ATOMIC SAFETY AND LICENSING APPEAL PANEL

11-17-88

Commissioner Curtiss, ✓

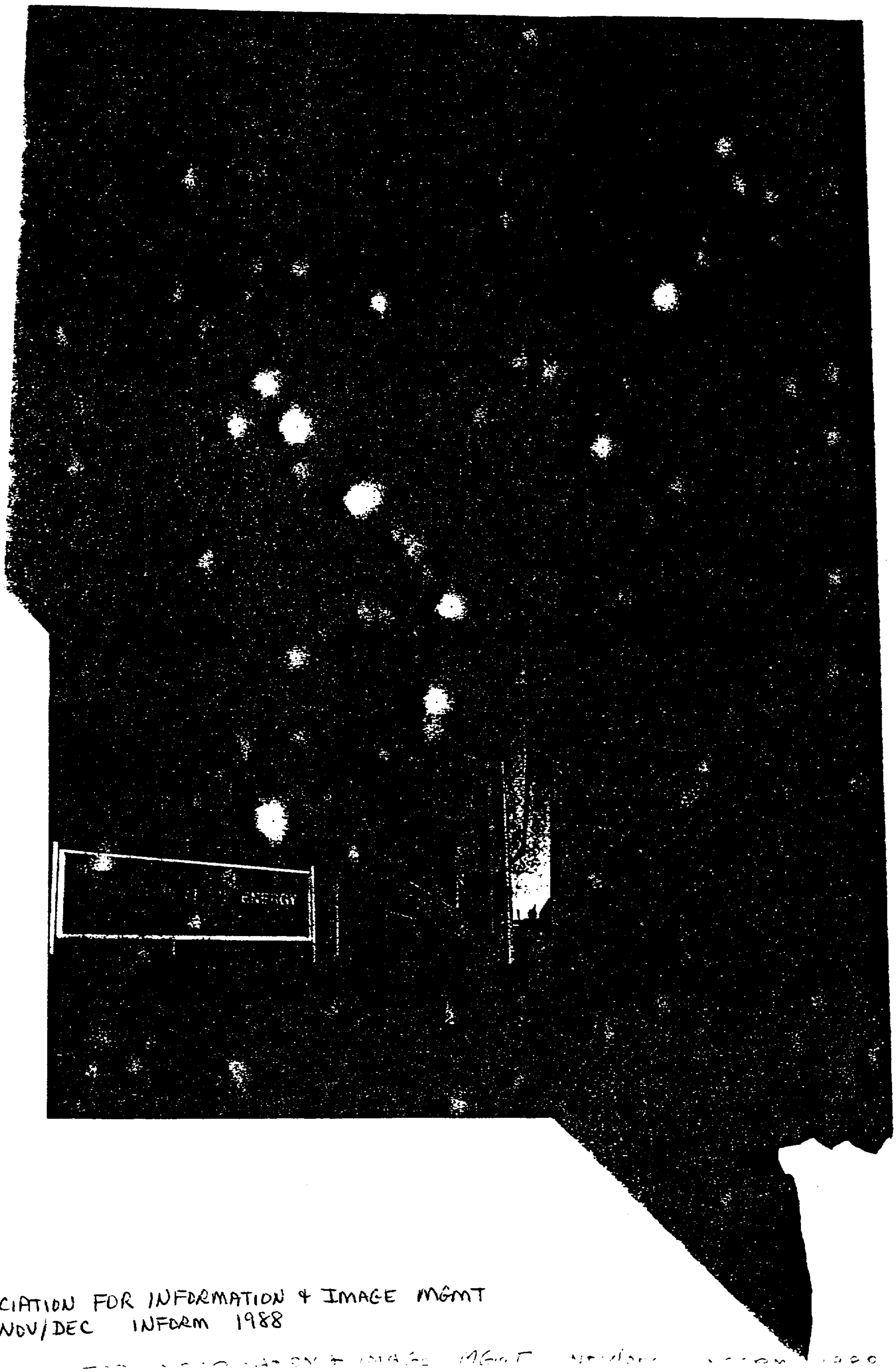
In light of our conversation on Nov. 14 where you indicated your interest in the LSS design, I have enclosed a magazine article I just received which provides a summary of the LSS. I thought that you may find it helpful in your evaluation of the proposed system.

