consider the areas in which classical studies has played a key role in pioneering technology and the impact that has had on the adoption of digital tools and methods. In The Digital Text we address the process of digitization itself, the texts that have been digitized, the ways they have been digitized, and the consequences following on from the choices made. This includes sketching out the new technologies that are enabling the publication of new material and increasing our textual corpus. In The Classics Ecosystem, we place the digital text in and against other openly accessible and easily available materials online, and assess how these enriched contexts can be brought to bear on textual analysis. We conclude by contemplating what the future might hold for the study of ancient Greek texts in Classics 2.0.

Throughout our aim is to address the transformative nature of this new medium. In particular we explore how the digital turn is reinvigorating the analysis of texts, through enhancing their dialogic potential, pushing the boundaries of what can be considered a text, and establishing frameworks for understanding that are constantly shifting and expanding. In other words, this article isn’t about the claims of modern technologies to provide us with more accurate information and more of it; instead it’s about how the digital can make those claims more transparent, test assertions of authority, and

¹ http://www.stoa.org/sol/history.shtml (last accessed 31.05.2015).
above all help us unpick the carefully woven “texts” of previous generations. Ultimately, we suggest, classical literature in the digital age has the potential not only to challenge dominant modes of thinking about antiquity, but even to disrupt traditional ways of conducting academic research.

1.2. Classics Computing

Classicists represent one of the foremost groups in the burgeoning discipline of Digital Humanities and are pioneering work in all fields of enquiry.² It should perhaps not be all that surprising that classicists have embraced digital technology: the study of Greco-Roman antiquity is by nature a data intensive research activity, which requires a variety of approaches to analyse and understand the different literary and material culture sources that remain available to us. Digital tools and methods are seen as an important complement to more traditional methodologies, given the affordances they offer in the searching, aggregation, visualization, and analysis of complex data.

In fact classicists were early adopters of digital technologies and have been at the forefront of technological innovation in research ever since. The first acknowledged “Humanities Computing” project, the Index Thomisticus, was a collaboration between the Italian Jesuit Priest, Father Roberta Busa, and IBM that started in 1946 and provides a “concordance of all the words of Thomas Aquinas, including conjunctions, prepositions, and pronouns, to serve other scholars for analogous studies” (Busa 1980: 84). By the late 1960s and early 1970s there were already concerted attempts to harness computational technology for developing electronic versions of classical primary source material—the creation and processing of electronic text being the most feasible application at this time given the limited processing power of computers. A number of projects that emerged during this early phase have adapted to changes in both the technology and scholarly environment to still be with us today.³ Foremost among these is the Thesaurus Linguae Graecae (TLG). Founded in 1971, the TLG built up a “computerized databank of Greek Literature”, a concept that was “extraordinary since no one until then had considered the marriage of classical scholarship with the rapidly emerging new technologies”.⁴ Building on this databank, David Packard’s Ibycus system (used to process, search and browse the TLG Greek texts) provided sophisticated search and retrieval capabilities which could, by 1986, search “the entire corpus of Greek Literature on the TLG… in just over 25 minutes”, making “a time-consuming, tedious, and mundane task as simple as pressing a few keys and waiting a few minutes for the results to be displayed on-screen or written to a disk file” (Hughes 1986: 1). Analysing words or phrases over single texts or entire corpora,

---

² For an up-to-date list of projects: https://wiki.digitalclassicist.org/Category:Projects (now curated every month).
³ Such as L’Année philologique, which has been cataloguing classical scholarly work in all languages since 1924, but which switched from print to an electronic database in 1969. For an overview of the history of the use of computing in classics, see: Brunner (1993); Crane (2004).
⁴ http://www.tlg.uci.edu/about/history.php (last accessed 31.05.2015).
which used to be the work of a lifetime—witness the concordances of major authors in nineteenth-century scholarship—could now be done at the press of a button. The TLG remains highly revered and commonly used by scholars, though the extent to which it will continue to dominate textual consumption in the rapidly evolving digital landscape is questionable, as we shall see.

Come the late 1980s advances on this basic “digitization” of texts were being developed. In 1982 The Duke Data Bank of Documentary Papyri (DDBDP) began compiling an “electronic corpus of published Greek and Latin documents written on papyrus, ostraca, or wooden tablets”,\(^5\) migrating to a website in 1996/7, and beginning the conversion to TEI compliant EpiDoc XML in 2007 (see section 2.1 below). Launched in 1987, the Perseus Digital Library was founded to collect and present materials for the study of ancient Greece, first published as two CD-ROMs and then in an online form in 1995. Since then, and massively expanding in scope to cover not only ancient Greek and Latin materials but also Arabic texts, Perseus promotes “a practical experiment in which we explore possibilities and challenges of digital collections in a networked world”.\(^6\) Classicists have also been interested in sharing their resulting research electronically. The “second oldest online scholarly journal in the humanities”,\(^7\) the Bryn Mawr classical Review (BMCR) began life as a listserv before becoming in 1990 an online, open access journal, in which form it still flourishes today.

