LA-UR-96-3680

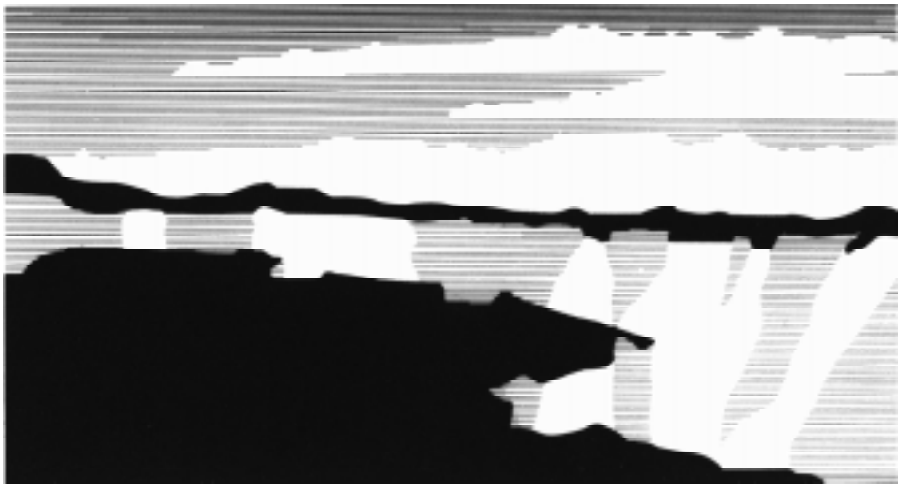
Title: **Providing Internet Access to Los Alamos National Laboratory Technical Reports:**

A Case History in Providing Public Access to Previously Restricted Documents

Author(s): Kenneth Alan Collins

Submitted to: **American Society for Information Science (ASIS)**

[for presentation at ASIS Mid-year Meeting, June 2-4, 1997, Scottsdale, Arizona]



Los Alamos
NATIONAL LABORATORY

Form No. 836 R5
ST 2629 10/91

Providing Internet Access to Los Alamos National Laboratory Technical Reports:

A case history in providing public access to previously restricted documents

Kenneth Alan Collins
LOS ALAMOS NATIONAL LABORATORY RESEARCH LIBRARY

Proposed Paper
for
ASIS Mid-year 1997
Scottsdale, Arizona

Abstract

The Los Alamos National Laboratory (LANL) Research Library recently fulfilled a strategic goal of providing worldwide desktop access via the Internet to full-image files of the complete unclassified holdings of Los Alamos technical reports in its Report Collection. This effort began in late 1994 with the scanning of paper and microfiche format reports. Concurrently, the Research Library helped to initiate shifting the model for publishing new technical reports from paper to electronic; the files could then be directly mounted on the Research Library's Web server. Providing desktop access to these reports was instrumental in expediting the development of internal policies that would better define what documents, previously restricted to the general public, could be publicly released. Undoubtedly, the most significant category of such reports were previously classified reports that had been declassified, but had not gone through a further review for public release. Collaboration with LANL's Classification Group led to approval for public release of 97% of these reports. The LANL Research Library's Web site now offers unique and unprecedented access to the world of a huge body of technical reports never available before anywhere in any form. This paper discusses the issues and steps involved in this achievement.

Introduction

The Los Alamos National Laboratory (LANL) was founded as the Los Alamos Scientific Laboratory in 1943 with a very specific mission: to design and develop an atomic weapon before the Germans and to be able to use it to assist in ending World War II. Through the massive undertaking known as the Manhattan Project, intellectual and financial resources were marshaled to make this happen. The world only learned, on August 6, 1945, what the Manhattan Project was about, when an atomic bomb was dropped on Hiroshima, Japan.

