# Cyberinfrastructure for Classical Philology

Gregory Crane  <gregory.crane@tufts.edu>, Tufts University
Brent Seales  <seales@uky.edu>, University of Kentucky
Melissa Terras  <m.terras@ucl.ac.uk>, University College London

**Abstract**
No humanists have moved more aggressively in the digital world than students of the Greco-Roman world but the first generation of digital classics has seen relatively superficial methods to address the problems of print culture. We are now beginning to see new intellectual practices for which new terms, eWissenschaft and eClassics, and a new cyberinfrastructure are emerging.

1

*For Cathy, Lincoln, Adrian and Russell*

The Athenians grew in power and proved, not in one respect only but in all, that equality is a good thing. Evidence for this is the fact that while they were under tyrannical rulers, the Athenians were no better in war than any of their neighbors, yet once they got rid of their tyrants, they were by far the best of all. This, then, shows that while they were oppressed, they were, as men working for a master, cowardly, but when they were freed, each one was eager to achieve for himself. (Herodotus 5.78, tr. after Godley)

I am no sculptor, to make statues fixed motionless on the same pedestal. Go, sweet song, on every merchant-ship and rowboat that leaves Aegina, and announce that Lampon's powerful son Pytheas [5] won the victory garland for the pancratium at the Nemean games. (Pindar Nemean 5.1-5, tr. after Diane Svarlien)

2

The first passage above follows a military encounter in which the Athenians show, for the first time, that terrible energy which would (at least according to our Athenian sources) fascinate and unnerve the rest of fifth-century Greece. Students of classical Athens have for millennia contemplated the energy that liberation released — Herodotus' wonder has echoed ever since and served as one motivation for human fascination with Athens and its achievements.[1]

3

The early years of the twenty-first century have seen a heroic age for intellectual life. Ideas have poured across the world and new minds have joined the professionalized academics and authors in grappling with the heritage of humanity. Often rough and unpolished, unconcerned with the niceties of convention, a new generation of digital entities has exploded across human society, creating wikis, blogs and millions of electronic resources. Plato's Socrates scorned writing itself on the grounds that the written word was as powerless to answer our questions as a mute painting (Plat. *Phaedrus* 275d). Now each question becomes a challenge as active readers probe relentlessly the sprawling information space beneath their fingers. Wikipedia has demonstrated a new form of intellectual production that challenges the assumptions that many of us internalized in graduate school about how knowledge can be described and ideas shared.[2] The scale of projects such as Wikipedia deserves serious reflection:

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

the English Wikipedia has, as of summer 2008, more than 2.4 million entries.[3] By one estimate, Wikipedia has absorbed 100 million hours of labor — put another way, Wikipedia has, if measured by the labor invested, become a billion dollar project [Shirky 2008].

4

Changes go beyond traditional academic channels. The 9/11 attacks in 2001 were the last major event owned by the centralized 20th century media. With the Tsunami and the 7/7 London bombings, we had shifted "from the broadcasters owning the story, to the people involved in the events owning their own stories and spreading it to those who they know and care about, using their own communication channels." [4]

5

Conventional streams of refereed publications (such as this collection) are necessary but insufficient — this introduction has already cited Wikipedia, a blog and the video for a presentation at a conference. We cannot make the decisions that we need to live in the world around us without constantly evaluating information that has no conventional academic pedigree. Every anxious editorial fretting about undomesticated ideas prowling through an internet jungle underscores the urgent need for that critical thinking that we in the liberal arts claim to instill. The internet may prove to be the best thing for humanities education since the rulers of early modern Europe found that classical training provided them with the administrators with whom to build strong nation states.

6

No field of study is poised to benefit more than those of us who study the ancient Greco-Roman world and especially the texts in Greek and Latin to which philologists for more than two thousand years have dedicated their lives. Our predecessors worked in Alexandria, Damascus and Baghdad as well as Berlin, Oxford and Venice. Many lived in states whose names we may never have heard. Most spoke languages like Syriac or the dialects of medieval Europe, which have themselves passed into history. They preserved the battered remains of the past in isolated monasteries and the libraries of aristocrats. They raised capital and set type, then sent Greek and Latin texts coursing through Europe and then the world. They convinced the powerful that the study of Greek and Latin would provide the supple and disciplined minds needed to fashion, maintain and expand the nation states of Europe and their empires across the world. And in the twentieth-century, as other disciplines emerged as filters to identify the promising and send them on their way to worldly privilege, classicists carried their field forward, opened up their curricula to those who had not learned Greek and Latin and, from the margins of the intellectual world, continued their researches on the texts that they loved

7

The papers in this collection reflect a new generation of classicists — entrepreneurial in their disruptive actions, impatient of convention, hunting for new methods to understand and to disseminate those ancient texts to which they, like dozens of generations before them, have dedicated their lives. It is hard to predict what the future holds for the intellectual practices and products of twentieth century print-based classical studies. In the opening of his fifth Nemean Ode Pindar reveled in the speed and reach of the written word: his songs could be copied and race across the known world in the largest ship and the smallest boat, while the grandest statues remain fixed and mute upon their pedestals. The texts of antiquity, freed from the tyrannical limitations of expensive print publication, preserved in

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

multiple servers across the globe, flash instantaneously anywhere that the internet can reach — hundreds of millions of desktops and mobile devices. Homer, Plato, Virgil, Cicero — they all reach more of humanity than ever was conceivable in the millennia since they set down their styli for the last time and passed into dust. And it is not just physical access — we already can, with simple links between source text and its commentaries, translations, morphological analyses and dictionary entries, provide a better reading environment than was ever conceivable in print culture. We know from the readers of our web sites that texts in Greek and Latin, of many types, now fire the minds to which twenty years ago they had no access. And if this reading environment now supports those proficient in English, we can already design libraries that will, within a reasonable period of time, support readers in the less commonly spoken languages of the European Union such as Croatian and Hungarian and widely spoken languages such as Arabic and Chinese.

# Terms and continuities

### Wissenschaft and Philology

As to the speeches which were made either before or during the war, it was hard for me, and for others who reported them to me, to recollect the exact words. I have therefore put into the mouth of each speaker the sentiments proper to the occasion, expressed as I thought he would be likely to express them, while at the same time I endeavoured, as nearly as I could, to give the general purport of what was actually said. [2] Of the events of the war I have not ventured to speak from any chance information, nor according to any notion of my own; I have described nothing but what I either saw myself, or learned from others of whom I made the most careful and particular enquiry. [3] The task was a laborious one, because eye-witnesses of the same occurrences gave different accounts of them, as they remembered or were interested in the actions of one side or the other. [4] And very likely the strictly historical character of my narrative may be disappointing to the ear. But if he who desires to have before his eyes a true picture of the events which have happened, and of the like events which may be expected to happen hereafter in the order of human things, shall pronounce what I have written to be useful, then I shall be satisfied. My history is an everlasting possession, not a prize composition which is heard and forgotten. (Thuc. 1.22, tr. Jowett)

8

The distinction between science and the humanities reflects particular traditions of the English speaking world. In German, for example, *Wissenschaft* includes all systematic intellectual work — we need to specify *Naturwissenschaft* or *Geisteswissenschaft* if we want to distinguish between the natural sciences and the humanities. The term *Altertumswissenschaft* describes the systematic analysis of the past, including both the textual and the material record.

9

We thus use the term Wissenschaft to describe the output of the systematic study of antiquity as it appears in material forms such as articles and monographs, plans and maps, images and diagrams, editions and reference works. Whether or not we believe that we can reconstruct aspects of the ancient world as they actually were, we develop our ideas on the basis of primary and secondary sources stored in material form.

10

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

For the purposes of this introduction, philology describes the production of shared primary and secondary sources about linguistic sources, while classical philology focuses upon classical Greek and Latin, as these languages have been produced from antiquity through the present. The famous passage from Thucydides, quoted above, is relevant for several reasons. First, Thucydides was one of the first to apply systematic methods to represent in textual form, as accurately as he could, the events of the past — his history of the Peloponnesian War has been a model for Wissenschaft. Second, Thucydides used writing as a medium to disseminate his ideas, but he drew upon every source available, including eyewitness interviews, archaeological remains, and the textual record. Third, Thucydides' words seek to represent an entire world — we cannot fully study Thucydides without engaging as well with the material record. Nor is this material record simply a source with which to illustrate the topics that Thucydides has included. We need to develop the fullest possible understanding of the material record in order to develop our own understanding of how Thucydides represents his subject.

11

The terms eWissenschaft and ePhilology, like their counterparts eScience and eResearch, point towards those elements that distinguish the practices of intellectual life in this emergent digital environment from print-based practices.[5] Terms such as eWissenschaft and ePhilology do not define those differences but assert that those differences are qualitative. We cannot simply extrapolate from past practice to anticipate the future.

**Classics and the Humanities**

Socrates: I heard, then, that at Naucratis, in Egypt, was one of the ancient gods of that country, the one whose sacred bird is called the ibis, and the name of the god himself was Theuth. He it was who [274d] invented numbers and arithmetic and geometry and astronomy, also draughts and dice, and, most important of all, letters. Now the king of all Egypt at that time was the god Thamus, who lived in the great city of the upper region, which the Greeks call the Egyptian Thebes, and they call the god himself Ammon. To him came Theuth to show his inventions, saying that they ought to be imparted to the other Egyptians. But Thamus asked what use there was in each, and as Theuth enumerated their uses, expressed praise or blame, according as he approved [274e] or disapproved. The story goes that Thamus said many things to Theuth in praise or blame of the various arts, which it would take too long to repeat; but when they came to the letters, "This invention, O king," said Theuth, "will make the Egyptians wiser and will improve their memories; for it is an elixir of memory and wisdom that I have discovered." But Thamus replied, "Most ingenious Theuth, one man has the ability to beget arts, but the ability to judge of their usefulness or harmfulness to their users belongs to another; [275a] and now you, who are the father of letters, have been led by your affection to ascribe to them a power the opposite of that which they really possess. For this invention will produce forgetfulness in the minds of those who learn to use it, because they will not practice their memory. Their trust in writing, produced by external characters which are no part of themselves, will discourage the use of their own memory within them. You have invented an elixir not of memory, but of reminding; and you offer your pupils the appearance of wisdom, not true wisdom, for they will read many things without instruction and will therefore seem [275b] to know many things, when they are for the most part ignorant and hard to get along with, since they are not wise, but only appear wise."

12

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

Those of us who grew up hearing that we should read more and that television had damaged our minds may smile when we hear Plato's Socrates two and a half millennia ago criticizing the written word for damaging our minds. In the early twenty-first century, complaints have emerged about the look-up culture of Google and ubiquitous connectivity.[6]

13

Nevertheless, the basic point remains valid, even if the media change. We must augment our biological memories by using material records, whether these are hand-written, printed or digital but external information can only augment internalized knowledge. We can only experience humor, for example, if we understand the joke as it happens. We can work our way through a Greek text, looking up every word in a dictionary and using modern translations to orient ourselves, but we will not understand the text in the same way as we would if we could understand the language fluently. And, even if we understand the Greek words and grammar, we will hear more from those words the more we have thought about Plato, the philosophical concepts that form the subject of his dialogues, and the culture in which he lived.