The technologies and working practices that these initiatives have helped develop or promote—above all annotation, networked collections and open access—are part of the story of where Digital classics is now that we explore in this article. It is a complex and fast changing picture. The increase in performance and availability of new technologies and the growing World Wide Web (Naughton 2000); the ubiquitous adoption and falling costs of personal computers and accessories; and the socio-economic climate that has encouraged and supported the preservation and dissemination of cultural heritage, have all contributed to a wave of digital experimentation and development (Terras 2010a). “Web 2.0”—the transition from a read-only to a co-creative experience—has enabled classicists (among humanists more generally) not only to digitize data but also to customise tools that allow searching, aggregation, visualization, and analysis. It has also helped to develop a shared academic infrastructure in the community. Community is the right word—young researchers work alongside esteemed scholars as members of a Digital Classicist network, whose range of activities include a discussion list,\(^8\) blog,\(^9\) wiki\(^10\) (containing supporting

\(^5\) [http://papyri.info/docs/ddbdp](http://papyri.info/docs/ddbdp) (last accessed 31.05.2015).

\(^6\) [http://www.perseus.tufts.edu/hopper/about](http://www.perseus.tufts.edu/hopper/about) (last accessed 31.05.2015).

\(^7\) [http://bmcr.brynmawr.edu/about.html](http://bmcr.brynmawr.edu/about.html) (last accessed 31.05.2015).

\(^8\) Archives are available at [http://www.jiscmail.ac.uk/lists/digitalclassicist.html](http://www.jiscmail.ac.uk/lists/digitalclassicist.html).
material describing tools, projects, reports, and research), and a seminar series hosted in London, Oxford, Berlin, and Boston. Together they are impacting on not only the development of digital resources, materials and tools, but also the very reading and understanding of the texts that underpin our subject.

2. THE DIGITAL TEXT

Our exploration of Greek Literature and the shifting technologies of reading in the digital age take us from the micro to the macro, from identifying a single ligature to conceptualizing vast literary traditions. In these ways it is not only the text that is being renewed and reinvented; ways of reading it too are coming under intense pressure to change. What those changes are, and what issues those changes in turn raise, is the subject of our investigation into the form and corpus of the digital text.

2.1 Annotation, engagement, re-use

The digital text has played a key role not only in positioning classics studies at the forefront of Humanities Computing endeavours, but even with regard to developing those technologies; in turn, those developments have impacted on research culture and practice. Some developments have resulted in a consensus about the text and ways of researching it; others have led to disputes and division whose repercussions are still being worked out. Either way, one side effect of the digitization agenda has been to question the very notion of a text and especially what one can do with it, or how one can analyse it. The debates around the digital text can be usefully brought into focus by considering the different paths taken by two of the longest running and far reaching projects: the TLG and Perseus.

As stated on its website, the goal of the TLG is to “create a comprehensive digital library of Greek literature from antiquity to the present era” and “combine the traditional methodologies of philological and literary study with the most advanced features of information technology”. Seen in this way the TLG has surely succeeded in its aims. Not only does it now host a comprehensive collection of texts written in Greek—ranging from the time of Homer to the fall of Constantinople in AD 1453—but, as we have noted, its digitization of texts has changed the way scholars and students interact with the material, through the development of word and phrase searches. The kind

---

9 http://www.stoa.org/.
10 http://www.digitalclassicist.org/.
12 https://www.tlg.uci.edu/about/ (last accessed 31.05.2015).
of low level “text mining”\textsuperscript{13} that the TLG’s combination of traditional methodologies with information technology has facilitated is now an ordinary part of scholarship.

Yet the TLG is under increasing strain as an outmoded kind of scholarship fit for the world of print, where the vigorous policing of content insured against the infringement of individual scholars’ copyright; this careful stewardship of the text can shade into asserting exclusive control over the ways in which it is to be consumed online. For example, while the TLG-team has “developed its own search engine” for its online corpus,\textsuperscript{14} other, better, search engines are available. Peter Heslin’s \textit{Diogenes} tool enables far more complex and nuanced searching of the TLG’s databases, including morphological analysis and regular expressions (patterns of text). Dedicated to the corpus of early Greek hexameter poetry, the \textit{Chicago Homer} uses computational search to make the distinctive traditional referential features of ancient Greek epic accessible to readers with and without Greek. Extending search capabilities even further, the \textit{eAQUA} project uses digital “text mining” methods to trace and explore textual “re-use” of ancient Greek corpora in order to ascertain the influence of individual texts (such as quotations of Plato) on later traditions (Büchler et al. 2013).

