CLASSICAL WORLD

July/August 1998
Volume 91, Number 6
Whole Number 1524
ISSN 0009-8418

SPECIAL ISSUE

COMPUTING IN THE CLASSICS

ARTICLES

PAPYRUS AND PRESERVATION

It will be just as well to get my confessions out of the way at the start.[1] First, the integrated information system that I am going to describe to you does not yet exist; it is in the very best traditions of vaporware. Second, although I was asked by the Commission on Preservation and Access[2] to represent them at this conference, I have very little claim otherwise to do so. I am neither an officer, nor an employee, nor a trustee, nor even a committee member of the Commission. My only official relationship with the Commission is that I am a current grantee. I have, however, been involved in library preservation efforts for more than a decade, including running a fairly large microfilming project on behalf of the American Philological Association, and I therefore have some sense of the historical context in which current work is set.

The project that I am currently involved in developing actually reflects quite neatly the Commission's basic mission. This is not exactly deliberate, but neither is it an accident. Let me first tell you something about the project, and then put it in the broader setting. The project is called Advanced Papyrological Information System; in keeping with the modern spirit of acronyms, we refer to it familiarly as APIS, in honor of the Egyptian bull god whose sanctuary in the Memphite necropolis was the source of some of the most important papyrological discoveries of the nineteenth century.

APIS is intended in the first instance to provide a seamlessly integrated system to support research and teaching in papyrology, but a considerable part of its raison d'être and appeal is the belief that it will mark a decisive step forward in opening up the rich material found in the papyri to those working in many areas of ancient studies, for whom papyrology often seems much as Egypt did to the Greeks, fascinating, exotic, and incomprehensible. We like to think that some features of APIS will make at least parts of it usable even by those with no specialized training in antiquity.

Papyrus was the most important writing material of the ancient world and perhaps ancient Egypt's most important legacy. On it was recorded everything from high literature to the myriad of documents and other communications of daily life. About one in ten of those studied to date is a fragment of literature, either a far more ancient witness to a work known otherwise from medieval manuscripts or a

---

[1] For publication I have made only minor changes to the text of my oral communication, keeping documentation to a minimum and preserving some elements aimed at making the subject intelligible to nonclassicists. As the project described here develops further we expect to publish more formal and detailed reports. Between the conference presentation and sending this version to press, this project received funding from the National Endowment for the Humanities and began work in July 1996. A pilot version is expected to be available in fall 1998. I am grateful to Alden Smith for the invitation to take part in the conference from which these papers stem and to Maxine Sitts of the Commission on Preservation and Access for several useful documents.

[2] Described below, 547.

text hitherto lost in antiquity. From the literary papyri the modern scholar both learns about the state of literary texts in antiquity before errors were compounded in the manuscript tradition of the Middle Ages and recovers such important works as the lyrics of Sappho and the paeans of Pindar, the comedies of Menander, the mimes of Herodas, the *Constitution of the Athenians* by Aristotle, and early Christian and Gnostic works.

The remaining nine of ten published texts are private letters or documents of every conceivable sort—legal and business papers, government regulations, property records and transactions, petitions to high officials, tax and rent receipts, bank deposits and payments, and farm and crop reports. As such these documentary papyri differ little from modern archival material; except for their usually fragmentary state and extreme antiquity, they reflect the quotidian affairs of government, commerce, and personal life in much the same way that modern records do. The papyri are thus the source of a large part of what we know about many aspects of antiquity, particularly those concerned with economic life, social relations, cultural interaction in a pluralistic society, and daily life.

