



# The Perseus Digital Library and the future of libraries

Gregory Crane<sup>1</sup>

Published online: 19 August 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

This paper describes the Perseus Digital Library as, in part, a response to limitations of what is now a print culture that is rapidly receding from contemporary consciousness and, at the same time, as an attempt to fashion an infrastructure for the study of the past that can support a shared cultural heritage that extends beyond Europe and is global in scope. But if Greco-Roman culture cannot by itself represent the background of an international twenty-first century culture, this field, at the same time, offers challenges in its scale and complexity that allow us to explore the possibility of digital libraries. Greco-Roman studies is in a position to begin creating a completely transparent intellectual ecosystem, with a critical mass of its primary data available under an open license and with new forms of reading support that make sources in ancient and modern languages accessible to a global audience. In this model, traditional libraries play the role of archives: physically constrained spaces to which a handful of specialists can have access. If non-specialists draw problematic conclusions because the underlying sources are not publicly available and as well-documented as possible, the responsibility lies with the specialists who have not yet created the open, digital libraries upon which the intellectual life of humanity must depend. Greco-Roman Studies can play a major role in modeling such libraries. Perseus seeks to contribute to that transformation.

**Keywords** Digital libraries · Translation alignment · Language technologies

## 1 Introduction

I write this piece both to offer my views about the future of digital libraries and to do so on the basis of my own past experiences and decisions. These experiences include 35 years of continuous development work on the Perseus Digital Library and 40 years of engagement in what might now be called the Digital Humanities. The motivations behind the development of what is now Perseus began, however, 50 years ago when I began, in fall 1972, to study Ancient Greek. My experiences and frustrations in the subsequent 10 years as I pursued this subject within the limits

of print culture shaped my goals from the earliest days when I embraced the digital turn in 1982 to present and still shape my aspirations for the future.

I write this piece also as an attempt to explore how libraries—and, at this point, I assume that those libraries will be digital—can evolve to better serve society. There is obviously a striking contrast between the day-to-day work and immediate goals of any one particular project and the larger future of libraries. But for me, the connection is natural and essential. The work that I have personally done or supported on Perseus reflects a model of an integrated library that goes beyond the capabilities that I see being integrated into any library infrastructure. In part that reflects my own training as a student of the past and, in particular, of textual sources that survive from the past in more languages than any normal human being could hope to master and for which no native speakers survive.

Let me begin with a clarification. In Europe and North America, Classics and Classical Studies have been used as shorthand to describe the study of Greek, Roman and (sometimes) Byzantine cultures. One reason why digital libraries are so important is that they can enable us to expand our intellectual range and to begin developing a

---

The work presented here builds primarily upon funding from the Mellon Foundation (“A Framework for Annotation Interoperability”: 1802-05569) and the National Endowment for the Humanities (“Beyond Translation”: HAA-266462-19, as well as upon a fall 2020 sabbatical from Tufts and fellowship for Crane at Harvard’s Center for Hellenic Studies).

---

✉ Gregory Crane  
gregory.crane@tufts.edu

<sup>1</sup> Perseus Digital Library, Tufts University, Medford, MA, USA

field of Classics/Classical studies that engages with traditions from around the world, including (but by no means limited to) China, the Indian subcontinent, the Arabic and Persian speaking worlds, and indigenous languages of the Western hemisphere and beyond.

We need to balance the need for a broad perspective against the need for a rigorous grounding in some particular sub-discipline precisely because such expertise will allow us to appreciate what we can and cannot do as we work with a wider range of cultural materials than we could ever hope to master. If digital libraries mainly aggregate PDFs (like this paper) and relatively static objects such as images or videos, we will not be able to study the past with the breadth and depth of analysis that meets the needs of both culturally complex societies such as that of the twenty-first century United States and of scholarly rigor.

In my view, although some projects have made progress in this direction, we still do not have any true digital libraries, and we are all still struggling toward an understanding, much less the creation, of a library that goes beyond models inherited from print culture. We call Perseus a Digital Library, but that label remains largely aspirational: Perseus has, over the past generation, constituted a series of experiments that explore the possibilities of libraries in a digital space. Instead, most of us still use digital methods primarily to enhance structures and practices that emerged in print culture. My colleagues and I used Edward Gibbon's *Decline and Fall of the Roman Empire* as a demonstration text in an introduction to Digital Humanities at Tufts University in fall 2021. Gibbon published the first volume in 1776 and the final volume thirteen years later in 1789 [1]. The text was challenging because of its size and complexity—we worked with two different digitized versions of a later edition of Gibbon from the early twentieth century [2–4]. What struck me most was how mature the conventions of scholarly publication already were in the eighteenth century: extensive footnotes point to secondary sources (which are cited by the page number of whatever edition Gibbon was using) and to primary sources (which, in many cases, use citation schemes that remain the same in the modern editions that we use today).

In publishing a digital edition of Gibbon, we could actually come much closer to a true digital library than is possible with openly licensed secondary sources that are published today. Harvard's Center for Hellenic Studies (CHS), for example, set an example for the humanities by making new publications available under a Creative Commons license and buying the rights to other key publications.<sup>1</sup> The openly licensed CHS publications cite articles and books that are almost never available via an open license, and so these documents are forced to remain disconnected from the network of publi-

cation that specialists would still explore in print libraries. All of Gibbon's own work and that of his sources has long passed into the public domain. A large number of Gibbon's sources, both primary and secondary, have been digitized and are freely accessible from sources such as the HathiTrust, the Internet Archive, and the various national libraries represented by Europeana. We could build a digital edition of Gibbon's history with dynamic links to much of the library upon which he based his work. We would not be able to perfectly replicate that library but we would provide access to more of Gibbon's sources than was available to the vast majority of his most privileged contemporaries. Only a handful of modern specialists on Gibbon would even imagine consulting the sources that Gibbon used and seeing for themselves how Gibbon used them.

This putative digital edition of Gibbon could transform how people can read his work by creating dynamic links between Gibbon and his sources, but those links also highlight the limitations of a digital library that simply converts citations to links: Gibbon writes in English but his primary sources are in Latin and Ancient Greek and a large proportion of the secondary sources that he cites are in French and Italian. If we are going to make these sources useful to a more general audience, we need to be able to provide translations into English. Machine translation (MT) from French and Italian to English has reached the point where it can provide a useful starting point (something I have tested by requiring that my students report on scholarship that they access via MT). MT for Latin is improving but lags behind, while no usable MT is available (as far I know) for Ancient Greek, but we can link from many Greek and Latin sources cited in Gibbon to English translations linked to the original source texts. Of course, the page images need to be converted to machine readable text if we are to use MT, but Optical Character Recognition (OCR) for older books has improved. Digital libraries need to integrate full pipelines that seamlessly lead from the images to text to translation. Books are not the black boxes in a true digital library. Digital libraries need to operate directly on the contents of documents—whatever their length or media (text, image, sound, video). Our current focus on manually produced metadata is excessive and reflects the limitations of print culture.

