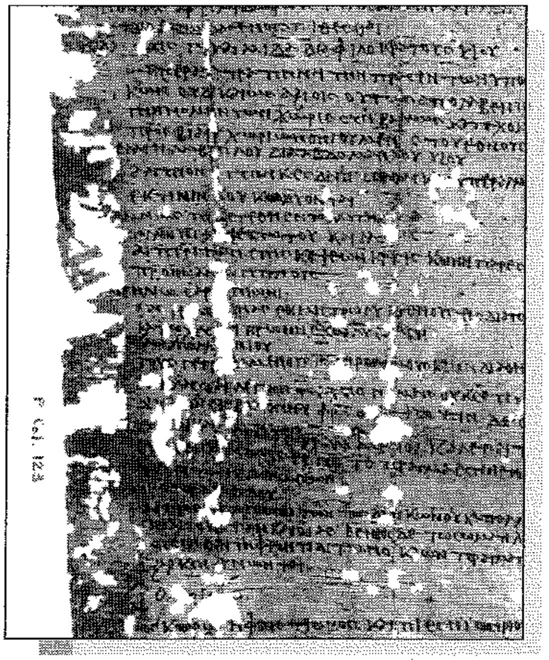
The Commission on Preservation & Access

Digital Imaging of Papyri

A Report to the Commission on Preservation and Access



September 1995

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.

Digital Imaging of Papyri

A Report to the Commission on Preservation and Access

Prepared by Roger S. Bagnall,
Professor of Classics and History

Columbia University

September 1995

Published by
The Commission on Preservation and Access
1400 16th Street, NW, Suite 740
Washington, DC 20036-2217

Commission Preface

This project, conducted under contract to the Commission on Preservation and Access, set about to document how to best scan papyri in anticipation of a major effort among papyrologists to digitize their collections. The report was prepared with the assistance of participants in the Advanced Papyrological Information System (APIS) and other expert consultants after a meeting in Ann Arbor, MI, in March 1995. APIS is an interinstitutional project growing out of a committee established in early 1994 by the American Society of Papyrologists to study imaging and other current technological developments and establish standards for the field.

The investigations have resulted in agreement on best practices for the capture and storage of digitized images for papyri that will be employed in the APIS effort to help ensure the usefulness of the created images for teaching and research.

Cover: *Apokrimata* Papyrus manuscript Alexandria, ca. March 14, 200 A.D.
Courtesy of Columbia University Rare Book and Manuscript Library

Additional copies are available for $10.00 from The Commission on Preservation and Access. Orders must be prepaid, with checks made payable to "The Commission on Preservation and Access," with payment in U.S. funds.

This paper has been submitted to the ERIC Clearinghouse on Information Resources.

The paper in this publication meets the minimum requirements of the American National Standard for Information Sciences-Permanence of Paper for Printed Library Materials ANSI Z39.48-1992.

ISBN 1-887334-44-0

Digital Imaging of Papyri

1. Purpose of the study

Fragments of the ancient writing material known as papyrus exist in at least a hundred collections in the United States and even more worldwide. Papyri pose significant challenges for both preservation and access: for preservation because of the damaged and fragmentary condition in which most papyri have survived, and for access because almost all research in papyrology involves studying pieces in many scattered collections. Sometimes fragments of a single papyrus may be found in several locations, and, still more commonly, related papyri are distributed in many libraries and museums. Published editions usually (for reasons of cost) include half-tone plates of only a selection of texts. Even with the growth of air travel in recent decades, visiting all relevant collections is rarely an option, and scholars are therefore faced with two unappealing choices: order conventional photographs usually slow to arrive and often prohibitively expensive, or use only published papyri and, even with those, be unable to check the editor's text. Color slides, experimented with in some European collections in recent years, have some advantages over black-and-white prints, but the quality and convenience of the viewing technology usually are deficient.

