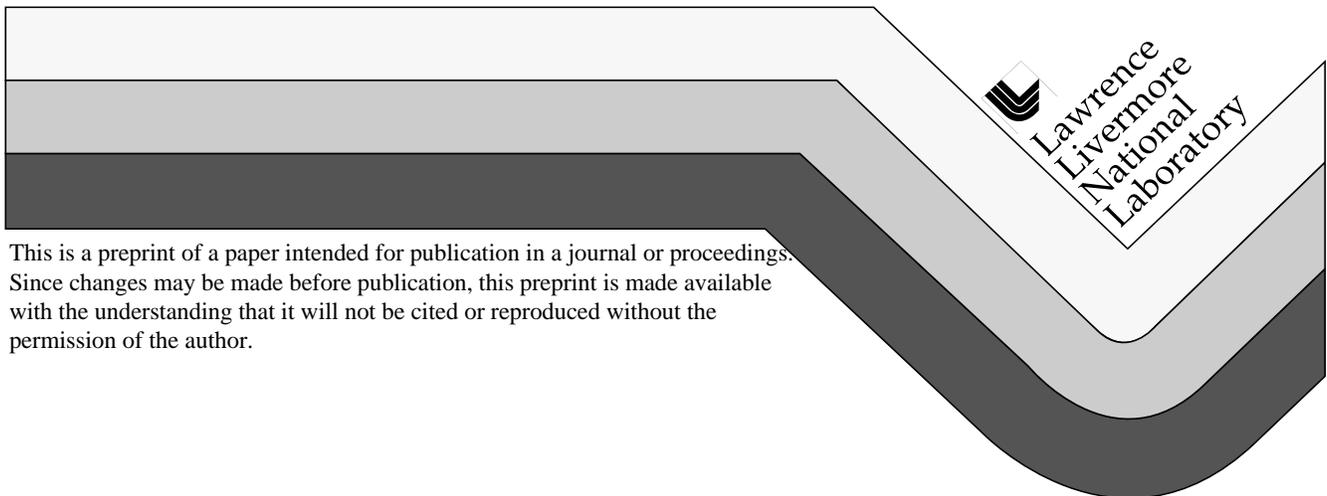


Lawrence Livermore National Laboratory's  
Library of the Future: The Future is Now

Isom Harrison

This paper was prepared for submittal to  
Inforum '98  
Oak Ridge, TN  
May 6-7, 1998

March 20, 1998



#### DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

# Lawrence Livermore National Laboratory's Library of the Future: The Future is Now

Isom Harrison, Library Division Manager  
Technical Information Department  
Lawrence Livermore National Laboratory

## ABSTRACT

The Library of the Future Project began in 1994 with the following objectives:

1. To develop the necessary infrastructure (e.g. network access, appropriate hardware, etc.) to support electronic transmission and processing from creation.
2. Official review and release through dissemination.
3. Storage of Laboratory-authored reports.

Authors submitted manuscripts in electronic format and all of the subsequent processing, including generation of the Library's catalog records and deposition of the full text on a Web-based server, would be done automatically. Currently, unclassified, unlimited reports that go through this process are available via the Internet. This paper will describe the steps taken to bring the Library of the Future into the present.

### Introduction:

We began the Library of the Future (LoF) project in 1994 in order to position ourselves to take advantage of the information superhighway. Also, we wanted to take advantage of the electronic information revolution to provide an environment where Laboratory scientists and engineers could access Library information from their desktops. In order to make this happen, we had to change the way we did business from a largely paper environment to an electronic environment. The plan was to:

- make unclassified, unlimited-distribution LLNL reports available on-line.
  - new reports first
  - old reports second

- make laboratory business guidelines available on-line.
- standardize cd-rom usage.
- create a unified interface to all library resources.
- establish collaborative links to organizations following a similar path.
- develop and retain a first class staff.

We had to develop the necessary infrastructure within the Library first of all, and throughout the Laboratory in order to encourage electronic transfer and processing of Laboratory authored reports. We also had to get buy-in from authors to send their documents to us in the desired electronic format (.pdf). Many authors were and are still reluctant to do the conversion prior to sending their documents to review and release. As an incentive, we requested they send us documents in the native application they were using (e.g. Word, WordPerfect, etc.) and we would convert the documents to the desired format.

