

Kirk M. Balcom, Esq.

*Route 2, Box 481
Marshall, Virginia 22115
(703) 364-3195*

May 17, 1990

Mr. John Hoyle
Chairman
LSS Advisory Review Panel
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555

Dear Mr. Hoyle:

Attached are the recommendations of the Header Working Group for the fields of information to be supplied by participants and the office of the LSS Administrator. The group consisted of myself as the representative of the State of Nevada; Eileen Tana, NRC, NMSS; Dona Mennella, SAIC, representing DOE; and Elizabeth Shelburne, NRC, LSSA.

If there are questions or possible revisions, please contact me as 466-4764.

Sincerely,



Kirk Balcom
Chairman
Header Working Group

RECOMMENDED FIELDS

FOR

LSS HEADER RECORDS

**Header Working Group,
NRC Licensing Support System Advisory Review Panel**

May 18, 1990

RECOMMENDED FIELDS FOR "HEADER" RECORDS

I. INTRODUCTION

The list of recommended fields for the LSS header (Section II and Appendix A) represents the consensus of the Header Working Group appointed by John Hoyle, Chairman of the LSS Advisory Review Panel, at the March 20 meeting of the Panel. The working group consisted of representatives of NRC, DOE, the State of Nevada and the Office of the LSS Administrator (LSSA), NRC.

As a starting reference point, the group used the list of fields prepared by the technical staff during the negotiated rulemaking, entitled "Draft Bibliographic Header Fields," Rev. 3, dated May 17, 1988. We also drew on the experience of SAIC, the DOE contractor, in cataloging 100,000 pages of the Site Characterization Plan and assorted other documents for the instrumented test bed ("prototype") and experience with other systems including those used in conjunction with NRC's records management (NUDOCS) and Public Document Rooms.

During our review process, several issues emerged which were important to our discussions but which we could not completely resolve. We believe they should be reviewed and deliberated by the Advisory Review Panel. Those issues are presented below in Section III with our recommendations. The issues did not prevent agreement of the number or use of the fields.

The recommended header fields apply to documents only. We do not have enough information at this time to determine treatment of the so-called "non-documents." This is discussed further in issue number 4 of Section III.

Participant is used throughout to mean "party" or "potential party" and submitter refers to the organizational units of the participant contributing documents to the LSS.

II. RECOMMENDED FIELDS

We agreed upon 28 basic fields and have grouped them by who, in our best judgment, should supply the information, the participant or the LSSA. The "bibliographic header" fields together with the remaining fields below constitute the "full header." They are listed with their attributes in Appendix A. A description of each field is in Appendix B.

Bibliographic Header (required to be supplied by participants):

Participant Accession Number

Submitter Center
Submitter Page Count
Title/Description
Author
Author Organization
Addressee
Addressee Organization
Document Date
Document/Report Number
Document Condition
Edition/Version
Event Date, Code
Protected Status
Related Documents
Special Class
Abstract/Summary (for non-documents)

Fields optional to participant, but completed by LSSA:

Document Type
Sponsoring Organization
Copyee
Copyee Organization
Publication Data
Descriptors

Fields optional to both the participants and the LSSA:

Identifiers
Comments
Abstract/Summary (for documents)

Fields not applicable to the participant, but provided by the LSSA:

LSS Accession Number
Number of Images
Pointers

III. ISSUES FOR ADVISORY REVIEW PANEL DELIBERATION; WORKING GROUP RECOMMENDATIONS

There are several issues that kept surfacing during the working group's meetings. They will undoubtedly come up for consideration again.

1. Multiple submissions for same document. This issue arises when two participants submit different headers for the same document and they characterize some of the information differently, for example, the title/description. Should all of the information be merged into one header or does the first header for that document prevail? We think this will happen frequently during

processing of the backlog.

Our recommendation is to append the subjective information from subsequent submissions that is different to the respective fields of the original header. In the title/descriptions field, subsequent descriptions would be separated in some fashion to differentiate multiple submissions. The participant's accession number would be carried in the header in order following the original submitter's accession number.