I do not have a degree in library or information science but I have spent much of my career exploring the possible form and functions that digital libraries might assume. I was the professor of Digital Humanities (DH) for six years at the University of Leipzig (2013–2019) but I find work in DH to be problematic in at least one regard. The goal of most DH research is to use digital technology to produce traditional publications. We may exploit sophisticated algorithms and create elegant datasets but the only lasting results are too often PDFs with figures that no one can check and pictures of visualizations that, even if once public, all soon go offline.

<sup>1</sup> <https://chs.harvard.edu/books/>.

A digital library should, insofar as it is a digital library, offer a network of nodes that are, and will for the foreseeable future remain, available to a global community. As the digital library grows, the nodes within it should be able to interact with each other and with their users over time in increasingly sophisticated, and often unpredictable, ways. If, for example, a group publishes a more effective model for syntactic analysis of Ancient Greek, online reference works should offer recalculated totals for frequency of various phenomena—and such recalculated totals may call into question earlier statements that human authors had made based on earlier states of the data. Readers should be able to customize these interactions, in ways that are transparent and ethical, to the needs of each user. Thus, if, for example, a text corpus grows and machine actionable models for named entity linking improve, users should be able to apply the new models to the new corpus. A developed version of such a digital library may not be feasible at present, as much because our libraries are still designed around metadata about books rather than the content of digital documents, around the protection of copyright, and around the practices of print culture as a whole. But, of course, library professionals must serve the patrons that they have. We will not see a next-generation library infrastructure until the faculty and—especially—the students require one.

At Perseus, we have always imagined our work as a long-term system that must evolve over time to survive. To some extent that seemed to be a necessary and natural constraint when I began work in 1985 on what would become Perseus: I was building on a continuous tradition of scholarship about the Homeric epics that was almost 200 years old (if I were to pick Friedrich Wolf's *Prolegomena ad Homerum* [5] as a starting point). The particular code that we have produced is relatively ephemeral. The data that we have collected, however, has a much longer life cycle—we still have sources that we digitized when full time work began in 1987. The goal has been to create a sustainable system that could become more sophisticated over time and that thus only added new features as we felt we could sustain them. The price for such caution has been steep: we have been working on core features of a modern reading environment since 2006 and 2009, respectively: the exhaustive annotation of morphology and syntax (known as treebanks [6,7], because the syntax of a sentence is typically visualized as an upside down tree) and translations that are aligned with the source texts at the word and phrase level. It took more than a decade before we could produce *the Scaife Viewer* [8] in 2018 as a new reading environment for Perseus with a fundamentally more scalable backend and only in 2021 where we are able to publish an initial version of *Beyond Translation* [9], a framework that could provide access to resources such as these.

## 2 Classical studies and Greco-Roman culture?

Most—though by no means all—of our work at Perseus has focused on the challenge of understanding sources for the Greco-Roman World. We at Perseus have, however, worked over the years on a number of other topics: the history and topography of London, Old Norse and Old English, Shakespeare and Early Modern English literature, extended work on sources from nineteenth century America, including newspapers, accounts of the American Civil War and a variety of cultural materials, exploratory work with Classical Arabic and growing efforts more recently with Classical Persian, Mandinka Oral Literature from West Africa and Arabic accounts from Timbuktu of the Mali and Songhay Empires [10–16]. Each of these efforts allowed us to produce content that others could use and enhance and each of these projects allowed us to explore new challenges (e.g., managing content in Arabic script and analyzing sources in Semitic as well as Indo-European languages). The shift between such projects and our focus on the Greco-Roman world has been a purposeful dialectic.

Experiences with projects from the vast range of topics outside of the Greco-Roman world can be overwhelming. The focus on Greece and Rome, however, already provides a very broad exploratory space. The study of the Greco-Roman world is big enough to test a range of techniques but sufficiently constrained so that our projects over the years have a cumulative impact, with digital work produced in the 1980s proving useful in research efforts undertaken a generation later in 2022. When I try to imagine how we can provide intellectual access to untranslated Serbian nationalist songs or performances of West African Epic or scripted television series in languages such as Turkish, Korean or Malay that are available on services such as YouTube and Netflix, I inevitably find my mind turning back to what we can do with Ancient Greek and Latin to make Greco-Roman culture more accessible. Greco-Roman culture has remained and remains an object of relatively stable interest over time—a long period of time, in fact. We can mine generations of print scholarship, now in the public domain, for machine actionable information with which to bootstrap a born digital reading environment. The leading practitioners of Digital Greco-Roman Studies have for more than a decade been committed to publishing the results of their work under a Creative Commons license.

Much of my work recently has centered around how to create a truly born-digital edition that makes the Homeric epics accessible to an audience that is as global as possible. Distribution channels such as YouTube and Netflix may seem to captivate the eyes and ears of their viewers but we remain forever tourists if we live in a world of subtitles. When I find my head spinning as I encounter video content online



**Fig. 1** Geographic coverage of the Pleiades Geographical Gazetteer

in Malaysian, Turkish, Korean, Latvian, Serbian and other languages that I will never have an opportunity to master, I return to Homeric Epic. But that return is not a flight from the eternal Tower of Babel but a retreat to a solid position in which I can address the problem that there are too many languages to learn. Homeric Epic is big enough (200,000 running words) so that quantitative methods can reveal interesting properties. It has also accumulated a dense collection of scholarly resources. Some of these (including grammars, specialized dictionaries, commentaries, and translations) are in the public domain and have been converted into machine actionable form. Other resources are born-digital and have been published under an open license. Homeric Epic has been an object of study for millennia and will continue to be studied for many years to come. A digital edition of Homeric Epic will thus probably not drop wholly out of fashion and will remain useful. And the audience for Homeric Epic is international: some elements of an edition (like translations) are tied to particular modern languages but many digital annotations can be efficiently localized into a range of languages. A born-digital edition of Homer thus provides a useful laboratory for ways by which we can make a source directly available to an audience from around the world and speaking many different languages. With Homeric poetry, we can begin to establish a pathway from casual exposure step by step to deep mastery, so that anyone can go as deeply into such a rich subject as their desire and ability allow.