Of greater import, however, is the extent to which these independent search tools and methods are critically limited in their application to the texts of the TLG by the latter’s licensing arrangements. The TLG’s closed access policy not only means that their texts are off-limits to anyone who hasn’t paid an upfront fee; leaving aside those researchers not fortunate enough to work within universities that have an institutional license, even those that do have access are unable to do anything with the digital texts other than search through the corpus. In this sense the TLG is a walled garden. Not only are its texts unable to make use of now-standard capabilities in the open-access world; by keeping them behind a paywall, the TLG makes it almost impossible for anyone to build services on top of its collection. (Peter Heslin has managed by carefully separating his search tool from the texts for which it is used.) The point here is as much about being open to allowing \textit{use} of a text, as it is about having open \textit{access} to it.\textsuperscript{15} In turn this points to a seismic shift in scholarly culture—from policing privilege to enabling fair use. In the digital medium resources and materials are generally optimized through other individuals or groups taking on development and taking it in new and interesting directions not envisaged by its creators or curators.

\textsuperscript{13} Text Mining is a process of deriving high quality information from the automated analysis of text through using statistical analysis to identify patterns and trends: Felman and Sanger (2006).

\textsuperscript{14} https://www.tlg.uci.edu/about/ (last accessed 31.05.2015).

\textsuperscript{15} A similar point has recently been made in Helma Dik’s considered \textit{classical Journal Review} of the Loeb (Digital) classical Library. As she puts it, “Instead of following the TLG’s lead, the Loeb Library could make its collection a \textit{research} library as well as a reading library” (our italics).
One example of a digitally-based open data/open systems model has been promoted by the Perseus Classical Library and is currently being energetically extended by its director, Greg Crane, in the Leipzig Open Philology project (Crane 2014). The open agenda is not without its own issues. Both projects are able to host open access texts only because their editions are out-of-copyright. As a consequence, they do not take into account either philological advances over the past century or indeed the discovery of more material (more on which below); for the latest scholarly editions, researchers must still rely on the curated world of books. Nevertheless, there are at least three ways in which, putting the open access issue to one side, Perseus and Open Philology are developing the use of primary sources qua digital texts that not only enhance but are transforming the research process. These are annotation, engagement and reuse.

Annotation is a fundamental marker of the difference between the analogue and digital text. The idea of adding notes or marginalia to documents goes back at least as far as the mediaeval manuscripts from which we reconstruct our corpus of ancient Greek literature (Jackson 2001). But in a digital context annotation refers to metadata (some kind of comment, explanation, or presentational “markup”), which can be attached to a text (or, for that matter, an image or other data). Often annotations refer to a specific part of the original data. Since 2004, Perseus has been employing a particular kind of textual encoding or markup known as TEI, or the Text Encoding Initiative, the community-standard encoding methods for machine-readable texts. A simple common schema provides additional information to the words on the page, including: in the header, author and text information (in the manner of a manuscript “incipit”); and in the text itself, both section demarcations (book, chapter, paragraph, etc.), which relate to scholarly canonical citations, and other information that might be of assistance for reading or interpreting the text, such as people- or place-name data. Information of this kind can give a text added value, whereby the traditional “hard copy” text (whether a manuscript or book) is provided with supplementary semantic information. With the text thus enriched, reading is enhanced, so that, where before one could search for single words or phrases through individuals works or large text corpora, now more complicated analysis is possible. (We’ll come to an example of how the named-entity field “place names” can be used in a moment.)

While TEI enables semantic annotation of texts, linguistic understanding can be promoted by using “treebanking” schema. Syntactic annotations record information about the relationships between the words and organize them into tree-like structures. The resulting “Treebanks” can let us see phenomena such as: the changing subjects and objects that a given verb takes over time; sentence structure (e.g., subject-verb-object vs. subject-object-verb); and the individual styles of particular
authors, genres, and periods. While Perseus has pioneered the morphological markup of ancient texts, the Historical Language e-Learning Project at Leipzig is currently extending the process to students of ancient languages, so that linguistic annotation becomes embedded in learning, and vocabulary and grammar can be grasped in a dynamic way.

In fact this initiative is one part of a more general shift towards using digital annotation to foster a more engaged kind of learning—often termed “citizen science”.16 Advances, say, in handwriting recognition technology are, as yet, unable to create automated annotations of high enough quality to not require further human intervention and correction: to make further progress, and to fully exploit the capacity of the digital text to carry enriched information, help is needed.17 Yet engagement is also seen as a value in and of itself. Given the costs attributed to hard copy texts, editing has always been the preserve of scholars: indeed, annotations—in the form of handwritten scrawls—on library books have always been (rightly) frowned upon. As well as being inherently annotatable, however, digital texts are also open to a much wider group of people to doing the annotation, with students taking on the role of editors and contributing early and often to the enrichment of semantic content. The Perseus-spawned Perseids project, for example, has developed a collaborative editing platform to enable both scholars and students to annotate quotations and text reuses of fragmentary authors, all within a framework of rigorous and transparent peer-review and credit mechanisms and strong editorial oversight. But Perseus is far from being alone in this change in practice. One of the most interesting examples is the Homer Multitext project, whose work even has repercussions for thinking about what a text is (Crane 2014). Unlike printed editions, which offer a reconstruction of an original text as it supposedly existed at the time and place of its origin, this web-based platform manages any number of textual variants and enables users to reconstruct the Homeric tradition as it was being textualized in the era of mediaeval manuscripts, including the transcription of important, previously untranslated, scholia. Fundamental to this transcription process are students, whose new generation of commentaries “address the actual problems that readers confront as they struggle with linguistically or culturally challenging texts”.18