2. Editing of headers by LSSA. How much latitude should the LSSA have in editing errors in submitter headers? If there are obvious errors, can the LSSA correct them during data capture or must the header be flagged for possible corrections by the submitter?

We recommend that the editing functions of the LSSA regarding individual fields be as follows:

A. For the fields submitted by the participant (required or optional), the LSSA staff will review the data against quality control standards. If submitted data is clearly wrong, e.g., the date or spelling of names, or if data is not formatted correctly, then the LSSA staff will correct the entries. In subjective fields, such as descriptors or title/description, the LSSA will not edit existing information. The LSSA will supplement with additional information as required to improve retrieval.

B. If the fields are optional to the participant and not completed, the LSSA will complete the following fields, if applicable: document type, copyee, copyee organization, sponsoring organization, publication data and descriptors.

C. With the remaining optional fields: identifiers, comments and abstracts, the LSSA will complete the information only if applicable in accordance with standardized procedures.

3. Abstracts. As it stands now, an abstract will be required for every unit that does not have searchable full text associated with it.

For other units, it is our recommendation to make selections for abstracting of documents ("searchable full text") based upon some consistent rules, such as length of unit or type of document. Because all documents would not be abstracted, the system should provide a warning message (e.g., "not all records contain abstracts") to users who use the abstract/summary field in their searches.

4. Fields for Non-Document Materials. Section 2.1003 (c) of the rule requires the LSS Administrator to develop "Access

Protocols" of information about materials that are not suitable for storage in either ASCII text form or bit-mapped image form. Information that such materials exists will be stored in fields in the LSS header. A code field to reference how the information can accessed should be in the header. The code will link with a table (to be updated as necessary) which explains how to access the item(s) referenced.

During the coming year, the LSS Administrator will be developing a plan for providing access to such "technical data." Part of that plan will be the development of these special header fields. It is expected that certain fields recommended here will be used, such as abstract, sponsoring organization, and pointers. In addition, there may be unique fielded information related only to non-documents technical data, such as storage facility, name of contact, access code or form of data. There will be, however, only one header and one data base for all materials.

Because the actual materials will not be available to the LSS Administrator's operations staff for "cataloging" and quality control, the header elements describing them will have to be provided by the participant organization. Certain fields, such as the Abstract, which are not required for documents, will be required in the participant's "technical data" header.

As the Access Protocols Plan is being developed, the LSS Administrator will keep the Advisory Review Panel informed and involved.

5. Miscellaneous Fields. It is probable that needs will arise for fields of information that have not been completely anticipated but which might need to be added without affecting the integrity of a submitter's coding. One such field is whether a copyrighted document has been approved from the source for distribution. We have not included such fields in our list but anticipate there will be some. We do not have a recommendation, but expect the issue will arise.

ENDIX A:

HIGH LEVEL WASTE - LICENSING SUPPORT SYSTEM
FIELDS FOR "HEADERS"

FIELD NAME	MULTI- VALUED	CONTROLLED AUTHORITY	FORMAT CONTROL	FREE TEXT SEARCHABLE
FIELDS REQUIRED BY PARTICIPANT:				
Participant Accession Number	Y	N	Y	NA
Submitter Center	Y	Y	NA	NA
Submitter Page Count	N	N	N	N
Title/Description	N*	N	N	Y
Author	Y	N	Y	N
Author Organization	Y	Y	NA	Y
Addressee	Y	N	Y	N
Addressee Organization	Y	Y	NA	Y
Document Date	N	N	Y	NA
Document/Report Number	Y	N	Y	NA
Document Condition	Y	Y	NA	NA
Edition/Version	N	N	Y	Y
Event Date, Code	Y	Y, code only	Y	NA
Protected Status	Y	Y	NA	NA
Related Documents	Y	N	N	NA
Special Class	Y	Y	NA	Y
Abstract (non-documents)	N	N	N	Y

FIELDS OPTIONAL TO PARTICIPANT, BUT COMPLETED BY LSSA:				
Document Type	Y	Y	NA	Y
Sponsoring Organization	Y	Y	NA	Y

