The Advanced Papyrological Information System

Rodney Ast, Stephen P. Davis

APIS is a collections-based databank containing images and metadata pertaining to papyrological objects primarily from ancient Egypt. These objects include papyri, ostraca, various kinds of tablets (bronze, lead, wood and wax), paper, some inscriptions, and a small number of fragments of parchment. Texts date from ca. 2000 B.C.E. to the 9th century C.E. and are inscribed in Hieroglyphic, Hieratic, Demotic, Greek, Latin, Aramaic, Coptic, Arabic and other languages. There are religious and literary texts, horoscopes, private letters, court proceedings, petitions, wills, birth and death certificates, household accounts and many other kinds of documents. Virtually no aspect of written culture is neglected, and the result is a wealth of evidence for ancient life and custom. The project currently hosts twenty collections from places in North America, Europe, and the Middle East. There are about 25,000 records, 15,000 images and 4,000 translations. Among them, one finds, for example, a famous codex from the 2nd or 3rd century containing the Letters of Paul, which is housed at the University of Michigan, a magical love charm in Oslo and an ostracon from the Roman military camp at Berenike. Objects can be found via both simple and advanced searches and by different browsing options.

The database serves both as a catalog and as a scholarly and pedagogical tool, and the range and extent of information can vary considerably. Some of the larger participating institutions employ full-time papyrologists who are capable of providing detailed, expert information about their collections and

can update their records regularly. Most of the larger places also maintain their own local databases and web sites and use APIS simply as a means of reaching an even larger audience. Other, smaller institutions have no local sites, and for them APIS serves as the sole catalog of their collections and their only exposure to the outside world. Involvement in APIS often gives these places the unique opportunity to discover what it is they have in their collections. After all, many places inherited their texts in the early part of the twentieth century from the Egyptian Exploration Society (EES), which organized and led archeological excavations in Egypt. The EES then distributed their archeological finds to the numerous museums and libraries around the world that supported its excavations. For decades, the objects lay in these places awaiting the kind of attention that APIS can now help institutions give.

APIS contains records of both published and unpublished material. Published texts generally receive much more extensive treatment, being accompanied by extensive scholarly information, including bibliography and parallel texts, as well as an English translation and full set of images that can be viewed in both low and high resolution. A typical record of a published papyrus includes a title, summary, inventory id, physical description, notes, translation, acquisition and custodial information, as well as an image link.

The project is primarily object-oriented and access to high quality images is central, thus we try to offer images to as many published pieces as possible. Participating institutions typically produce 600 ppi color archival images in the TIFF format, and from these lower resolution JPEG derivatives are served up for display on the internet. As part of a new initiative, APIS has taken the first steps towards creating a repository of archival TIFF images, in order to ensure the long-term preservation of the project's digital content and to allow future experimentation in innovative image display techniques. As for images of unpublished material, we leave it to individual participants to decide whether to display legible images or to withhold them from public view. This is one way in which participating institutions exercise control over

the dissemination of information pertaining to their holdings.

The Technical Side of APIS

APIS operates on a distributed contribution model, where each new partner is responsible for preparing they own cataloging and contributing images from their collections, following guidelines established by the APIS directors. Columbia University Libraries has been the central technology host for APIS since the project began in 1997.

In 1998, project staff developed an APIS "contribution format" consisting of approximately 50 data elements. The format is based on MARC but has a number of extensions needed to accommodate papyrological cataloguing practices. These extensions consist chiefly of more specific note types to allow for the separate display, sorting and retrieval of specialized papyrological descriptors. For example, the APIS 510_dd element for encoding DDBDP citations is based on and backwardly compatible with the MARC 510 (Citation / References Note) field.

The APIS contribution format also accommodates structural metadata and links to locally or centrally hosted image derivatives. It also provides the option to include a "linkback" URL to the institution's corresponding locally-mounted version of the catalog record.

Two PC-based cataloging applications were developed early on in the project at Berkeley and Michigan (written in MS Access and Filemaker Pro, respectively). By 2000, these programs had been modified to export cataloging in the APIS contribution format and were being made available to new APIS partners as data collection tools. APIS partners also may use whichever local cataloging systems they wish so long as they can export their records to the central database in the specified APIS contribution format. Before the end of 2006 an XML-based version of the contribution format will also be made available.