The geographic and chronological boundaries of the Greco-Roman world are substantial. A visualization of places covered in the Pleiades gazetteer (Fig. 1)<sup>2</sup> provides one view of the geographic scope of the Greco-Roman world. On the one hand, major sections of modern Europe fall outside of this space - Scandinavia, the Baltics, most of Eastern Europe—while the southern boundary tapers out over the Sahara desert. The eastern scope of the Greco-Roman world extends not only through Turkey, the Levant, Modern Iran and Afghanistan but even, to some extent, into the Indian subcontinent.

<sup>2</sup> <https://pleiades.stoa.org/>.

The study of the Greco-Roman world typically begins with the bronze age cultures that are relevant to the Homeric epics. The later bound should extend, as it did for Gibbon in the eighteenth century, through, at least, the fall of Constantinople to the Ottoman Turks in 1453. For practical reasons, our work at Perseus has focused on the period through c. 600 CE, shortly after the death of the Byzantine emperor Justinian, the last Roman emperor to control territory in the eastern Mediterranean as well as Italy and the West, with particular attention to the first thousand years of the written record for Ancient Greek, roughly 700 BCE through 300 CE.

Linguistically, the advanced study of Greco-Roman culture assumes not only an understanding of Ancient Greek and Latin but also of modern scholarship published in English, French, German and Italian. The field as a whole has been, of course, Eurocentric. More significantly, the study of the Greco-Roman world can and, at some times, has served as a practice to establish a cosmopolitan culture that transcends local identities and the many differences in language and culture across Europe. Viewed from an American perspective, European cultural identity may seem too narrow because nations such as the USA must fashion a culture that integrates contributions from many other parts of the world. Indeed, if academic departments in the USA are to use terms such as Classics and Classical Studies, they must broaden their coverage well-beyond ancient Greece and Rome. But we should not underestimate the positive value that transnational European identity has been able to exert. The European Union has, for all of its issues, been one of the great achievements of human history. As I write these words in March 2022, as Russian troops are invading Ukraine, European identity is particularly poignant. I, like many others, view this as an attack based on fear that Ukraine, in moving toward a liberal, European identity may prove too powerful a model for the Russian people as a whole.

If I am working to interpret the lyrics of a song on YouTube in Latvian, I do not have ready access to native speakers and, even if I did, I would still want to push beyond what I am told or the translations offered to me and to engage, as best as I can, with the source text in its original language, whether or not I have had the privilege of studying Latvian. Of course, we can study a small number of languages but there will always be more languages to learn for many large-scale problems.

Many, many people have worked on Perseus over the decades and others produced the vast majority of what most people encounter when they work with the web version of Perseus. For me, however, the Perseus Digital Library is, and has been from the beginning, an on-going experiment with the form of the library. Who has access and to what? What can they actually do with the library if they do have access? How do libraries—or, at least, libraries that allow patrons to push to the frontiers of our understanding in a given subject—advance the public good?

This last question may seem so general as to have little meaning but it may fall into the class of questions that are so obvious that they are rarely asked. Indeed, given the fact that resources are scarce and academic libraries always feel stressed to serve their immediate communities, discussion of the public good rarely takes place. We focus upon maintaining the collections and services for the patrons—mainly tuition-paying students, paid employees and outside researchers who can be said, on some official form, to advance the mission of the bill-paying institution. Where policy allows members of the public to use the library, they generally have access to a resource designed for others and offered as-is. There are, indeed, grim moments where members of an institution view their access to their particular library as an advantage and an instrument of power and consciously seek to restrict access. I can recall one particular conversation decades ago with a senior colleague of mine when I was a junior faculty member at Harvard, which has traditionally maintained the finest university library in the world. My senior colleague made it quite clear that he did not want any outsiders in the collection. He had gained tenure at a very early stage of his career and he felt proprietary ownership over the privilege he had to find virtually any publication that he needed in the Harvard collections. But such attitudes are, in my personal experience, rare—or at least rarely articulate. Most advanced researchers and library professionals warm to the idea of opening up access and serving a broader community. And, to be fair, I suspect that my unsettling senior colleague probably has often had more generous feelings than he happened to have on that particular day.

But if we ask first and foremost how our work within academia and within the humanities as a whole advances the public good, I, at least, find the consequences to be profound and quite tangible. This question has disrupted my fundamental attitude toward my own work and thrown me into a space that is very uncomfortable. Open Access is a necessary, but insufficient, condition if we in the humanities are to serve the public good: making our traditional primary and secondary sources and our reference works freely available in a format that augments a print model is an essential first step but only the beginning of a larger, more fundamental transformation. The question is not how to convert print resources into machine actionable form—and that is a major question in and of itself. The question is how we imagine people from around the world interacting with the human record as broadly and deeply as possible, from anywhere in the world, thinking in as many different languages and from as many cultural backgrounds as possible. To some extent, this is a more verbose version of Google's mission "to organize the world's information and make it universally accessible and useful". In practice, however, a philologist trying to support, for example, American readers engaging with the Persian Song of Kings and Persian-speaking readers

engaging with the Homeric *Odyssey* will have very different priorities and questions than the carefully selected engineers at Google.

Before exploring how I struggle to make the human record play the broadest and most constructive role in human intellectual life possible, I will provide some background to explain how I became interested in developing Perseus. I do so, in part, to shed light upon some of the motivations behind the particular ideas that I present here so that others can more clearly see my assumptions and decide for themselves the extent to which they share them, if at all. I also do so because the historical conditions that shaped my priorities were largely formed in the 1970s and 1980s. Some of those conditions would be foreign—and, indeed, require a leap of concentrated historical imagination—for any of my students and many of my younger faculty colleagues.

### 3 Challenges to the student of the past in print culture

I include here an explanation for some of the strategic decisions that have driven my contributions to the transition from a print-based to a primarily digital infrastructure for the study of the past in general and ancient cultures in particular. This information serves a second and historical purpose, as it reflects the perceptions of someone who began as the product of an entirely print culture and who seized upon digital media to solve problems with which few of those born in recent decades will have experienced so keenly. The autobiographical elements thus serve to place into the record personal experiences about a transformation that may still have long to go but that has been underway for a generation.