It has been clear for some years that digital imaging offers the promise of more satisfactory solutions to these problems. Digital imaging provides an opportunity to create a worldwide virtual library of images, freeing scholars and students no matter where they are located to study all relevant papyri in any collection. Such a 'library' would encourage wider use of papyrological texts in study and teaching about the ancient world and greatly improve the quality of research. At the same time, the ready availability of images could help reduce the physical handling of the original objects and thus help prevent further damage. These benefits also are relevant to the other writing-bearing objects generally classed with papyri, like potsherds (ostraca) and wooden tablets. Classes of manuscript material have similar characteristics, and findings about digital imaging may have relevance for them as well.

Before major projects get underway, however, it is important to determine how to accomplish imaging so that research and teaching objectives will be achieved. Early in 1994, the American Society of Papyrologists established a committee charged with studying imaging and other current technological developments and establishing standards for the field. Out of this committee grew an interinstitutional project for an Advanced Papyrological Information System (APIS).[1] A contract with the Commission on Preservation and Access supported a study of the imaging component of the planned system that would answer the following questions:

(1) Is electronic imaging now capable of serving as the main means of capturing the images of papyri and similar objects for research access and for preservation?

(2) What objectives must be met in such imaging?

(3) What technical standards should be established to accomplish those objectives and preserve the interchangeability and permanence of data?

(4) What are the limits of present technology, the controls and safeguards needed to ensure data quality and integrity, and the means of preventing obsolescence of the product?

(5) How well can currently available equipment satisfy the standards and objectives?

2. Method of the study

(a) Building on APIS experimentation. Three of the institutions cooperating in the development of APIS have acquired relevant expertise in earlier work, two with papyri (Duke University and the University of Michigan) and one with color maps (Columbia University). This previous work has served as the basis of the present study. It may be briefly summarized as follows:

Duke: As part of a conservation, cataloging, and imaging project supported by the National Endowment for the Humanities, Duke adopted 300 dpi color scanning instead of black-and-white photography as its main imaging technology. A few of the resultant images have been put on a World Wide Web (WWW) server in combination with cataloging records for public examination. Several hundred papyri have been scanned in this fashion so far.

Michigan: About 500 images were initially scanned using a grayscale scanner at 300 dpi; they were given extensive use in teaching as well as research, with generally satisfactory results. In the summer of 1994 a period of intensive experimentation with both grayscale and color technology began, including scanners at densities up to 600 dpi working from originals, photographs, and slides. When a Kontron digital camera (see Endnote 5) became available in October 1994, it was added to the test, scanning at 3000 x 2000 pixels. Some of these images have been made available on the WWW for public study and comment.

Columbia: A project over the past year has studied the problems inherent in capturing geological and topographical maps. The problems of color differentiation and resolution addressed in this project are relevant to those encountered in papyri. In particular, this project studied the utility of full-frame color microfiche for recording such large-format color media. One of these fiche was then captured with the Kontron digital camera to see how completely digitization captured the data recorded on the fiche.

All of this work had developed a general sense that digital scanning generally offers substantial advantages over silver halide photographic technology. In particular, the electronic manipulability of images more than compensates for any drawbacks, including the lower density of information provided by scanning.

Flatbed scanners have some limitations that cameras do not have. Although they can handle originals that are not entirely flat, they cannot capture very curved ostraca. In addition, it is difficult to work accurately with many small fragments on scanners because the written surface is face down. Also, scanners cannot handle larger papyri satisfactorily.[2]

Digital cameras are superior in allowing underlighting where that is useful, and they allow special lighting or filtering techniques. They also avoid the potentially harmful ultraviolet light emitted by flatbed scanners. The digital camera thus seemed to have the advantages of both photography and scanning to a considerable degree.

(b) Drawing on experts from leading-edge imaging projects. Representatives of the institutions cooperating in APIS met at the University of Michigan in early March 1995, with experts from several imaging projects chosen both for technological characteristics and for the significance of the materials as they relate to papyri:

Gregory Bearman (Jet Propulsion Laboratory, Caltech) and Sheila Spiro (Ancient Biblical Manuscript Center, Claremont School of Theology) represented the Dead Sea Scrolls imaging project. This enterprise has pioneered the use of multispectral imaging on manuscript material. Its results and relevance for the papyri are described in the next section.