After World War II the central mission of LANL evolved into developing nuclear weapons for deterrence during the Cold War. While "reducing the nuclear danger" remains LANL's core mission today, the Laboratory has diversified considerably in recent years into civilian research areas. Nevertheless, Los Alamos National Laboratory has amassed a unique knowledge of all aspects of nuclear weaponry. Much of this knowledge is critical to a broad range of basic sciences and a great deal of it lies in the technical research reports published by the Laboratory. Initially, almost all of these reports were classified, and many of them remain classified to this day, because of the information they contain about the most crucial elements of weapon design, testing, effects, evaluation, and disabling, information that could still be invaluable to rogue nations and organizations committed to nuclear terrorism.

Declassification: necessary but not sufficient

Over the decades, authorized derivative classifiers at LANL determined that many of its reports could be declassified. Initially declassification of a report was tantamount, in most instances, to its being unclassified and without further restriction on public access. All the same, in a paper-based world, these reports were essentially inaccessible, because the dissemination systems were not in place to make their existence known. They were not usually provided, upon declassification, to the National Technical Information Service (NTIS) or its predecessor agencies for announcement. Online catalogs and large citation databases accessible through local or wide-area networks, much less the Internet, were unknown.

In addition, given the aura of closely guarded secrets that was such a part of the ethos of LANL, there was little motivation to call attention to these reports.

But over time, U.S. law and U.S. Department of Energy (DOE) orders increasingly began viewing information in several different ways. Some information was considered to be "born unclassified," that is, inherently unclassified and never susceptible to restricted access. Other unclassified information could be restricted from open release because of certain "caveats" that might apply. As various caveats were created, some documents, previously publicly releasable, were later determined to fall under one of the caveats, and thus controls placed on who could access them.

Probably the most complex category of information was declassified information. With greater sophistication, it was realized that though some information might not fit into the existing definition of classified matter, it could nevertheless cause harm to the national security if made freely available. Over the decades, beginning in the 1970's, the determination that information was declassified was no longer sufficient within the DOE community for making a public release decision. A further review for public releasability was imposed. Although numerous Los Alamos reports had been declassified over the years,

they had, by and large, never undergone a further review to determine if they could be publicly released.

In effect, a wealth of technical information was locked up in these declassified reports, with no opportunity for scientists outside Los Alamos or the public at large to benefit from it. While the public release status of these declassified documents could not be predicted for certain, there was a strong suspicion on the part of those who were familiar with them that most of them would probably be determined to be publicly releasable, once reviewed.

The LANL Research Library's Report Collection

One component of the Los Alamos National Laboratory Research Library is the Report Collection, responsible for managing and providing patron access to an enormous and unique repository of technical reports. Amongst the approximately 1,400,000 reports, both classified and unclassified, held by the Collection, are a virtually unbroken run of Los Alamos "formal" reports, Los Alamos reports in many less formal series, and broad runs of reports from other Department of Energy laboratories and plants, from agencies of the Department of Defense (DoD), from DOE and DoD contractors, universities, and foreign governments and research institutes.

One of the unique features of the LANL Research Library, is that though it is a special library, it is physically accessible to the general public and provides service to the public, albeit limited compared to the service offered its primary customer base, those affiliated with LANL (e.g., employees, contractors, official visitors). Even though the Report Collection itself is physically inaccessible to individuals who do not have a current DOE security clearance, members of the public who come to the Research Library in person, can use unclassified, publicly releasable technical reports on the premises, though outside the Report Collection itself. In addition, the Library has served as a place of last resort for those who could not obtain Los Alamos reports from any other source. Thus, the Library has had a venerable tradition of making its holdings available to those who are not members of its primary community.

The Library Without Walls

In 1994 a new project at LANL was initiated, the Library Without Walls (LWW). This project was committed to research and development that would explore new approaches to providing access to the Research Library's resources electronically and from each customer's desktop. This would thus vastly expand the concept of access to a library's resources. Another key element was to add value to these resources in electronic form, to permit users to explore and take advantage of the information at the desktop in ways that would be impossible with traditional formats.