Tom Scarbrough

T. Scarbrough

cc: Joe Gray
Steve Sohinky

RECEIVED
1988 NOV 17 PM 12: 08



ASSOCIATION FOR INFORMATION & IMAGE MGMT
NOV/DEC INFORM 1988

SYSTEM:

MORE WITH LSS

BY
GREGORY E. KAEBNICK

If plans formulated in the last few years come to fruition, the Department of Energy will convert a dry mesa known as Yucca Mountain into a new type of environmentally safe repository for storing highly radioactive waste. The plans will not be carried out easily, however. Construction of the site is contested by the State of Nevada and a range of environmental groups, Indian tribes, and other organizations, and if past experience is any indication, the opposition could be fierce: similar sites in two other states have already been dropped from consideration following intense political pressure and congressional action.

The plans could run into still other difficulties due to restrictions placed on the Nuclear Regulatory Commission (NRC), which is responsible for licensing the plans. According to the Nuclear Waste Policy Act (NWPA) of 1982, NRC is allowed only three years, with a possible one-year extension, to review a license application for the site. If opponents don't defeat the plans outright during the licensing, they may be able to string the licensing out long enough to achieve the same purpose.

Yucca Mountain vs. The Paper Mountain Shortly after the passage of the NWPA, NRC became convinced it couldn't possibly meet the constraints of a three-year license review if it conducted a traditional hearing process. Identifying and producing documents for the groups reviewing and protesting the license application was too overburdened and too slow. A typical case before the NRC licensing board might generate a record of 10,000 documents and take seven years to review; the Yucca Mountain case might generate 300 to 1,000 times as many documents and take even longer.

In 1985, NRC proposed a two-pronged solution:

- installation of a computerized information storage and retrieval system for handling the case documents, and
- negotiated rules, accepted by all parties to the licensing proceedings, that would define what documents would be entered into the system and how the system would be used and administered.

That solution is now taking shape. By 1991, one of the most ambitious optical disk-based systems yet conceived should be up and running. The system, known as the Licensing Support System (LSS), is of truly grand proportions: It will span the country, with chief nodes in Washington, D.C., Las Vegas, Nevada, and San Antonio, Texas, and access also possible in ten other cities. By the year 2009, 40,000,000 pages could be placed on the system, and the entire database will feature full-text indexing. The total budget for the system based on a ten-year life cycle is estimated at about $200 million. The project has been given presidential priority.

A $5½-million contract for design and implementation of the system was awarded in October 1987 to Science Application International Corporation, and a detailed preliminary analysis of LSS was completed this summer under that contract. A pilot system will soon be installed, and acceptance tests and loading of the database for the full-scale system should begin in 1990 and continue to 1995, when the licensing process begins.

Loaded with Options If the conclusions of the cost benefit analysis completed in July 1988 are accepted, LSS's features will include:

- a centralized search system and online optical disk image storage system, probably in Las Vegas;
- headers and full text in ASCII of all documents;
- bit-map images of all documents with reproduction at a central location and online printing at special local workstations;
- multiple document capture systems, allowing scanning, text conversion, correction, and cataloging;
- workstations capable of displaying headers and ASCII text;
- specialized workstations capable of displaying images;
- retrieval by means of indexes or through software for full-text searching;
- electronic mail; and
- microform archives for permanent storage.

This design is known as the "base conceptual design." Some variants to the design were considered as part of the evaluation process. These included:

- replication of the search and image system in both Washington, D.C., and Las Vegas;
- full-text searching with specialized hardware instead of with software programs;
- elimination of workstations capable of displaying images;
- microform digitization instead of optical disk storage of images;
- offline microform storage instead of the online bit-mapped image system; and
- rekeying text for ASCII storage instead of converting it from a scanned bit-map.