*First*, I doubt that anyone who was born in the twenty-first century will be able to imagine how dependent we were upon textual information and how scarce still images were, much less video. When I was in graduate school in the 1980s, the books on literature and history were largely in Widener while those on art and archaeology were, for the most part, a ten minute walk away in the Fogg Art Museum library. Anyone trying to combine the textual and material record of Greco-Roman culture had to move back and forth between separate spaces. Worse, the books on art and archaeology could only contain a handful of images, almost always black and white and relatively low in resolution. We spent much of our energy in the opening years of planning and developing Perseus (c. 1985–2000) looking for ways to collect new photography—even when high-resolution photographs existed, they tended to focus on one or two details and were designed to support slide lectures and plates for print books. There were no high-resolution images, and three-dimensional virtual spaces existed only for specialized applications such as training pilots. I cannot imagine how a generation that grew

up immersed in *Assassin's Creed Odyssey* (which actually focuses on the opening of the Peloponnesian War, c 431–422 BCE) will conceptualize people and spaces as they work with Thucydides or Athenian drama. While we still lack the curated archives of art and archaeological sites that we really should have, anyone can use a web search engine to summon at least representative images and iconography to make many topics visually concrete. Scholarly thinking has not yet caught up with this reality and we have barely begun to exploit the potential of what is already possible.

As the Web grew more mature and the amount of visual information about the Greco-Roman world began to expand, I shifted my focus away from collecting and commissioning photography and drawings. The rest of the world was addressing that in a heterogeneous but powerful fashion. My one regret is that a shift to a more textual focus distracted me from continuing an earlier commitment to ensuring coverage of both the textual and material record. From a pragmatic perspective, I decided that the best approach was to build out an open textual infrastructure that others could integrate with the material record.

*Second*, few who grow up with smartphones embedded in a global network can imagine the absolute and utter limits of a purely print culture. Physical books were either available or they were not. Great libraries of print books did exist and I experienced one when I became an undergraduate and then PhD student at Harvard but I had already experienced a sense of desperate and frustrated isolation. When I was a child, different topics captured my attention and I searched for everything that I could read on various topics such as early human evolution, Genghis Khan and the Mongols, or the early twentieth century performer Harry Houdini. I lived in Greenwich, Connecticut, one of the wealthiest cities in the USA and had access to as well-funded a public library as anyone in a medium sized city could expect. I was able to read everything that our library collected on each topic that captured my attention. I can remember thumbing through a glossy propagandist description of 1960s Mongolia, filled with pictures of beaming workers in tractor factories, because I had run out of other materials relevant to the Mongol empire. I remember identifying each and every article relevant to my topic in the *Encyclopedia Britannica* and quickly realizing how few topics received coverage in the glossy, multi-volume general encyclopedias of print culture. And I can remember a librarian tossing her head back in exasperation as she caught sight of the persistent child, determined to try yet again to ferret out some new source for whatever I was studying at the time. When I was in high school in Rhode Island, I even got permission to ride the bus to Providence so that I could scour the Rockefeller Library at Brown for medieval French epics (chanson de geste) and found far less than I could absorb. When I first visited England at the age of 15, I drew a talking to from my older brother because all I talked about was

going to Blackwell's bookstore at Oxford, the only place on earth with which I was familiar where I could find printed editions for many—but by no means all—the standard Greek and Latin authors. If I had access to Widener Library or the New York Public Library or some similarly grand institution, I could have spoken with the ennui of generations about the overwhelming number of sources to read. But I did not have access and I was starved for knowledge. I felt that I was suffocating intellectually. I developed a deep and still driving consciousness that access to sources is the start for intellectual life. And I have never, ever lost the sense of desperate isolation that I felt when I was cut off in the luxurious intellectual desert of a transcendently prosperous American suburb of the 1960s and 70s.

*Third*, fewer today, and even fewer going forward, will experience the feeling of desperation that readers in a purely print culture experienced when they confronted pages of text in languages that they did not know. The printed page was a door locked shut. Only years of labor could begin to unlock it. When most readers of this piece think about what has changed, they will probably, at least in the years immediately following the composition of this article (early 2022), think of machine translation. Active readers can now expect that they can work directly with sources in many languages that they do not know. When readers encounter posts online in social media, for example, they increasingly expect that machine translation will give them at least some insight into what has been written. Machine translation into English from many languages has reached a point that I began in spring 2021 regularly assigning articles on Greco-Roman culture to my students for which they had to rely on machine translation—when a student knew one typical language of classical scholarship (e.g., French), I assigned another that they did not know (e.g., German, Italian or Spanish). None of my US students in recent memory was familiar with all three of these languages and, for the first time, any student of Greco-Roman antiquity can begin work with the publications not only in the modern languages supported by graduate programs but in Arabic, Croatian or Persian—we can begin to engage with voices from a far wider network of cultures than was thinkable. Even if machine translation was never to improve and we only digested the implications of what can be done in 2022, the implications would be transformative.

The problem that I addressed when I decided that I wanted to study ancient Greek as an eighth grader in 1971 was not the lack of a translation. An encounter with W. H. D. Rouse's prose translation of the *Iliad*, in a two week mini-class on Homer, had unexpectedly captured my attention and fired my imagination. I decided that I wanted to learn how to read the ancient Greek in the original because I believed—quite correctly as the past fifty years have shown me—that translation could only convey a pale shadow of an original text that would fascinate and teach me new things for the rest of

my life. Direct translations from Greek and Latin exist for virtually any text that English speakers ever read, and far more compelling Greek and Latin literature exists in English translation than most human beings will ever read. (The same is not true, however, for most languages—Iranians, for example, typically read indirect translations into Persian of Herodotus and Xenophon that are derived from English and French translations.) While the rise of machine translation may be revolutionary in that it could provide a workable translation for any source text, for me, the formative challenge and my motivation to learn Greek and Latin was the immediate conviction that no translation, certainly not from languages and cultures as alien as ancient Greece and Rome, could do justice to the original.

