

Clay Balls and Compact Disks: Some Political and Economic Problems of New Storage Media

Gregory Crane
Harvard University

I would like to begin with a Theognidean $\alpha\nu\theta\omicron\varsigma$ that illustrates some aspects of the way we use computers, not only in Classics, but in most fields of the humanities. I recall a theory about the beginning of writing that I learned during a protracted and unsuccessful attempt to learn Sumerian. Originally, in what would later become Sumer, a shepherd would take as a receipt for his flock a hollow clay ball within which there were a number of tokens, one for each of the sheep that had been put in his care. In the end you would break the hollow ball, count the tokens and see if the shepherd still had as many sheep as you found tokens. At one point, some sly fellow figured out that you could mark the outside of the clay ball with symbols to represent sheep. That way you didn't have to break the clay ball to know how many sheep the shepherd was supposed to have. A little bit later, somebody figured out that you didn't really need the tokens inside the ball anymore: you could just write symbols on clay. The rest is, as they say, history, or rather, in this case perhaps, the beginning of history.

I am not so much interested in the fact that people developed writing as in the transition from clay balls to tablets. The transition may seem obvious now — I remember the amusement that we felt in our Sumerian class when we learned this theory — but it took quite a while for someone to figure out that you didn't need the clay balls any more. Doubtless radical Sumerians suggested doing away with the clay balls for some time, and were naturally buried alive, impaled on stakes, or denied tenure by their

colleagues. In the end, several generations lived and died, inscribing marks on the outside of a round object, presumably pushing off, by an equal amount, the wonderful world of freeways and sushi bars in which we now have the privilege to live.

Every time I complete a manuscript and ship it off to the tender mercies of some journal, I think of those clay balls. Here I have created a structured document, footnotes marked, a bibliographic database linked in, macros for everything, and what happens? In the surprising event that said manuscript gets accepted, someone will retype the entire thing, sprinkling typos throughout the text, quietly changing punctuation and doing random damage. Suppose, however, I can just hand an electronic copy of my manuscript off to someone with a compatible system. If I submit something to *HSCP*, it can be incorporated into the volume almost immediately, shipped over to Gary Bisbee of Chiron Inc. electronically, and reprinted with the change of only a few parameters. What happens then? The Electronic manuscript gets tossed aside and the paper printout gets distributed. Of course, we all say — why else did I hammer the document into the machine? But this is a little like throwing away the banana and eating the peel. What can you do with a book? You can read it. What can you do with an on-line document? You can read it on the screen, or you can print it out, and then read it. In fact, if you use a moderately priced laser printer, the document that you print out will probably look as good as most journals when they are actually printed.

But, of course, you can do more than just print something out. You can search around in the text, looking for references to Wilamowitz or Sophocles or $\sigma\phi\rho\sigma\acute{\upsilon}\nu\eta$. How much time does the standard classicist spend, hunting back and forth in an article or book, looking for a passage that he or she read, now needs, but can't find?

Now I could go on indefinitely talking about what you could do with books stored on-line that you cannot do with them if they are printed. The last time I had the opportunity to speak here at UCLA, I outlined some of the things that you could do with an on-line LSJ. People here have since begun preliminary work on such project, and at least two subsequent talks will touch upon it. So this morning, I will not worry about the qualitative differences between how you structure an on-line text and how you would structure a book. Instead, I will focus upon some of the more external logistical and political challenges that clay balls and tablets present.

First of all, an on-line document may, of course, go off-line, but it need not be checked out, nor need the storage medium be physically present. Thus, if a document is on-line at a central facility, its contents can be transferred electronically somewhere else, whether you are two miles away working at home or two thousand miles away working at a different institution. Further, compact disks now allow us to store almost infinite amounts of textual information — a shoe box would probably store most of the information that we want. The continually delayed, but ultimately inevitable, appearance of CD WORM's, CD on which you can write, but

not erase information, means that you could tailor your library. Rather than using a CD with all of *HSCP* or *AJP*, you could extract those documents that were relevant to the topic at hand, and then keep them on a smaller number of disks.