From the start, the Report Collection's archival holdings of Los Alamos technical reports was identified as an area where an LWW R&D focus could have an

impact. Los Alamos technical reports constitute the Laboratory's own research memory and because of its location in a secure vault facility, do not get optimum use. Yet a 1993 survey[1] of customers familiar with the holdings indicated they were an invaluable resource that could never be replaced. Furthermore, and very important, there was no copyright issues involved in providing electronic access to these reports. While the reports have an implicit copyright, it is a non-restrictive government copyright. The only request made of those who use the information is that they acknowledge the source.

Beginning in the fall of 1994, a project was developed under the Library Without Walls banner, to scan the entire holdings of unclassified Los Alamos formal technical reports with the aim of providing Internet access to them as full-image electronic documents, both for the Laboratory community and for anyone outside the Laboratory who wanted to be able to use them. "Formal" reports refers to reports with the report series prefix LA- . For research reports to be published in this series, they must receive an (internal) scientific review and undergo careful editing and preparation by professional technical editors and illustrators.

Scanning Process

The initial approach to scanning these reports was to scan from microfiche copy, when available. Equipment had been acquired on the premise that it could be programmed to scan from microfiche in an automated fashion. Given the number of page images that would need to be created (ultimately, in excess of 600,000), it was felt this offered the greatest hope for high productivity. In addition, scanning from microfiche meant no damage to the paper originals. Unfortunately, the equipment did not perform as expected, in part because of quality control problems with the microfiche format.

After less than 1,000 reports had been scanned from microfiche, the Library made the decision to scan from paper. Fortunately, the Report Collection's policy had been to retain four paper copies of all LA- series reports. While in some cases, there were fewer copies, this was not a common phenomenon. A formal process and checklist were developed for preparing reports for scanning, for tracking them through the scanning process, and for returning the reports to the Collection. Included in the tracking effort were notes on any unusual features of reports, or any problems that were identified. Special categories of reports were identified for which scanning was postponed, because they presented problems for which a solution was not, at that time, available. These included both physical format problems (e.g., reports containing oversize attachments) or potential security problems (usually locally-imposed restrictions whose current status could not be determined).

The project ultimately went on to scan all Los Alamos reports, regardless of report series, held in paper in the Report Collection

The scanned images were created as TIFF files, and a careful postprocessing effort was developed to ensure that the final product would be of the highest possible quality considering the condition of many of the source documents. The TIFF files were enhanced using Seaport software and then all page images from a single report bundled into a PDF "wrapper." The PDF file, representing a single report, was passed through Adobe Capture to provide textword search capability when the file was read using Adobe's Acrobat software.

While the archival scanning project was proceeding, a related effort was begun to work with authors and the technical editors to obtain new reports in original electronic file form, so that it would not be necessary to scan from paper. It requires fewer resources at lower cost than archival scanning, with resulting higher quality (including color images, when embedded in the report). The Laboratory is now well along in meeting its goal of moving from traditional publishing to electronic publishing. The Research Library is an integral part of this new process, so that reports are made available through its Web server within a short time of the final copy being prepared.

The public releasability issue

The initial scanning effort called for great attention to determining if a report was publicly releasable. It had to have been "born unclassified." This was usually determined by its availability from NTIS. However, in the course of preparing the entire run of Los Alamos formal reports from 1943 on, numerous markings were noted that had never been systematically examined before, much less documented. Since the aim of the project was to provide maximum access to unclassified Los Alamos technical reports, the Library wished to balance careful concern over security issues with not unnecessarily restricting legitimate access.

From the wide variety of issues that emerged out of both operational needs and the scanning effort, staff of the Report Collection met over a period of almost one year (from summer 1994 to summer 1995) with members of the Laboratory's Classification Group. The latter is responsible for reviewing all scientific and technical information generated at the Laboratory that is intended for dissemination beyond the authoring group. The Classification Group thus has expert knowledge of not only classification issues, but of the various categories of caveats that may cause unclassified information to be restricted or "controlled."

Those involved, both from the Research Library and the Classification Group quickly realized that there were some imposing obstacles. A few of the more significant ones were:

- many of the caveats and restrictions found on reports were local and undocumented;
- there was no longer a good institutional memory of what might have been the original motivation behind the local restrictions;

- some early classifications corresponded to what were now caveats on unclassified information;
- a crash government declassification effort in the first part of the 1970's had serious flaws, and new procedures needed to be worked out to review those results and determine whether a technical report could be declassified and made publicly releasable.