I spent decades studying various languages, balancing the desire to read not only with fluency but also with a growing awareness of the cultural context. My first exposure to Latin, as a first year student in high school, was rocky: I had no idea how such a language worked and I only passed my first semester because the instructor raised my grade from a failing 57 to a “circle 60” (meaning I had really failed). The logic of a highly inflected language clicked in the spring and I took to first year Greek the following year (fall 1972) with energy and pleasure. In the following spring, I finished the Crosby and Schaeffer text book [17] on my own and threw myself into my first text, Plato’s *Apology*, after the school year. Alone, with nothing but the Intermediate Greek Lexicon [18] and the Burnet edition and commentary [19], I struggled to make sense of the Greek. I can remember, for example, encountering a verb, *diaballô* (Plato, *Apology* 19b), for which my lexicon offered the initial definition “to throw over or across, to carry over or across,” a literal rendering of the verb (*ballô*, “to throw”) and the preverb (*dia*, “through”). The actual meaning in the *Apology* was “to slander,” a meaning that appears later on in the entry and that I only noticed after reading and rereading that passage without comprehension several times. Over and over I missed the obvious or could not make connections that a more experienced reader would have made without thinking. I ultimately had the opportunity to find a local Latin teacher, Donald Connor, then of Greenwich Country Day School, who would work with me several times a week. With direct answers to my questions, I could make rapid progress. I spent the three summers of 1973, 1974, and 1975 intensively reading Greek and Latin. When I arrived at college, I had considerable fluency and took the graduate survey of Greek in my first semester. I had very little contextual understanding and less sophistication in interpreting what I read, but relative youth was, if anything, an advantage for linguistic analysis and I could parse the sentences. I internalized the general attitude (never explicitly expressed but constantly communicated by comments by most of my teachers and fellow-students about non-specialists) that only

those who could understand Greek and Latin with fluency could speak with any authority about these sources.

My linguistic complacency crumbled, however, in graduate school. As I prepared to write my dissertation, I came to believe that Homeric epic had emerged as part of a larger, transnational culture that stretched from Greece to Egypt and into the Near East. Figures such as Inanna and Ereshkigal seemed to have strong links to figures such as Persephone, Circe and Calypso. The *Iliad* and *Odyssey* themselves seemed to me as if they each could, in their own ways, been influenced by the Gilgamesh epic. Driven by philological ardor, I began studying Sumerian, Hittite and Akkadian in my third year of graduate school. I realized that I could never achieve mastery of these languages comparable to what I enjoyed with Greek and Latin. Even if I could, the tools for these languages were much less developed—the final volume of the first complete Akkadian lexicon (Von Soden’s *Akkadisches Handwörterbuch*) [20] came out as I was studying the language. For Sumerian, there was no dictionary at all—we used index cards with signs and definitions that Thorkild Jakobsen had left in his office after he had retired. I had to produce the first translations for several texts in my dissertation and I can remember spending weeks returning again and again to one particular, very short word. I finally broke down and asked my professor, William Moran, for help. He took one look and, if memory serves, declared that it was a late Babylonian form of *nadânu*, “to give.” Whatever the actual answer, the effect of this exchange was profound. I realized that if I only worked with those languages that I had studied intensively, I would never be able to explore larger questions, such as reconstructing the grand cultural continuum in which Archaic Greece, Mesopotamia, the Levant, and Egypt all participated. There were just too many languages and none of them had the developed scholarly infrastructure available for Ancient Greek and Classical Latin. There may be no substitute for the expertise that we can acquire over years and even decades of work on a particular language and culture, but I decided, in 1984 as I struggled with languages of the Ancient Near East, that digital services could provide for others working with Greek and Latin the basic services that I lacked when I labored to understand sources in Akkadian or Sumerian.

Multilingual sources, with the same text juxtaposed in different languages, have existed for millennia. The Rosetta Stone, carved in 196 BCE, preserves the same text in Ancient Greek and Ancient Egyptian using hieroglyphic and demotic scripts respectively. In 1822, J.F. Champillon was able to use the parallel texts on this stone to decipher Egyptian hieroglyphs for the first time. Given enough time, readers with texts in Greek on one page and English translation on the other would be able to decipher the Greek. In normal circumstances, a reader with no Greek can do nothing at all with the Greek text and looks only at the English. And most

of those who have studied Greek will not be able to recognize every inflected form or know the meaning of every word in the Greek. I focused initially on creating a system that could provide linguistic explanations for each inflected form and then offer machine actionable links to online lexica for readers. Instead of having several physically distinct and inert books—bilingual edition, grammar to explain the forms, and dictionary to explain the meaning—we were developing an interactive system composed of different components, each of which adapted its behavior in relation to the other—the dictionary, for example, would check to see if the reader had been looking at a passage in the *Iliad*, whether the dictionary had something to say about what that word meant in that particular passage.

In 1982, at the end of my third year in graduate school, I began to rethink the study of the past in light of an emerging digital age that would transform our intellectual practices and challenge us to rethink our fundamental goals in light of new possibilities and challenges. That beginning was not gradual—I was transfixed by the opportunity to participate in a revolution that would, I believed, take place throughout the course of my career. I began working on July 1, 1982, as a student researcher (a US category called “work study”). I had no experience programming and began with the Kernighan and Ritchie “white book” guide to the C programming language [21]—a book written for professional programmers who were migrating from assembly language to C. To illustrate features of the language (such as call by reference vs. call by value), it showed how to implement algorithms such as shell sort and bubble sort (if memory serves), under the assumption that the readers knew the algorithms well. I had no clue and could do nothing for months—until suddenly, I could do anything I wanted. I could not do it well or write elegant code. It took me a long time. But I could get things done. Drawing inspiration from Michael Lesk’s source code for the Unix Refer application (which used hashing to look up bibliographic entries through the use of key words [22,23]), I had developed what turned into a 10,000 line package to search the first collection of machine readable Greek—a series of digitized texts distributed on magnetic tapes for a fee and under a license.

This early work shaped my thinking in several ways. First, the license attached to the texts was burdensome. I updated the format of the texts to facilitate searching and analysis. Other groups became interested in what we began to call the Harvard Classics Computing Project. I could share my source code but I could not share the work that I had done on the Greek sources for fear of violating the license. And fear was in fact something that affected users—I often was warned not to do anything that would cause the licensor to cancel the license. That fear addressed not only the particulars of the agreement but the anxiety that the person who controlled the license would act out of spite or malice if he

felt in any way unhappy. The license did stipulate that individuals could use individual texts as the starting point for manually produced scholarly editions but the license carefully established monopoly control to the texts as a whole. There was no way to use this collection of digitized texts as the foundation for new scholarly projects. Years of struggling with the limitations of this license drove home to me the foundational need for what we now would call open data and for which we now have standardized licenses such as those from Creative Commons.