14

Thucydides set out to express in material, written form a record of the past that would last forever. Plato questions the value of any written record except insofar as that record finds full expression in human minds. We already live in a world where the books have begun to talk with each other.[7] When data mining systems detect fraudulent activity on our credit cards, they do a better job of finding significant patterns than could human analysts alone — if there were human analysts to sift through trillions of transactions. Financial institutions do not care how they identify fraud because fraud detection is a means to an end.

15

Text mining can detect words and phrases that are unusual in Plato.[8] We can even imagine syntactic analyzers that can not only parse every surviving Greek and Latin word but that might at some point be better able to justify its decisions by pointing to other similar patterns in that vast corpus than has ever been possible for any human reader. But such information would only realize its full value if it becomes knowledge in a living human mind and allows a reader to see something that would not otherwise have been visible.[9]

16

For the purposes of this discussion, we use the terms classics and the humanities to describe that focus upon internalized knowledge and intellectual practices designed to help us perceive new connections and increasingly sophisticated patterns not only in the texts that we read but in the images that we see and the sounds that we hear. Human beings are the measure of all things in the humanities. Philology truly matters insofar as it serves classics and its goal of bringing classical Greek and Latin to life in the minds of human beings.

**Infrastructure**

Tell me now, you Muses that have dwellings on Olympus— [485] for you are goddesses and are at hand and know all things, whereas we hear but a rumor and know nothing—who were the captains of the Danaans and their lords. But the common folk I could not tell nor name, no, not though ten

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

tongues were mine and ten mouths [490] and a voice unwearying, and though the heart within me were of bronze, did not the Muses of Olympus, daughters of Zeus that bears the aegis, call to my mind all who came beneath Ilios. Now will I tell the captains of the ships and the ships. (Homer, Il. 2.484-493, tr. after A. T. Murray)

17

Infrastructure provides the material instruments whereby we can produce new ideas about the ancient world and enable other human beings to internalize those ideas. Infrastructure includes intellectual categories (e.g., literary genres, linguistic phenomena, and even the canonical book/chapter/verse/line citation schemes whereby we cite chunks of text), material artifacts such as books, maps, and photographs, buildings such as libraries and book stores, organizations such as universities and journals, business models such as subscriptions, memberships, and fee simple purchases, and social practices such as publication and peer review. Our infrastructure constrains the questions that we ask and our sense of the possible. Thus, the Homeric narrator rules out the idea of representing the names of every hero who participated in the Trojan War. The twenty-first century fan of American baseball can, by contrast, locate not only the name but the basic statistics recorded for every person who ever threw a pitch or swung the bat in a major league game. By the classical period, we begin to find lists of citizens, office-holders, temple dedications, tribute paid and similar categories.

18

Thucydides drew upon textual, archaeological and verbal sources and he could leave behind a written text to which he had attached his own name, but there were no libraries in the modern sense. He could not cite transcripts of public speeches in a congressional record or even a *New York Times* article. He could not footnote official documents in a classical Greek equivalent to the *Official Records of the Union and Confederate Armies* [United 1880].[10] There were no recordings of those who survived to describe civil war in Corcyra or the Sicilian Expedition. He could not publish pictures or even expect that diagrams would be faithfully reproduced over time. A stream of words was the only medium by which he could represent his chosen subjects.

19

Infrastructure is so fundamental that it may become invisible to us but the resulting blindness makes us confuse the limits that we face with our larger goals. In periods where our infrastructure advances incrementally, we may take it for granted. Infrastructure does not simply affect the countless costs/benefit decisions we make every day — it defines the universe of what cost/benefit decisions we can imagine.[11]All the tribute of the Athenian empire could not have paid for one color photograph of Pericles. Rarely, if ever, can we predict the full implications of relatively modest technological change. Gutenberg did not think that, in using movable type to print a Latin bible, he was creating a technology to make translations of the bible ubiquitous, enable new forms of Christian worship and facilitate revolutionary change.

20

But even if we cannot foresee the future with perfect clarity, we must constantly reexamine the goals that we choose to pursue today in the light of what is already possible. Before shifting to the digital infrastructure already taking shape and its implications for current practices in classical philology, we should review what has and has not changed for classical philology as the core information

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

infrastructure of human life as a whole has shifted, decisively and irrevocably, from atoms to electrons.

**Classics in 2008**

I shall begin with our ancestors: it is both just and proper that they should have the honor of the first mention on an occasion like the present. They dwelt in the country without break in the succession from generation to generation, and handed it down free to the present time by their valor. [2] And if our more remote ancestors deserve praise, much more do our own fathers, who added to their inheritance the empire which we now possess, and spared no pains to be able to leave their acquisitions to us of the present generation. [3] Lastly, there are few parts of our dominions that have not been augmented by those of us here, who are still more or less in the vigor of life; while the mother country has been furnished by us with everything that can enable her to depend on her own resources whether for war or for peace. (Pericles' *Funeral Oration*: Thuc. 2.36.1-3)

21

Classicists can identify with the Athenian audience of Pericles' *Funeral Oration* — at least, the oration that Thucydides presents to us. We do not, like the Athenians, like to say that our ancestors were sprung from the dirt and our ancestors have not inhabited the same small rocky peninsula since they were sprang from the earth — classicists have come from countries and periods far beyond the experience of any classical Greek. Our field has an ancient history but we have begun to expand, like the Athenians of fifth century Athens, into a much larger space than we ever could occupy before. The digital world has become our sea, but our empire offers freedom, and the natural borders that will contain our field are nowhere to be seen. Much as we may have achieved, we are still as a field in the incunabular phase of development, more focused upon the problems of the past than the opportunities of the present.[12]

22

Classicists were among the first humanists to exploit digital technologies and enjoy a reputation as being arguably the most digitally advanced field. Certainly, classicists were, as a field, early adopters. If one includes the study of any Greek and Latin texts under Latin, Father Busa's famous concordance of Thomas Aquinas, produced with the help of IBM in the late 1940s, would constitute the start of digital classics (see [Busa 1974] and [Busa 1980]).

23

If we restrict ourselves to the Greek and Latin authors commonly taught in classics departments of the 20th century, then we must move twenty years forward to the late 1960s. Full professors of classics today have been born after David Packard, who working in the basement of the Harvard Science Center digitized the text of Livy. There are classics majors who received their undergraduate degrees in the spring of 2008 who were born after the Perseus Digital Library began serious work in the late spring of 1987. Not only are virtually all publications — whether distributed in print or not — produced digitally, but digitized textual corpora, digital versions of printed secondary sources, electronic reviews, bibliographic databases, and web sites are all standard elements of our work.[13] Two leading departments of classical philology have even discovered the value of the preprint servers on which some of the most demanding areas of research have depended for more than fifteen years.[14]

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

24

The early use of digital tools in classics may, paradoxically, work against the creative exploration of the digital world now taking shape. Classicists grew accustomed to treating their digital tools as adjuncts to an established print world. Publication — the core practice by which classicists establish their careers and their reputations — remains fundamentally conservative. While we may congratulate ourselves on the innovative content of what we write and while we will always need publications that articulate particular arguments at a particular point in time in a particular voice, the format of our publications is essentially the same as that which Gibbon used in the 18th century.[15]

25

While the documents were digital in form, almost none of their content was machine actionable: strings such as "Thuc. 1.38.2" had not been analyzed and converted into machine actionable links to the text of Thucydides, book 1, chapter 38, section 2; a reference to Thucydides did not have associated with it any information whereby an automated system could reliably determine whether this Thucydides was the historian or one of the various other figures by this name; quotations of Greek and Latin authors were not dynamically linked to multiple online editions, nor did they carry with them links to any linguistic apparatus (textual notes, dictionaries, grammars, commentaries, translations) not offered by the author of the articles. While these articles may be online, the main bibliographic resource for classical studies, *L'Année Philologique*, still relies upon manual summaries to index and disseminate these articles in its digitally disseminated bibliography. Nor can the reader, of course, see what later articles cite earlier publications.

26

We can add each of the features listed above to existing documents automatically with reasonable accuracy — simple text search provides functionality that is increasingly comparable to the manually produced indices on which we had to rely in print culture. [16]Google has already popularized the ability to identify and disambiguate place names and to find quotations embedded in unstructured text — automatically generated maps became a standard feature of Google Books in 2007 and frequently quoted passages soon followed.[17] Particular domains may need to adapt general services to their needs: classicists need Optical Character Recognition (OCR) systems that can not only provide useful results for classical Greek but can also recognize Latin and do not helpfully convert *t-u-m* (a Latin word for "then") into English *t-u-r-n*.[18] Scholarly disciplines need page layout analysis systems that can recognize and parse not only general document formats such as notes at the bottom of the page, and the individual entries of indices, encyclopedias, and lexica, but also specialized document formats such as the commentary and textual notes.[19] Scholarly disciplines such as classics need specialized named entity searches: we need to determine not only whether "Th. 1.38" is a citation to a primary source but also, if so, whether it designates Thucydides, book 1, chapter 38, Theocritus, *Idyll* 1, line 38 or some other text.

27

The production of these services is the most important task for classics and for any scholarly discipline which does not focus solely upon the contemporary English-language, mass market American culture which the Web of 2008 primarily serves. While we may need to support less and less software, we will then only shift our efforts to the production and refinement of the knowledge sources which support general services: we need machine actionable reference works that can help

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

general services run by giants such as Google to distinguish one Antonius or one Alexandria from another.[20]

28

Classicists of the 20th century built their work upon a foundation that took shape in the 19th century. In the last decades of the twentieth century, ambitious classicists began to shift their efforts away from infrastructural tools such as editions and commentaries. Instead they turned towards articles and expository monographs on topics often derived from their colleagues in the Modern Language Association. The Pax Stereotypica of the 20th century has, however, collapsed. We live in a digital age in which we need to rethink our most fundamental resources -- we are reinventing the forms and functions of our editions, lexica, encyclopedias, commentaries, grammars, bibliographies and every other textual category that evolved in a print ecosystem. And as we feel our way forward, we need to rebuild our entire infrastructure. In a primarily print world, we can turn to digital tools for documents that contribute at the margins — e.g., digital scholia for a major classical author. In the digital world, we want the scholia but we also need editions of our canonical authors. The *Editiones Principes Electronicae* for every major author are still waiting to be produced. A new generation of editors spreads across a new and uninhabited world in which they can acquire for themselves the digital *kleos aphthiton* ("undying fame") that the pioneers of Hellenistic Alexandria and early modern Europe earned for themselves.

29

The greatest barrier that we now face is cultural rather than technological. We have all the tools that we need to rebuild our field, but the professional activities of the field, which evolved in the print world, have only begun to adapt to the needs of the digital world in which we live — hardly surprising, given the speed of change in the past two decades and the conservatism of the academy.