Papyrology embraces both the editing of texts, the labor of the editor faced with a papyrus or photograph, and the use of these texts to study the societies, economies, and cultures of the regions from which the papyri come. Most of them were found in Egypt, but recent discoveries of papyri and similar documents have greatly broadened the geographical range of papyrology. The Latin tablets from Vindolanda in Britain and the Greek papyri from the upper Euphrates in the Roman province of Mesopotamia mark the extremes of this range, and they give fair warning that it is not possible any longer to think of the papyri as something safely quarantined to Egypt, nor yet of papyrology as strictly limited to papyrus. For that matter, recent scholarship has shown that Egypt was far less different from other parts of the ancient Mediterranean world than people have liked to believe.[3]

Working with papyri is above all a matter of texts—texts especially in Greek, but also in Latin, Demotic Egyptian, Coptic, Arabic, Persian, Aramaic, and various other languages. The Greek documents are already available in electronic form through the Duke Data Bank of Documentary Papyri, but much work remains to be done in adding nondocumentary papyri—literature and what are usually but infelicitously called sub-literary texts—and the texts in other languages to the data bank. Texts are, however, only the beginning. They are normally published with translations, introductions, and detailed commentaries. Once they have been published, people write about them in articles and monographs, producing bibliography. Papyrological bibliography is only partially covered in *L'Année philologique*, but fortunately we have a special bibliography, published in Brussels, the *Bibliographie papyrologique*. Since 1932 it has been published

[3] A more detailed discussion of the character of papyrological documentation can be found in my *Reading Papyri, Writing Ancient History* (London 1995).

on 3 × 5 inch index cards, and no library that I know has succeeded in keeping up with filing the six hundred cards issued each year. The *BP* is now in the process of being converted to electronic form; starting with 1995, the old cards have been replaced by a choice of an electronic version or a printout from it, and the retrospective bibliography is in the process of conversion. The entries from 1976 to 1990 are already available, and a complete 1960–97 bibliography should be available by the end of 1998.[4] We are not so fortunate with the repertory listing all proposed corrections to published texts, which exists only in manual form and is generally six to eight years behind.[5]

So far, so good. The traditional use of multiple types of information by papyrologists is steaming ahead into the modern world. But the process leaves some things trailing. Suppose you need to check to see if a reading on a papyrus is right, or you want to see if the handwriting is similar to that on a piece you are working on; what do you do? Traditionally, you wrote and asked for a photograph. After a few weeks, or months, or even years, and sometimes with a hefty invoice, it may arrive. You might study it intensely for months or discover in thirty seconds that it did not help.

For the nonpapyrologist, the problem might be simply to discover in the first place what exists. Suppose you are studying wills: how do you find those surviving on papyrus? The papyrologist's answer is that you find the most recently published—by luck or by searching in the data bank for suitable terminology—and hope that the editor has been conscientious enough to tell you where a list has been published, and what has been added since. For a nonspecialist, this entry hurdle is too high. What one wants is a systematic, searchable catalog with subject classification and a controlled vocabulary of document types. Such a thing now exists for one collection, that of Duke University, where the National Endowment for the Humanities has supported in the last few years the development of a prototype catalog.[6]

Even where electronic tools have developed, work on the papyri is still dauntingly complex. What APIS seeks to do is to bring together the various needed resources in a way that is easy to use. It will combine electronic catalog, digitized text, scanned publications, electronic bibliography, and digitized color images in a single system, in which the user can switch back and forth from one part to another with a click. Let us say you do want to study wills using

[4] The 1976–90 version was published by the American Society of Papyrologists through Scholars Press; the 1960–97 bibliography will be published by the Fondation Egyptologique Reine Elisabeth in Brussels.

[5] This is the *Berichtigungsliste der griechischen Papyrusurkunden aus Ägypten*, now (1997) in nine volumes plus one index volume, and published by Brill. The most recent volume (1995) covers publications of 1987–90. Although the *BL* remains to date a manual tool, the process of enhancing the Duke Data Bank with corrected readings is underway.

[6] See Peter van Minnen, "Introducing the Online Catalogue of the Duke Papyrus Collection," *BASP* 31 (1994) 159–70.

this system. A search of the catalog brings up a list of such papyri; you click on one, and up comes the text. You want to see the translation and commentary, and another click brings that. You are interested in the physical layout of a will, and another click brings up its image. And so on.

Those who have used net browsers like Mosaic, Netscape, and Explorer will of course recognize that what I have been describing is essentially a variation, an enriched version, of such browsers. And in fact APIS is intended to be constructed using the hypertext standard, HTML, prevalent in the World Wide Web files now proliferating at a great rate around the Internet.