Michael Ester (Luna Imaging) described two specific projects and general lessons derived from two projects. One was a study of image quality for works of art and viewer perception of that quality commissioned by the Getty Art History Information Program. The other was the application of digital imaging to a collection of Frank Lloyd Wright's architectural drawings.

Frederick Mintzer (IBM Watson Research Laboratory) reported on the IBM/Vatican library manuscript project, the test phase of which is now in operation. This project is capturing a wide variety of manuscripts of the medieval period, especially illuminated manuscripts. He also described work done by IBM on the digitization of art works for Andrew Wyeth.

(c) Joint discussion by imaging experts, papyrologists, systems experts, librarians, and curators. The March 3-5 conference allowed for intensive discussion. All six APIS institutions were represented, with a mix of preservation specialists, rare book and manuscript librarians, conservators, and digital library systems analysts. All questions listed above were considered from all angles in light of meeting the objectives both of papyrological research and of sound library and information management practices.

Discussion was informed by reports from the above imaging projects and from APIS institutions. Although there was a considerable range of opinions about many points, joint analysis allowed the emergence of a clear consensus about best practices for papyri, taking into account the way they are used and their particular characteristics.

(d) Visual inspection and experiment following a period of public display and comment. As indicated above, Duke and Michigan have made images available over the Internet in the last year. Michigan incorporated into its study a questionnaire on the quality and usability of the images. User response was generally in line with the observations of the papyrologists at Michigan and helped to confirm internal judgment about the relative merits of the different imaging technologies and sources.

At the meeting in Ann Arbor, the papyrologists took advantage of the availability of a digital camera to experiment further with image density, looking in particular for the point where the image has a high enough resolution that it remains sharp when subjected to the kind of enlargements needed for research use. The five papyrologists were thus able to compare impressions of the same image at the same moment. The recommendations in the next section are based in part on that common visual inspection and discussion.[3]

(e) Ongoing research. Imaging technology is changing rapidly, and this report obviously cannot settle all questions. APIS will build in mechanisms for public feedback, and project papyrologists at all institutions will keep systematic records of observations about image usefulness and quality. The group expects that technical advances will affect current recommendations, especially for those objects least satisfactorily captured by standard color digital imaging.

3. Results of the study

(a) Physical attributes to be captured. Two points served as the essential grounding of the group's thinking. The first is that no single level of image resolution and dynamic range will be appropriate for every application; these levels will vary according to the objectives of a particular application. The second is that multiple levels of quality can have different roles in a particular application.

In the case of papyri, it is clear that legibility of the writing on papyrus is the most important objective in imaging. Legibility includes clarity of resolution when the image is blown up considerably larger than life-size. Images should not begin to dissolve into individually visible pixels at any magnification likely to be encountered in research use. It is not possible to give a simple measure for this characteristic, but it includes magnification sufficient to allow the viewer to see the ancient pen-strokes, even small dots of ink, and all corrections with complete clarity. In the test during the meeting, a small papyrus (about 9 x 15 cm) scanned at 600 dpi was blown up to fill a standard color monitor, then magnified to twice that size — to about eight times original size. Image integrity at that point was used as the rule of thumb for adequate resolution.

In the art world, by contrast, other features occupy a central role in assessing image quality; these include highlights, shadows, depth, color overlay, and translucency. These features are all desirable in working with papyri, but are of secondary importance compared to writing legibility. In fact, at times they even may be at odds with legibility. For example, a higher contrast between writing and background material than that offered by the original would usually be seen as desirable by papyrologists, even at the cost of fidelity to the actual appearance of the artifact.

Apart from the readability of letters, however, other features of papyri that affect capture are: fiber structure, joins of sheets, sharpness of margins, remains of gesso or similar substances, seals, erasures, folds, color, and ink color. Capture of some of these features sometimes may be at odds with the capture of maximum contrast between the medium and the writing.