Thus we see two interrelated phenomenon emerging that are changing our relationship to the text. First, the idea of the text itself as a stable, unchanging artefact is being challenged. In truth the text

---

16 A term used interchangeably with “crowdsourcing” to describe the use of online activities and behaviour to aid in large-scale ventures such as tagging, commenting, rating, reviewing, text correcting, and the creation and uploading of content in a methodical, task-based fashion. See Terras (Forthcoming) for an overview of the use of citizen science and crowdsourcing in cultural and heritage contexts, and Ridge (2014) for a range of projects operating in this space.
17 Though there are projects in the Digital Humanities working on this particular method: e.g. Transcriptorium, which is using crowdsourced transcriptions from Transcribe Bentham to train Handwriting Recognition Technology. 
18 http://www.perseus.tufts.edu/hopper/research/current (last accessed 31.05.2015).
has always been far less written in stone than the impression our books tend to give, as a cursory
glance at any apparatus criticus reveals. But the new digital medium brings to the surface and
makes apparent the extent to which just how far the text is a living organism, dependent on the
medium of the time (leaving aside the issue of interpretation, which is another matter entirely). In a
ture digital edition of fragments, such as that being developed by the Leipzig Open Fragmentary
Texts Series, dynamic links lead directly back to the source text from which each fragment is
drawn, as well as to citations and annotations about them. As we are beginning to see, the text is
more open to a user and the persona of that user is changing, as more people have access to it and,
importantly, have the capacity to annotate it and even help edit it (Crane 2014, Hutton 2014).

The digital text, with its new accretions and enriched semantic context, is not only usable but can
also be re-used. That is to say, open access digital texts can be used in ways not only not undertaken
by the data provider but also not even envisaged by them. In fact it can be argued that, whereas the
printed book relies on its longevity and inviolability on being kept in a pristine, unused state, the
digital text must be used and re-used—be useful—in order to survive the transformations in rapidly
changing technologies.

A case study that demonstrates some of the possibilities of textual re-use, and also touches upon the
other issues raised in this section about annotation and collaboration, is the Hestia project. Inspired
by the idea that Herodotus uses the new medium of his age—writing—to represent the geopolitical
world in ways that had not been possible before (Goldhill 2002, Barker 2009), Hestia uses (among
other strategies) the new medium of our time, digital technology, to investigate the cultural
geography of the ancient world through its representation in the Histories. Hestia’s digital strategy
has largely been experimental—to explore the potential of digital resources for studying the
geospatial structure of Herodotus’s narrative, and develop the means of reading text and maps
alongside each other (Barker et al. 2013). Yet, the extent of that digital experimentation was greatly
facilitated by not having to start from scratch: a digital text was already available from Perseus and
crucially open to re-use (under a Creative Commons Attribution-Sharealike (CC-BY-SA) license).
Furthermore, that text was enriched by annotation—the vast majority of place references were
already tagged within the English translation. Various issues had to be addressed, including:
updating the TEI markup; aligning the Greek text with the English translation being used; and
cleaning up the data associated with the place names (Barker, Isaksen and Ogden, Forthcoming).
Nevertheless, (re)using the Perseus text enabled the Hestia team to move more or less directly to
analysis.
Moreover, the kind of research that the Hestia team was able to undertake was facilitated, if not engendered, by virtue of working with a digital text. Being able to extract and aggregate the place-name information from the Perseus text represented more than just having a database of place names; it was now possible to try out different visualizations to help inquiry and understanding. In effect this meant creating maps out of the text, not so much to illustrate the places that Herodotus mentioned in some kind of cartographic form as to enable their interrogation; maps here are understood as a means of analysis, not the end product (Bodenhamer 2015). The theory behind this use of maps can be traced back to the spatial turn of the 1980s, specifically the counter-cartographic movement that challenged the object of mapping to “produce a ‘correct’ relational model of the terrain” (Harley 1989: 4). Current digital technologies are well placed to exploit the spatiality of texts, though familiar webmapping applications such as GIS and GoogleEarth tend to reproduce the dominant Cartesian modes of spatial representation (Harris et al. 2011). Instead, literary maps, such as those developed by Franco Moretti in his analysis of the modern European novel, offer the potential to make visible the “ortegebunden, place-bound nature” of a text, with “its peculiar geometry, its boundaries, its spatial taboos and favorite routes”; maps, then, that could “bring to light the internal logic of narrative: the semiotic domain around which a plot coalesces and self-organizes” (Moretti 1998: 5).