All of this is very nice, you may be thinking at this point, but papyri are scattered around the world in hundreds of collections. How on earth will it be possible to realize such a system without gigantic computers and huge administrations? Can it be done at all?

APIS began as a collaborative project of three major universities with papyrus collections, Columbia, Duke, and Michigan. Before we were nine months into the planning, strong sounds of interest started coming from other universities, and we have added Berkeley, Princeton, and Yale to the group. In addition, the Université Libre de Bruxelles is a limited partner, contributing to the development of the electronic *Bibliographie papyrologique*. Getting six universities, with their own practices, traditions, and egos, to cooperate on something like this is by no means simple; it is like "herding cats," as one participant called it.

Just what is at stake here? It is not, as one might think, the access tools for such a system. Rather, it is the set of standards for the collection and storage of data. The catalog, for example: Duke adopted a special archives and manuscripts version of the standard library digital catalog record format called MARC. When you use any library electronic catalog, the database supplying the information is in this standard format. Using such catalog records for papyri has great advantages; the records can go into standard utilities like OCLC and RLIN, for example, and the subject fields can be searched by librarians who know nothing about papyri. Creating MARC records is, however, rather time-consuming and thus expensive. Not everyone is persuaded that the cost is worthwhile, especially for small fragments with little information. One solution, which we have decided to adopt, is to create simpler finding-list aids for groups of smaller items, with a MARC record representing each group rather than each item. The Bancroft Library of the University of California at Berkeley has been developing such finding aids for its manuscript collections under a Title II-C grant.

Even these lists, however, require standards if they are to be created in six places and usable in a single system, and this has been a central interest of the Berkeley project. Here we have adopted the SGML standards (for which consult Susan Hockey's paper in this issue). These allow software that knows SGML to identify all sorts of textual units, like chapters, sections, and paragraphs, and types of text like italics. What is critical is the adoption of a single

method of marking. As SGML is at the base of such efforts as the Text Encoding Initiative, it should ensure the compatibility of APIS with the rest of the electronic resources on the Internet.

When we come to electronic images, the challenge is greater. There are many types of electronic image capture systems now in existence, and the technology is changing rapidly. How can a project like APIS adopt a standard that will prevent what it does from becoming obsolete in a short time? This is the point at which the Commission on Preservation and Access entered the picture, and this is the moment for me to back up from APIS and say a bit about the Commission and its objectives.

The Commission set out several broad goals when it was founded, and it describes itself briefly as a "private, nonprofit organization acting on behalf of the nation's libraries, archives, and universities to develop and encourage collaborative strategies for preserving and providing access to the accumulated human record." Institutional collaboration is thus a central part of the Commission's approach. In its earliest form this collaboration was deeply concerned with consortial approaches to preservation microfilming, particularly in making it easy for one library to determine if anyone else had already filmed an item it was considering filming. As the technical environment has changed in the last few years, the Commission has increasingly turned its attention to newer technology. It has, for example, fostered the Digital Preservation Consortium, which happens to include five of the six APIS partners, seeking to "advance the use and utility of digital technology for the preservation of and access to library materials."

A second major goal of the Commission is to help make preservation work supportive of scholarship. This might sound like apple pie, but anyone who has tried to persuade scholars to use microfilm knows that library preservation efforts have often been seen by scholars as an enemy, a dragon that eats books and spits out an inconvenient film that must be used on a special machine—special, that is, in usually being broken, causing eyestrain, and making it difficult and expensive to get decent copies of wanted pages. The Commission's name embodies the seriousness of its concern that preservation not mean making the past less accessible but making it more accessible. Digital technology clearly has the potential to do this, but it is a potential and not an automatic outcome.

A third critical element of the Commission's identity is its focus on research and development. It is not interested in becoming a large organization operating major projects, a kind of industrial establishment of the preservation world. Such things are needed, but the Commission is more interested in fostering the development work that will make them possible. There is an immense amount still to be learned about how technology can be used for preservation, what protections and safeguards are needed, and how libraries can use it to the greatest benefit of the scholars and students for whom they exist.