It is also desirable to be able to see large papyri in their entirety, not only as collections of images of segments. With current technology, a level of resolution chosen to meet legibility objectives may not allow the capture of an entire large papyrus in a single image. One must therefore use different levels of resolution for some images. In short, archival capture of a papyrus in digital imagery may sometimes require multiple images with differing characteristics.

(b) Archival images and delivered images. It was agreed that archival images should seek to capture a level of quality that provides the maximum degree of satisfaction of scholarly requirements. That does not mean, however, that the same level of quality would actually be delivered to the user under all circumstances. Ester's formulation is useful: "Delivery quality places the premium on satisfying the needs and constraints of specific applications. Archival quality lays emphasis on the investment for initial image capture and the long-term value of images. Looked at as alternatives, these contrasting perspectives exist in obvious tension. Both sets of interests can be addressed without inherent contradiction, however, provided that archival quality determines the quality of scanning and that archival images become the reservoir of quality that is reduced and modified to suit the requirements of delivery quality."[4] The group anticipated rapid changes in the ability of the Internet to transmit large images, of screens to display them, and of printers to produce hard copies of them. But these are for our purposes secondary to the question of the required archival level.

(c) Methods of capture. The results of the discussions and viewings were clear: digital cameras[5] produce the best results for most purposes. New models of digital cameras are coming out with great frequency. It does not seem useful to specify particular models or vendors, as there are many sources for cameras capable of achieving the standards set out in the section below. But there was agreement that standard color imaging was for most papyri the preferred method. Such images, because they are susceptible to manipulation with a variety of computer tools, are far more useful than the traditional black-and-white photographs used in papyrology or even the color photographs sometimes found.

Most papyri are relatively small, at least as they have survived to the present, and there is no need to create photographic versions; digital images created with current equipment will meet the group's specifications with no difficulty.

Dgitial images are as useful for large papyri, but there are simply more problems in working with them. Digital cameras have significant limitations, since current models do not make it possible to meet the standards below when capturing the entirety of papyri larger than (roughly) a standard sheet of writing paper. It is likely that advances in imaging technology will enable us to deal with such larger items. For larger papyri, the experience of the art projects and the Columbia map project came into play: Color microfiche using the entire fiche for a single map image yielded outstanding results to the digital camera. The group therefore recommended that for materials over roughly 7 x 10 inches (17.5 x 25 cm), electronic imaging[6] be supplemented with 4 x 5 inch color photographic transparencies. Experience in reproducing paintings and other fine arts has shown that these transparencies are sufficient for digital imagery as its capabilities improve.[7]

Further, some papyri and other objects have special problems of legibility that standard color imaging may not be able to address based on the criteria set out above. These objects include pieces with overlays of papyrus or of gesso plaster, palimpsests, ostraca with poor ink-background contrast, and objects where dark medium can barely be distinguished from dark ink. For these challenges, there is promise from multispectral imaging (MSI), a technology pioneered by Bearman and Spiro. Developed for remote sensing in the U.S. space program, MSI is now in use in several other scientific fields, including oceanography, geology, and environmental studies. It "relies on the unique spectral signature of different parts of the target image (... e.g., the ink versus the writing surface). MSI acquires images simultaneously in many narrow, contiguous spectral bands over a spectral range."[8] MSI is thus capable of picking up very slight differences between ink and medium and enhancing contrast and legibility dramatically. It is particularly useful in the far infrared spectrum but could also be used for the ultraviolet, where some late antique inks can usefully be captured.

MSI is a developing technology. The cameras that use it are not interchangeable with those supporting standard color imaging, and the files it produces are very much larger than ordinary image files. MSI may have considerable use for the categories of objects mentioned above, but all agreed that it had no advantages for objects where such factors do not come into play. MSI therefore will be an important tool for only limited classes of material.

