

## Preface

More than fifty years have passed since 1968, when Harvard University Press published the Concordance to Livy (*A Concordance to Livy* [Harvard 1968]), the first product of what we might now call Digital Classics. In the basement of the Harvard Science Center, David Packard had supervised the laborious transcription of the whole of Livy's *History of Rome* onto punch cards and written a computer program to generate a concordance with 500,000 entries, each with 20 words of context. Fourteen years later, when in 1982 I began work on the Harvard Classics Computing Project, technology had advanced. The available of Greek texts from the Thesaurus Linguae Graecae on magnetic tape was the impetus for my work – the department wanted to be able to search the authors in this early version of the TLG on a Unix system. There was also a need to computerize typesetting in order to contain the costs of print publication. Digital work at that time was very technical and aimed at enhancing traditional forms of concordance research and print publication.

When I first visited Xerox's Palo Alto Research Center in 1985, I also saw for first time a digital image – indeed, one that was projected onto a larger screen. As I came to understand what functions digital media would support, I began to realize that digital media would do far more than enhance traditional tasks. As a graduate student, I had shuttled back and forth between Widener, the main Harvard library, and the Fogg Art Museum library, a five or ten minute walk away. That much distance imposed a great deal of friction on scholarship that sought to integrate publications about both the material and the textual record. It was clear that we would be able to have publications that combined every medium and that could be delivered digitally. My own work on Perseus began that year with a Xerox grant of Lisp Machines (already passing into obsolescence and surely granted as a tax write-off).

A generation later, the papers in this publication show how far Digital Classics has come. When I began my own work on Perseus in the 1980s, much of Greek and Latin literature had been converted into machine readable texts – but the texts were available only under restrictive licenses. The opening section of the collection, *Open Data of Greek and Latin Sources*, describes the foundational work on creating openly licensed corpora of Greek and Latin that can support scholarship without restriction. Scholars must have data that they can freely analyze, modify and redistribute. Without such freedom, digital scholarship cannot even approach its potential. Muellner and Huskey talk about collaborative efforts to expand the amount of Greek source text available and to begin developing born-digital editions of Latin sources. Cayless then addresses the challenge of applying the methods of Linked Open Data to topics such as Greco-Roman culture.

*Cataloging and Citing Greek and Latin Authors and Works* illustrates not only how Classicists have built upon larger standards and data models such as the Functional Requirements for Bibliographic Records (FRBR, allowing us to represent different versions of a text) and the Text Encoding Initiative (TEI) Guidelines for XML encoding of source texts (representing the logical structure of sources) but also highlights some major contributions from Classics. Alison Babeu, Digital Librarian at Perseus, describes a new form of catalog for Greek and Latin works that exploits the FRBR data model to represent the many versions of our sources – including translations. Christopher Blackwell and Neel Smith built on FRBR to develop the Canonical Text Services (CTS) data model as part of the CITE Architecture. CTS provides an explicit framework within which we can address any substring in any version of a text, allowing us to create annotations that can be maintained for years and even for generations. This addresses – at least within the limited space of textual data – a problem that has plagued hypertext systems since the 1970s and that still afflicts the World Wide Web. Those who read these papers years from now will surely find that many of the URLs in the citations no longer function but all of the CTS citations should be usable – whether we remain with this data model or replace it with something more expressive. Computer Scientists Jochen Tiepmar and Gerhard Heyer show how they were able to develop a CTS server that could scale to more than a billion words, thus establishing the practical nature of the CTS protocol.