#### The Process:

I want to devote this section to discussion of how we digitized our reports collection. However, I need to mention some of the initial steps taken to get the project going. We were faced with transitioning from a mostly paper-based environment. In order to make the transition, we had to establish some guidelines. We put together two teams (Information Architecture and the Electronic Document Input Team) to develop necessary business processes and guidelines. The Information Architecture Team was tasked with determining the conceptual architecture of the Digital Library Document System. They were to determine file structure, directory structures, naming conventions, and customer user interfaces. These structures and interfaces would be used to support the conversion of 77,000 unclassified, unlimited-distribution paper reports to electronic format. The Architecture Team also looked at the types of repositories needed for electronic document storage and user interface integration (e.g. WAIS, Gopher, Mosaic, WWW, etc.) for browsing and searching.

The Electronic Document Input Team was responsible for developing an electronic review and release process (F1),

documenting conversion processes, establishing printing capabilities (print-on-demand), and determining the necessary interactions with Classifications and Patents offices.

After a few months, we found the teams needed input from each other to accomplish their work. As a result, we decided to combine them into one team.

### Document Conversion:

We determined we had 77,000 documents eligible for conversion. We purchased three scanning workstations to begin the project. Initially, staff in the Circulation section would scan documents when they had some extra time. This created a disaster because there was no consistency from one person to the next. We then dedicated two full-time people to scanning in order to reduce the number of errors generated when several people were scanning. The scanners were to spend six hours per day preparing documents (e.g. destapling reports, checking light print, etc.) and scanning. The last two hours were spent running ScanFix (deskewing and despeckling software) on the reports scanned that day and checking scanned files against a log for file and page count before transferring the files to a non-local server for conversion. We archived reports in TIFF format but chose to put our full-text documents in .pdf format because .pdf retained the original look and feel of the document.

After scanned documents were sent to archiving, the scanners used a software package called Alchemy to convert documents to .pdf. The conversion process was done in batch mode overnight. The programmer had to write several scripts to ensure the process ran uninterrupted during the night. The TIFF images were retained and stored on magnetic tape while the converted .pdf files were stored on CDs. Eventually the .pdf files were removed from CDs and placed on a server where individuals would review the documents using Acrobat Exchange. The documents were reviewed for such things as missing pages, pages out of sequence, disclaimers, readability, and classified citations. If the reviewer found any problems with the document, it would be placed in a rejects folder with an electronic sticky note attached to the front page stating why the document was rejected. If the document passed inspection, it was placed in a

prerelease folder (F2) which was eventually released to the Internet or Intranet, based on whether it was intended for limited or unlimited distribution. This whole process was very labor-intensive and time-consuming. We needed a better method to reduce the number of steps and people involved in the process.

We tested Adobe's Capture software to see if it would meet our needs for creating TIFF image, converting them to .pdf, and reviewing documents all at the same workstation. After testing the software, we decided it met our needs, so we purchased it. Capture allowed us to create TIFF files and .pdf files simultaneously. This cut the conversion process in half. We were able to use the same people doing the scanning to create both TIFF and .pdf files, review the document, and release the documents to the server. Document release to the Internet or Intranet take place only after Publication Services has performed the review and release function to determine the classification and distribution limitations of a document (see F3). The final step for a document is our Documents On-Line database.

Documents On-Line is a full-text database of Laboratory-authored technical reports. The database is searchable by author, title, report number, document Accession Number (DOCACCN), keywords, conference, year, and subject category. Searching any of these fields leads to a page with report numbers, DOCACCN number, and titles. The title field also gives the size of the file, number of pages, and the option to view the report in .pdf format. If the user wants more bibliographic detail, he can click on the report title. A new page appears listing DOCACCN number, title, author, report number, corporate author, conference (if applicable), and year. The user also has the options to view the document or order printing. The print option tells the user the printing cost for a document. The user can decide if it's more cost effective to print the document on a local printer or have the Print Plant do it on their high speed printer. The minimum print cost for a document is \$15.00. The Print Plant option is the best option where large documents are being printed or several prints of the same document are requested or special paper or binding is necessary.

Currently the Documents On-Line database contains 13,000 reports on the Intranet and 9,950 reports on the Internet.