There will probably be other specialized tools, and projects creating electronic images of papyri need to monitor new developments to see what new technology might solve remaining problems. An example is a project underway in Oxford to deal with the special problems of incised texts (stylus tablets, lead tablets, and the like), using special lighting and image-processing software that allows the user to combine images taken with different lighting to eliminate much of the background and to differentiate the strokes from the background much more clearly.[9]

(d) Technical standards and specifications. We recommend the following basic technical specifications for digital images of papyri and similar objects. These are intentionally chosen to include industry standard practices and to avoid experimental, proprietary, or rarely found technological characteristics.

1. Full continuous-scale color; color calibrated with software;
2. Minimum resolution of 600 dpi for primary archival images; some images at 300 dpi may be useful for recording large pieces, and some at higher resolutions where the density of information is unusually high;
3. Image files in TIFF format, 24 bits (8 bits per channel), conforming to the TIFF Specification Revision 6.0;[10]
4. Archival files stored without compression; files using ISO-standard JPEG (1.02) File Interchange Format compression to be provided over the Internet;[11]
5. Inclusion of standard textual identifying material in images (ruler, color scale, digital watermark);
6. Provision of management data in separate data base files (identity codes, method of capture, resolution, equipment used, color specifications, operator, date, any compression, original format, any film intermediary, digital version, quality, consistency);[12]
7. System design to minimize damage to originals from heat and light during imaging.

An important issue for archival files is image integrity; that is, the ability of the user to tell if the file has been altered, either accidentally or deliberately. This subject has recently been discussed by Peter Graham, and we have adopted his conclusion that electronic files (in this case, image files) should bear digital time stamps (DTS) that reveal any alterations.[13] The Dead Sea Scrolls project has been sensitive to this issue, given the controversial history of the scrolls over the past half-century, and it has also decided to adopt DTS technology to guarantee its images. All archival files of images should adopt some protection plan. One public domain program available for this purpose is Pretty Good Protection, PGP. Another type of image protection is the secure visual identification of the image as coming from a particular collection. Such protection allows a proprietor of intellectual property the ability to demonstrate the origin of an image and thus protect the property rights in it. At the same time, it serves as an authentication of the image. A digital watermark method has been developed by IBM as part of the Vatican library manuscript project, and such a watermark (produced by altering the brightness of pixels to produce the seal of the library) appears on each image from that project made available over the Internet. IBM plans to make this technology widely available.

(e) Quality control, migration and refreshment issues. Archival images deserve that term only if they are both checked for quality after they are created and have a high likelihood of surviving future generations of technology. The group recommends that both human and mechanical methods of quality control be built into any project to capture papyrological images electronically. The human control will necessarily include a scrutiny of the image by a trained papyrologist to determine if the functional resolution standards described above have been met and if the document is otherwise in order. It will also require a check of the image against the management data record to be certain that they are correctly linked. The mechanical control will include a machine check of color using embedded color bar targets. It was recommended that a Q60 target be included at the start and end of batches of images.

The protection of the image data from destruction or obsolescence cannot be provided once and for all by any single set of measures. Against data loss, the group recommends that multiple copies of the data be created. In the case of APIS, this will mean providing each of the six partner institutions with a complete set. The usefulness of this approach will, naturally, depend on the longevity of the APIS consortium and its continued unity on matters of standards. Questions of migration are less clear due to the inherent unpredictability of technological change. During the active phase of the APIS project data, will be updated frequently. In the longer term, data might best be protected from obsolescence through inclusion in a larger body of digital library material managed by universities, rather than managed separately.

(f) Practicalities. This study has not concerned itself with costs and operating concerns for three reasons:

(1) Even with APIS, and even more when one moves to a larger universe of users of this technology, it is likely that different approaches to implementation of digital imaging will be taken by different institutions. Our concern is with ensuring compatibility of the results of these implementations.

(2) The technology and its costs are changing rapidly, and any details will be obsolete within a short time.

(3) The bulk of the costs of any extensive digital imaging project are likely to be comparable to those of microfilming projects, because most of the cost lies in staff time and the actual operations performed are very similar; operating a digital camera is, if not quite as routine as operating a microfilm camera, only modestly more complicated.