If there were a Nobel Prize for Classics, my nominations would go to Blackwell and Smith for CITE/CTS and to Bruce Robertson, whose paper on Optical Character Recognition opens the section on *Data Entry, Collection, and Analysis for Classical Philology*. Robertson has worked a decade, with funding and without, on the absolutely essential problem of converting images of print Greek into machine readable text. In this effort, he has mastered a wide range of techniques drawn from areas such as computer human interaction, statistical analysis, and machine learning. We can now acquire billions of words of Ancient Greek from printed sources and not just from multiple editions of individual works (allowing us not only to trace the development of our texts over time but also to identify quotations of Greek texts in articles and books, thus allowing us to see which passages are studied by different scholarly communities at different times). He has enabled fundamental new work on Greek. Meanwhile the papers by Tauber, Burns, and Coffee are on representing characters, on a pipeline for textual analysis of Classical languages and on a system that detects where one text alludes to – without extensively quoting – another text.

At its base, philology depends upon the editions which provide information about our source texts, including variant readings, a proposed reconstruction of the original, and reasoning behind decisions made in analyzing the text. The

section on *Critical Editing and Annotating Greek and Latin Sources* describes multiple aspects of this problem. Fischer addresses the challenge of representing the apparatus – the list of variants traditionally printed at the bottom of the page. Schubert and her collaborators show new ways of working with multiple versions of a text to produce an edition. Dué and Hackney present the Homeric Epics as a case where the reconstruction of a single original is not appropriate: the Homeric Epics appeared in multiple forms, each of which needs to be considered in its own right and thus a Multitext is needed. Berti concludes by showing progress made on the daunting task of representing a meta-edition: the case where works exist only as quotations in surviving works and an edition consists of an annotated hypertext pointing to – and modifying – multiple (sometimes hundreds) of editions.

We end with a glimpse into born-digital work. *Linguistic annotation and lexical databases* extends practices familiar from print culture so far that they become fundamentally new activities, with emergent properties that could not – and still cannot fully – be predicted from the print antecedents. Celano describes multiple dependency treebanks for Greek and Latin – databases that encode the morphological and syntactic function of every word in a text and that will allow us to rebuild our basic understanding of Greek, Latin, and other languages. Passarotti's paper on the Index Thomisticus Treebank also brings us into contact with Father Busa and the very beginning of Digital Humanities in the 1940s. With Boschetti we read about the application of WordNet and of semantic analysis to help us, after thousands of years of study, see systems of thought from new angles.

I began my work on (what is now called) Digital Classics in 1982 because I was then actively working with scholarship published more than a century before and because I knew that my field had a history that extended thousands of years in the past. Much has changed in the decades since, but the pace of change is only accelerating. The difference between Classics in 2019 and 2056 will surely be much greater than that between 1982 and 2019. Some of the long term transformative processes are visible in this collection.

One fundamental trend that cuts across the whole collection is the emergence of a new generation of philologists. When I began work, few of us had any technical capabilities and fewer still had any interest in developing them. What we see in this collection of essays is a collection of classical philologists who have developed their own skills and who are able to apply – and extend – advances in the wider world to the study of Greek and Latin. This addresses the existential question of sustainability of Greek and Latin in at least two ways.

First, I was very fortunate to have five years of research support – 1.000.000 EUR/year – from the Alexander von Humboldt Foundation as a Humboldt

Professor of Digital Humanities at Leipzig. I also have been able to benefit from support over many years for the Perseus Project from Tufts University. Both of those sources contributed to a number of these papers, both directly (by paying salaries) and indirectly (e.g., by paying for people to come work together). But what impresses me is how rich the network of Digital Classicists has become. We were able to help but the system is already robust and will sustain itself. We already have in the study of Greek and Latin a core community that will carry Digital Classics forward with or without funding, for love of the subject. In this, they bring life to the most basic and precious ideals of humanistic work.

Second, we can see a new philological education where our students can learn Greek and Latin even as they become computer, information or data scientists (or whatever label for computational sciences is fashionable). Our students will prepare themselves to take their place in the twenty-first century by advancing our understanding of antiquity. Our job as humanists is to make sure that we focus not only on the technologies but on the values that animate our study of the past.

Gregory R. Crane  
(Perseus Project at Tufts University and Universität Leipzig)