

**A DOCUMENT IMAGING PROJECT
AT MOSS LANDING MARINE LABORATORIES'**

Joan Parker
Moss Landing Marine Labs
P.O. Box 450
Moss Landing, CA 95039
E-mail: parker@mlml.calstate.edu

The Institution

Moss Landing Marine Laboratories (MLML) is a teaching and research institution of the California State University System. Students from campuses located in San Jose, San Francisco, Hayward, Sacramento, Fresno, and Stanislaus apply to attend the graduate program at MLML. The Labs offer a comprehensive program of study in the marine and ocean sciences. Research interests of faculty and students range from topics in invertebrate zoology, ichthyology, benthic ecology, phycology, and marine birds and mammals to biological, chemical, geological and physical oceanography.

In 1989 the Loma Prieta earthquake destroyed the Laboratories, located in the town of Moss Landing. As a result, we have been dislocated in separate communities some twenty miles apart for six years. Temporary quarters in Salinas are the primary site for teaching, the administrative offices, the Library and most research labs. Some classroom and laboratory space remains in Moss Landing. Construction of the new facility in Moss Landing is scheduled to begin in 1996 with projected completion in two to three years.

Geographically, the Labs are centrally located in the Monterey Bay region. And, as might be expected, resources and problems unique to this area are a focus for research. But with the Pacific Ocean as our backyard, the scope of MLML's research interests are far more cosmopolitan in nature.

The Library

The Library that supports the dual mission of Moss Landing Marine Laboratories consists of approximately 9000 books and eighty-three current periodical subscriptions. Until 1995, monographs were cataloged by the Clark Library of San Jose State University (SJSU), added to their online catalog and ultimately to OCLC's cataloging database under SJSU's holdings symbol. However, MLML has no separate online catalog so the card catalog in the Library at the Salinas facility continues to be the primary access point for the MLML collection. Furthermore, most documents and reports are cataloged only at the local level.

Some of our most valuable and irreplaceable resources are the approximately three hundred reports and theses completed by our students. Access to this collection is complicated by the fact that graduate degrees are actually conferred by each home campus and thus our theses are not comprehensively included in the SJSU online catalog. Nonetheless, approximately thirty percent of this collection is in circulation at all times, primarily, but not exclusively to, students and scientists throughout the Monterey Bay region.

Why Document Imaging

Using the Internet to bring information out from behind the walls of an institution is by no means a new idea. Part of the value of this project lies in the ability of MLML to demonstrate that a small marine science institution can contribute tremendous amounts of information relative to its size. Document imaging provides a technology to complement this process.

One of the risks we take by opening up access to our collection is that demand from outside MLML will strain, and ultimately, break the system. For example, creating a bibliographic database of citations to our student theses and placing this on the Internet will force those who want to borrow a thesis to contact the Library. Instead of bringing the brunt of what could be a staggering workload to the Library, this system of providing access to image files of the full-text of these documents will allow users to access this information without the direct involvement of the Library.

Authors are legitimately concerned with the potential loss of intellectual property rights and copyright protection that might arise from placing files on the Internet. Storing documents containing data and information as image files limits the amount of manipulation possible.

The Methodology

Prior to beginning the project former students will be contacted for permission to include their work. The Lab maintains an active list of alumni addresses and many former students remain in the Monterey Bay region; therefore few problems are anticipated in the permissions process. And, because the total number of theses completed at MLML is under three hundred, the project is of a manageable size.

Each thesis and student report will be scanned as an image file, stored on a local server, and indexed using Wide Area Information Server (WAIS) software. Access will be read-only to protect copyright. Only those documents with written permission from the author will be included as full text. All others will be included in the index with pointers to a brief abstract. Implementing this project will involve several steps and a variety of hardware and software.

The Technology

First, each document will be page-scanned at a resolution that will provide good readability. The resulting files will then be compressed into a .GIF or JPEG image file format. The original scanned files will be archived on magnetic optical disks and the compressed files will be placed on 2 gigabyte hard disk drives that are attached to a file server. At a scanned resolution of 300 dpi, a 100 page thesis or report will occupy approximately 5 megabytes of digital storage in compressed form. Once the documents are stored in a digital format, they can be linked to a WAIS search gateway program that will be implemented on a web server.

The Moss Landing Marine Labs, though geographically separated, are connected via a frame-relay and fractional T1 Wide Area Network (WAN). This WAN is then connected through our main campus gateway to the California State University Network (CSUnet) and then into the Internet. This network accessibility will allow easy access to the imaged documents via a web browser such as Mosaic or Netscape.

The Cost

The critical funding elements for this project are personnel and hardware costs. Our captive labor population, graduate students, will be hired to scan the documents. With an anticipated 80% affirmative response rate, an average of 100 pages per thesis and a minimum scanning rate of three minutes per page the project will require nearly 1,500 hours of student time. At an hourly rate of \$9.00 plus fringe benefits, approximately \$20,000 is needed to complete the scanning portion of the project .

While the preferred scenario is to identify and win funding for this project, a lack of success will not terminate the project. As the Labs have scanning and optical storage technologies available the purchase of dedicated equipment is more critical for the time needed to completion than the entire life of the project. However, if funding were secured, a Sun Sparc file server (~\$12,000) and two 2-gigabyte hard drives plus an optical drive for archiving data (~\$6,000) would be purchased to facilitate access to the imaged documents.

The Future

A grant based on this project was put forward to a national funding agency and we are awaiting word of its success or failure. In the meantime, a few sample theses will be selected as a pilot project and the scanned images will be stored on the MLML web server. Providing better access to our invaluable collection of Moss Landing Marine Labs theses using document imaging technology is an important project and every attempt will be made to find a source for full funding.