30

Perhaps the most important point of continuity — and the greatest reason why publication in classics has adapted so little to the digital world — appears before we even begin reading publications. An informal survey reveals that forty of forty-one classics publications available online from Johns Hopkins University Press (97.5%) are products of a single author — the only exception was an archaeological publication in *Hesperia*, the journal of the American School at Athens.[21] While expanding this survey would provide greater statistical certainty, the conclusion would be the same: classicists in 2008 devote most of their energies to individual expressions of particular arguments.

31

An even more problematic issue is that the editions, commentaries, grammars, lexica, and other elements of scholarly infrastructure have not adapted in any significant way to the digital world.[22] In the five centuries since the first printed editions of classical texts began to appear, print culture assembled an immense amount of intellectual capital with which to support thinking about Greek and Latin texts. This knowledge must, however, be converted into a machine actionable form.[23] Converting this intellectual capital from human readable print to machine actionable knowledge is both fundamental and complex: we need to convert statements such as " *facio, facere, feci, factum* " into something that a morphological analyzer can use to recognize a form fecisset as the pluperfect form of the verb *facio*; we need to mine from a set of encyclopedia articles the data that will allow us

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

to search primary and secondary sources alike for one among dozens of historical figures named *Antigonus*; we need grammars and lexica that provide not only a handful of examples but that can also locate the phenomena that they describe in any corpus of Greek or Latin; we need editions that can tell us precisely, how and how often they differ from another and which previous editions and/or manuscript witnesses they follow most closely.

32

More than fifteen years ago the Text Encoding Initiative (TEI) was circulating methods with which to create machine actionable editions that can support advanced services and, more importantly, can be updated and maintained over time [Sperberg 1994].[24] The process was an open one that invited participation from scholars in Europe and North America. Any editor developing a capital resource such as a text, designed to serve an intellectual community for decades to come, had an opportunity to learn how to design a digital edition that could be printed in the short term and then maintained — and even updated — over time.[25] In the fifteen years that have passed since the TEI documented how to produce digital editions, a new generation of scholars has passed from secondary school to the faculty, but all of the new editions of classical authors still appear as static print documents, the rights sold to commercial publishers.[26] If the electronic files were freely available, they would be of limited use because their authors did not follow the guidelines that the TEI published. Classicists have relied for the most part on the *Thesaurus Linguae Graecae* (TLG) to provide searchable versions of the reconstructed texts that have appeared — without the introductions, textual notes, indices or other scholarly apparatus available in any digital form.

33

Converting digital editions to print is a particularly messy task. Editors often do not repeat in the textual note the precise passage to which the textual note applies — they assume that their human readers will be able to make these connections themselves. In a recent study, Federico Boschetti applied a range of techniques with which to associate the notes in a textual apparatus with the appropriate place in the text. He found that these techniques could correctly associate only about 80% of the textual notes with the text to which they referred [Boschetti 2007]. This does not even address the task of analyzing the content of the textual notes so that we can then pose queries such as "where does MS P differ from V by using the same grammatical form but P and V use different dictionary words," "visualize the evolution of the text of Aeschylus, allowing me to see how each edition differs from those which precede it, which editions are most closely related to one another and which editions have been most influential," or "which variants have the biggest apparent impact on the text based on a range of criteria."

34

The articles in this collection reflect the most recent stage in the evolution of digital classics and point to the future, but to appreciate that future, we need to review major developments on which that future builds. These articles point forwards to an emergent Cyberinfrastructure, but this Cyberinfrastructure builds upon three earlier stages of digital classics: incunabular projects, which retain the assumptions of print culture, knowledge bases produced by small, centralized projects, and digital communities, which allow many contributors to collaborate with minimal technical expertise

**Digital Incunabula: the *Thesaurus Linguae Graecae* (1972)**

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

35

Digital incunabula are forms that replicate the established forms of print. Thus, the TLG was, in the early 1970s, designed as a gigantic, infinitely flexible concordance. Its texts capture the basic page layout and canonical citations of the original editions, and a sample search of it is illustrated in Figure 1. The Bryn Mawr Classical Review has been successful because it used forms such as email and then the Web to produce traditional reviews that any classicist could produce and read. The digitized publications in JSTOR, Project Muse, and Google Books provide new methods by which to search and disseminate knowledge, but the ultimate objects of exchange are facsimiles of exchange. These projects tend to require either very large or very small capital investments. They focus on producing, as quickly as possible, the same intellectual objects to which their communities are already accustomed. In this stage of work, catalogues may grow far more elaborate — the TLG and JSTOR allow us to search all the words in primary and secondary sources, while Google dynamically generates maps of places and lists of frequently quoted passages automatically extracted from its image books. All of these projects provide, in effect, a new generation of catalogues where the books remain unchanged. The system designers do not want to get bogged down in the specifics of any particular domain, while the domain experts do not want to get bogged down in the technology.

Figure 1.

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

A search of the TLG digital library containing 100 million words of classical Greek texts. First begun in 1972, the TLG provides word searches of various types that deliver excerpts of text that mirror print sources — even the hyphens are retained. The most important contributions of the TLG are (1) very accurate transcriptions of the text (without textual notes, introduction, indices etc.) and (2) encoding one canonical citation scheme by which scholars cite these sources. Incunabular systems have themselves evolved. Storage has grown so much less expensive (by one measure, at least 300,000 times cheaper [27]) that more recent systems assume page images of the original are available. The representative of one national library asserted that it would not even accept collections of transcribed text without images of the original pages.

36

Incunabular systems have been under development for a long time — there are tenured professors of classics who were born after the TLG began work in 1972. Figure 2 illustrates the generation of incunabular systems that emerged in the 1990s with a sample text from the Open Content Alliance (OCA), whereas Figure 3 illustrates a sample from Google Books. Where the TLG provides a fully transcribed version of source texts, the OCA, Google Books and other projects provide only scanned page images and such text as OCR software can generate. These projects provide noisier — and, in the case of Greek, no — searchable text, but they index all of the text on the page, and their accuracy will increase as OCR software becomes more sophisticated.[28] Also, projects such as the OCA provide open-content licenses and encourage third parties to download and repurpose the scanned page images. Thus, the Mellon-funded *Cybereditions* Project is creating within the OCA an open source library of Greek and Latin critical editions, on which advanced services can be built. The scanned editions, though simple in form, provide a foundation on which more sophisticated digital objects can be built: no license will later pull these image books out of circulation and no license restricts the ways in which they can circulate.

Figure 2.

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

Twenty-first century incunabular publications such as the books digitized by the OCA are designed not only to provide useful services in the present but to be integrated into more sophisticated services over time. The digitized collection of fragmentary Greek historians above will be joined by a digital edition that builds upon, precisely references and extends the content of the print edition. Such composite editions are part of the fourth-generation collections described in Classics in the Million Book Library (Crane et al., in this collection).
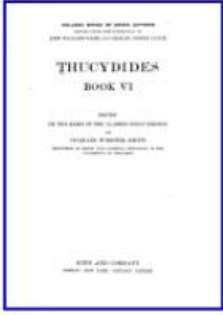
37

In the incunabular stage, if you retrieve a book in a language that you cannot read or on a topic that you cannot understand, then it is your responsibility to find a translation and any other background information you may need to make sense of what is before you. In the incunabular stage, the center of computation is external to the document, emphasizes general algorithms and depends upon little, if any, domain specific machine actionable knowledge. In incunabular projects, the physical distance between readers and publications dissolves.

Figure 3.

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

| About this book | Read this book | **Thucydides** By Thucydides, Charles Forster Smith |
| --- | --- | --- |

**Subjects**
Greece

**More details**
*Thucydides: ed. on the basis of the Classen-Steup edition*
By Thucydides, Charles Forster Smith
Published by Ginn, 1913
Original from the University of Michigan
Digitized Dec 13, 2006
250 pages

Add to my library
Write review

[ Read this book ]
[ Download PDF ]

**Buy this book**
Abebooks
Alibris
Amazon
Barnes&Noble.com
Google Product Search

**Borrow this book**
Find this book in a library
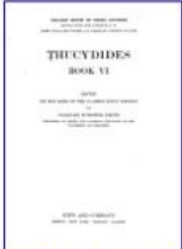
**Contents**

**Selected pages**

Page 253 | Title Page | Index

**Search in this book**
[                    ] [ Search ]

**Popular passages**

Λεοί'τά'ους τε ἀναστήσαντες ατιμώρητοι γενήσονται και τους λοιπούς ἔτι ξυμμάχους αυτών διαφθείροντες αυτοί την ἅπασαν δνναμιν της Σικελίας σχήσουσι, κίνδυνον είναι μη ποτέ μεγάλη παρασκευή Δωριής τε Δωριεῦσι κατά το ζυγγενες καί ἅμα ἅποικοι τοις εκπέμψασι Ἱλελοποννησίοις... - Page 15
Appears in 2 books from 1913-1996

Ισονομεῖσθαι ; καὶ πῶς 39 δίκαιον τους αυτούς μη των αυτών

**Other editions**

Thucydides book VI
by Thucydides, Kenneth James Dover - Literary Criticism - 1999
This is an edition of Book VI of Thucydides.
No preview available - About this book - Add to my library

Thucydides
by Tucídides, E. C. Marchant, Thomas Wiedemann.

A commentary on Thucydides as seen in Google Books in July 2008. Note that the general OCR engine has begun to provide output for Greek print that, while still far from perfect, is searchable and comprehensible to an expert reader. Google Books does not, of course, understand the citation scheme by which scholars can cite Thucydides but it has recognized the title page and the index, and it has recognized a page with a map as something of interest.

**Machine-actionable knowledge bases: the Perseus Digital Library (1987)**

38

These kinds of projects, unlike incunabular projects, set out to create knowledge about a particular domain that machines can manipulate and that begin to move beyond the forms of print. In classics, the Perseus Project provides an example of such systems. Perseus set out, in the middle 1980s, to build an environment where knowledge about the ancient world, including both the material and textual record, could be dynamically recombined to support new forms of inquiry. Figure 4 illustrates

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

a sample text as it appears in the Perseus Digital Library.The focus of Perseus was to create resources that were in print either impractical in print (e.g., producing dozens or hundreds of high resolution color images of for thousands of Greek vases) or impossible (e.g., interactive tours of archaeological sites and searching/browsing services based on automated morphological analysis of Greek and Latin).[29] Semantic text markup is a characteristic feature of such projects: rather than simply recording that a word is, for example, in italics, these systems try to interpret the content and thus to record whether the italics indicate rhetorical emphasis, the title of a literary work, a word quoted from a foreign language, or some other category.[30] As these systems grow more intelligent, they convert an increasing portion of the content inside the books into well-structured information that machines can process. These systems depend upon individuals who understand the evolving relationship between the possibilities of technology and the needs of the discipline.[31]

Figure 4.