It should by now not need a lot of drumbeating for it to be obvious that APIS seeks to embody all three of these goals. Given

the Commission's interests, our need to develop standards for electronic imaging fitted well into the Commission's desire to foster research into collaborative efforts that would support scholarship, and they have funded a current study that included a meeting at the University of Michigan just four weeks ago. The objectives of this study fit very closely with the interests of the Digital Preservation Consortium that I have already mentioned as a project of the Commission. The Consortium seeks, among other things, "to define and promote shared methods and standards for the production, storage, and distribution of digital images," to foster collaboration among institutions, and to study the relationship between traditional preservation on film and the use of digital technology. These specific goals dovetail even more closely with APIS's objectives.

Just the experience of two of the partners in APIS, Duke and Michigan, gave us some sense of the problems that needed to be confronted. Michigan had begun scanning papyri a few years ago, with a grayscale flatbed scanner producing images at 300 dots per inch—roughly the resolution of almost all laser printers until a couple of years ago. They were very pleased with the results: it was possible to print out fairly usable images on a laser printer, and on the screen it was possible to use the whole bag of tricks of programs like Adobe Photoshop to enhance the image, blow it up, and so on. In teaching, the image could be projected on a screen so that the professor could point out problems or characteristics to the whole class at once. It was a real revolution.

Hardly two years later, Duke started experimenting with color scanning from the papyri, at roughly the same resolution. Not only did this have all the advantages of the grayscale scanning, it allowed more sophisticated color manipulation to enhance contrast and give a much better sense of such details as fiber color, useful in trying to join fragments. But technology produces leapfrogging, and last year Michigan decided to try grayscale and color scanning of originals, plus color scanning of photographs and slides, in a kind of papyrological image bakeoff. The images were mounted on a server at Michigan, allowing users all over the Internet to look at them and register their opinions about image quality and usefulness. Then Michigan was able to borrow a digital camera, made by Kontron in Germany, and added images from that to the test. The results of the digital camera were considerably superior, they thought, and they suggested adopting this as our method.

But things are not that simple. There are other digital cameras on the market, and we had read also about the work in Claremont using multispectral imaging with the Dead Sea Scrolls. Was that relevant to us? And what about the other projects we sometimes read about, like the IBM project at the Vatican Library, or the efforts by Getty to develop standards for imaging of artworks? Columbia was studying the use of microfiche to record large-format color maps; could these then be successfully transferred to digital form? And so our meeting in Ann Arbor was born, out of a sense that we needed

to discover what was essential and what peripheral, what relevant and what not, what useful in some contexts but not for papyri. The most basic conclusion was that it was of no use to decide on equipment; that changes constantly and is not a standard. Rather, we needed to define resolution, color control, file format, and other basic technical standards. With these in place, it would be possible for institutions anywhere to work independently and produce interchangeable parts for the great engine of research and teaching we hoped to create.

This is not an appropriate forum for discussing the more technical aspects of this, but a few of our conclusions are worth mentioning. The first is that one should adopt widely known formats and procedures, not peculiar or proprietary ones. This insight came, interestingly enough, from the representative of IBM Research who talked about the Vatican Library Project. Even though this project uses an IBM internally developed electronic camera that you can't even buy from IBM's marketing division, the project uses it to produce files in normal formats. Second, the appropriate standard for papyri is not the same as for artworks, where it is essential, for example, to be able to discern translucence and precise color differences. For papyri, the characteristics of the underlying medium are of secondary importance to the clarity and density with which the writing is recorded. Third, it seems that a standard of about 600 dpi is both achievable now and sufficient to meet any foreseeable need. With that resolution, the image can be blown up considerably larger than life, to a size exceeding anything that one would need in research, with no significant loss of sharpness or density in the characters, even in small fragments of letters. An electronic camera with a 3,000 x 4,000 pixel recording capability, like that in use in the Vatican, can capture a papyrus at this density as long as it does not exceed 5 × 7 inches; beyond that point, segmenting is needed. But we know that cameras with 4,000 x 6,000 pixel capability will soon be readily available, and almost all papyri of normal format will be capturable at 600 dpi using such a camera.[7]

A fourth conclusion that deserves brief mention is that multispectral imaging would be extremely useful for hard cases, texts where contrast between medium and ink is poor—where, for example, there is a thin film of plaster over the letters, as often happens with papyri used to wrap mummies, or with texts written on dark red-brown pottery. For ordinary papyri, however, multispectral imaging has no advantages.