## On-Line Public Access Catalog (OPAC):

When we started the LoF project, the OPAC was a vital part of the infrastructure that needed serious attention. It was slow (access was via Telnet) and difficult to use. We had numerous complaints about its speed and structure. In order to do an exhaustive subject search, a user had to search three separate databases (books, journal, reports). Users had to log in, do their search in that database, log out and perform the same ritual for the next database. Our programmers were tasked with integrating the OPAC databases and increasing the speed of accessing and searching the OPAC. The result was an integrated database where the users were able to do one search that covered all the material types in the OPAC (books, journals, reports, videotapes). The OPAC was further refined when Information Dimensions, Inc. (IDI) developed a Web interface for Techlib. We integrated the Web interface with our existing OPAC. Users could now access the OPAC via Telnet and the WWW. Users can now search the entire OPAC by author, title, keyword, subject, corporate author/conference, call number, report number, year, and material type. Material type defaults to all types (e.g. article, book, CD-ROM, journal, report, tape, video). However, the user can restrict his search to any material type he/she chooses. When a search is conducted, the results screen gives the accession number, call number/report number, and title. When the user clicks on a particular title, the bibliographic data for that title is displayed. It lists the item location (hypertext links), title, alternate title (for journals), title information, language, document title, and frequency. Document On-Line and the OPAC now reside under the same Web interface which gives users access points for the full-text reports data from two different locations. It also eliminates the serious problem of multiple storage locations for any given record which required multiple edits and updates to insure data consistency. Reports can be downloaded, printed at the user's workstation, or in the case of large documents or multiple documents, can be sent to a high speed printer in the Print Plant.

## Desktop Access:

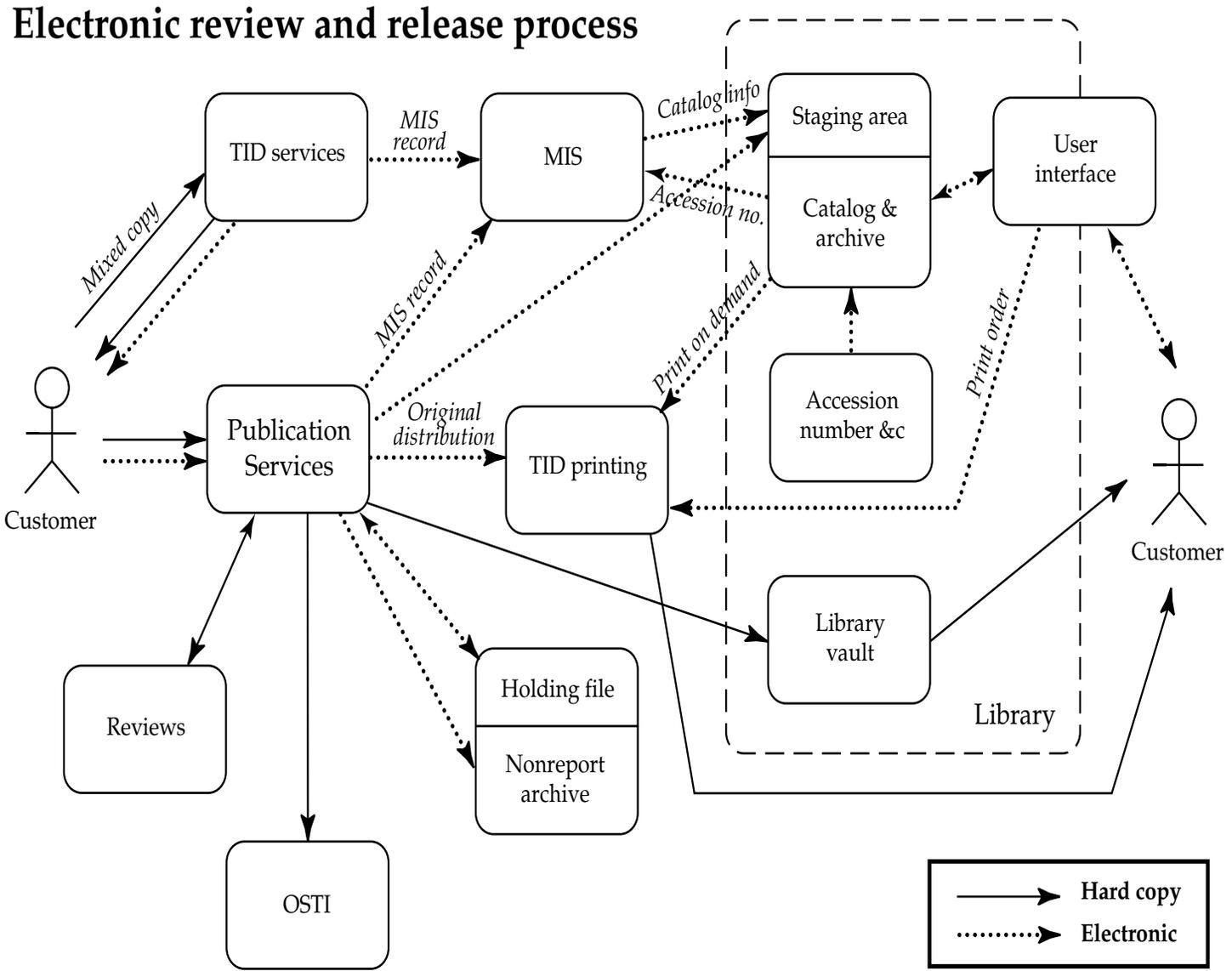
The Library of the Future (LoF) project has reached maturity and has been integrated into our daily business practices. Laboratory scientists can access the Library and most of its holdings from their desktops with the exception of stand-alone CD-ROM databases. Users can log on to the Library's Home Page to access on-line catalogs (OPAC, Documents On-Line, University of California MELVYL system), on-line forms for requesting Library products and services (photocopies, ILLs, books, journals, reports), electronic reference products (literature searches, journal table of contents, current awareness, full-text library bulletins). Users can also access the following electronic resources from their desktops:

- Site-licensed full-text resources (LLNL only)
- Journal tables of contents
- General interest electronic documents
- Health and Safety manual
- UC Contract 48
- USGPO Home Page
- Other national laboratories home pages
- CFR, CBD, DOE Orders, Federal Register
- Directory of Library Services

As you can see, "the future is now." The proliferation of information technology tools such as the WWW and sophisticated software packages have allowed us to accomplish most of our goals well before we expected. The review and release process can take place completely in an electronic environment (creation to storage, including dissemination to OSTI), and most scientists have access to the Library from their desktops. We will continue to add electronic information resources to our collection to assist the Laboratory's scientists with their research efforts.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

Figure 1



## Reviewing Process for Scanned PDF Files

All PDF files must pass through this review prior to being released on the Web Server. The flow of the PDF file is illustrated below.

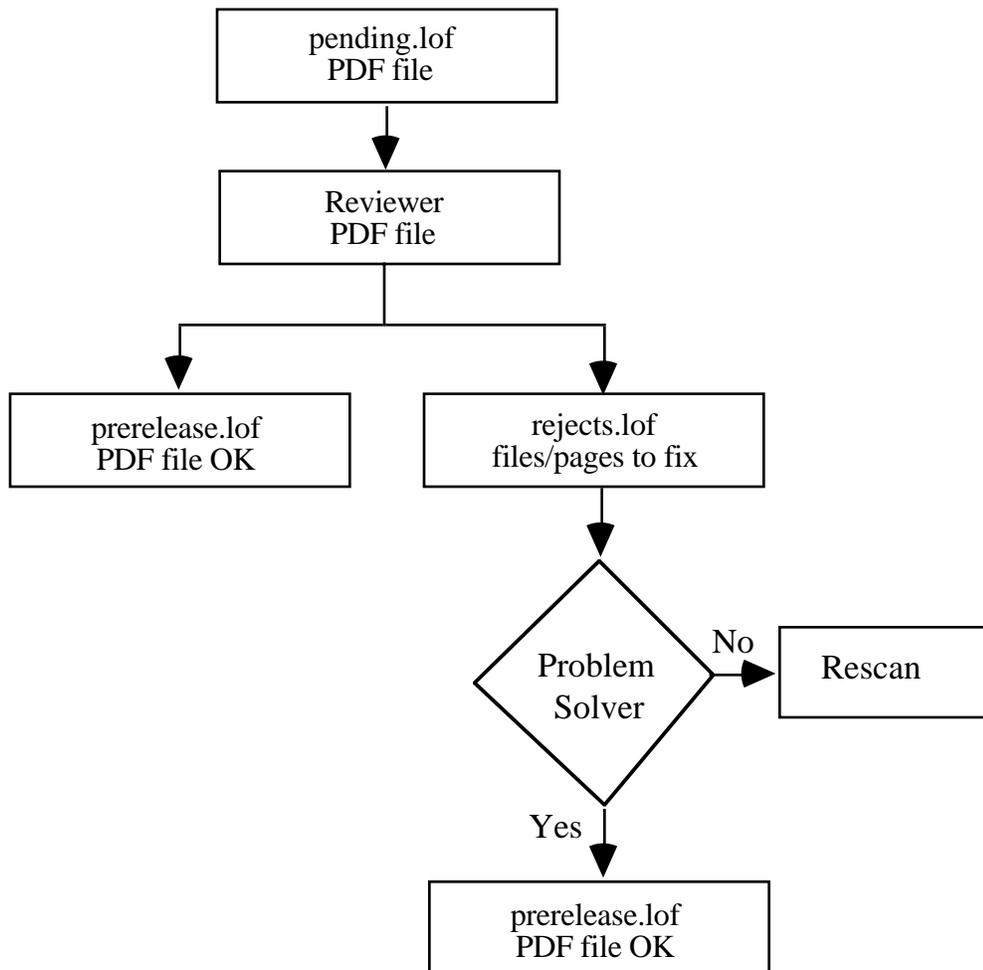


Figure 2

The workflow for internal & external updates .