The figure above illustrates some of the information about the opening lines of the *Odyssey* available in the *Perseus Digital Library*. First, documents in this collection have markup illustrating their logical contents: thus, where incunabular systems can only recognize the physical page divisions, the knowledge base allows the digital library system to recognize for dictionary words in the LSJ Greek-English lexicon the many separate entries that appear within a single page or that begin on one page and end on another. Second, the primary source citations have been automatically analyzed and encoded. Thus, the system can take a chunk of Greek, recognize what lines it contains, and then locate dictionary entries (or commentaries, encyclopedias, articles etc.) that refer to the lines in the chunk displayed. In a mature digital library, citations from one text to another become bi-directional links, allowing readers not only to follow the documents that a particular work cites, but also to find works that subsequently cite the document that they are viewing. Third — and perhaps most importantly, morphological knowledge has been represented in machine actionable form. Thus, an automated

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

system is able to recognize that the string ἔννεπε is a form associated with the dictionary entry ἐνίσπω.

39

Reference materials, in particular, are structured to support automatic systems (e.g., the morphological analyzer learns Greek and Latin morphology from a machine actionable grammar) and to be decomposed into small chunks and then recombined to provide dynamic commentaries. If you retrieve a book in a language that you cannot read or on a topic that you cannot understand, the system can find translations where these already exist, machine translation and translation support systems, reference works, and general background information suited to the general background and immediate purposes of the reader. In knowledge bases, the boundaries between books begin to dissolve.

**Digital Communities: Stoa Publishing Consortium (1997)**

40

Knowledge bases such as Perseus were (and, to a large extent still are), produced by small teams of experts who bridge the gap between the technology and individual disciplines to make documents and the ideas within them intellectually as well as physically more accessible. Digital communities enable more people to participate in more ways and in on-going, dynamic forms. New forms of publication such as wikis, blogs, and various websites open up new instruments with which individuals and groups can contribute in an on-going, dynamic fashion.[32]

Figure 5.

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.



The Suda On Line (described in Anne Mahoney's essay in this collection) illustrates a digital community that emerged in the late 1990s before the rise of Wikipedia. The Suda is a 625,000 word, 30,000 entry Byzantine encyclopedia that offers a great deal of information not otherwise preserved about the classical Greek world. A group of classicists in Europe and North America organized a collaborative project to create the first comprehensive English translations of this resource. Progress has been steady and solid: in April 2000 1,500 entries had been translated, by July 2008 that number had increased by 23,000, with 24,500 entries translated and vetted (see http://www.stoa.org/sol/about.shtml.

41

The Stoa Publishing Consortium, founded in 1997 with a grant from the Fund for the Improvement of Postsecondary Education, has done more than any single effort to foster the rise of digital communities in classics. Stoa.org provided support in a variety of ways to most of the major projects and classicists who emerged over the following decade. One such project, the Suda On Line, is illustrated in Figure 5. The papers in this collection provide an imposing, and still partial, account of the impact which the Stoa has had.

42

If you examine a digital object in a digital community, you can not only find the background information that you need to interpret that object, but you can also make your own contributions by creating annotations directly, producing a blog linked to the object, or in some other fashion. In digital

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

communities, the distinctions between author and reader and between reading and writing begin to dissolve (as the very act of reading becomes a statement of at least initial interest and thus a contribution).[33]

## Cyberinfrastructure

From the anvil Hephaestus rose, a huge, panting bulk, halting the while, but beneath him his slender legs moved nimbly. The bellows he set away from the fire, and gathered all the tools with which he was building a silver chest; and with a sponge wiped he his face and his two hands, [415] and his mighty neck and shaggy breast, and put on a tunic, and grasped a stout staff, and went forth halting; but there moved swiftly to support their lord servants wrought of gold in the semblance of living women. They possessed understanding in their hearts, and speech [420] and strength, and they knew cunning handiwork by gift of the immortal gods. These busily moved to support their lord. (Homer, *Iliad* 18.411-421, tr. after A. T. Murray)

43

The three classes of digital project outlined above reflect three different sources of energy: the industrialized processes of mass digitization and of general algorithms, the specialized production of domain specific, machine actionable knowledge, and the generalized ability for many different individuals to contribute, in ways large and small. When these three sources of energy begin to interact with one another, the resulting environment is qualitatively different not only from print culture but from any of the three digital environments taken in isolation. Having reviewed some developments in the previous generation, we can now begin to consider the implications for ePhilology (primary and secondary sources relevant to classical Greek and Latin), eClassics (ancient Greek and Latin as they work within human minds), and Cyberinfrastructure (the material systems whereby we exchange the objects of our intellectual labor and ourselves internalize these objects). The following sections describe ePhilology and eClassics. The conclusion to this collection returns to the Cyberinfrastructure towards which the individual articles point.

## Producing new knowledge: ePhilology

Any one can discourse to you forever about the advantages of a brave defence, which you know already. But instead of listening to him I would have you day by day fix your eyes upon the greatness of Athens, until you become filled with the love of her; and when you are impressed by the spectacle of her glory, reflect that this empire has been acquired by men who knew their duty and had the courage to do it, who in the hour of conflict had the fear of dishonor always present to them, and who, if ever they failed in an enterprise, would not allow their virtues to be lost to their country, but freely gave their lives to her as the fairest offering which they could present at her feast. (Pericles' *Funeral Oration*, Thuc. 1.43.2)

44

If we think only in terms of word searches, the production of camera-ready copy, image management, the ability to generate basic maps, and manually produced format such as wikis and blogs, increased storage and computational power may seem relatively unimportant. For anyone whose career extends more than a decade, current technologies are astonishingly powerful. In 1982, it cost the Harvard Classics Computing Project $34,000 to purchase a 660 megabyte disk drive to store early versions of the TLG: the disk was the size of a washing machine, arrived in a wooden crate, needed a special disk

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

controller, took two days for the technicians to install and required modifications to the version of the Unix operating system then available. The maintenance contract cost c. $4,000/year and was essential. As this introduction is written, $100 buys a terabyte of storage — more than 1000 times as much storage as its 1982 predecessor for 300 times less money, a decrease in cost of more than 300,000 in one quarter of a century. We can now take for granted storage that was previously unimaginable, collecting huge digital images as well as texts and datasets with little regard for the costs of storage or computation. A generation ago, only a few of the wealthiest departments could raise tens of thousands of dollars to provide the storage to search a few million words of Greek and support the first generation of digital publishing. In 2008, many cell phones have more than enough storage and computational power to do much more.

45

All of us in the academy and in society as a whole, of course, already depend upon general services, such as Google, that require stunning amounts of storage and computational power — even academics who may proudly dissociate themselves from the web of digital services depend completely upon those services for the paper publications that arrive in the mail and the catalogues by which they find books on the shelf. And, of course, we already depend upon digital infrastructure for the paychecks, medical treatments and other fundamental components of material life. Within classical studies, it is easy to see the need for vast networked storage and high performance computing for the analysis and visualization of quantitative and visual evidence from the material culture.[34]

46

Consider the basic problem of reading Greek and Latin. The machine-actionable Liddell-Scott-Jones (LSJ) Greek-English and Lewis and Short Latin-English lexica developed by the Perseus Project contain 422,000 and 303,000 tagged citations to 800 Greek and 80 Latin authors. In LSJ, half of the 422,000 citations are to a half dozen canonical authors. For Lewis and Short, the top dozen authors account for more than two-thirds (215,000) of the citations.

47

Not all lexicographic projects have such narrow focus, but extensive lexicographic coverage is extraordinarily labor intensive. The *Thesaurus Linguae Latinae* (TLL) is building a lexicon that covers Latin from earliest times through AD 600 and bases its work on an archive of 10,000,000 slips with information about particular words. The TLL in 2008 boasts a staff of twenty Latinists, began work in 1894, published its first fascicle, and has been an international project since 1949. Its official website promises that the TLL will during 2009 "reach the end of the letter P, at which point more than two thirds of the complete work will have appeared".[35]

48

The ten million or so words of ancient Latin may require more then a century of labor, but they constitute, of course, a relatively small corpus. The TLG had accumulated 99,000,000 words in 2007.[36] An individual Latinist, Johann Ramminger, had accumulated a wordlist of later Latin from Petrarch up through 1700 that was based on 200,000,000 words of text already available in digital form. Semi-automated methods involving computerized data but still dependent upon manual analysis of each form may increase productivity by a factor of two or three, but simply enhancing traditional approaches would require centuries to provide us with truly comprehensive lexica of Greek and Latin.

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

49

No branch of scholarship is probably older than lexicography, but our traditional methods do not scale up to the challenges of representing textual materials in Greek and Latin. We have no choice but to exploit, as vigorously as we can, automated methods. The essay by Bamman and Crane in this collection describes some of these methods as they exist today. The essay by Finkel and Stump illustrates how automated methods can reconfirm — but place on a profoundly new foundation — ancient analytical instruments such as the reduction of Latin verbs to a four dimensional space defined by the traditional principal parts.

50

Ultimately, automated and manual methods reinforce one another. Decisions embedded in print reference materials such as lexica, indices, and grammars can be, at least in part, extracted and converted into machine actionable data. In effect, human annotators provide the examples and rules from which automated systems learn. The automated systems present the results of what they learn when they work with new materials. Human readers then correct and augment the automated results. The automated systems recalculate their statistical models and then recalculate.[37] In a mature system, we separate training data from test data so that we can automatically measure the impact that our changes have upon performance.

51

Complex algorithms can be computationally demanding even when we are working with small corpora. In preliminary work on sense detection in 2005, we found that by comparing five different translations with the 150,000 Greek words in Thucydides we can identify words with many senses in Thucydides: e.g., passages where the Greek word archê corresponds to "beginning" or to "empire". It took days of processing power from a single CPU to identify clusters of word senses in five translations of the 150,000 words in Thucydides.[38] Even if we shift to these algorithms, analyzing millions of words and thousands of translations in a half dozen languages would require more computational power than any desktop system could readily deploy.

52

The infrastructure of 2008 forces researchers in classics and in the humanities to develop autonomous, largely isolated, resources. We cannot apply any analysis to data that is not accessible. We need, at the least, to be able gather the data that is available today and, second, to ensure that we can retrieve the same data in 2050 or 2110 that we retrieve in 2010.[39] We need digital libraries that may be physically distributed in different parts of the world but that act as a single unit: we need to be able to pose queries such as "find all Greek editions and modern language translations of Aeschylus, *Persians*, lines 1-40" and retrieve machine actionable results from a variety of sites.[40]

53

There are two components to this problem. First, we need libraries that can preserve collections in the digital world as they have preserved them in the print world. The institutional repository movement is slowly addressing this challenge.[41] Thus, the publications in this collection are a part of a long-term institutional repository that can manage static expository prose with very general features such as sections, footnotes, bibliography etc.