Along these lines, the Hestia team produced a series of network graphs that brought to the fore the underlying ways in which Herodotus constructs space in terms of relations between peoples and places rather than according to topographic ‘reality’. Figure 1 provides an example of the rich relatedness of places (and their peoples) in Book 5 of the Histories. The size of each node relates to the frequency of a place’s mention; its centrality shows a place’s role in the network—those places mentioned more often in relation to others are more central. The line (‘edges’) indicate relationships, arrows the direction of the relationship—the thicker the edge, the stronger the relationship. By visualizing the data in this form, alternative spatial relationships become discernable, primarily those built around action and influence rather than geographic proximity. On the one hand, the graph suggests the beginning of a radical split between east and west, as indicated by the clustering of east and west entities. On the other, it challenges that very division by locating the centrality of Miletus as the hinge between (or gateway through) these two clusters (suggesting Greek involvement, if not culpability, in the separating out of these two spheres), and draws attention to internal conflict (such as the overwhelming influence Sparta exerts on Athens).  

19 For a fuller account, see Bouzarovski and Barker Forthcoming. These data, while aggregated and visualized in a digital form, relied on close textual study. At the same time, the identification of places (and proxies as places) and the categorization of the relationships to each other owed much to computational approaches tried out elsewhere as part of the Hestia project.
diagrammatic form offers alternative pathways through Herodotus’s space that subvert our long-standing cartographic tradition and suggests ways of reimagining ancient spatial understanding (cf. Purves 2010).

Figure 1. A network graph showing the relatedness of places in Herodotus’s Histories 5

Of course, additional challenges remain, not least the problem in using maps to represent complex (humanities) data: the graph (especially in its non dynamic print form) is difficult to read. Nevertheless, work on the Hestia project suggests one potential new use of the text in a digitized form and, concurrently, new ways of thinking about visualizations, as the means and starting point of analysis, not the end product. Digital representations of texts can be used to create different, more challenging maps that can disrupt our inherited notions of geopolitical space. In turn this suggests the potential of not only new research questions and knowledge (counter-cartographic maps based on texts) but also a new method, based on the dialogic encounter of text and image, map and space. It is important to note here that the methodology employed in the Hestia project was consciously multifaceted, where the textual and digital, geographical and classical, were brought together in some sort of tense relationship. Using the digital medium does not mean ditching the text. Far from it: it’s even more imperative to keep it in view (Barker and Pelling Forthcoming).
In this section we have discussed the transformation of the text into a digital form, explored its limitations and boundaries, as well as indicated some of the new paths taken to make optimum use of this new medium. In the next section we turn to the new textual material that is coming to light through the use of digital technology.

2.2 Expanding the Corpus

It is tempting to think of classical literature as a bounded corpus, with the manuscript tradition of extant texts well documented and understood. But in fact the range of texts and their precise form is constantly in flux, with new material coming to light from a variety of sources. Renovation work on a mummy led to the discovery of our longest, and most lively, fragment of Archilochus (Merkelbach and West 1974). Apart from one poem,20 most of our knowledge of Sappho is pieced together from physical remains, supplemented by quotations in earlier authors, with new, contested, finds still emerging (Obbink 2014). Many of these discoveries come from the sands of Egypt: the Sackler Library at Oxford houses millions of papyri fragments unearthed in the late nineteenth century from an ancient rubbish dump at Oxyrhynchus in Egypt (Parsons 2007), witnesses whose evidence still needs sifting through. A recently transcribed fragment contains thirty previously unknown lines by Archilochus, describing a pre-Trojan War encounter between the Greeks and Telephus, king of Mysia (Obbink 2006). As new sources multiply, digital technologies can enhance our reconstruction of texts in a variety of ways: formats that allow new forms of publication and knowledge sharing; platforms that allow new transcriptions and editions; and imaging technologies that allow the reading of previously indecipherable material.

A prime example of sources brought together, documented, transcribed, and made available to the scholarly community, is the Epigraphic Database Heidelberg. This freely available database presently contains some 68,500 Latin and bilingual (Latin-Greek) inscriptions from the Roman Empire. Importantly its inscriptions are described in a standardized text-encoding framework, derived from TEI, specifically designed for epigraphical sources, known as Epidoc. Providing guidelines for encoding scholarly ancient documents, Epidoc enables both interoperability (information exchange with other online resources) and reuse (by encouraging publication to the widest possible audience under an open license).21 As a collaborative endeavour by epigraphers and papyrologists, Epidoc is changing the traditional print model of publication for inscriptions, while also facilitating digital tools for searching across the corpora, encouraging more complex research

20 The Hymn to Aphrodite is available online at http://www.stoa.org/diotima/anthology/vandiver.shtml (translated by Elizabeth Vandiver, last accessed 31.05.2015).