Of course, the scholar cannot yet, with his computer, his modem and his box of disks, retreat to his cabin in Newfoundland or Fiji. Surprisingly, technical issues, adequate hardware and software, are not the primary barrier. William Johnson will be showing the baby Ibycus with a TLG compact disk during this conference. As soon as we can get an interface board working for the Phillips CD ROM, my colleagues at Brown and I will have similar capabilities. If you have an IBM PC and a CD ROM, you can, for about \$200, purchase a CD with the Grolier encyclopedia and the software that you need to use it. Two years ago few people were interested in dealing with massive textual databases, but the CD has begun to change that attitude, and there are plenty of people out there grappling with the problems that we, who wanted to use the TLG, had to solve by ourselves. Software and hardware will only become cheaper and more useful over the next few years. Taking advantage of new tools will pose endless problems, but the first wagon trains of Classicists have already begun to move. Some of us will certainly fall by the wayside, perhaps none of us currently involved will succeed, but someone will soon enough. Californians don't need to learn about Manifest Destiny from a New Englander.

Nor is data entry really the major barrier. Certainly, we can't just march into the Classics collection of UCLA Library, and start entering books. If given such a task, I myself would beat a premature retreat to that cabin in Newfoundland or Fiji, leaving behind no forwarding address and taking with me an assumed name. The UCLA-Library-on-a-Disk approach is obviously impracticable. It is also backwards.

I mentioned above that a single disk could easily store all ninety odd volumes of *HSCP*. How many people and institutions could, today, use all of *HSCP* on a CD? Not many, perhaps, because few have CD ROM's, but how many will have the proper equipment in a year? in three years? in five years? How many would be willing to pay \$1000 for such a disk? Not many, perhaps. What about \$100? Assume it costs \$10 to master each CD, \$5 to mail if off, and \$20 for a software license.¹ That leaves \$65 to pay for putting the journal on-line. How many CD's could you sell? Suppose you charge institutions more than individuals? Or, put another way, we are trying to establish how much money the field, as a whole, would be willing to spend to acquire a major journal on-line.

¹ This figure is quite low (the smallest per unit cost to the OEM for such software would probably be about \$50 today). On the other hand, the educational market, especially the lower end that serves junior high schools, demands inexpensive products, but provides a large market. This area will hopefully drive down the price of software.

Let's turn now to the other side of the issue: how much would it cost to enter *HSCP*? Roughly speaking we have a hundred volumes with about 300 pages each, or about 30,000 pages total. Retypesetting this corpus would cost us between \$300,000 and \$400,000. This figure should represent an upper limit, since it covers not only manual data entry in the US (much more expensive than if contracted abroad), but much extraneous specialized work involved in tweaking the printed page. Speaking then in very rough terms, 10,000 CD's, each netting \$50, would generate \$500,000. This gives us some idea of how large a market we would need to pay for the expense of entering the data.

At the same time, you are not going to fill a warehouse with *HSCP* CD's. Each year you ought to create a new disk with the most recent volume of *HSCP*. This in turn opens new options: suppose you don't have the money or time to put all ninety odd volumes on-line. You might still be able to put the past ten years on-line. You might then, the next year, put another ten years on-line, releasing a CD with 22 volumes. The third edition could slide back ten more years and contain 33 volumes, and so on. In such a scheme, you could impose a onetime initiation fee (\$100?, \$500?, \$1000?), and then sell each year's edition for cost plus 20% or some such formula.

Now, of course, much as I love my department, I realize that *HSCP* on a disk would not allow everyone to leave the library forever. There are actually days when people do not even look at a volume of *HSCP*. But what if we had a number of the major American journals, *TAPA*, *AJP*, *CP*, *CA*, *GRBS*, etc? That would start putting a noticeable dent in our work. If you can put *HSCP* on-line, you can handle just about anything, and the more journals available, the more attractive a CD version of any one journal would be.

More importantly, we have begun to attack the problem of archiving printed material in a manageable way, dividing the tasks up between the different parties who are responsible for disseminating a given set of documents. Thirty thousand pages of *HSCP* or *AJP* is a lot, but it isn't Widener Library. Of course, such progress will be uneven: who will take care of defunct journals? How long will it take our European colleagues to embark upon similar projects? Many important documents in out of the way publications will initially be left out, some will never make the transition.

What I have said so far, however, covers journals and monograph series, but not books. Many publishing houses have vanished as utterly as Sumer, more so in fact. The volume is staggering, and the problems are numerous, but here I am not so much concerned with the archival problems. I am much more interested in how living publishers will react.

First, consider the traditional limits on distribution. The compact disk increases by more than two orders of magnitude the quantity of textual information that a single "item" can conveniently contain. If a single compact disk can store most of the Oxford University Press's commentaries and texts, why would you want to purchase the works piecemeal?