According to the cost benefit analysis, the initial design and its alternatives would vary little in cost. One alternative, the use of offline microform storage instead of optical disk storage, was found to be less expensive, but it was also found to be less useful. As has been found with other optical disk-based systems, more money was wrapped up in the document capture process than in any other element of the system. Document capture was predicted to consume approximately 62 percent of total costs, due primarily to the huge backlog of documents that has already been created by the program and that must be entered into LSS. Entering the backlog is extremely labor-intensive, as documents must be collected and prepared for scanning manually. Overall, labor would account for 70 percent of the total system costs. Only 16 percent of the total costs would go to hardware, and the rest would be spent on facilities, telecommunications, hardcopy production, and software.

The study found that the costs of document capture could not be significantly altered by increasing the rate at which documents were processed; the cost was primarily a function of the total number of documents. Increasing the rate from 18,000 to 20,000 pages per day would allow the backlog to be loaded one year earlier and would increase the costs only about $1 million. DOE plans to begin the document capture process in January 1990, at which time the backlog is expected to exceed 10 million pages, and hopes to have the backlog loaded into the system by October 1994, six months before the license application process is scheduled to begin.

Once the backlog has been loaded and the system installed and running, the document capture process should be considerably streamlined. Instead of being scanned into the system, most documents will be entered directly in digital form; they will either be transferred from a compatible digital storage format or pulled directly from word processing terminals.

Emergence of a Giant LSS is a case study of the maturation of the optical digital field. The system will owe much to its predecessors. Although it will push current optical disk technology to its limits, and although the optical disk-facilitated licensing process is itself unprecedented, much of the system will have been presaged in earlier programs and will grow naturally out of the results of earlier groundwork.

The chief step towards LSS will occur in 1989, when DOE installs a prototype to test the fundamental components of the system, especially the database. The prototype, itself an expansion of still other systems, will help to determine how the users will really use such a system and how the hardware and software components interface. Thus, definition of the final system functions will be the fruition of seeds planted years earlier.

The pilot will contain a sampling of the LSS's "core program information." The site characterization plan, upon which the entire licensing process is based, will be contained in its entirety on the pilot, plus all documents referenced in the plan, the administrative record of the licensing process, selected correspondence, and some samples of raw scientific data.

Even the pilot will be a sizable system. The site characterization plan alone runs 7,000 pages (and weighs 28 pounds in hardcopy). Adding the references extends this to 70,000 pages. As the search was conducted for program documentation that would create a meaningful set, the size of the pilot database grew from 100,000 to nearly 200,000 pages.

The primary ancestor of the LSS pilot is known as the Information Management System Bridge Program and includes equipment for capturing, displaying, and storing images; converting the images to ASCII text; indexing the text; and retrieving text and images. Equipment already installed and operating in the bridge program includes a PC-based scanning workstation built around a 400-dot-per-inch (dpi) Ricoh IS-400 scanner; a Palantir Compound Document Processor; a PC-based retrieval and display workstation featuring a 300-by-150-dpi LV-700 monitor and a Ricoh LP 3080 laser printer; and a LaserDrive 1200 optical disk drive connected to a MicroVAX II by a TECEX optical disk controller. The bridge program also features full text storage on magnetic disks and serves both images and full text. All equipment is connected via an Ethernet local area network.

Central to the prototype, and thus to the eventual system, is the database and indexing. Indeed, the process of searching for information will *be* the system as far as the user is concerned. Every bit of information will be recorded in the index, although the form of it may vary.

All documents will be stored in ASCII form except those whose information would be meaningless in anything but image form, such as charts, figures, graphs, and formulas. The documents will be displayed in textual form as the result of full-text searching and will be linked to their images. The image-based information includes not only the original images of all documents but also raw data such as results from geologic, hydrologic, seismic, and other studies and the analyses performed on them. The database will be segmented according to document type—correspondence, memos, raw data, among others—and indexed with bibliographic headers to supplement the full-text retrieval.