21 Epidoc: http://sourceforge.net/p/epidoc/wiki/Home (last accessed 31.05.2015).
on these materials (Cayless et al. 2009). Building such online collections under best practice standards provides an opportunity for gathering and federating content in one central space: EAGLE, The Europeana network of Ancient Greek and Latin Epigraphy, is developing a single portal that will make available the vast majority of surviving inscriptions from the Greco-Roman world, collecting millions of multi-lingual digitized items from museums, libraries, and archives across twenty-five EU countries, as well as providing essential accompanying information and translation into English.

Online infrastructure has also allowed wider communities to contribute to the transcription and editing of ancient materials. Crowdsourcing—the harnessing of online activities and behaviour to aid in large-scale ventures such as tagging, commenting, rating, reviewing, text correcting, and the creation and uploading of content in a methodical, task-based fashion (Holley 2010)—seeks to engage a wider audience with tasks that have traditionally fallen within the purview of specialist academic activity (Terras Forthcoming). A core aim has been to facilitate the reading of damaged and fragmentary texts, with hundreds of projects now using volunteer time and effort to sort, explore, and transcribe archival materials (ibid). Above we mentioned the scholarly outsourced Suda Online and the student-based Homer Multitext: crowdsourcing has the potential to open up classics to an even more general audience. In the Ancient Lives project anyone can decipher, measure, and transcribe previously unstudied papyri fragments from Oxyrhynchus, the results of which then feed back into research. The use of distributed online platforms for increasing access to un-transcribed material and for assisting in transcription turns online digitization into an act of co-creation, which is helping to expand both our corpus of ancient texts and our understanding of them.

Many of the ancient sources that remain unread are, to all intent and purposes, unreadable by the human eye, due to damage, abrasion, fading, and elision. Advances by computational and engineering sciences in imaging and image processing, however, are revealing texts once presumed lost. Multispectral imaging—a method for acquiring image data over a series of wavelengths across the light spectrum—of a thirteenth-century prayer book famously revealed the Archimedes Palimpsest—a treatise copied from a tenth-century Byzantine manuscript, which had been erased (so that the vellum could be reused), as well as new speeches by the Athenian orator Hyperides, and a lost commentary on Aristotle’s Categories (Netz and Noel 2007, Netz 2011; see Figure 2).

22 Epidoc continues to be developed and adopted by a core use community, and underpins a variety of online epigraphic resources including Inscriptions of Aphrodisias, i.Sicily, and projects developed at the University of Oxford’s Centre for the Study of Ancient Documents, including the Corpus of Ptolemaic Inscriptions, and the Ashmolean Latin Inscriptions Project.
23 http://www.eagle-network.eu/about/who-we-are/ (last accessed 31.05.2015).
Figure 2. Spectral image and recovery estimate images from k-means clustering, PCA, ICA and LSMA, indicating different inks on the Galen Syriac Palimpsest. Processing done in Giacometti (2013), original multispectral images from the Archimedes Palimpsest Project (Walters Art Museum 2006, Toth et al 2010).

The technology of multispectral imaging continues to mature and become less expensive, prompting the need for best practice guidelines and approaches, if its potential is to be optimized (Giacometti et al, Forthcoming). Other emerging techniques for cultural heritage and text recovery include infrared imaging (Workman and Weyer 2007), 3D imaging (Baumann et al 2008), 3D laser scanning (Barnett et al. 2005), and Reflectance Transformation Imaging (RTI) (Earl, Martinez and Malzbender 2010). Still in their infancy, the results are often much hyped: X-Ray Phase Contrast Imaging has recently been used to detect individual letters on the blackened, carbonized scrolls from Herculaneum (Mocella et al 2015), but it is far from the case that all its contents can now be read (as claimed in the media: e.g. Jaggard 2015). Nevertheless, many other works, thought to be
lost, could soon be recovered from damaged source materials or from palimpsests held in library collections.

Literature’s twin cousins, papyrology and epigraphy, are both benefitting from and helping in the development of digital technologies, allowing us to simultaneously read and understand volumes of material that had been lost from view, while creating digital surrogates that increase access to the texts themselves and the information held within them. Common standards for information interchange, such as the Epidoc framework, allow for “linked data” searches of material physically held in different collections, and encourage ever more varied ways of analysing ancient literature, as we shall now see.

3. CLASSICS 2.0, OR THE EXPANDING ECOSYSTEM
Reading Greek literature has always been an interdisciplinary venture. Whether one is researching Homeric debates, the verbal form of curses, or second sophistic rhetorical treatises, it is vital to read literature in and against various contexts informed by archaeological, philosophical, and historical evidence. At the same time, the ways in which classics is taught and delivered at university tends to militate against holistic approaches, particularly as scholarship expands ever larger and becomes ever more specialized. These issues are even more acute in the digital world, free of the physical constraints of library cataloguing systems, where data of different kinds can seemingly float without formal context, and where serendipitous discovery is much harder. On the other hand, it is precisely the unboundedness of the digital text that can potentially revolutionize the way it can be contextualized. In this final section we consider this broader network—the ecosystem, to borrow a metaphor from the sciences—and reflect on how the very form and idea of the text is being transformed in an online environment, as well as the ways of interpreting it.