Of course, no publisher will want to sell you so much information at a price that you would be willing to pay. To stay in business, the publisher would have to find some other way to extract a comparable amount of money. It will have to make up for hundreds or thousands of dollars, distributed over smaller purchases, spread over a number of years. The compact disk has made more than one publishing house nervous. Already, there are plans to produce less effective storage media, to contain, artificially, the amount of information that an individual can acquire.

On the other hand, should the cost of a CD be the sum of the individual publications carried on that CD? Hardly, I think. When you spend X dollars for a text or commentary, very little, if any, of that money goes back to the author. Instead, you are paying for a hidden infrastructure that takes a manuscript and distributes it as a book. First you pay for camera ready copy, then printing and binding, then you rent space for the book to sit in a warehouse for several years, you pay for the salary of the people who put the book in the warehouse and then ultimately take it out again. While the book sits on the shelves, you pay for it to be advertised so that you will know that it exists. All of this works very well if you are producing a book that will sell in large quantities, because mass production is usually more efficient. But if you can only expect 500 or 1000 copies, then each book must generate more income to pay for the system behind it.

Each of the costs that I have described above can, however, be simplified. Consider camera ready copy. Ninety per cent of the work for ninety per cent of the papers published is practically automatic: someone who prepares a document with a powerful formatter such as TROFF or TEX, and who has ready access to a laser printer, can almost produce their own camera ready copy. Sometimes there are technical obstacles (Gary Bisbee of Chiron Inc., for example, recently set a nightmarish book on Greek meter), but the major obstacles here are political: publishers often do not want to accept the format that you are using, and often have their own arrangements for typesetting that they do not want circumvented. Printing and binding, on the other hand, lose much of their significance if a document exists primarily on-line. With inventory, however, the real payoff begins. The approximate dimensions of a compact disk, in its plastic case, are 6 by 5 by 1/2 inches, or 15 cubic inches. The first volume of my Thucydides OCT contains less than 1 megabyte and has dimensions 8 by 6 by 1, or 48 cubic inches. Even if your compact disk is only half full and contains 250 mbytes of data, you can store 750 mbytes where you previously could only store 1/2 a megabyte, or, in other words, 1,500 times as much. The storage issue becomes so minor, that you will often not need anything more elaborate than a closet to act as your warehouse. What about advertising? You advertise scholarly publications so that people will know that some new book is available and then put up their money to buy it. Each book needs its own place in the Publisher's advertisement, and trails with it certain costs. The whole process would be cheaper and simpler if you dealt with larger volumes of material.

Now I am not attacking the idea of books. I am quite fond of reading things outside of our field, and I depend upon Cambridge book stores to present a selection of readable titles from which I may choose. But I rarely spend more than ten dollars on such books, nor do I want thousands of these in my home. I am rather speaking now as a professional classicist, whose primary task is to reconstruct the ancient world and whose raw material is information. From an economic perspective, academics are suppliers of raw materials. If you recall my comparison, a bit earlier on, of books and bananas, you might say that we are the inhabitants of a classical banana republic, negotiating with the friendly American fruit companies. I do not believe that the people who run, for example, the American Fruit Co. are malicious or actively malign. I do, however, think that they deserve much of the credit for making Central America what it is today.

You may think this facetious and be shocked at comparing your local university press to the people who bring you Chiquita Bananas. But when we (or our libraries) pay several hundred dollars for a text, or when an American university press charges more than fifty dollars for a "student" commentary, we are paying a well-developed bureaucracy to distribute our ideas, we are paying more than we can afford, and we are strangling our own work. You can never escape this bureaucracy if you are distributing books: they are so bulky, and you have to maintain so many of them in stock that no academic would want to cope with the overall mess. But efficient new storage media such as compact disks change the variables in the equation. It may be preferable for a department to maintain and distribute publications with its own personnel.

To sum up, I suspect that our most fundamental challenge may lie not in the technical, but in the administrative area. We need to reevaluate and revise the economic factors that govern the way we exchange information in our field. Tools such as the compact disk do not represent a simple, linear extension of our previous capabilities, but lift us up and carry us, in a single gigantic motion, from one plateau to another. The basic forces remain unchanged: we need to maintain a system that can store and disseminate scholarly information, but the external factors are different: what will be the relationship between printed and on-line material, how much money will the scholarly community need to produce to fire up the system and keep it running? These are problems that will inevitably be solved, but solutions will be all the more effective and quicker the more we grapple with the underlying forces at work.