Back now to more central concerns. The overall point is that if all six institutions adopt common standards for the capture and representation of data, be that text, catalog, or images, then the results will be usable in a single system. This will be true even if each institution devises its own user interface. Because the development of digital libraries is proceeding at such astonishing speed, we simply cannot predict what any one institution will have on-line in five

[7] Postscript (May 1997): Cameras with a 7,000 x 7,000 pixel capacity are now on the market.

years, or how it will provide access to it. At an extreme, one could imagine a situation where each institution's files were on its own server, and each institution had a totally different user software, although we hope it will not come to that. Even so, Columbia's browser, provided with the necessary addresses, could move effortlessly from one collection to another, so that I would hardly even be aware that one papyrus and its electronic offspring were at Duke, another at Michigan, and a third at Berkeley.

This sort of sweeping claim is likely to raise two—well, at least two—questions in your minds. The first is what this means in terms of the software used to provide access to these files: what programs am I talking about, and are they vaporware too? The other is how far the principle can be extended: today six universities, tomorrow the world?

The software mostly exists already. Web browsers like Explorer and Netscape have the ability to move among files in just the way I have described, as long as they are properly marked with the hypertext links between the text, bibliography, image, and so on. Serious work on images requires separate software, but this too is offered by several commercially available packages. The big gap is just where you would expect: support of nonroman alphabets, Greek first but also Coptic, Arabic, and so on. There is no great difficulty in working with Greek nowadays in word processors, and there are various programs for different platforms which search the Thesaurus Linguae Graecae, the Duke Data Bank, and other collections of electronic text. But there are several different standards for coding Greek in use. The databanks are recorded in what is called Beta Code, made up of Roman characters with various marks for accents, breathings, and the like. But no word processor that I know of operates with beta code; these all have proprietary methods of representing Greek. If you have tried to put together a publication with disks from the authors, you will know what I mean: Nota Bene, Microsoft Word, and WordPerfect do not talk to each other. Even more, there is no means of having your microcomputer—the client—understand what is Greek on a server.

A solution to this problem exists, an international standard for representing characters. In its 32-bit version it is catchily referred to as ISO 10646, but the 16-bit equivalent is called by the somewhat sexier name Unicode. In principle, if server software and microcomputer software all understand Unicode, no alphabet should be a problem. The only catch is that there is almost no software out there that uses Unicode. There probably will be one of these days, but this is something that has been awaited for some time now and hasn't yet happened. We will probably be forced to use beta code and create a module to work with Netscape or another browser that interprets it, at least as a stopgap. But work is underway in various places on this problem, and it may well be solved by the time APIS is operational.[8]

[8] Postscript (May 1997): A successful implementation of beta code on the Web has been carried out with the mounting of the Duke Data Bank at Perseus. Microsoft Office 97 for Windows supports Unicode, and ability to use Unicode in APIS is moving closer.

There will be some other work needed to give maximum utility to APIS's front-end software. For example, it would be helpful, especially to the nonpapyrologists, to incorporate into it a handy program developed by the Belgian scholar Willy Clarysse that converts the dates in the papyri into Julian form. For example, if you came across a text dated to the sixteenth regnal year of Antoninus Pius, the month of Pachon, and the third day, probably only five people in the world could tell you without looking it up that this was A.D. April 28, 153. With Clarysse's converter, you would simply pull up a window into which you would type the emperor, the year, the month, and the day, and up would come the answer. Over time, of course, it is easy to imagine other such tools that would be handy. For starters, how about incorporating digitized versions of the maps of Egypt being prepared for the *Atlas of the Greek and Roman World*? Then you could type a place-name encountered in the papyrus into the pull-down dialogue box and have a map of Egypt pop up showing where it is. This dream will take a little longer; but it is realizable.