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

54

We need, however, more than digital preprints. A second component is the need for sophisticated citation and reference linking services. Smith's paper in this collection, "Citation in Classical Studies", describes the system of canonical text citations by which classicists identify precise chunks of text within the surviving corpus of classical Greek and Latin. The Canonical Text Services (CTS) described in this piece begin where library catalogues end and provide furthers layers of granularities essential for classical scholarship: the CTS provides a common language whereby we can aggregate information about particular lines in the *Iliad* or a numbered section from a chapter in Thucydides.[42]

55

The TEI has developed a shared language whereby humanists can describe the same phenomena in similar ways so that we can more readily combine documents produced by different groups. The TEI has many different methods, however, and it is possible to represent the same phenomenon in many different TEI-compliant ways. Cayless et al. describes how experts in Greek inscriptions as a community adapted the very general TEI framework to their needs, allowing classicists to create documents that are increasingly interoperable and easy to maintain over time. Robertson documents research in methods to describe historical events in a format that is not only machine actionable but language independent, contributing to the production of multilingual scholarship. Dué and Ebbott describe editorial standards for a new generation of dynamic digital editions. These new editions do not simply provide a single best attempt at reconstructing a single text but can dynamically represent multiple versions of the text as it has appeared over time and provide databases of variants, conjectures, testimonia and other materials. Elliott and Gillies look more generally at how we can then build on these and other services to manage geographic information about the ancient world in new ways. Wikipedia has provided a famous and famously successful model for distributed authorship, but classicists had already begun pioneering such systems in the 1990s. Mahoney's article describes the infrastructure for the Suda On Line project, which has produced translations for more than 24,000 entries of a fundamental reference work about the classical Greek world produced in 10th century Byzantium. At the same time, Finkel and Stump illustrate how methods from computer science can manage such fundamental structures as Latin morphology.

56

And, of course, only a small part of the printed record relevant to classical Greek and Latin has been — or will be — carefully transcribed and edited. If we begin to consider the challenge of extracting and analyzing information about classical Greek and Latin scattered throughout very large collections of books available as scanned page images, the challenges of storage and computation become daunting. The collection of essays thus ends with articles about converting print materials into a form that can support the kinds of services that the previous articles have articulated. Rydberg-Cox describes the issues involved in trying to convert early printed scholarship into a machine actionable form. Later publications lend themselves much more readily to automated analysis. Crane et al. consider the problems and opportunities that emerge for classics as whole research libraries become available in digital form.

Figure 6.

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

An early element of cyberinfrastructure for philology: In this display, a reader has inquired about the form ἐξίτηλα. The morphological analysis system has, as it has since the 1980s, forged a link between this form and the dictionary entry ἐξίτηλος, but two elements have been added. First, a simple machine learning system has analyzed morphologically unambiguous words in the Greek database to rank the probability for each possible analysis in this context. It has, however, chosen accusative, the wrong alternative in this case, but one of the readers has added a vote for the correct analysis (the adjective is, in fact, nominative). This figure thus includes (1) a simple transcription of a print source, (2) the output of knowledge-driven systems, and (3) feedback from a digital community which will, in turn, affect subsequent automatic analyses.

57

Infrastructure includes not only data, services and physical systems but the social practices as well. Figure 6 illustrates some of the particular elements of the cyberinfrastructure needed for philology. The papers in this collection illustrate shifts in the practices of classicists as a new cyberinfrastructure develops:

- Expository argumentation: While new forms of scholarship and new intellectual practices are taking shape, we should emphasize that the collection published here reflects the on-going need for expository arguments that articulate particular points of view constructed at a particular time. Nevertheless, even when the superficial form of argumentation remains largely traditional in form, the substitution of dynamic links for static citations can exercise a major impact upon the content and the audience that publications can reach. Stoa.org was founded in 1997 to support, among other things, new forms of publication that would provide rich links to original sources while bringing classics to a broader audience. Thomas Martin's

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

> Overview of Classical Greek History in the *Perseus Digital Library* and Ross Scaife's *Diotima*, an electronic publication on gender in antiquity, did much to inspire this goal. All of the publications associated with the Stoa illustrate forms of publication that were not feasible a generation ago. Christopher Blackwell's *Demos: Classical Athenian Democracy* illustrates how a publication that is traditional in form can exploit online evidence and publication to provide better documentation on a major subject to a wider audience than was feasible in print.

- Collaboration: While the final form of the papers in this collection may be familiar, their production and content reflects a fundamental change in scholarly practice: the majority of the papers published here have multiple authors, while the single-author papers either report on group projects or on general methods whereby classicists can create interoperable data.

- Open access and open source production: All of the scholars who have contributed to this collection depend upon open access and open source production. In contrast, Figure 7 illustrates an example of a much more closed form of access. In cases where authors are making particular arguments at a particular point in time, open access allows third parties to locate and automatically analyze what they have produced: search engines such as Google can index and then deliver their arguments to anyone online; more specialized text mining systems could analyze what has been written to search for trends in scholarship or to apply specialized services designed for classics (e.g., the ability to recognize strings such as "Thuc. 1.86" as citations to primary sources).

Figure 7.



Twentieth century infrastructure in the digital world. Business models are a core component of every intellectual infrastructure. When information circulated on physical books through a thin network of research libraries, subscription models evolved to generate revenue. In a digital environment, such subscription models lead to situations such as that pictured above, where a digital copy of a two page

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

review, produced by a scholar to reach the widest possible audience, distributed by a non-profit organization (JSTOR) would cost $19US. The medium sends a strong message to the general public.

58

The authors of these papers represent, however, a greater advance than the work that they have produced so far. In part, this reflects the hope that they will produce even more in the future. They also represent a new community, one large enough to foster junior scholars within the field, and in this way they may indirectly spawn far more productive work than all of them could in the aggregate produce during their own careers. But more significant than any output is the sense within this community that the field of classics is being reborn and that limitations with which many of us grew up are no longer relevant. This new digital world not only changes what we can do but who can do what. The collection of essays thus opens with Blackwell and Martin's article about undergraduate research. Before introducing that discussion, we need return to the broader topic of classics and the humanities in a digital environment that has begun to increase the intellectual reach of humanity as a whole

**Extending the intellectual reach of humanity: eClassics & eHumanities**

In short, I say that as a city we are the school of Hellas; while I doubt if the world can produce a man, who where he has only himself to depend upon, is equal to so many emergencies, and graced by so happy a versatility as the Athenian. (Pericles' *Funeral Oration*, Thuc. 2.41.1)

59

We look to a new digital infrastructure not only so that we can increase the body of published information about classical Greek and Latin but so that these languages can play an increased role in the intellectual life of humanity. We can do this in two ways. First, we can create environments that more fully engage those already working with Greek and Latin — we have already begun to address this by creating searchable corpora of Greek and Latin, by making secondary sources available online as PDF files or by adding links between inflected words in a text and their dictionary entries and thus reducing time spent flipping large dictionaries. These all reduce the time between when we pose a question and when we receive an answer. It would be hard to overstate the degree to which cost-benefit decisions, often unconscious, shape the directions that we take in our intellectual lives. Classicists have for millennia understood the difference between being in a small, poorly organized collection and a large collection in which it is easy to find what we want. Cyberinfrastructure provides new threads that we can follow through the vast body of published information.

60

The second way to increase the role of classical Greek and Latin is to engage more people in reading and thinking about these languages. Anecdotal evidence suggests that this began to happen as soon as substantial bodies of Greek and Latin became available to the general public. Perseus quickly received letters from students in isolated locations such as rural homes and naval vessels at sea who were using online lexica and texts. Even more interesting, people who had studied Greek and Latin decades before found that the reading support tools available online gave them the support that they needed to begin reading Greek and Latin again.

61

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

The first paragraph in the opening "Call to action" of the National Science Foundation's 2007 "Cyberinfrastructure Vision for 21st Century Discovery" calls for "an individualized health model of every human being for personalized health care delivery" ("Cyberinfrastructure Vision for 21st Century Discovery", March 2008: page 5). Such models would open up new methods where doctors and patients could not only determine the best courses with which to treat disease but also to identify potential problems and predispositions in advance. Health records that include decades of medical tests and case histories clearly raise daunting issues of confidentiality, but the potential benefits are enormous. Emergent cyberinfrastructure for health care includes thus both methods to represent our particular background in great detail and a major investment in maintaining personal privacy.

Figure 8.



Customization of Latin vocabulary[43]

62

The same instruments developed for health care can be adapted for our intellectual backgrounds. We can begin to devise ways for us to keep track of what we have learned so that we can receive background information customized for our particular needs when we confront a new object of study.[44] Figure 8 illustrates a system that compares an arbitrary text of Latin against a model of the vocabulary that a particular reader has encountered, then calculates which words have been seen before and which are new. Seen words can then be associated with the places where they have been seen in the past, while unseen words can be ranked by their importance according to various criteria (e.g., numerical frequency, relevance to a particular theme etc.) The implementation is conceptually simple but represents the first stage at an open-ended process. As our data sources improve, we can look for more complex linguistic phenomena such as syntax and semantics (e.g., a new sense of a seen word). As our learning models grow more sophisticated, we can begin helping readers identify areas of weakness on which they can focus to enhance their ability to read with fluency.

63

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

Even small advances in our ability to work with multiple languages can be important if they open up historical languages to new audiences, whether these audiences are professional researchers using more linguistic sources or members of the public reading Greek poetry that they would not otherwise have experienced. The biggest benefits are likely to come when we open up linguistic materials to audiences with little or no training in the language. None of us has the opportunity to become familiar with more than a handful of languages. None of us can, in print culture, work with un-translated sources in dozens of languages.

64

Classics can, however, show how knowledge about an ancient culture can be designed to serve the speakers of multiple languages. The traditional method is for communities to choose a lingua franca — Akkadian, Greek, Latin, French, German, and now English have all served as common languages of diplomacy and scholarship. The speakers of an unbounded set of local languages communicate by learning one of these linguae francae — thus, the Chinese businessman in a Damascus hotel will probably carry on his business in English. Classicists are more broad-minded but generally expect scholars to publish materials in English, French, German and Italian. Speakers of Croatian or Modern Greek must learn these languages if they are to gain access to most information about the Greco-Roman world.

65

Classicists can, however, design their cyberinfrastructure from the start to be as portable as possible across multiple languages. There are at least three basic strategies, the third and most important of which is peculiarly suitable to historical fields where primary sources are finite and heavily studied.

66

First, we need to be able to optimize machine translation for the field of classics.[45] We can develop statistical models that capture the idiosyncrasies of documents about Greco-Roman culture. We develop these models by adding markup, using a combination of manual and automated methods, to finite bodies of material as training sets. Machine learning systems then scan these bodies and recognize that Alexandria usually refers to the city in Egypt and almost never to the suburb of Washington, DC, by that name. An ambiguous word such as "case" probably designates a grammatical case in a Greek grammar and a display case in a museum catalogue. These domain specific features, once identified, can help general machine translation systems avoid many of the worst problems they face and improve the quality of their output.

67

Second, we need to include as much basic information as we can in forms from which they can be converted into multiple languages. Thus, if we represent birth and death dates in a generic form, we can then develop modules to represent that knowledge in multiple languages. [46] Some ontologies such as the CIDOC-CRM for museum objects and FRBR for books have been under development for years and can represent a great deal of basic background information.[47]

68

Third, canonical literary texts attract very large amounts of labor. We can use that labor to create databases of linguistic annotations that describe syntax (e.g., the subject and object of a verb), co-

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

reference (e.g., which person is the subject of a particular verb), semantics (e.g., where does *oratio* correspond to "prayer" rather than "oration" or some other concept). These annotations stored in treebanks and other linguistic databases not only allow us to put our understanding of Greek and Latin on a wholly new, quantifiable foundation but can resolve the ambiguities that bedevil machine translation and can ultimately support higher quality machine translation.[48] Such annotations are expensive but are, in effect, the digital successors to print editions. Where print editors labored to resolve ambiguities and problems in the textual tradition, digital editors provide machine actionable annotations that resolve where possible ambiguities in the reconstructed texts.