What emerged from this collaborative examination of problem reports and the issues they raised was a lengthy guidance memorandum from the Classification Group to the Report Collection that made it possible to unambiguously determine whether a report could be treated as publicly releasable or not. A paper presented at DOE's Office of Scientific and Technical Information provides considerable detail on the specific issues.[2]

As a result, many reports for which scanning had been postponed were determined to be publicly releasable and thus were added to the pool of reports that qualified for worldwide Internet access.

Public release review of declassified reports

Inspired by the resolution of what had begun as a diverse set of confusing information security questions, the Report Collection was emboldened to take on the single biggest subset of problem reports: those that had been declassified but never approved for public release.

The Classification Group was willing to commit scarce personnel resources to a special project to review all identified declassified Los Alamos reports with the aim of making a public release decision. Between August 1995 and August 1996, 1,280 declassified reports were reviewed and 1,247 (97.4%) approved for public release.

As this massive review was proceeding, the Report Collection developed a careful process to properly remark all those reports that were reviewed, regardless of the decision. For the reports that were determined to be publicly releasable, brief descriptive cataloging records were created for the Library's online catalog. While these records have their limitations (at this time, there is no personal author entry), they are at least identified by report number and title.

Unique issues in providing electronic access to declassified reports

Two issues emerged from the effort to provide electronic access to declassified reports that had been approved for public release.

First, the Library wished to identify these reports as declassified, publicly releasable documents in the catalog records. A careful review of the MARC fields led to the selection of the 540 field (*Terms Governing Use and Reproduction Field*) as the logical tag to place this information.

A sample 540 field, from an actual record, is structured as follows:

540 BB a Declassified
b LANL Classification Group
d Approved for public release: 20 Nov 1995

The record, in the OPAC, appears as follows:

TERMS OF USE :Declassified. Approved for public release: 20 Nov 1995

Once the format of the 540 field was developed for declassified, publicly releasable Los Alamos reports, formats were subsequently developed for declassified, publicly releasable reports from other agencies and for unclassified restricted reports for which the existing caveat was subsequently canceled. Figure 1 shows a sample bibliographic record display of a declassified Los Alamos technical report.

The second issue was to clearly identify each page of the electronic full-image document as being publicly releasable. When documents are declassified, the classification level (e.g., SECRET), stamped at the top and bottom of every page must be lined out. However, due to the vagaries of stamp inks, both in the classification stamp and the lining out, it was necessary to ensure that every page, in its electronic form, would be clearly indicated as unclassified. In addition, for paper documents, the stamp indicating public release of the document only appears on the title page. One can easily print a single page from the electronic file with no clear identifier that it is part of a declassified, publicly releasable report.

The solution developed, in conjunction with the Laboratory group that handled the quality control and postprocessing of the image files, was to electronically "stamp" every page, top and bottom, with the statement "Approved for public release." This statement thus always appears on the electronic image and on any print copies of every page in the file. Different approaches were tested in terms of font size, position of the "stamp" and whether it would let any text show through or would block out the text it overlaid (normally, there were adequate top and bottom margins but on some early reports this was not the case).

The electronic stamping solution was demonstrated to both the Classification Group and Information Security Group at the Laboratory and met their requirements to ensure there was no confusion about the status of the information. Figure 2 shows a sample page display illustrating the solution.

Accessing electronic reports

Public access to this unique body of information, as for all Los Alamos scanned unclassified technical reports is via the Library's online catalog (http://lib-www.lanl.gov/opac). As of November 1996, this catalog exists only in a character-based form, but a Web-based interface is expected to be available in the near future.

Once the bibliographic record for a Los Alamos technical report that exists in electronic form is retrieved, the holdings information identifies the electronic "copy" and provides, as its location, the electronic address or URL (found in the MARC 856 field). In the character-based system, one can "copy" the URL and "paste" it into the one's Web browser to open the location. This then launches the PDF file.