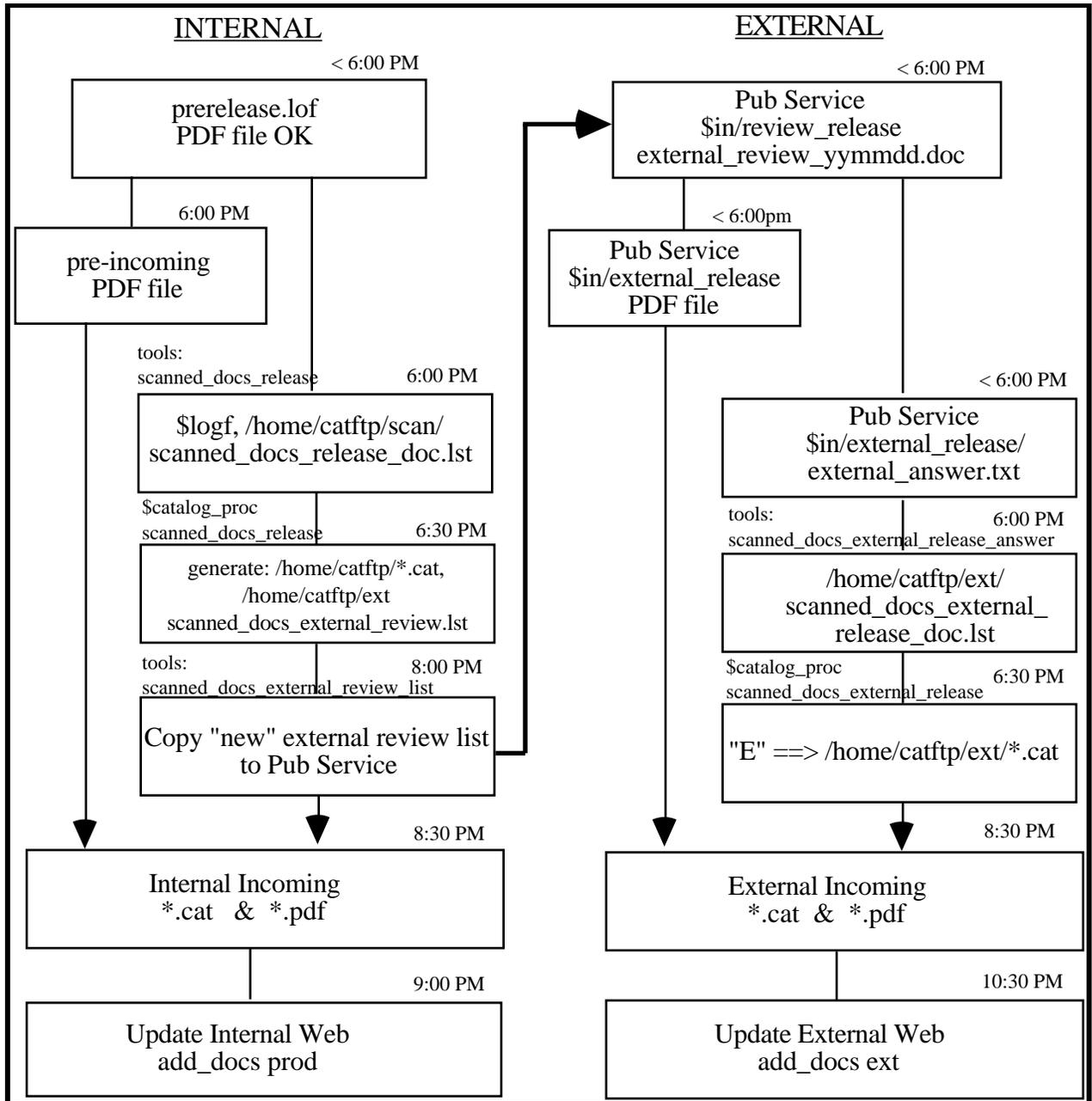


Figure 3

*Technical Information Department · Lawrence Livermore National Laboratory*  
University of California · Livermore, California 94551