69

The problem of multilingual knowledge thus breaks down into language independent and language dependent phases.

70

Knowledge bases (e.g., basic propositional statements) and linguistic annotation can be created by speakers of any language. The tag sets of ontologies and annotation schemes are relatively contained and can themselves be translated, allowing authors to work entirely with Greek, Latin and their own primary languages: the birthdate of a given author may be uncertain but that uncertainty can be represented in a general form by the speaker of any language. We may differ in how we construe the syntax of a sentence, but anyone who knows Greek, regardless of their native language, can decide which word depends on which and represent this in a common format.

71

Communities that want to make publications in their own languages accessible to wider audiences will have to develop the training sets for documents about classics. The results will not be perfect but readers can then use dictionary lookups and other translation aids to more closely study the original language. Each language needs its own training sets but this approach will not only make publications in the traditional languages of publication accessible to wider audiences but will also open up publications in less widely read languages (e.g. Croatian and Dutch) to much larger audiences.

72

Communities that want to be able to read basic knowledge about the Greco-Roman world in their own language will need machine translation that can be optimized for classics and language specific drivers that can convert the basic knowledge from ontologies into their language, and systems that can exploit the dense linguistic annotations available for major canonical source texts.

73

The creation of knowledge bases designed from the start to flow from language to language would be a radical change from traditional scholarly practice. Nevertheless, there are profound strategic reasons for this new form of scholarship in the two major classes of society that produce scholarship about the Greco-Roman world.

74

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

Classical Greek and Latin are the foundational languages of Europe and were the languages of high culture and trans-European discourse until relatively recent times — in fact, Turkey, whatever its religious background, would only restore to Europe a region that had been lost to it from the past. The European Union has a commitment to make the cultural heritage of its nations intellectually accessible to the widest possible audience. This implies an infrastructure that maximizes what can be learned not only in English, French, German, and Italian, but in all of the other official languages of Europe.[49]

75

The United States, Canada, Australia, New Zealand, and South Africa are, however, not only geographically distinct from Europe but are fashioning themselves into cosmopolitan societies, European in origin but creating new identities with roots from every civilization of humanity. The United States has in particular identified Chinese and Arabic as the two strategic languages on which it will concentrate its resources. While Europe concentrates on making its cultural heritage accessible to the speakers of its official languages, American scholars can take the lead in making classical antiquity increasingly accessible to speakers of Chinese, Arabic and other languages. Ultimately, the increased distribution of Greco-Roman cultural materials into many other languages will speed the complementary process of opening up materials in classical Chinese, Arabic, Sanskrit and other languages to speakers of English and other European languages. Our larger goal must be to make the record of humanity accessible to everyone regardless of linguistic and cultural background.

76

While a linguistically and culturally portable knowledge base about the Greco-Roman world may seem daunting, the tools already at hand allow us to rethink not only who can read and consume primary and secondary sources but who can contribute substantively to the field. Blackwell and Martin's essay opens this collection by describing how the practices of undergraduates have begun to change. The rise of undergraduate research is arguably the most important and promising development for classics as a discipline since classics lost its privileged position. Before we can appreciate the possibilities of the technology now available but not yet fully exploited, we need to see how much classicists have already begun to accomplish.

77

Before turning to the prospects for undergraduate and more general non-specialist research in classics, we should emphasize that the collection of essays published here themselves illustrate the greatest achievement of classical philology in this digital world. We now have a critical mass of classicists who are committed to building and exploiting the evolving digital infrastructure upon which all scholarship and teaching in our field will depend. While discussions of digital humanities still revert to the problem of tenure and promotion, several of the contributors to this collection have already earned tenure by pursuing digital projects. All of the authors here are able to review innovative forms of digital scholarship on its intellectual merits, neither penalizing or rewarding the use of digital technologies per se but assessing the degree to which the new work advances our ancient and unchanging goals to bring the Greco-Roman heritage in general and ancient Greek and Latin in particular ever more fully to life in the minds of the broadest audience possible.

78

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

No one showed more vision and patience to create this community than our colleague and beloved friend, Allen Ross Scaife. He showed the way with his own pioneering work on *Diotima*, a digital representations of women in antiquity. As director of the *Stoa* from its founding until his death ten years later, Ross always understood that the greatest resource for any field was the people whom it attracted. Ross supported, fostered, encouraged, and advanced careers that will continue now for decades and will shape other careers as well. "Do not lament," the Pericles of Thucydides (1.143.5) tells the Athenians, "houses and land but people, for it is not houses and land that acquire people but people who acquire them." The passing of Ross Scaife wounds the field of classics more deeply than would have the loss of everything that the field as a whole has produced. But the community that Ross fostered with intelligence, patience and love and that produced these essays is greater than any single achievement that their authors could ever produce.

**Notes**

[1]The publications in this collection were the production of a workshop, sponsored by the National Science Foundation, on the subject of Cyberinfrastructure in the Classics in September 2007 (NSF GRANT INFO). We gratefully acknowledge the support that we received from the University of Kentucky and the Department of Classics.

[2]For some evaluations of the Wikipedia phenomenon and the challenges it has offered to more conventional forms of intellectual production, see for example, [Lally 2007] and [Rosenzweig 2006].

[3]Statistics retrieved from http://www.wikipedia.org, accessed August 2, 2008.

[4] http://wealthofnetworks.wordpress.com/, a blog by Margaret Gold that contains summaries of John Dartington's remarks at a conference entitled, "The Wealth of Networks: Digital Economies and the Next Generation Internet," held in the UK in July 2008.

[5]For a discussion of ePhilology and its role in the larger cyberinfrastructure environment, please see [Crane 2007]; also, [Dimitriadis 2006].

[6]For example, Jeffrey Garrett discusses whether the use of Google and full text searching is being substituted by some for deeper reading and analysis, see [Garrett 2006]; a recent report by the British Library and JISC has explored how Google and the internet has influenced the younger generation of searchers, see *Information Behavior and the Researcher of the Future*. January 11, 2008. Joint Report funded by the BL/JISC. http://www.jisc.ac.uk/media/documents/programmes/reppres/gg_final_keynote_11012008.pdf.

[7]For more on this theme, see [Crane 2005] and also [Kelly 2006].

[8]Text mining within the humanities and within classics has received a fair amount of attention in recent years, for example, see [Plaisant 2006]; [Don 2007]; and [Hyman 2008].

[9]Matthew Kirschenbaum has offered a useful overview of how text and data mining are reshaping reading in the digital environment, see [Kirschenbaum 2007].

[10]Cornell University has published electronic versions of this series on-line as a part of the Making of America Digital Library: http://cdl.library.cornell.edu/moa/browse.monographs/waro.html (last accessed August 12, 2008).

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

[11]Several recent reports have called for expanding our ideas of infrastructure in order to create a larger cyberinfrastructure, see [Arms 2007] and [ACLS 2006].

[12]For more discussion on this topic, please see [Crane 2006a].

[13]For an overview of how many classicists use digital materials as evidenced by citations, see [Dalbello 2006].

[14] [Pritchard 2008]. The papers for a 1995 workshop by the American Physical Society, online at http://publish.aps.org/EPRINT, include talks from the previous year about preprints. The ArXiv.org server founded in 1991, contained (as of June 29, 2008) 484,758 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology and Statistics.

[15]Classics is not the only field that has been challenged to modernize its publication system, the issue of scholarly communication and the need for major change has been the subject of much discussion, it recently served as the topic for the Winter issue of the Journal of Electronic Publishing.

[16]While manually created indices such as back-of-the-book indexes are still considered essential by many, the automatic creation and remodeling of such indices is a growing research area, see [Csomai 2006] and [Chi 2007].

[17]For further information on the Google Books system, see [Kolak 2008] and [Schilit 2008].

[18]The adaptation of commercial OCR systems for Greek and Latin as well as the development of other text recognition systems have been explored by several research groups; see for example [Gatos 2006] and [Moalla 2006].

[19]Specialized document layout analysis systems for historical documents has been an active research field for years, for a recent overview see [Sankar 2006] and for some recent work in this area involving texts digitized by the Open Content Alliance, see [Lu 2008].

[20]We have reported on our own work in historical named entity recognition in [Crane 2006c] and [Smith 2001]; for several examples of the growing research in this area, see [Geleijnse 2007] and [Borin 2007] and [Tobin 2008].

[21]This informal survey examined the articles in sample issues that Johns Hopkins made publicly available for marketing purposes. Where there was not a public issue, the most recent online issue was examined. Seven single author articles in http://muse.jhu.edu/demo/american_journal_of_philology/: vol. 126 (1) 2005; five single-author articles in http://muse.jhu.edu/demo/arethusa/: vol. 38 (1) 2005; four single-author articles in http://muse.jhu.edu/demo/classical_world/: vol. 99(1) 2005; http://muse.jhu.edu/demo/helios/: vol. 34 (1) 2007; nine single-author articles in http://muse.jhu.edu/journals/journal_of_late_antiquity/toc/current.html: vol. 1 (1) 2008); two single-author articles in http://muse.jhu.edu/journals/mouseion_journal_of_the_classical_association_of_canada/toc/mou.7.1.html: vol. 7(1) 2007; ten single-author papers in http://muse.jhu.edu/demo/transactions_of_the_american_philological_association/: 135(1) 2005; three single-author papers in http://muse.jhu.edu/demo/hesperia/: 71(1) 2005. By contrast, there was only a single multi-authored paper in this group: [Kraft 2005].

[22]For further discussion of this issue, see [Crane 2006b].

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

[23]Research into how reference works can be made machine actionable has been investigated by [Veltman 1999] and [Buckland 2007]. Other interesting work has examined how less traditional reference sources such as Wikipedia can be turned into knowledge bases, see [Ponzetto 2007].

[24]Early versions of these guidelines were circulating at least as early as 1990. For an example of current technology available to manage properly structured textual data, see [van den Branden 2007].

[25]A variety of approaches to designing digital editions have been developed over the years, many based on the TEI, for several (but by no means exhaustive) examples, see [Audenaert 2008]; [Dekhytar 2006]; [Riva 2005].

[26]As often in the history of scholarship, New Testament scholars have, by contrast, pioneered the use of information technology, see P. Robinson's work for example [Robinson 2000]; [Robinson 2005].

[27]See the discussion of storage costs in 1982 below. The TLG was founded ten years earlier, in 1972, when disk storage itself had just begun to emerge.

[28]Google has sponsored development of OCRopus, an open-source document analysis and OCR system in order to promote development of more sophisticated OCR technologies, http://code.google.com/p/ocropus/.

[29]For a list of publications describing this work, please see http://www.perseus.tufts.edu/hopper/about/publications/.