Much of my personal energy in subsequent decades went into creating openly licensed digital data that anyone could freely use, modify and redistribute. Leipzig University, Harvard’s Center for Hellenic Studies, the Harvard Library, the University of Virginia Library, Mount Allison University, and Perseus at Tufts University, as well as volunteers from around the world, have contributed to the Open Greek and Latin Project (OGL) since its founding in 2017 [24,25]. As of spring 2022, we have made 48 million words of Greek and Latin, along with 20 million words of translations into English and other modern languages, available under an open license. Much remains to be done to make all surviving textual sources produced through 600 CE are available to support a new generation of scholarship that is fully transparent and not dependent upon proprietary data. For complete coverage, we would need to add roughly another hundred million words of Greek and Latin. Enough has, however, been released to support many research projects and, certainly, to transform student work. The existing collection produced by Perseus and more recently by OGL offers materials from a range of periods, including later works such as the Byzantine Encyclopedia known as the Suda, but our current focus is to provide comprehensive coverage for the first thousand years of Greek (roughly everything produced from the Homeric Epics through c. 300 CE). And because the collections are open, anyone can contribute new materials and incorporate what we have done. If those studying the Greco-Roman world are committed to producing open scholarship that draws upon openly licensed data, the community of paid professional researchers could finish what we started and provide comprehensive coverage for all published Greek and Latin sources that survive from the ancient world.

The autobiographical notes from the preceding section provide a narrative explanation for a fundamental principle that has driven my outlook in that forty years of work. I reorganized my life so that I could learn Greek and Latin when I chose a high school where I could study as much Greek and Latin as I chose. I had been planning to remain at a local school and upended my plans when a passion to learn about the Greco-Roman world seized me. Years would pass before I would have access to a major library. My attitudes were shaped by the years that I spent before I was in higher education and had a faculty of experts from whom I learned. I





**Fig. 2** Entrance to the Boston Public Library, with the motto “Free to all” above a bust of the goddess Athena

derived from that a deep conviction that academic publication, in and of itself, had no direct, larger value whatsoever beyond specialist networks. The value of academic research is entirely potential. That potential is only realized insofar as it fires the intellectual lives of those who do not belong to institutions of higher learning—who do not take a subscription to JSTOR for granted or assume that their true audiences should be familiar with current scholarly jargon (Fig. 2).

#### 4 Principles behind the design of the Perseus Digital Library

The experiences described above led to very concrete and far-reaching decisions. *First*, libraries are, to quote the words gracing the entrance of the Boston Public Library, “free to all.” In the nineteenth century, that meant that physical libraries were to be open to anyone who walked in the door. In a digital age, however, libraries are not spatially constrained—indeed, they can only close themselves off by imposing technological barriers to restrict access. In a digital age, neither physical buildings nor the gated communities of subscription services can claim to be libraries in the fullest sense. These resources can still contribute to the public good but they are, instead, equivalent to print-culture archives, i.e., locations where a limited number of specialist researchers can explore foundational primary sources. Researchers in an archive know that they have privileged access to materials that virtually no one among their audiences will ever themselves see. The materials in the archive are not part of the published record and the job of the scholar is to make the contents of those archives intellectually accessible through their own publications. Access is a privilege and imposes responsibility. If the wider audience draws incorrect inferences about the subject because the researcher has failed to provide adequate and practical documentation of what they found in the archive, the researcher is responsible and feels that responsibility. Errors among the wider audience weigh

upon the researcher and provoke attempts to provide clarifying information.

Among humanists, the conventional point of view, however rarely articulated or examined, tends to be the opposite: access to information, as well as to the training by which to make good use of that information, confers authority. Specialists answer to no one but their peers. Those who are not their intellectual colleagues cannot pass judgment and may be viewed as tiresome. At best, outsiders simply need too much explanation and, even if they do grasp the basic points, they have nothing interesting to add. At worst, fields deliberately cultivate specialist terminology for often mundane ideas in order to separate themselves from outsiders. When I asked one specialist why the quotations in her talk were so short, the answer was that the audience was expected to know the context already. Quoting enough to make the point clear to a general audience would, I was told, have alienated other experts and reduced the authority of what was being said and of the speaker.

The rise of digital libraries, however, made it possible, decades ago, for humanists to begin rethinking the role that published primary and secondary sources could play in the intellectual life of humanity. Publication under an open license is a necessary, if not sufficient, condition for true publication. Researchers that allow commercial entities to restrict their publications behind subscription firewalls have not truly published anything. They have only added to the murky archive upon which other specialists may depend—assuming, of course, they belong to an institution rich enough to pay for the particular subscription gateway behind which a new publication lies. It is possible that the publication hidden behind a paywall may contribute indirectly to the wider intellectual life of humanity but only if someone else reads that publication, extracts data or ideas (which are not subject to copyright) and makes them possible in some new, openly licensed publication.

When we first began planning for Perseus in the fall of 1985, we licensed textual and visual materials from publishers and museums. A decade later, we realized that this would not scale—we had too many agreements with too many licensees, each of whom had their own policies and agendas. We needed to be able to build the collection over many years and move it freely from one platform to another. The transition in 1995 from physical CD ROMs and videodiscs to the World Wide Web brought this home and we made a conscious decision only to build on materials that were fully in the public domain or (when Creative Commons licenses took shape a few years later) that were available under an explicit, irrevocable open license. In 2007, we formally shifted to a policy of using a CC license throughout.

There are, of course, cases where competing needs (health records and the need for privacy being one salient example) do require that information be restricted. Most of us also

benefit from cultural productions that are produced because copyright and payments make their production possible. Our job as researchers, however, is to make our ideas and conclusions public as fully as we can without violating such restrictions.

The shift to a fully open ecosystem requires a major change in orientation. In traditional scholarship, we may ask what are the most recent materials on a given subject. Traditional scholarship places the interests and the desires of the researcher first. In some cases, this academic independence bases itself on historical narratives such as Galileo resisting and then capitulating to the Catholic Church. I cannot recall many instances of academic freedom enabling such heroic stances. Far more often I have seen this heroic narrative of academic freedom used to avoid having to explain ourselves to tiresome outsiders.

When we build a system with open data, the question is whether we have materials that are sufficiently interesting, sufficiently rich, and sufficiently well-structured to set alight intellectual chain reactions of examination, reflection, interpretation, reexamination, and more reflection among non-specialists. We can choose to build such an open system on any subject but we may have to work around licensing constraints. We cannot directly annotate copyrighted detective novels or Netflix series, but we can create arguments that cite pages or time sequences within a book or video. With YouTube, we can create hybrid documents that not only cite a particular video but define how many seconds into the video your citation should point. The major risk here is that the video that you have annotated may vanish, often because the YouTube post violates a third party's rights. But if you choose content that a stable institution has made available, then the risk can be acceptable. Rights holders have made an enormous amount of music and television, as well as a range of films, available on YouTube. Much of this will be online for the foreseeable future. There is no reason we could not have a new generation of scholarship articulating the methods at play in music and film to engage both a specialist and a general audience. The discipline of addressing both at once can benefit each audience.

