Imaging of Papyri: A Strategic View

ROGER S. BAGNALL

Abstract

In the summer of 1992 Roger Bagnall outlined a proposal for the creation of an Advanced Papyrological Information System (APIS). In this paper he describes the strategic considerations on which the project is based and the advantages of providing on-line images of original papyri from several different collections.

The digital imaging of papyri is part of a larger set of developments. This is true both in the sense that papyrological use of imaging is part of a broader introduction of imaging into scholarship in the humanities and in the sense that images of papyri are only part of a larger change in the ways that papyrologists use information and do their work. It is this latter context that is discussed here.

Virtually the entire textual corpus of Greek documents on papyri, potsherds, and tablets is now available in electronic form through the Duke Data Bank of Documentary Papyri; much work remains to be done in adding non-documentary papyri—literature and what are usually, but infelicitously, called sub-literary texts—and the texts in other languages to the data bank. However, even in a textual field, texts are only the beginning. They are accompanied by translations, introductions, and detailed commentaries. Once they have been published, people write about them in articles and monographs, producing bibliography. Papyrology is fortunate in having a special bibliography, published in Brussels, the Bibliographie Papyrologique (BP). Since 1932 it has been published on 3 × 5 inch index cards, and no library that I know of has succeeded in keeping up with filing the 600 cards issued each year. The BP is now in the process of being converted to electronic form, and a version encompassing 1960–95 is available in electronic form. The first version of a repertory of documentary papyri developed in Heidelberg is also now available, allowing access to the material through a rich set of keywords in various fields. Also, a project at Duke University in the last few years has, for the first time, catalogued a papyrus collection according to the cataloguing standards of the international library world, providing another complex, rich form of access.

The traditional use of multiple types of information by papyrologists is thus steaming ahead into the modern world. The process had, however, left some things trailing. Suppose you need to check to see if a reading in a papyrus is correct, 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 might arrive. You might study it intensely for months or discover in thirty seconds that it did not help.

Two things were thus evident several years ago. One was that the rapid transformation of papyrological scholarship by the development of electronic tools was leaving the image of the physical object behind. The other was that we increasingly were faced with an array of tools that did not interconnect, which could be used separately but not together. These concerns led in two directions: first, to some experiments in the use of imaging technology, with flatbed scanners (see the paper by Gagos in this issue); secondly, to the concept of an integrated information system, sparked by the emergence onto the scene of the first Web HTML-based browsers. For some time now, a committee of the American Society of Papyrologists has been working to bring these two strands together.

The result was a consortium of six American universities (Columbia, Duke, Michigan, Princeton, California-Berkeley, Yale), which applied in 1995 to the National Endowment for the Humanities (NEH) for support for creating the Advanced Papyrological Information System, or APIS. Despite NEH's reduced means, they have made us a similarly reduced grant, with which work has begun in earnest.

Because the consortium sees imaging of papyri not as a self-standing activity but as part of an information system based on HTML links, it has been thought from the start that standards and interoperability were the most critical elements of its work. It is not necessary for APIS to be a single unitary system running on a single machine; indeed, in the long run, this is a most unlikely outcome. Rather, it will eventually become a little web of resources, located in many places. However, we believe that a uniform user interface, coupled with universally accepted standards, can lead to APIS's appearing to the user to be a seamless single system.

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 catalogue, for example: Duke adopted a special archives and manuscripts version of the standard library digital catalogue record format called MARC. When you use any library electronic catalogue, the database supplying the information is in this standard format. Using such catalogue 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

Correspondence: Roger S. Bagnall, Department of Classics, Columbia University, 606 Hamilton Hall, New York, NY 10027, USA.

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 Berkeley has been developing such finding aids for its manuscript collections.

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. 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 and most WWW pages, it should ensure the compatibility of APIS with the rest of the electronic resources on the Internet.

Standards seem to us particularly important for imaging—not standards in the sense of imposing a single way of doing things on everyone, an approach that we thought would not be indefinitely sustainable as other parts of the world joined in the task of creating APIS, but standards in terms of seeking a common quality and technical description of the outcomes. There are, after all, many different 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? The Commission on Preservation and Access took a strong interest in this side of our work, for similar issues face them in developing national and international preservation initiatives with all sorts of materials. They therefore commissioned us to produce a study on the problems and appropriate standards for imaging papyri; the study in turn underlay the application that we made to the Endowment. It was founded on a meeting in Ann Arbor at which we brought papyrologists together with a variety of experts in areas of imaging and information management, to see what in the real world the best achievable practice might be.

The technical aspects of this are discussed in more detail by Gagos (this issue), but a few of our general 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 cannot even be bought 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 art works, where it is essential, for example, to be able to discern translucence and precise colour 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 d.p.i. 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.

A fourth conclusion that deserves brief mention was that multi-spectral 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. However, for ordinary papyri, multi-spectral imaging has no advantages.