3.1 Linking to and reading out from the text
Being able to link—to go from one document to another—is one of the great facilitating features of the World Wide Web. It is how you are able to find out more about or get a different perspective on something that interests you at the click of a mouse. But how are these linkages enabled, when researchers working independently from each other upload their data, such as digital texts, to the Web? How can we build the systems or infrastructure that enable these isolated materials to be read in relation to each other? How, in short, can the mentality and practice of working within separate disciplinary spheres be overcome?
One strategy has been to use semantic web ontologies to structure data and align schema. However, ontologies either tend to be so complex (in order to create meaning) that their extension to other datasets is seriously compromised, or else they are so simple (in order to be used by the optimum number of people) that they have very little semantic worth (cf. Isaksen 2011). An alternative method has been to find easy and simple ways of facilitating connections. The Linked Open Data approach is in essence much more concerned with openness and discoverability than with formal collaboration and inference. As such it avoids complex ontological challenges—quality matters but utility matters more. This is due to the recognition that, while computers can do the hard graft of aggregating data, humans do the interpreting. The hook that can enable that essential process of linking to take place is the act of annotation that we mentioned above. By annotating a text, we allow for a much greater range of information to be held about it—not only the author, subject, date and so on, but the people, places and periods or categories used within it.

For the ancient world, the project that has been laying down the foundations for this kind of simple and pragmatic linking is Pelagios. Rather than getting everyone to agree on how to represent their data, Pelagios provides a set of lightweight conventions for how to express links between the things described in it, based on the use of common gazetteer references when referring to places (Simon et al., Forthcoming). Such place references could mark up a piece of literature or a research article, as well as express the find spot of a particular item in an archaeological database, identify a toponym in a digitized old map, or record the location of a historic site depicted on a photograph. By expressing place references in the form of Uniform Resource Identifier (URIs) from a shared authority (a gazetteer), otherwise isolated datasets are implicitly joined up to an interconnected graph, with the gazetteer as its central backbone. The metaphor of annotation is not only appropriate for the act of identifying (or “tagging”) a place reference in arbitrary digital content. It also suggests that the identification (or “tag”) is not to be regarded as fact, but rather that someone (a human editor, an automated geo-parsing script) is making a claim of a relationship between document and place.

Since 2011 Pelagios has fostered a growing community of some forty ancient world projects, which, at the time of writing, have collectively interlinked their resources through approximately

---

24 Phrased in more technical terms, the first Pelagios convention states that whenever you refer to a place in your data, you should do so using a gazetteer Uniform Resource Identifier (URI). For the ancient world, these URIs are provided by the Pleiades gazetteer. The second Pelagios convention is that the resulting place metadata must be published online as open data, according to a common technical serialization format: for this purpose the Open Annotation Data Model is used.

25 Another significant aspect is that the place metadata can be stored separate from the source data. This approach is sometimes referred to as “standoff markup”, and helps to avoid the data management problems which can arise when annotations have to be natively incorporated into an existing data model.
one million annotations (cf. Isaksen et al. 2014), in an instantiation of the idea that the geography of
the past is as every bit as interconnected, interactive, and interesting as the present (Elliott and
Gillies 2009).26 Regarding texts, there are two primary reasons why these combined annotations
could potentially transform our reading and understanding of them. On the one hand, we now have
the ability to link to any text from any other Web document through the common reference to a
place. This instantly makes relevant information for thinking about places, such as texts, more
discernable. An example of what this might look like is shown in Figure 3, which depicts the
range of documents, including texts, that contain information relating to Delos. Finding out key
information in the text is now as easy as clicking on a link (though of course what one does with the
link remains the purview of the scholar).27 On the other, it is now possible to read out from, or
through, the text to other Web documents. Reading this other information (literally) in and against
the text can provide enriched context for our interpretation of it.

What is more, there are also consequences for the way we write up our research.28 Referencing is a
defining feature of academic discourse: it isn’t enough to have a bright idea or discover something
remarkable; we have to relate our ideas and discovery to a wider body of research and locate them
in the scholarly debate. But in the print world referencing is limited to pointing backwards to
previously written material and outwards to the additional information that someone else provides.
Pelagios annotations help provide both future footnotes and reverse references. It allows us (anyone
in fact) to create hooks that connect our texts to both old and new material as it becomes available,
without having to manually update those links ourselves. Furthermore, by annotating content we are
simultaneously contributing our information to a wider cloud: i.e. it becomes directly available to
other people through their own annotations. Annotation of the kind promoted by Pelagios enables
the cross-referencing that underpins academia in a way that has never been possible before.