Half of the prototype database will be created in-house by scanning the documents with a Palantir Compound Document Processor OCR scanner, which converts documents first to a bit map form. Two service bureaus will create the other half.

A New View of Information One of the most striking aspects of the LSS is its reflection of changing attitudes towards information. The planning behind the LSS reveals first of all a recognition that information must be handled through very comprehensive programs—embracing all information generation and processing systems and uniting them in the overarching structure of information resource management. But beyond this simple recognition, the planning for the LSS reveals confidence that such systems can, in fact, work.

Information resource management was necessary because LSS needed to draw upon information stored and generated in many different sources. Prominent among these sources were computer systems and programs already in existence and designed and operated in isolation from the proposed LSS. Also critical, though, was the information generation that takes place directly at desks throughout the organization due to the proliferation of office automation and microcomputers. Proper management of LSS required that relevant information from all these sources be entered directly into LSS.

Not surprisingly, establishing information resource management turned out to be a difficult undertaking. Managing information as a resource is a fashionable topic, but the complexity of such an endeavor makes it a challenge to put into practice. Two requirements for any attempt to establish information resource management were soon identified: information must be viewed as supporting the organization's mission or business, and the information resource manager must have sufficiently high rank to be able to work with the organization's decision-makers. This done, the information resource manager can provide the glue between systems and applications, understand the underlying similarities and connections for requirements coming from diverse parts of the program, and be able to draw upon program-wide

Continued on page 46

DOE'S PROPOSED

Continued from page 14

information. The manager can bring together data administration, office automation, records management, telecommunications, technical data management, program management, software management, and documentation. In information resource management, information generation and information storage are intertwined. Everything is grist for the mill.

Clearly, LSS needed to cut across all information in the organization and build a framework that would bring everyone on board. With this as a platform, LSS could be defined as (1) a set of procedures for gaining access to whatever information was needed in examining the proposed repository site and (2) equipment for automating these procedures.

Determining what these procedures would be and what this information would include began with the system's eventual users. In addition to DOE and NRC, these included environmental groups, Indian tribes, state and local governments, and other organizations who may become involved in the licensing process. These groups met for over a year to agree upon rules for entering and locating documents to support or challenge the license application, as well as upon procedures for using LSS during the "document discovery" phase of the license hearing.

The very fact that the groups agreed to rule-making at all represented a significant step forward for the process of information resource management because it revealed a need for and confidence in improved means of managing information. By agreeing to the rules, the groups gave up some of their traditional rights to document discovery, but they sacrificed these rights because they felt the system would help them build a better case during the license hearing process. Their actions suggested that users are acquiring faith in the technology of handling information and recognition that the information age has matured.

The rule-making process also forced DOE and NRC to reconsider how they would process their information. In order to provide the users the information that they wanted when they wanted it, DOE and NRC agreed not to use a traditional microfilm-based system. Early in the project, microfilm had been considered a perfectly adequate and less expensive format for storing and retrieving some of the backlog of documents. During the rule-making, however, the need for the most timely retrieval of information became apparent. In the cost benefit analysis, DOE evaluated offline microfilm as an alternative to online digital storage but eventually rated it lower than the optical disk-based proposal because of image quality and the time needed for both image return and image capture.

Similarly, DOE considered and abandoned the possibility of cutting down either the total volume of documents placed in the system or the percentage that were indexed by full text. The system as it is now perceived by its users has all relevant program documentation stored and retrieved in these forms.

Clearly, the cost benefit analysis does not follow a traditional approach. DOE decided to evaluate the perceived usefulness of the system to meeting the requirements as well as the system's actual costs. The justification was the need to design a system to suit its users. Only by optimally providing what the users wanted would all parties agree to use the LSS for document discovery, and only with the LSS could DOE and NRC hope to meet the three-year time limit. Indirectly, though, the approach still cost-justified the system. Each year that can be pared off a license review saves the government an estimated $195 million. ■