*Second*, digital libraries in the sense described above (i.e., libraries built on open data) enable, I would say require, that researchers must labor to make the full stack of information upon which they base their conclusions visible to their readers. Of course, existing publications cite sources that are, and will probably remain for the foreseeable future, inaccessible to a global audience, either because they exist only as physical artifacts or because they are only available in digital form behind a paywall. These will play a role comparable to publications that cite archival sources that are only accessible in a particular location. Making primary or secondary sources accessible online with an open license may be a necessary, but it is not a sufficient, condition to support the intellectual

Odyssey Book 5: Aligned Translation  
Amelia Parrish, Gregory Crane

1. Dawn [Ἥως], out of[ἐκ] her[0] bed[ἄσχετον] from beside[παρ'] noble[ἀγαθῶν] Tithonus[Τιθωνοῖο].
2. began to rouse herself[3rd sg imperf ind mp: ἄρυσθ'] , so that[ἵνα] she[0] might carry in[3rd sg pres opt act: φέροισι] light[φάος] to[ἀθανάτων] the[0] undying[ἀθανάτων] gods[0] and[ἠδὲ] to[βροτοῖσιν] the[0] mortal men[βροτοῖσιν].
3. The[οἱ] gods[θεοί] started to take their seats[3rd pl imperf ind act: καθίζουσιν] in their places[θρόνονδε], and[δ'] among[ἐν] them[τοῖσι] then [ἄρα] was[0]
4. Zeus[Ζεὺς], who thunders on high[ὑψηλῆμετης], and[τε] whose[οῦ] power[κράτος] is[3rd sg pres ind act: ἔστι] the[0] greatest[μέγιστον].

Fig. 3 Opening to a born-digital aligned translation of book 5 of the *Odyssey*

life of humanity. Documents composed with the assumption that they would only reach specialists with a shared, advanced understanding of terms and background knowledge were not designed to be intellectually accessible to a wider audience. We need to think about how we can make the human record as comprehensible as possible to the widest possible audience.

We have chosen to focus on subjects for which we have an opportunity to make the full stack of data fully visible and thus to work toward scholarship that is as transparent as possible. Digital editions have been a central theme in the first generation of Digital Humanities scholarship. Traditional scholarly editions have been designed to serve specialists and encode within that design assumptions of advanced knowledge within the field. We need a generation of editions that can not only serve specialists but that can make their contents comprehensible to the widest possible audience. The great research challenge for students of the human record is, in many fields, not to produce new ideas about the past—we have already in the published record more primary sources and secondary sources about the Greco-Roman world than any human brain could hope to process. The great challenge is to make that full stack of information, from the observed sources through to the conclusions that we draw from them, as comprehensible as possible to as many human beings from as many cultural backgrounds and thinking in as many different languages as possible.

*Third*, we need a new generation of digital libraries that can collect hybrid publications that combine narrative text with machine actionable data and that can reuse and recombine with other sources that may not yet exist to support uses that we may not yet have imagined. Figure 3 provides one example of a new, born-digital, machine actionable publication that requires accompanying expository prose. Hundreds of people have translated the Homeric *Odyssey*. Some of these translations have, like the one above, been composed to reflect the structure of the Greek as closely as possible. There have even been interlinear translations of Greek and Latin sources that are almost as closely aligned as the one above. But this translation, produced by Amelia Parrish (Tufts '21) and with some help by me, is designed for subsequent computation.

The expression “her[0]” in the opening line, for example, indicates that there is no corresponding pronoun “her” in the Greek—it was added because English speakers expect possession to be marked where Greek allows listeners to infer this. We can thus calculate that in 10 out of 14 times where “her” appears in the English, the possessive pronoun has been added. Such figures help us quantify differences between Greek and English. The translation “began to rouse himself” in line 2 emphasizes two features of the verb: (1) as an imperfect, it describes an event that takes place over time; (2) as a “middle” voice, it describes an action that affects the subject. The goal is for learners to be able to generate textbooks based on the machine actionable annotations for the corpus that interests them. When they learn the imperfect or the middle, they can see how often each of these features appears and they can see each instance of both features. Our goal is that the translation will reinforce the grammatical features that they are learning. We expect readers to compare our work with more literary translations.

We cannot, however, fully publish a translation such as the one above without providing a general explanation of what we are doing and discussions of the decisions that we made at particular points of the text—there are a number of points where we had to decide among sides of a scholarly argument or we were making a particular point that should be explained. Although the aligned translation itself includes natural language, it constitutes a machine actionable data set. But that data set is incomplete without accompanying explanations.

Put another way, digital scholarship must be a superset of traditional scholarship; in that it includes expository prose as well as machine actionable components such as code or interpretive annotations.

We are very much in the midst of building out environments in which we can not only exploit machine actionable data but also read about the various decisions made along the way. Nevertheless, we can see that such a system is needed.

## 5 Conclusion: classical studies for the twenty-first century

We hope that through our work on the Perseus Digital Library, we have helped others not only use the collections and services that we have created, but also helped others think about how we can begin developing true digital libraries. I will conclude by suggesting what I view as requirements for the collections, associated data, and services needed to advance the study of the human record. I will use the traditional label Classical Studies. I am not attached to this term, but I want to emphasize what a Classical Studies should be if we are to keep this term. I am assuming that we have an integrated reading environment for a range of languages that includes:

- A library of primary sources. These may be largely available as page images with a subset available as curated TEI XML.
- A library of translations aligned closely enough to the source texts that readers can compare the two easily. A subset of these translations should be born-digital, aligned translations designed to help readers learn how the source language works. Roughly, 5,000 running words of text would be enough for a first year course or equivalent. Learners would then aspire to create their own translations to augment the initial seed corpus. This should be a collaborative effort with learners working together and experts reviewing the results of their joint work.
- As many types of linguistic annotation as are available, with part of speech tags and syntactic analysis as a start. A subset corresponding to the aligned translations should be hand curated. Learners should use the existing annotations to learn the language and then collaboratively apply what they have learned to curate more annotations.
- A collection of reader-driven discussions about questions and possible answers. These should be periodically summarized.
- As many types of visualization as can be produced to help readers see larger patterns. These include topic models, automatically generated maps and social networks, timelines, etc.