26 The spectrum of organizations extends beyond the academic research community to include: museums (the British
Museum, the Ure museum), a national database (Arachne), a national library (the British Library), a specialist learned
society (Nomisma), a digital open access library (Perseus), an aggregator service (CLAROS), a government agency
(Portable Antiquities Scheme), a collaborative global online knowledge base (Wikidata), and an equally diverse array of
voluntary partners and interest groups (e.g. the Ancient History Encyclopedia or http://vici.org/).

27 Because the Pelagios graph is open and distributed, it is possible to create services that allow for seamless
interlinking between online resources. In other words, you don’t have to go to a centralized portal or search service to
discover relevant material. You simply discover it naturally, through hyperlinked references and footnotes in online
books or articles (or webpages, or pictures, or maps, or songs or videos...and so forth).

28 For the argument here see the post by Leif Isaksen: http://pelagios-project.blogspot.co.uk/2014/05/future-footnotes-
reverse-references-and.html.
3.2 Back to the Future

Just as the canon of texts from the Greco-Roman world continues to expand and develop, so does the use of digital technologies within this space. There is yet more work to be done to understand how best to prepare for a linked-data world, to forge links between separate datasets and different projects, and to enable ever more refined text-mining techniques and visualization analysis.

Currently Pelagios is extending the reach of its network beyond the ancient world to encompass different mapping traditions, including early Christian Mappae Mundi and pilgrimages, Portolan charts, and Islamic and Chinese maps. The open data service technology it has championed is now the de facto international standard for open linked geospatial data concerning the ancient world, and is being used by other group initiatives (e.g. Standards for Networking Ancient Prosopographies; PeriodO: a gazetteer for period assertions). Community projects such as Pelagios and Epidoc are creating a World Wide Web of antiquity, whereby scholars, students, cultural heritage providers, or general enthusiasts can not only make their resources more discoverable and usable, but also enable others to find out about the cities (or peoples) of antiquity and explore the rich interconnections.
between them. Here we have digital classicists at the forefront of activities that are providing the tools and resources for studying the ancient world, establishing best practice models for adoption across the humanities, and helping to shape the information environment and its analysis.

Not that it is necessary for all classicists to be Digital classicists: the techniques and technologies described here intersect with more traditional close-reading methods, and both approaches are needed within the community. It is worth noting too that undertaking research which crosses disciplinary boundaries between computing and classical studies brings its own logistical, practical, and personal issues, such as negotiating different academic cultures; finding appropriate publication venues; developing robust promotion and tenure cases; developing adequate skills; and recruiting, training, and maintaining staff who can straddle both camps (Terras, 2010b). Nevertheless, those working in this area are increasing both the usability and visibility of classics-related information in the networked environment, and changing the scope and reach of questions we can ask of the sources.

4. **CONCLUSION**

This review article has shown how classical philology has always been at the forefront of digital technological innovation, and continues to adopt and adapt appropriate methods that can be used to understand the classical world. Digital methods are an established part of a classicist’s arsenal, routinely present in quotidian research to be used appropriately and alongside existing research methods to allow as broad an understanding of the Greco-Roman world as possible. Classics has always been an interdisciplinary endeavour, and many of these digital approaches are experimental methods that continue to push the boundaries of philological study.

The range of digital methods in use across the community presented here, including text encoding, geospatial analysis, and advanced imaging technologies, are both expanding the range of texts we have access to, and expanding the information we have for understanding the ancient literary landscape. These mechanisms are breaking down barriers to working across related disciplines such as linguistics, history, and archaeology, while also questioning the distinction between learning and teaching, as students contribute to the development and management of scholarly texts. Crowdsourcing further questions the traditional distinction between academia and the wider public, involving volunteers in the study of the past that can help support and facilitate specialist research. Digital technologies are constructively disrupting established notions of the limits and remit of Greco-Roman study.
Since 1998, two hundred contributors from twenty countries and five continents have carried out the seemingly impossible annotation and translation of the *Suda’s* 31,000 entries, the results of which are available freely to anyone with a connection to the Internet. The cooperative aspect of the Digital Classicist community, working with distributed resources within a technological framework, indicates much of what underpins the use of digital methods within the study of Greek and Roman texts—shared goals to use the changing information environment to explore and expand the way that we can read the past. Moreover, it is an on-going endeavour (Hutton 2014). A translation of the last of the *Suda’s* 31,000+ entries was submitted to the database on 21 July 2014 and vetted the next day, but the process of improving and annotating the translations continues.
CITED REFERENCES


http://www.digitalhumanities.org/dhq/vol/003/1/000025/000025.html#d3903e113


doi:10.1038/ncomms6895. 20th January 2015,
http://www.nature.com/ncomms/2015/150120/ncomms6895/full/ncomms6895.html