The international dimension is more challenging. Even if we can develop a set of standards for data that all significant collections in this country will adopt—and this is by no means assured, even with the participation of the six most important collections—can we really hope for cooperation from the dozens of European collections? Despite the grandiose term "European Union," local particularism is alive and well in Europe, and the Not Invented Here syndrome in matters of computer technology is widespread. Some European scholars remain profoundly suspicious of the electronic world and resist its spread.

All the same, looking back five or ten years helps to give us a sense of just how far things have changed even on the Continent. I do not think that anti-technology sentiment will in the end do much to affect the course of things. But will we be able to agree on a list of standard terminology for describing types of documents, just to use one example? In my view the key to success will be the enlisting of a few large collections at a relatively early stage. The first phase of APIS, launched in the summer of 1996, has attracted European interest already. The University of Heidelberg is cooperating in linking its typological database of papyri and its digital images to the Duke Data Bank (and thus to APIS), and the Oxyrhynchus Papyri have begun digital imaging to our standards and will become a full part of APIS during phase 2. I presented a sketch of the idea of APIS to the papyrological congress of 1992; most of my hearers seemed to think it was fantasy, but times change.

For scholars, two themes seem to me to emerge from all this. First, scholarship will be best served by involving scholars in the design of the systems they will use. With APIS we have started from the premise that papyrologists know how they do *their* work, and that it is above all their definition of what is and what is not an adequate set of resources that should guide the project. This view is shared by the Commission, which supported our study not just to be helpful but because it thought that scholarly evaluation of the use-

fulness of different types of images was the ultimate test of what would constitute a standard for preservation.

Equally, however, our experience suggests the limits of what scholars can bring to the process. The results of scholars' decision-making about appropriate standards must then be integrated into a wider context in order to avoid obsolescence as technology changes. What we found, actually, was that we had not necessarily even framed the questions correctly, and that we had to broaden and generalize what we were considering in order to take account of what had already been done by others. The consultants we brought in—Frederick Mintzer from IBM, Michael Ester from Luna Imaging, and Gregory Bearman and Sheila Spiro from the Dead Sea Scrolls project—were vital in forcing us to enlarge our view and see that it was standards, not ways of meeting them, that was our real subject. We are, of course, still very much at the beginning of our work. Along with the technical glitter will go a tremendous amount of labor of a very traditional sort: physical conservation of papyri and intellectual work of deciphering and describing the texts on them. These things remain in many ways unchanging. But not entirely: the intellectual work of editing papyri has already been dramatically transformed by the availability of the Duke Data Bank. In teaching I find that the most laborious steps of beginners have been greatly shortened by their ability to do electronic searches on the first few words they are able to read. As the bibliographic and image resources of papyrology grow, other parts of the task will become easier and more rewarding. In particular, I think the entire physical or archaeological side of the discipline will be changed. Where now the papyrus as artifact takes a back seat because one must hunt for a limited number of published plates in order to observe the format of documents and the development of handwriting, in the future the tangible reality of the papyrus will be brought home by omnipresent and readily available color images. Those without large collections of their own will in this way be able to be at home with the full range of hands and layouts, and even experienced papyrologists will for the first time be able to transcend on a daily basis the limitations of their own collections.

Despite a great deal of rhetorical flourish about how technology will change scholarship, most scholars' use of the computer has up to the present simply made it possible to do faster and more easily what they always did. That has begun to change under the influence of the large textual data banks, but only slowly. With APIS we face the advent of a worldwide virtual papyrus collection and what I believe will be a true transformation of scholarly work; the same will be true of other such projects in other disciplines. It will take at least a generation to get there, but the time is ripe to set out on the journey.

Columbia University
CW 91.6 (1998)

ROGER BAGNALL
bagnall@columbia.edu