The value added to electronic reports appears in several ways:

1. Thumbnails that can be used both for navigation and to identify pages that contains figures, photographs, tables, etc.;
2. Bookmarks for the table of contents (when there is one) so that one can hyperlink to individual chapters within a report;
3. Textword searching of the entire report to locate key words or phrases that might be extremely difficult and tedious to locate in the hardcopy format.

To date, there is little hard data on how much demand there is for Web access to declassified Los Alamos reports. The Research Library publicizes their availability through the (electronic) Library newsletter (http://lib-www.lanl.gov/libinfo/news/news.htm), through our recently added Technical Report Resources page (http://lib-www.lanl.gov/infores/reports/reports.htm) and through several other places from its Web page (http://lib-www.lanl.gov).

Based on the historical experience of the Report Collection, we believe that once there is greater awareness of this resource, which is available nowhere else, interest and high use will follow. We are actively collecting data on usage of declassified reports identified through the online catalog, and are actively meeting with (internal) customers to learn more about how they use electronic technical reports. Over the next several years, we expect to learn a great deal more about the value and problems of these Web-accessible documents.

One early possible indicator is the dramatic increase of contacts from external customers requesting technical reports as a result of the information about Los Alamos reports made available through the Research Library's Web site. The Report Collection is now regularly receiving requests for assistance from users throughout the world, most of whom communicate via e-mail.

The Report Collection also sees making these reports publicly available through the Library Web site as very much in keeping with the Department of Energy's Openness Initiative, one component of which is to give very high priority to the declassification and public release of hitherto classified or otherwise restricted information.

Related issues

Some other interesting issues have emerged in the process of providing worldwide electronic access to Los Alamos technical reports, that relate to questions of who controls access to electronic documents and in what ways, if any, do such documents differ from their hardcopy siblings.

One question is author control versus institutional control in terms of decisions whether information is or is not to be disseminated via the Internet. A second involves copyright. This is largely uncharted territory and no generally agreed to guidelines have yet emerged.

At Los Alamos National Laboratory, the emerging guidelines for these "new frontier" issues are coming out of the Information Architecture project. The mission of the Information Architecture (IA) project is:

"The IA project is the Laboratory-wide effort to develop the framework for the selection, implementation, and support of computing and communications products and services."

"The overriding goals of the project are to promote

sharing of information
connectivity and interoperability among computer systems
improved employee productivity
high return on investment (ROI)
effective management of information assets
responsible use."

Out of the IA project (go to http://www.lanl.gov/projects/ia/hello.html for more information) have emerged a series of standards and guidelines covering the full spectrum of the project's sphere of activity. Extensive guidelines for publishing on the Web have been issued[3] but the focus has been, for now, on what authors cannot publish. If certain criteria are satisfied, then an author may publish on the Web. However, the Report Collection has encountered the flip side of this issue. All the criteria are met -- the information does not fall into official restricted categories (patentable, proprietary, compromise of national security) but the author does not want the report available through the Internet (even though it may be for sale from NTIS). For now, formal policy does not exist that addresses questions of institutional control versus personal author control of what appears on the Internet when none of the formal restriction categories apply. A search of the recent literature has not turned up any case histories of organizations that have faced and resolved this issue.

