UNITED STATES
NUCLEAR REGULATORY COMMISSION
WASHINGTON, D. C. 20555

August 7, 1989

Mr. Daniel J. Graser
Information Resources Management
Office of Civilian Radioactive Waste Management
U.S. Department of Energy
Forrestal Building
1000 Independence Avenue
Washington, D.C. 20006

Dear Mr. Graser:

As part of the NRC efforts to review the design of the LSS, I am enclosing NRC comments on the SAIC reports "LSS Prototype Header Design," and "LSS Prototype Cataloging Manual." Although these reports focus on the LSS Prototype, our comments will need to be considered in establishing the header design and cataloging manual for the final LSS. Most importantly, the manner in which the NRC adjudicatory record has been incorporated into the Document Type and Detailed Document Type fields fails to ensure the completeness and the unique identification of this critical set of documents. In this regard, we would be interested in discussing the resolution of our comments and questions at your convenience.

A continuing dialogue on the issues related to header design is particularly important in light of the NRC upgrade of its document control system (NUDOCS). Part of the upgrade process is a re-evaluation of the headers, indexing manuals, and authority files for NUDOCS. We would like to ensure consistency between the LSS and NUDOCS headers and would encourage coordination of these header design efforts. In this regard, we would invite you and your contractor to evaluate the next version of the NUDOCS header design which will be ready for review in October 1989.

If I can provide any further information on our comments, please feel free to contact me.

Sincerely,

Chip Cameron

Francis X. Cameron
Chairman
LSS Internal Steering Committee

Enclosure: As stated

7/26/89

COMMENTS ON SAIC REPORTS ENTITLED

"LICENSING SUPPORT SYSTEM PROTOTYPE HEADER DESIGN"
March 7, 1989 version

and

"LICENSING SUPPORT SYSTEM PROTOTYPE CATALOGING MANUAL"
March 14, 1989 version

I. GENERAL COMMENTS AND QUESTIONS:

1. The Cataloging Manual (CM) states that 120,000 pages of documents will be captured. We understand that this represents about 2,600 documents including the SCP, its references, some of the "administrative record" and some handwritten notes. We are concerned that these documents are not a representative sample of the document types that will populate the system later on. At a minimum, this will affect the validation of the Document Type authority files. Also and more important, it will limit the ability of the various classes of searchers to fully evaluate the prototype in the "test phase". In what areas do you expect the header might change as the true makeup of the database evolves?

2. What is the source or basis for some of the specific format requirements in the Cataloging Manual? Is it patterned after any existing system, such as the DOE's ARS? NRC has provided SAIC with the NRC's NUDOCS header record layout, indexing manuals, and authority files. What, if any, are the reasons why some of the NRC conventions (such as Document Type structure and Affiliation codes) were not adopted?

3. How are numeric and alpha-numeric fields structured so as to allow for sorting and listing? Will indexers have to "zero-fill" or will the software justify appropriately?

4. What procedures are envisioned for the modification and update of the authority files based on submitter's suggestions and needs?

5. There needs to be much more discussion internally within NRC and between the parties about the following issues:

   A. One issue that is not addressed to any degree in the header design document is the extent to which the submitter's authors or authoring offices as opposed to the submitter's catalogers will complete portions of

the header, specifically the title and/or abstract. From the experience with NRC's NUDOCS, it may be better for the submitting office to at least "propose" as much of the subjective information as possible in order to limit the number of errors or misrepresentations committed by the catalogers unfamiliar with the context or the subject matter. From the description of the bibliographic fields, it appears that most can be completed by the submitting office with party's catalogers performing review functions for quality control and for format or classification consistency. There are cost/benefit issues to debate.

B. Abstracting
   * the need and purpose of an Abstract
     - for all documents or
     - for just selected documents by type
       - if so, which types of documents.
   * who, personally and organizationally, will prepare the abstract, depending on document type.
   * When will the abstract be prepared.
   * Any differing considerations on above issues between "backfit" phase versus the "real-time" phase when the timeliness of entry requirement will compete with the requirement for the quality of indexing for long-term retrieval.

C. More work must be done on the Document Type classification scheme. See Fields 5 and 6 for more information.

6. Section 3.1 of CM. Windows and pull-down menus are high-tech. They will certainly help new indexers and eliminate inconsistent entries. However, experienced indexers may want faster entry. Will it be possible for such authorized indexers to bypass windows and enter directly. Entries in specific fields could then be automatically checked against authority files at "end" before record is 'closed out'.