[30]The importance of semantic markup for digital library texts has been discussed for many years, particularly the issue of potential semantic interoperability of such markup or metadata, for two examples see [van 2006] and [Elings 2007].

[31]We have previously described this role as that of corpus editors, see [Crane 2000].

[32]The phenomenon of digital communities and the new ways in which individuals can contribute to them has been extensively explored, for some recent work, see [Cosley 2006]; [Krowne 2003].

[33]For some interesting efforts to create digital reading/writing environments that allow for the creation and sharing of annotations and also support other types of more sophisticated scholarly communication, see [Bradley 2008]; [Fitzpatrick 2007]; [Schroeter 2007].

[34]The need for support for grid level computing for digital humanities projects has been discussed by [Gietz 2006]; [Blanke 2006].

[35]See http://www.thesaurus.badw.de/english/index.htm/, accessed August 3, 2008.

[36] http://www.tlg.uci.edu/, accessed August 3, 2008, lists August 12, 2007 as its last modification date.

[37]For some examples of this process, please see [Ganchev 2007], [Vlachos 2006], and [Culotta 2005].

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

[38]Work, still unpublished, conducted by D. Sculley, a PhD student in Computer Science at Tufts University. This preliminary work led to the subsequent funded research described by Bamman and Crane in this collection.

[39]This need for long term data curation of the scholarly record has recently been discussed by [Gold 2007] and [Luce 2008].

[40]For further discussion on the need for distributed digital libraries that can be searched seamlessly and the issues involved, please see [Simeoni 2007], [Trnkoczy 2006], and [Lagoze 2006].

[41]For more on digital preservation and the need for institutional repositories, see [Marshall 2008]; [Cantara 2006] and [Hockx-Yu 2006].

[42]For more on the potential of CTS, see [Romanello 2008] and [Porter 2006].

[43]Reprinted from [Crane 2007].

[44]Adaptive systems that customize themselves automatically to what a user has already learned have been in development for a number of years now, for some recent work see [Heilman 2008].

[45]Some cultural heritage projects have conducted research into how machine translation can be customized for more resource-poor languages, see [Jones 2007].

[46]Various research has explored the potential of either translating semantic markup into multiple languages or mapping between languages, see [Monroy 2007] and [Bia 2006].

[47]For a specific look at how CIDOC-CRM is being used with multilingual texts, see [Genereux 2006].

[48]The Perseus Project has recently begun work on a Greek treebank, and work on a Latin treebank has been ongoing for over three years, for more on the Latin treebank, see [Bamman 2007] and [Bamman 2006].

[49]The challenges of supporting multi-lingual access to Europe's cultural heritage through the European Digital Library have been discussed in [Agosti 2007].

**Works Cited**

ACLS 2006  *Our Cultural Commonwealth: A Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*, 2006. http://www.acls.org/uploadedFiles/Publications/Programs/Our_Cultural_Commonwealth.pdf.

Agosti 2007 Agosti, M. et al. "Roadmap for Multilingual Information Access in the European Library." In *Proceedings of the ECDL 2007*: 136-147.

Arms 2007 Arms, W. and R. Larsen. *The Future of Scholarly Communication: Building the Infrastructure for Cyberscholarship*. Report on a NSF-JISC Workshop, April 17-19 2007. http://www.sis.pitt.edu/~repwkshop/SIS-NSFReport2.pdf.

Audenaert 2008 Audenaert, N. and R. Furuta. "Annotated Facsimile Editions: Defining Macro- Level Structure for Image-Based Electronic Editions." *Digital Humanities 2008 Abstracts.* http://www.ekl.oulu.fi/dh2008/Digital%20Humanities%202008%20Book%20of%20Abstracts.pdf.

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

Bamman 2006 Bamman, D. and G. Crane. "The Design and Use of a Latin Dependency Treebank." *TLT 2006: Proceedings of the Fifth International Treebanks and Linguistic Theories Conference*: 67-78. http://dl.tufts.edu/view_pdf.jsp?pid=tufts:PB.001.002.00005.

Bamman 2007 Bamman, D. and G. Crane. "The Latin Dependency Treebank in a Cultural Heritage Digital Library." In *Proceedings of the Workshop on Language Technology for Cultural Heritage Data (LaTech 2007)*: 33-40. http://dl.tufts.edu/view_pdf.jsp?pid=tufts:PB.001.002.00002.

Bia 2006 Bia, A. et al. "A Multilingual Markup Translation Web-Service. An Entry Level Solution to Internationalize XML Markup Vocabularies." *WEBIST 2006*. http://cio.umh.es/ES/publicaciones/ficheros/CIO_2006_06.pdf.

BL 2008 BL/JISC. *Information Behavior and the Researcher of the Future*. January 11, 2008. Joint Report funded by the BL/JISC. http://www.jisc.ac.uk/media/documents/programmes/reppres/gg_final_keynote_11012008.pdf.

Blanke 2006 Blanke, T. et al. "Digital Libraries in the Arts and Humanities Current Practices and Future Possibilities." *INSCIT 2006*, http://www.slideshare.net/inscit2006/digital-libraries-in-the-arts-and-humanities-current-practices-and-future-possibilities.

Borin 2007 Borin, L. et al. "Naming the Past: Named Entity and Animacy Recognition in 19th Century Swedish Literature." In *Proceedings of the Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2007)*: 1-8. http://www.aclweb.org/anthology-new/W/W07/W07-0901.pdf.

Boschetti 2007 Boschetti, F. "Methods to Extend Greek and Latin Corpora with Variants and Conjectures: Mapping Critical Apparatuses onto Reference Text." In *CL 2007: Proceedings of the Corpus Linguistics Conference*, http://www.corpus.bham.ac.uk/corplingproceedings07/paper/150_Paper.pdf.

Bradley 2008 Bradley, J. "Pliny: A Model for Digital Support of Scholarship." *Journal of Digital Information*, 9:26 (2008), http://journals.tdl.org/jodi/article/view/209/198.

Buckland 2007 Buckland, M. "The Digital Difference in Reference Collections." *Journal of Library Administration*, 46:2 (2007): 87-100.

Busa 1974 Busa, R. *Index Thomisticus*. Stuttgart: Frommann-Holzboog, 1974.

Busa 1980 Busa, R. "The Annals of Humanities Computing: The Index Thomisticus." *Computers and the Humanities*, 14:2 (1980): 8390.

Cantara 2006 Cantara, L. "Long term Preservation of Digital Humanities Scholarship." *OCLC Systems & Services*, 22:1 (2006): 38-42.

Chi 2007 Chi, E. H. et al. "ScentIndex and ScentHighlights: Productive Reading Techniques for Conceptually Reorganizing Subject Indexes and Highlighting Passages." *Information Visualization,* 6:1 (2007): 32-47.

Cosley 2006 Cosley, D. et al. "Using Intelligent Task Routing and Contribution Review to Help Communities Build Artifacts of Lasting Value." *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*: 1037-1046.

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

Crane 2000 Crane, G. and J. A. Rydberg-Cox. "New Technology and New Roles: The Need for Corpus Editors." *Proceedings of the 5th ACM Conference on Digital Libraries* 2000: 252-253. http://perseus.mpiwg-berlin.mpg.de/Articles/corpused.pdf.

Crane 2005 Crane, G. "Reading in the Age of Google: Contemplating the Future With Books That Talk to One Another." *Humanities*, 26:5 (2005), http://www.neh.gov/news/humanities/2005-09/readingintheage.html.

Crane 2006a Crane, G. et al. "Beyond Digital Incunabula: Modeling the Next Generation of Digital Libraries." In *Proceedings of the ECDL 2006*: 353-366, http://dl.tufts.edu//view_pdf.jsp?urn=tufts:facpubs:gcrane-2006.00002.

Crane 2006b Crane, G., and A. Jones. "Text, Information, Knowledge and the Evolving Record of Humanity." *D-Lib Magazine*, 12:3 (2006). http://www.dlib.org/dlib/march06/jones/03jones.html.

Crane 2006c Crane, G. and A. Jones. "The Challenge of Virginia Banks: an Evaluation of Named Entity Analysis in a 19th-Century Newspaper Collection." In *JCDL '06: Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*: 31-40. http://dl.tufts.edu/view_pdf.jsp?pid=tufts:PB.001.001.00007.

Crane 2007 Crane, G., et. al. "ePhilology: When the Books Talk to Their Readers." In *A Companion to Digital Literary Studies* (New York, London: Blackwell Publishing, 2007): 29-64. http://dl.tufts.edu//view_pdf.jsp?urn=tufts:facpubs:gcrane-2006.00003.

Csomai 2006 Csomai, A. and R. Mihalcea. "Creating a Testbed for the Evaluation of Automatically Generated Back-of-the-Book Indexes." In *Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*, 2006. http://www.cse.unt.edu/~rada/papers/csomai.cicling06.pdf.

Culotta 2005 Culotta, A. and A. McCallum. "Reducing Labeling Effort for Structured Prediction Tasks." In *Proceedings of AAAI 2005.* http://www.cs.umass.edu/~mccallum/papers/multichoice-aaai05.pdf.

Dalbello 2006 Dalbello, M. et al. "Electronic Texts and the Citation System of Scholarly Journals in the Humanities: Case Studies of Citation Practices in the Fields of Classical Studies and English Literature." In *LIDA 2006: Proceedings of Libraries in the Digital Age*, http://dlist.sir.arizona.edu/1638/.

Dekhytar 2006 Dekhytar, A. et al. "Support for XML Markup of Image-Based Electronic Editions." *International Journal of Digital Libraries*, 6:1 (2006): 55-69.

Dimitriadis 2006 Dimitriadis, A. et al. "Toward A Linguists WorkBench Supporting eScience Methods." In *E-SCIENCE '06: Proceedings of the Second IEEE International Conference on e-Science and Grid Computing*: 131-9. http://www.lat-mpi.eu/papers/papers-2006/escience-sketch-final2.pdf/view.

Don 2007 Don, A. et al. "Discovering Interesting Usage patterns in Text Collections: Integrating Text Mining with Visualization." In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on Information and Knowledge Management*: 213-222. http://hcil.cs.umd.edu/trs/2007-08/2007-08.pdf.

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

Elings 2007 Elings, M.W. and G. Waibel. "Metadata for All: Descriptive Standards and Metadata Sharing across Libraries, Archives, and Museums." *First Monday*, 12:3 (2007), http://firstmonday.org/issues/issue12_3/elings/index.html.

Fitzpatrick 2007 Fitzpatrick, K. "CommentPress: New (Social) Structures for New (Networked) Texts." *Journal of Electronic Publishing*, 10:3 (2007). http://hdl.handle.net/2027/spo.3336451.0010.305.

Ganchev 2007 Ganchev, K. et al. "Semi-Automated Named Entity Annotation." *Proceedings of the Linguistic Annotation Workshop. ACL, Prague, Czech Republic*, 2007: 53-56. http://www.aclweb.org/anthology-new/W/W07/W07-1509.pdf.