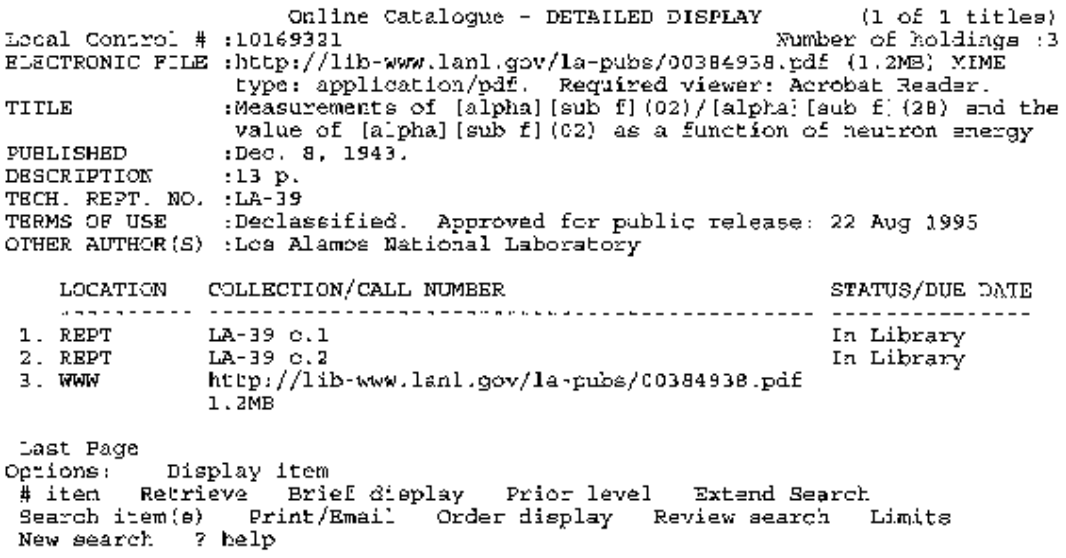
Another complex area is copyright. Many Los Alamos publications that are assigned a technical report number are actually not a technical report, strictly

speaking, but intended for journal publication or presentation at a conference. They are assigned a technical report number for the purpose of internal review and release. Often these reports are cited, either in writing or through personal communication, and as a result, the Laboratory receives requests to provide the reports. As a LANL report it is copyrighted, but it is a U.S. government non-exclusive (i.e., royalty-free) copyright. However, many journal publishers will not accept the paper for publication without the author agreeing to give the publisher an exclusive copyright. Furthermore, the author is typically prohibited from putting the same paper on the Web, regardless of whether it is identified only by the technical report number or by the report number with a cross reference to the journal citation[4]. While U.S. government-supported published research (which includes research performed by government-owned contractor-operated organizations) falls under the government's non-exclusive copyright and should permit publication on government Web sites, the reality is somewhat different. This remains, for now, a hazy area where the issues need to be worked out. Another of the other Department of Energy national laboratories, Lawrence Livermore National Laboratory, has begun to investigate how to resolve the copyright issue[5].

Conclusion

The Los Alamos National Laboratory Research Library's effort to provide desktop access to a broad range of information resources has resulted in, as one of its first concrete products, worldwide Internet access to the entire unclassified archive of Los Alamos technical reports in full-image form. The initial project to scan archival holdings from paper identified a number of potential information security problems that needed to be resolved, including determining the public releasability of reports that had been declassified, many of them decades earlier. Lacking the public release step, these reports, though declassified and containing invaluable research information, were essentially unavailable to the scientific community and the public at large.

By working through the issues with the LANL Classification Group, the Report Collection succeeded in making the vast majority of these declassified reports widely available in convenient electronic form to the world. The verdict is still out on how valuable this ultimately will prove to be, but preliminary indications are that these reports will get heavy use.



```
                    Online Catalogue - DETAILED DISPLAY        (1 of 1 titles)
Local Control # :10169321                                  Number of holdings :3
ELECTRONIC FILE :http://lib-www.lanl.gov/la-pubs/00384938.pdf (1.2MB) MIME
                 type: application/pdf.  Required viewer: Acrobat Reader.
TITLE           :Measurements of [alpha][sub f](02)/[alpha][sub f](28) and the
                 value of [alpha][sub f](02) as a function of neutron energy
PUBLISHED       :Dec. 8, 1943.
DESCRIPTION     :13 p.
TECH. REPT. NO. :LA-39
TERMS OF USE    :Declassified.  Approved for public release: 22 Aug 1995
OTHER AUTHOR(S) :Los Alamos National Laboratory

    LOCATION    COLLECTION/CALL NUMBER                          STATUS/DUE DATE
    ----------- ------------------------------------------------ ---------------
 1. REPT        LA-39 c.1                                       In Library
 2. REPT        LA-39 c.2                                       In Library
 3. WWW         http://lib-www.lanl.gov/la-pubs/00384938.pdf
                1.2MB

 Last Page
Options:    Display item
 # item   Retrieve   Brief display   Prior level   Extend Search
 Search item(s)   Print/Email   Order display   Review search   Limits
 New search   ? help
```

Figure 1: Sample bibliographic record display for a declassified, publicly releasable Los Alamos technical report. Note TERMS OF USE and LOCATION information.