7. In Section 3.1.3 of CM, the Query function as discussed seems cumbersome and unsophisticated. Can it handle multi-parameter searches? If so, how will it handle embedded Boolean statements within statements? How does it differentiate between: ([A and B] or C) versus (A and [B or C]) ?? Is it thought that the prototype cataloguers did not need such sophisticated search capabilities? In the full system, both cataloguers and searchers will need such a capability.

8. Section 3.2.1 of CM implies that documents come to the station with LSS Accession numbers already assigned. What are the pre-indexing procedures and rules? Who defines and determines the "cataloging units"? When and how are accession numbers assigned? Who and how are duplicates searched?

9. In Section 3.2.3 of CM, "Deleting a Record", it is stated that the phrase "Delete Number?" will appear before an entire record is deleted. This could be confusing to the cataloger in that he or she may assume that only the LSS accession number will be deleted rather that the entire record. Also, is it possible to archive these "deletions" at least temporarily instead of erasing them so that they can be recovered if needed?

10. In Section 3.2.4 of CM, "Using Query," on pages 17 and 19 of CM, the method of performing a search is described. It is assumed that the described method is only for the use of catalogers or other individuals who have extensive experience with the LSS. The search software for most LSS users must be much more helpful.

11. Section 3.3.2 of CM -- How are "batches" defined? What if more than one batch is done in a day? or if one batch spans more than one day? The command "After what date (YYMMDD) does not seem to allow for this.

12. Section 4 of CM - Quality Control -- there definitely should be more than one level of QC. Also, the initials of the QC persons should also be carried on the data record. Each submitting party will have their own Quality Control procedures. However, QC should be given a lot of attention and the responsibility should be a major line function, not relegated to a committee.

II. COMMENTS ON SAIC PROPOSED FIELDS AND ASSOCIATED CATALOGING RULES:

Field #1. LSS Accession Number:

NRC places their Accession Number in the lower left corner. It would be of interest to know if there was a reason for your decision to place the LSS number in the upper right corner. Many NRC documents have notations in the upper right corner. One alternative placement would be the lower right corner; although some organizations place page revision numbers there. Another alternative would be to place the number vertically in the middle of the left margin.

We are confused as to how the "Package" header will differ from the header of the "parent" document. Or will the parent records just carry two Accession Numbers? In the NRC systems, the Accession Number of the Parent or Mother is carried on the data record of all the "children" and the Parent document carries a flag to denote the existence of "children". How will your method effect the hit counts, the sorting, and printouts? More explanation and some examples are needed here.

Field #2. Title/Subject:

The "subject line" on correspondence is usually a very broad characterization with little thought given toward long term retrieval or distinguishing it from other documents. It is acknowledged that brief abstracts (NRC NuDocs has 4 lines) prepared by catalogers are time consuming and not always the best. It is also acknowledged that with the full-text of documents available for on-line searchers, this short abstract may not be critical for search and retrieval purposes. However, for the purpose of listings, bibliographies, announcements, court certifications, and for scanning large "hitlists" to determine the relevant documents for further review, something more than the "subject line" will be required. Remember, not all end-users of the LSS will be on-line. Also, most letters do not have a "subject line" like Memoranda. Maybe this is the purpose of the LSS "Abstract Field #22. If so, it is not clearly stated.

On page 5 of SAIC LSS Prototype Header Design Report (HD), the last two lines of the discussion of Field 2, Title, state that the title of an encompassing work will be in the "Bibliographic Citation" field. It is not clear to which of the fields this statement is referring.

Cataloging Rules:

- What is the proposed length of this field?
- Why are the format rules so very specific?
  Given the number of varied catalogers from different parties overtime, will it not be a real burden on the indexers to follow these strict professional-type cataloging rules and on LSS staff to assure compliance and consistency? For what end? If it is to do sorting (and filing) by title alphabetically, then couldn't some software routine be written to ignore preceding articles?

- p. 27 of CM, 2nd paragraph -- there should be more explicit rules about what to cover in title descriptions. Phrases like "meaningful" and "reflecting the content" are too vague. The NRC system has more specific rules on what aspects of the document content should be covered varying by document type. While they may not be perfect, at least they should be reviewed.

- Shouldn't the same convention as with Abstract Field be used to denote actual wording of the Title versus indexer-composed description.

Field #22 Abstract Field is discussed here due to its interrelationship with the Title/Subject Field.

It is unclear in the SAIC Prototype reports which documents will be abstracted.

The CM states that a "brief description on the content" will entered. In the final LSS design, much more must be decided and said about the Type of Abstract required or accepted. Also, what is the proposed length of this "brief" description?

Field #3. LSS Pointer

p. 30 of CM, 2nd paragraph of "instructions" -- how will be cataloger know that a 'revision' is already in the system? Can't or shouldn't that be caught in the pre-indexing review (duplicate check)?