Upon receipt by the central technology host, new and updated APIS records

are validated, converted into an internal version of the APIS data format and merged into the existing data file. For purposes of simplicity and quality control APIS uses a full-file replacement scheme rather than record-based addition and deletion.

Until recently APIS data was loaded into, managed in and published from an SQL database (DB2 / Solaris) that had been developed at Columbia in the late 1990s as a "master metadata file" for digital projects. In 2006 APIS was migrated to an entirely XML / Lucene / Java-based platform to provide more flexibility, improved response time and a richer environment for future development.

The APIS search system provides extremely fast and efficient searching of both normalized elements in the record -- such as dates and identifying numbers -- and text. Keyword / Boolean searches may be restricted to certain elements or executed over all elements in the record. Searches may include or exclude translations. Browsable, alphabetical lists by subject, physical format, genre and language are also generated as static pages and made available. In addition to standard HTML displays, APIS records may now also be accessed as XML documents for deep-sharing of content with other systems.

Future areas for development include: creating a centrally hosted image store to allow for more effective image management and more innovative image display options; developing different aggregate "views" of the APIS data; functional integration of APIS data with other systems such as the forthcoming Papyrological Workbench ("APIS Plus"); additional modes of record contribution such as direct input into the central system and automatic record harvesting; implementation of a handle-based identification system to provide "permanent URLs" for APIS data. Planning efforts are already underway to accommodate the long-term digital archiving of cataloging and images.

The Future of APIS

The field of papyrology is fortunate to be supported by several digital projects that complement each other in important ways. In addition to APIS, there exist, for example, the Duke Data Bank of Documentary Papyri (DDBDP) and the Heidelberger Gesamtverzeichnis (HGV). The former is a database offering searchable texts found on papyrological objects. The project focuses strictly on published Greek and Latin documents such as wills, legal proceedings, tax documents, leases, imperial edicts, private letters, etc., texts which reflect the administrative procedures and daily life of Greco-Roman Egypt. Like the Thesaurus Linguae Graecae (TLG), the DDBDP allows users to enter a search term or set of terms and returns results for papyri containing the desired character string. One can then view a text in its entirety, which is very useful when the published edition is not readily available.

While the DDBDP deals with the actual texts, HGV concerns itself with information pertaining to the text. It tracks changing scholarly opinion about published Greek and Latin documents, offering online the same basic information that is provided in the printed edition: the publication number, title, provenance, and information about images. HGV is also the primary authority for the dates of Greek and Latin documents, which often are very precisely dated, down to the exact month, day, and year. The editors of HGV verify the dates that scholars propose and suggest improvements to those texts which have been incorrectly or imprecisely dated. The project has the advantage over printed editions of allowing scholars' improvements to be documented almost as soon as they are voiced.

APIS, DDBDP and HGV have, each in its own way, dramatically enhanced conditions for papyrological research. In order to maximize the benefits of all three initiatives, APIS has drafted plans to integrate the three into a single search and retrieval system that will allow users to search across the projects and display results from all three within a single interface. To realize

this integrated system, we have undertaken a pilot project, tentatively called APIS Plus or the Papyrological Workbench, whereby we will take each of the three existing databases and create an interface in Java. Within the single interface, users will be able to search for and view text (from the DDBDP), image (from APIS) and scholarly and/or collection-related information (from APIS and HGV) pertaining to any given object. The advantage will be not only a more efficient system that eliminates the need to deal individually with each project, but also the potential to present text, image and metadata in new and interesting ways. For example, a user will be able to view the Greek text of a papyrus alongside an image of the object, or will be able to call up images and texts of related objects and view them simultaneously.

The technical challenges to creating an integrated system are various. One critical component is the development of a set of unique numerical identifiers that will permit easy identification of a single object as it appears in each project. Much progress has already been made on this front. Colleagues at Duke and Heidelberg have begun assigning numbers to objects represented in both the DDBDP and HGV, and once they have compiled these identifiers, they will share them with APIS. After the development of numerical identifiers has been completed, there can be significant progress towards the more complicated kinds of exchange that we are hoping to achieve. Ultimately, we would like to have a reliable system in place that will be flexible enough to accommodate any other projects that may wish to integrate their information.