Bath volatility can be reduced by employing fluorides or other salts of high boiling points, such as the fluorides of calcium, magnesium, barium and aluminum. Of these, the most desirable from the viewpoint of contamination are the salts of barium. Accordingly, the major effort was directed toward the development of a satisfactory barium-base electrolyte, although exploratory experiments were made with the others.

B. Experimental

The following table (Table I) summarizes the results of the electrolytic experiments. Current densities are necessarily approximate, and the voltages include line and contact drops. Temperatures were measured at the bath surface with a Leeds and Northrup optical pyrometer. Unless otherwise specified, the cathode was a water-cooled steel pipe 0.54 inches in diameter with closed end, and was quenched in kerosene containing solid carbon dioxide. The salts used were C.P. or reagent grade. The uranium deposited was almost invariably in a matrix of red-brown material and was recovered by crushing, sieving and washing.

Examination of the results indicates that while uranium may be electrodeposited from simple mixtures of sodium and uranium fluorides, the proportion of tetrafluoride is high and evaporation is excessive. The metal can also be successfully deposited from similar mixtures with calcium fluoride but there is a marked tendency for current fluctuation, and the tetrafluoride content was high in order to reduce the melting point. It is likely that lower UF_4 concentrations would result in more unsteady conditions. A simple mixture of barium and uranium fluorides behaved in similar fashion. However, mixtures of barium fluoride, barium chloride, uranium dioxide and uranium

[redacted]

Figure 2: Sample page of a scanned declassified technical report, showing the electronic stamping of each page with the phrase "APPROVED FOR PUBLIC RELEASE."

Footnotes/Bibliography

[1] Library Services Group, Los Alamos National Laboratory. *Classified Report Collection Study: Final Report* (Internal Document), August 1993.

[2] Kenneth Alan Collins, Los Alamos National Laboratory. LA-UR-96-690, *Public Releasability of Technical Information: The Experience of Los Alamos National Laboratory*, February 1996. (http://lib-www.lanl.gov/la-pubs/00285837.pdf)

[3] IA-6301: *Guidelines for Publishing on the Laboratory Internet/WWW*. (http://www.lanl.gov/projects/ia/stds/ia63011a.html)

[4] Private conversation with Carol Duncan, Publication Services Group Leader, Technical Information Department, Lawrence Livermore National Laboratory (LLNL), Livermore, CA, April 1996. In early 1996, Ms. Duncan conducted an informal survey of journal publishers. She queried them on what their policy is if the author of a paper submitted for publication in one of their journals also published the same paper (identified only as a LLNL pre-print) on the LLNL Web site. Responses fell into three main categories: (1) no pre- or post-journal dissemination permitted anywhere else, including on the Web; (2) dissemination on the Web permitted only prior to publication in the journal with removal required after publication in the journal; (3) no concern about the co-existence of the document on the Web.

[5] Private conversation with Carol Duncan, Publication Services Group Leader, Technical Information Department, Lawrence Livermore National Laboratory (LLNL), Livermore, CA, October 1996. Ms. Duncan followed up her initial survey effort (see [4] above) by contacting publishers again and taking the position that the U.S. government's non-exclusive, royalty-free copyright entitles it to disseminate its publications on the Web, provided that it cross references the pre-print (i.e., internal report) with the corresponding journal citation. Ms. Duncan is beginning to have some success with journal publishers. Some publishers with the most restrictive policy indicated that they would accept this approach.