Garrett 2006 Garrett, J. "KWIC and Dirty? Human Cognition and the Claims of Full-Text Searching." *Journal of Electronic Publishing*, 9:1 (2006), http://hdl.handle.net/2027/spo.3336451.0009.106.

Gatos 2006 Gatos, B. et al. "An Efficient Segmentation-Free Approach to Assist Old Greek Handwritten Manuscript OCR." *Pattern Analysis & Applications*, 8:4 (2006): 305-320.

Geleijnse 2007 Geleijnse, G. and J. Korst. "Creating a Dead Poets Society: Extracting a Social Network of Historical Persons from the Web." In *Proceedings of the Sixth International Semantic Web Conference and the Second Asian Semantic Web Conference (ISWC + ASWC 2007)*: 156-168. http://iswc2007.semanticweb.org/papers/155.pdf.

Genereux 2006 Genereux, M. and D. Arnold. "Preserving Meanings in Multilingual Text Mining for Cultural Heritage." In *ICS-Forth Workshop: Exploring the Limits of Global Models for Integration and Use of Historical and Scientific Information*,2006 http://cidoc.ics.forth.gr/workshops/heraklion_october_2006/genereux_arnold.pdf.

Gietz 2006 Gietz, P. et. al. "TextGrid and eHumanities." In *E-SCIENCE '06: Proceedings of the Second IEEE International Conference on e-Science and Grid Computing*. http://www.textgrid.de/fileadmin/TextGrid/veroeffentlichungen/TextGrid-Amsterdam-2006-final.pdf.

Gold 2007 Gold, A. "Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries." *D-Lib Magazine,* 9 (2007). http://www.dlib.org/dlib/september07/gold/09gold-pt2.html.

Heilman 2008 Heilman, M. et al. "Retrieval of Reading Materials for Vocabulary and Reading Practice." *Proceedings of the Third ACL Workshop on Innovative Use of NLP for Building Educational Applications,* 2008: 80-88. http://aclweb.org/anthology-new/W/W08/W08-0910.pdf.

Hockx-Yu 2006 Hockx-Yu, H. "Digital Preservation in the Context of Institutional Repositories." *Program: Electronic Library & Information Systems,* 40:3 (2006): 232-243.

Hyman 2008 Hyman, M. D. "Term Discovery in an Early Modern Latin Scientific Corpus." *Digital Humanities 2008 Abstracts*: 136-137. http://www.ekl.oulu.fi/dh2008/Digital%20Humanities%202008%20Book%20of%20Abstracts.pdf.

Jones 2007 Jones, G. J. F. et al. "Multilingual Search for Cultural Heritage Archives via Combining Multiple Translation Resources." In *Proceedings of the Workshop on Language Technology for*

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

*Cultural Heritage Data (LaTeCH 2007)*: 81-88. http://www.aclweb.org/anthology-new/W/W07/W07-0911.pdf.

Kelly 2006 Kelly, K. "Scan This Book!" *New York Times Magazine*, May 14, 2006: 42+. http://www.nytimes.com/2006/05/14/magazine/14publishing.html?ex=1305259200&en=c07443d368771bb8&ei=5090.

Kirschenbaum 2007 Kirschenbaum, M. "The Remaking of Reading: Data Mining and the Digital Humanities." In *NGDM 07: National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation*. http://www.cs.umbc.edu/~hillol/NGDM07/abstracts/talks/MKirschenbaum.pdf.

Kolak 2008 Kolak, O. and B. N. Schilit. "Generating Links by Mining Quotations." In *HT '08: Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*: 117-126.

Kraft 2005 Kraft, J. C., Rapp, G., Gifford, J. and Aschenbrenner, S., "Coastal Change and Archaeological Settings in Elis", in *Hesperia* 74 (2005): 1-39.

Krowne 2003 Krowne, A. "Building a Digital Library the Commons-Based Peer Production Way." *D-Lib Magazine*, 9:10 (2003). http://www.dlib.org/dlib/october03/krowne/10krowne.html.

Lagoze 2006 Lagoze, C. et al. "Metadata Aggregation and Automated Digital Libraries: a Retrospective on the NSDL Experience.". In *JCDL '06: Proceedings of the 6th ACM/IEEE-CS joint conference on Digital Libraries*: 230-239.

Lally 2007 Lally, A. M. and C. E. Dunford. "Using Wikipedia to Extend Digital Collections." *D-Lib Magazine*, 13: 5/6 (2007). http://www.dlib.org/dlib/may07/lally/05lally.html.

Lu 2008 Lu, X. et al. "A Metadata Generation System for Scanned Scientific Volumes." In *JCDL '08: Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*: 167-176.

Luce 2008 Luce, R. E. "A New Value Equation Challenge: The Emergence of E-Research and Roles for Research Libraries." In *No Brief Candle: Reconceiving Research Libraries for the 21st Century*. CLIR 2008: 42-50, http://www.clir.org/pubs/reports/pub142/pub142.pdf.

Marshall 2008 Marshall, C. C. "From Writing and Analysis to the Repository: Taking the Scholars' Perspective on Scholarly Archiving." In *JCDL '08: Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*: 251-260.

Moalla 2006 Moalla, I. et al. "Image Analysis for Palaeography Inscription." in *DIAL 2006: Document Image Analysis for Libraries*: 303-311.

Monroy 2007 Monroy, C. et al. "A Multilingual Approach to Technical Manuscripts: 16th and 17th-century Portuguese Shipbuilding Treatises." In *JCDL '07: Proceedings of the 2007 conference on Digital libraries*: 413-414.

Plaisant 2006 Plaisant, C. et al. "Exploring Erotics in Emily Dickinson's Correspondence with Text mining and Visual Interfaces." In *JCDL '06: Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*: 141-150.

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

Ponzetto 2007 Ponzetto, S. P. "Creating a Knowledge Base From a Collaboratively Generated Encyclopedia." In *Proceedings of the NAACL-HLT 2007 Doctoral Consortium*: 9-12. http://acl.ldc.upenn.edu/N/N07/N07-3003.pdf.

Porter 2006 Porter, D. et al. "Creating CTS Collections." *Digital Humanities 2006*: 269-274. http://www.csdl.tamu.edu/~furuta/courses/06c_689dh/dh06readings/DH06-269-274.pdf.

Pritchard 2008 Pritchard, D. "Working Papers, Open Access, and Cyber-infrastructure in Classical Studies." *Literary and Linguistic Computing,* 23:2 (2008): 149-162. http://ses.library.usyd.edu.au/handle/2123/2226.

Riva 2005 Riva, M. and V. Zafrin. "Extending the Text: Digital Editions and the Hypertextual Paradigm." In *HYPERTEXT '05: Proceedings of the sixteenth ACM conference on Hypertext and hypermedia*: 205-207.

Robinson 2000 Robinson, P. "The One Text and the Many Texts." *Literary and Linguistic Computing*, 15:1 (2000): 5-14.

Robinson 2005 Robinson, P. "Current Issues in Making Digital Editions of Medieval Texts or, do Electronic Scholarly Editions have a Future?" *Digital Medievalist*, 1:1 (2005). http://www.digitalmedievalist.org/journal/1.1/robinson/.

Romanello 2008 Romanello, M. "A Semantic Linking Framework to Provide Critical Value- Added Services for E-Journals on Classics." In *ELPUB2008. Open Scholarship: Authority, Community, and Sustainability in the Age of Web 2.0 - Proceedings of the 12th International Conference on Electronic Publishing*. http://elpub.scix.net/cgi-bin/works/Show?401_elpub2008.

Rosenzweig 2006 Rosenzweig, R. "Can History be Open Source: Wikipedia and the Future of the Past?" *Journal of American History*, 93:1 (2006): 117-146. http://chnm.gmu.edu/resources/essays/d/42.

Sankar 2006 Sankar, K. et al. "Digitizing a Million Books: Challenges for Document Analysis." in *Document Analysis Systems VII* (2006): 425-436, http://cvit.iiit.ac.in/papers/pramod06Digitizing.pdf.

Schilit 2008 Schilit, B. N. and O. Kolak. "Exploring a Digital Library through Key Ideas." In *JCDL '08: Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*: 177-186.

Schroeter 2007 Schroeter, R. et al. "Annotating Relationships Between Multiple Mixed-Media Digital Objects by Extending Annotea." In *Proceedings of ESWC 2007*: 533-548. http://espace.library.uq.edu.au/view/UQ:151380.

Shirky 2008 Shirky, C. "Here Comes Everybody." Retrieved 08/02, 2008, from http://blip.tv/file/855937/.

Simeoni 2007 Simeoni, F. et al. "A Grid-Based Infrastructure for Distributed Retrieval." *Proceedings of the ECDL 2007*: 161-173.

Smith 2001 Smith, D.A. and G. Crane. "Disambiguating Geographic Names in a Historical Digital Library." In *ECDL '01: Proceedings of the 5th European Conference on Research and Advanced Technology for Digital Libraries*: 127-136, http://perseus.mpiwg-berlin.mpg.de/Articles/geodl01.pdf.

Crane, G., Seales, B., Terras, M. (2009). "Cyberinfrastructure for Classical Philology". In Crane, G. and Terras, M. (eds) (2009). "Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure". Digital Humanities Quarterly, Volume 3 Number 1, Winter 2009.

Sperberg 1994 Sperberg-McQueen, C. M. and L. Burnard, Eds. *Guidelines for Electronic Text Encoding and Interchange*. Chicago and Oxford: Text Encoding Initiative, 1994.

Tobin 2008 Tobin, R. et al. "Named Entity Recognition for Digitised Historical Texts" in *Proceedings of the Sixth International Language Resources and Evaluation Conference (LREC'08)*. http://www.ltg.ed.ac.uk/np/publications/ltg/papers/bopcris-lrec.pdf.

Trnkoczy 2006 Trnkoczy, J. et al. "A Grid-Based Architecture for Personalized Federation of Digital Libraries." *Library Collections, Acquisitions, and Technical Services*, 30:3-4 (2006): 139-53.

United 1880 United States. War Dept., United States. War Dept. War Records Office., et al. *The War of the Rebellion: a compilation of the official records of the Union and Confederate armies*. Washington, Govt. Print. Off., 1880.

van 2006 van Gendt, M. et al. "Semantic Web Techniques for Multiple Views on Heterogeneous Collections: A Case Study." In *Proceedings of ECDL 2006*: 426-437.

van den Branden 2007 van den Branden, R. and E. Vanhoutte. 2007. "Through the Reading Glass: Generating an Editorial Microcosm Through Experimental Modelling." *Digital Humanities 2007*. http://www.digitalhumanities.org/dh2007/abstracts/xhtml.xq?id=182.

Veltman 1999 Veltman, K. "Digital Reference Rooms: Access to Historical and Cultural Dimensions of Knowledge." *INET 99*. http://www.isoc.org/inet99/proceedings/2b/2b_1.htm.

Vlachos 2006 Vlachos, A. "Active Annotation." In *Proceedings of the EACL 2006 Workshop on Adaptive Text Extraction*. http://acl.ldc.upenn.edu/W/W06/W06-2209.pdf.