Endnotes

[1] APIS was planned to take advantage of recent developments in digital technology to integrate images, texts, cataloging, and bibliographical resources in a single system for research and education. The six U.S. partners in the first phase of its development are Columbia University, Duke University, Princeton University, the University of California at Berkeley, the University of Michigan, and Yale University. In addition, the Université Libre de Bruxelles is cooperating in the bibliographic part of the project in Phase 1.

[2] Cf., e.g., Y. L. Yao and F. C. Mintzer, "A TDI-CCD Colorimetric Scanner and its Applications," IBM Research Report RC 17554 (#76393), 10/11/91. Michael Ester noted during the Ann Arbor meeting that flatbed scanners have been thoroughly unsatisfactory in reproducing art works.

[3] The project on Frank Lloyd Wright's drawings on which Ester reported similarly used panels of scholars to rate resolution and dynamic range.

[4] "Image quality and viewer perception," Leonardo 23 (1990) suppl., 52.

[5] This term is used here to include both products described as scanners in camera form (like the Pro/3000 scanner developed by IBM Research) and products described as digital cameras (like the Kontron).

[6] This is the largest size that can be captured at 600 dpi using a digital camera with a 4000 x 6000 pixel frame, the largest reasonably available frame at present.

[7] The preservation experts on our panel noted that there are no archival standards for color negatives, and that color is not stable in the long term. They agreed, however, that because of the rapid development of imaging technology even a relatively short time horizon for such color negatives could be helpful in preserving information.

[8] G. Bearman, B. Zuckerman, K. Zuckerman, J. Chiu, "Multi-spectral digital imaging of Dead Sea Scrolls and other ancient documents," Abstract of paper at the 1993 Annual Meeting of the American Academy of Religion and Society of Biblical Literature.

[9] Image-enhancement of ancient documents. Collaborators: Alan K. Bowman (Christ Church), Michael Brady (Engineering Science), R.S.O. Tomlin (Wolfson College) and Andrew Zisserman (Engineering Science), all at the University of Oxford.

[10] It was the group's consensus that using more than 8 bits per channel offered no significant gain in accuracy for papyri, where color is secondary.

[11] We note that the IBM/Vatican library manuscript project makes only compressed and watermarked files available over the Internet, keeping all archival uncompressed files on systems not accessible to the Internet in order to enhance protection of the archival files.

[12] The management data system developed by Luna Imaging, Inc. has been a useful guide.

[13] Peter S. Graham, "Intellectual Preservation: Electronic Preservation of the Third Kind," Commission on Preservation and Access paper (March 1994).

Conference on Digital Imaging of Papyri

University of Michigan, March 3-5, 1994

List of Participants

Roger S. Bagnall, Columbia University (papyrologist; President, American Society of Papyrologists)

David Barber, University of Michigan (technology librarian)

Gregory H. Bearman, Jet Propulsion Laboratory, Caltech (physicist; MSI technology)

Anthony Bliss, University of California at Berkeley (rare books curator)

Robert Cartolano, Columbia University (academic information systems specialist)

Paul Conway, Yale University (library preservation officer)

Peggy Daub, University of Michigan (librarian, special collections and arts)

Michael Ester, Luna Imaging, Inc. (imaging of art works)

James Fox, University of Michigan (special collections curator)

Traianos Gagos, University of Michigan (papyrologist)

Janet Gertz, Columbia University (library preservation officer)

Brenda Johnson, University of Michigan (librarian)

Ludwig Koenen, University of Michigan (papyrologist)

Paul Mangiafico, Duke University (library technical services)

Frederick C. Mintzer, IBM Thomas J. Watson Research Center (imaging scientist; IBM/Vatican library manuscript project)

John F. Oates, Duke University (papyrologist)

John Price-Wilkin, University of Michigan (Humanities Text Initiative librarian)

Don C. Skemer, Princeton University (curator of manuscripts)

Sheila Spiro, Claremont School of Theology (Dead Sea Scrolls manuscript scholar)

Peter van Minnen, Duke University (papyrologist)

Jack von Euw, University of California at Berkeley (library technical services)