Students should develop as much mastery as possible for at least one historical and one modern language. The historical language challenges students to think about how they can understand sources where no living speakers survive and we have to infer all of our conclusions by observing the evidence. The modern language challenges them not only to develop their ability to write, speak and listen, but also to engage with speakers from a different country who may have a very different perspective on the historical sources that the student is learning.

Equally important, students should learn how to exploit the full range of automated tools available to them, using linguistic annotations to push beyond the surface of translations (whether those translations are generated by machines or human beings). For historical languages such as Ancient Greek and Latin, these tools are particularly important since it is no longer practical for most of those who wish to major in Greco-Roman studies to learn the languages and read extensively in them if they begin study at the university level and take traditional classes.

Students should use automated tools to focus on at least one historical and one modern language that they are not in a position to master. While some may, for example, choose Ancient Greek and French to learn and then practice language hacking on Latin and German, departments should

be designed to support non-traditional combinations (e.g., Ancient Greek and Classical Persian as historical languages, modern Persian and German as modern languages).

Students should demonstrate their ability to combine close reading with larger-scale methods of textual analysis. This includes demonstrating that they understand the limitations of the larger-scale methods as well as of close reading of a relatively small number of passages.

A final capstone project should include curated data and/or code as well as expository prose. Students should demonstrate that they communicate their research questions and conclusions clearly to both advanced researchers and a general audience.

## References

- Gibbon, E.: *The History of the Decline and Fall of the Roman Empire*, pp. 1776–1789. Strahan and Cadell, London (2002)
- Gibbon, E.: *The History of the Decline and Fall of the Roman Empire*. Fred de Fau and Company, New York (1906)
- Gibbon, E. *The History of the Decline and Fall of the Roman Empire*. <https://www.ccel.org/g/gibbon/decline/index.htm>. Accessed 23 Apr 2022
- Gibbon, E.: *The History of the Decline and Fall of the Roman Empire*, 12 Vols. | Online Library of Liberty. Online Library of Liberty. <https://oll.libertyfund.org/title/lecky-the-history-of-the-decline-and-fall-of-the-roman-empire-12-vols>. Accessed 23 Apr 2022
- Wolf, F.A.: *Prolegomena ad Homerum: sive de operum Homericorum prisca et genuina forma variisque mutationibus et probabili ratione emendandi*. Libraria Orphanotrophi (1795)
- Bamman, D., Passarotti, M., Busa, R., Crane, G.: The annotation guidelines of the Latin dependency treebank and index thomisticus treebank: the treatment of some specific syntactic constructions in Latin. In: *LREC 2008: Proceedings of the Sixth International Conference on Language Resources and Evaluation* (2008)
- Bamman, D., Crane, G.: The ancient Greek and Latin dependency treebanks. In: *Sporleder, C., van den Bosch, A., Zervanou, K. (eds.) Language Technology for Cultural Heritage, Theory and Applications of Natural Language Processing*, pp. 79–98 (2011)
- Scaife Viewer | Home. <https://scaife.perseus.org/>. Accessed 23 Apr 2022
- Beyond Translation: Iliad (Greek Text of Munro & Allen). <https://beyond-translation.perseus.org/reader/urn:cts:greekLit:tlg0012.tlg001.perseus-grc2:1.1-1.7>. Accessed 23 Apr 2022
- Crane, G., Babeu, A.: The challenge of Virginia Banks: an evaluation of named entity analysis in a 19th-century newspaper collection. In: *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 31–40 (2006). <http://hdl.handle.net/10427/42691>
- Crane, G., Wulfman, C.: Building a hypertextual digital library in the humanities: a case study on London. In: *Proceedings of the First ACM+IEEE Joint Conference on Digital Libraries*, pp. 426–434. ACM Press (2001)
- Crane, G.: Designing Documents to Enhance the Performance of Digital Libraries: Time, Space, People and a Digital Library of London. *D-Lib Magazine*, vol. 6, no. 7/8 (2000)
- Crane, Gregory: Greco-Roman studies in a digital age. *Daedalus* **145**(2), 127–33 (2016). [https://doi.org/10.1162/DAED\\_a\\_00383](https://doi.org/10.1162/DAED_a_00383)
- Crane, G.: Greco-Roman studies and the future of Europe. *Class. Contin.* **102**, 53–68 (2022)
- Crane, G.: Thoughts on classical studies in the 21st century United States. *Class. Inq.* (2021). <https://classical-inquiries.chs.harvard.edu/thoughts-on-classical-studies-in-the-21st-century-united-states/>
- Crane, G., et al.: Toward a cultural heritage digital library. In: *Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 75–86 (2003)
- Crosby, H.L., Schaeffer, J.N.: *An Introduction to Greek*. Allyn and Bacon, Boston (1928)
- Liddell, H.G., Scott, R.: *An Intermediate Greek-English Lexicon: Founded upon the Seventh Edition of Liddell and Scott's Greek-English Lexicon*. Oxford University Press, Oxford (1882)
- Burnet, J.: *Plato Euthyphro; Apology of Socrates; and Crito*. Clarendon Press, Oxford (1924)
- von Soden, W., Meissner, B.: *Akkadisches Handwörterbuch: unter Benutzung des lexikalischen Nachlasses von Bruno Meissner (1868–1947)*. O. Harrassowitz, Wiesbaden (1965)
- Kernighan, B.W., Ritchie, D.M.: *The C Programming Language*. Prentice Hall, Hoboken (1978)
- Lesk, M.: Some Applications of Inverted Indexes on the Unix System. *UNIX Programmer's Manual: Supplementary Documents*, vol. 2, 7th edn. Holt, Rinehart and Winston, New York (1982)
- Refer (Software). Wikipedia, 28 Mar 2022. [https://en.wikipedia.org/w/index.php?title=Refer\\_\(software\)](https://en.wikipedia.org/w/index.php?title=Refer_(software))
- Muellner, L.: The free first thousand years of Greek. In: Berti, M. (ed.) *Digital Classical Philology*, pp. 7–18. De Gruyter Saur, Berlin (2019). <https://doi.org/10.1515/9783110599572-002>
- Open Greek and Latin: An International Collaboration Committed to Creating an Open Educational Resource for Ancient Greek and Latin. <https://opengreeklatin.org/>. Accessed 1 June 2022

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.