Using this field and maybe others, how will marked up copies of the same document be handled, i.e. reviewers handwritten comments and editing on a report or "pen & ink" changes ? How will such a document be indexed? Also, how will a cover letter forwarding various/selected replacement pages to a previously-submitted document be handled? How will drafts, revisions, errata, etc. be linked together? The listed codes in the "controlled vocabularies" do not seem to cover such a case.

This information must be captured somehow and the search software must utilize it to notify searchers passively that previous or later versions and/or errata exist and are on the system.

Field #5. Document Type

This should be a repeating field!

Is this and the "detailed document type" scheme already in use at DOE? As you know, NRC has their own scheme based on their own terminology. Somehow a mutual interagency list should be devised.

!!!> There is a major concern regarding the instruction for completion of this field on page 34 of the CM. There, it is stated that the cataloger will select the first [and no other] document type that matches the form of the document from the provided list. The eighth document type in that list is Legal Materials which (as discussed in the description of Field 6, Detailed Document Type, on pages 36 to 40) includes those documents associated with the NRC adjudicatory record. Under the current instructions for the catalogers, a document that is part of the adjudicatory record may not be identified as such if the cataloger finds a document type in the list that matches the document prior to reaching the Legal Material document type. This is a serious problem and must be corrected as soon as possible.

THIS FIELD CAN NOT BE USED AS THE DELIMITER FOR THE HLW ADJUDICATORY FILES!! See Section III for proposed new field.

A "legal" document in some other proceeding may be submitted to LSS. It should get a "legal" document type BUT may or may not be part of the HLW adjudicatory record.

Any non-legal document type, i.e.. drawing, journal article, letter, etc, at first may be entered as they are. Then later that document becomes an exhibit. It then must also carry the Legal document type while keeping its original document type code.

Field #6. Detailed Document Type

This also should be a repeating field. Some of the elements here are not mutually exclusive within one document.

These codes should be mapped to NRC document type codes. If for no other reason than to test clarity of both systems. Specifically, more work must be done on the legal document types.

Field #7. Document Date

How will transcripts and minutes of meetings spanning multiple days be coded?

Field #8. Document/Report Number

Is this field only for numbers of the specific document being cataloged, i.e. (1) contract number for actual contract and amendments, not reports done under that contract; (2) the USGS or NUREG report and revisions, not other documents, memos, letters about the USGS or NUREG report? The description appears this way. Assuming this is true, how will documents commenting on or 'about' such reports be coded?

What is the purpose to preceding alpha codes listed on p.45 of the CM? This appears redundant to the Document Type Code. Does this not put a burden on the searcher to know what kind of number he/she has been given to search? If final retrieval software has a 'wildcard' character, then this problem could be eliminated, but I think the classification is not justified.

Rule 7 on page 47 of the CM states that common abbreviations should be used where possible. While this suggestion is acceptable in theory, the examples provided are not common to all LSS users. It may be best to refrain from using abbreviation except where their meaning is obvious and unambiguous. NUDOCs has attempted to keep an authority files of accepted abbreviations and it is not always up to date or used. The problem will get much worse given the multiple parties contributing to the LSS over long period of time.

Field #9. Edition, Version/Revision

This field will require more detailed instructions to handle:

- -- selected pages submitted as Amendment 9 of looseleaf document such as this indexing manual or the application !versus!
- -- whole indexing manual or application including revised interfiled pages thru Amendment 9

Shouldn't this field be linked to occurrences in the previous field? Might there not be cases where Rev. 6 to a Sandia report then becomes NRC NUREG-####, which is later supplemented 6 times? Sad but true.

In reference to the use of this field "for describing computer codes and code manuals" , more explanation is required. As one reviewer of these reports stated "I think I know what this means, but surely this needs to be spelled out better so we are all singing from the same hymnal"

Field #10. Author Name

Should concurrences, either by name or organization, be picked up if they appear on the document?

Field #11. Author Organization

How will the authority file rules handle organizational name changes, subsidiaries, reorganizations, etc?

Must develop rules for authors who write in two of more capacities, i.e. letterhead says ACME utility, but author is writing as the head of the utility owners group. OR lawyer works for DEWY, CHEATEM & HOWE but is representing EXXON. OR NMSS staff chairing inter-agency or intra-agency review group? How handled? -- will you pick up both?

In the NRC system, the Affiliations (and the Document Type Codes) have hierarchical scheme to classify document authors and recipients,

NRC AFFILIATION SCHEME

First level - E for external vs N for internal. This would not be appropriate in this system

Second level - type of organization"
i.e. SG = state government
UT = utility
LO = local government
US = Federal agency
LG = legal firm
MV = manufacturer or vendor
etc

DOCUMENT TYPE CODE SCHEME

CLUTN = correspondence/letter/utility to NRC
TRUTIN = text/report/utility inspection report

POINT: These codes can be very powerful in searching, especially along with the Boolean "Not Equal" to narrow scope of searches to their essence. Many times after searching known parameters, the resultant hitlist is still too large to be useful. At this point, the

searcher may not know what he/she wants or be able to positively select a narrowing concept, but he/she knows what he/she does not want. Then, using the 1st (and second) level Affiliation codes [and/or Document Type codes] truncated, he/she can exclude classes of documents by type of author, by type of recipient or by type of document.

Field #12. Recipient Name

Proposed instructions state that this field would include attendees at a meeting as recipients. Some meetings may have a long attached attendance list and it may not be feasible or beneficial to list all of them in this header field. More specific rules must be developed to narrow the scope and intent of this data capture. Consideration: if smaller number of attendees (i.e. less than twelve) are listed at the first of meeting minutes or meeting summary and it was a "participatory"-type meeting, then such persons should be captured. In this case, one could argue that such persons are more "authors" than "recipients. Better yet, have another field for "attendees". The requirement to complete this different field could be triggered for all records having certain document types.

Field #17. Publication Data

Instructions state that an entry is required. From the description of this field, however, it is not clear whether an entry will be appropriate in all instances.

Field #18. Subject Term

Please provide more information as to the intent of this field. The broad nature of the terms may cause this field to be of little value in searching for particular documents. Is it to be used to segment the database? If so, there may be problems because many documents may address several of the listed terms such that the submitter and cataloger would have difficulty in assigning a single term to a document. Also the searchers may take issue with the view of the cataloger. It will be hard to make the segments mutually exclusive by the subject scheme. Page 63 of the CM states that this field will not be used in the Prototype. Therefore, it will be impossible to test the usefulness of this item.

Field #21. Special Class

More discussion is required on this field because it appears that this field and the Document Type fields are being used in combination to "segment" the Adjudicatory Record file for the adjudicatory Boards.

In this field or in the "Project" field, documents related to rulemakings and documents referenced/cited in other documents should be captured.

In the proposed list of "special classes", it is not clear what documents will be encompassed by the following terms: EA-AR (Part of the Environmental Assessment Administrative Record), LA-AR (Part of the License Application Administrative Record), and Lit (Part of EA Siting Litigation). Are these DOE-specific classes? If not, it would be difficult for others to assign such codes. Other parties will have their own "special" codes. The LSS Administrator will have to maintain authority list.

The description of NRC evidence in the special class list should be revised to read "Unit is evidence in an adjudicatory proceeding" because evidence may be oral or written.

Field #22. Abstract.

See comments in section on the Title/Subject field (#2).

Field #25. QA Level Code.

Please provide more information on the scope and usage of this field.

Field #27. Page count.

How will the page count for package records be handled? This has been a sticky issue in the NRC's NUDOCS, especially when an enclosure in the new "package" is a document already indexed earlier and therefore is a "duplicate" which must be tagged to this new package for completeness.

III. PROPOSED ADDITIONS.

The following elements of information were not included as separate fields but may be of value in performing search tasks:

- Date docketed
- HLW adjudicatory document "tag" -- see comments on fields #5, #6 and #21 for more information.
- Concurrence Names
- Reference Affiliation/Organization (use same Controlled Vocabulary as used for Author or Recipient Organization.)
- Referenced Documents and/or regulations (parts of CFR)
- Event Date -- dates of meetings, inspections, "incidents".
- Alternate availability -- other sources of same document, i.e. NTIS, GPO, ORNL and/or location and contact of core samples, data tapes, maps, travel vouchers, etc.

In addition, there are certain elements of information that are captured in more generic fields which might warrant their own specific field:

- Witnesses & Speakers (currently in 'author field')
- Attendees (currently in 'recipient field. If kept as part of more generic field, I could debate that attendees should go in 'author field', especially for small meetings - less than ten people.)
- Contract numbers (currently in the 'Report field')

There is an argument which states that some of the above listed information could be found by searching the full-text. Also, some of this information could be loaded into more general fields. However consistency of capture and format would argue for a specific field. The existence of such fields would trigger indexers to capture the information in a standard format. This would relieve the burden on the searchers.

A paragraph or so explaining and justifying the exclusion of such data capture and alternate retrieval methods should be provided. It would be helpful to those of us who follow (advisory committees) and wonder "why not?". Further, weren't there other fields proposed or discussed during the negotiations? If so, what was their disposition? Those fields that were considered but not included should be listed and discussed <u>somewhere</u>.