FIELD NAME	MULTI-VALUED	CONTROLLED AUTHORITY	FORMAT CONTROL	FREE TEXT SEARCHABLE
Copyee	Y	N	Y	N
Copyee Organization	Y	Y	NA	Y
Publication Data	N	N	Y	Y
Descriptors (Thesaurus)	Y	Y	NA	Y

FIELDS OPTIONAL TO BOTH PARTICIPANT AND LSSA:

Identifiers	Y	N	N	Y
Comments	N*	N	N	Y
Abstract/Summary	N*	N	N	Y

FIELDS NOT APPLICABLE TO PARTICIPANT, BUT SUPPLIED BY SYSTEM OR LSSA:

LSS System Accession Number	N	N	Y	NA
Number of Images	N	N	Y	NA
Pointers	Y	N	Y	NA

 In each of the four columns, Y = Yes, N = No, NA = Not Applicable

Multi-valued = multiple entries, e.g., authors, allowed.

Controlled Authority = list of accepted entries with which all participants must comply, such as organization names or document types.

Format Control = whether the entry must follow guidelines or cataloging rules.

Free Text Searchable = the ability to perform phrase or single-word searches of the header fields .

* = Only one variable-length text field. Multiple entries just appended to previous text. See Section III, Issue #1.

APPENDIX B:

DESCRIPTIONS OF RECOMMENDED BIBLIOGRAPHIC HEADER FIELDS

Participant Accession Number: a unique identification code required by 10 CFR 2 to be assigned by the participant to each unit submitted for entry into the LSS. This code assists the submitter in locating documents they have submitted and assists the capture operation in verifying the identity of the document received and matching it with the image and ASCII file. This field should include a specific alpha code identifying the participant organization, e.g., DOE, NRC, NEV, and any other alpha/numeric scheme which the submitting organization might be using to control their own documents.

Submitter Center: a coded field for the name and location of the participant or its subdivision submitting material for inclusion in the LSS. This field provides a contact point for material that is rejected by the LSS capture facility. It provides a contact point for notification that the header, image, and ASCII text have been loaded into the search and image system and are ready for review and verification by the first submitting participant.

Submitter Page Count: the number of pages identified by the submitter as representing the length of the document. This assists the capture station in determining whether the document is complete.

Title/Description: a brief description given to a unit (usually by the author) to distinguish the unit from other units. The complete identifying title of the unit, including any subtitle, is stored in this field. If there is no title, a description of the unit is entered. A mechanism should be available to distinguish those instances where a title does not exist on the unit but has been created by a person other than the author.

Author: the names of all persons listed as responsible for the creation of all or part of a unit. This includes editors or compilers (identified by "ed(s)" or "comp" after the name on the title page) but not those who merely concur or approve. Only personal authors are entered in this field. Corporations or organizations as authors are entered in the Author Organization field and are linked to their respective authors for report and search purposes.

Author Organization: the name of the organization with which each the author was affiliated at the time the unit was created is stored in this field. This field is also used for the name of the organization when a unit has no personal author, for the corporate source of a report, or for the corporate author of a book. In

those cases where both an author and an author organization are entered, a mechanism will be available to link the author with his respective organization. If an author works for one organization but is representing another, then both affiliations should be captured, e.g., an attorney using a law firm's letterhead stationery but representing a client organization, or a scientist for a lab chairing a formal working group or task force.

Addressee: the names of all persons to whom the unit is addressed (correspondence only). It is linked to the Addressee Organization for report and search purposes. See also Copyee field.

Addressee Organization: the organizational affiliation of each recipient, if indicated (correspondence only).

Document Date: the date on which the unit was published or created. If the date is unknown, information in the unit will be used to determine a likely date. A mechanism should be available to distinguish those instances in which a date does not exist on the unit but has been created by a submitter or cataloger.

Document/Report Number: any identifying numbers that have been assigned to a unit and appear (typed or handwritten) on the unit itself are considered to be control numbers for that unit. This field contains these control numbers, which are usually assigned to a document by the issuing agency or organization. Examples are report numbers, contract numbers, public law numbers, and any other identifying numbers on the unit.

Document Condition: the condition of the unit at the time of entry into the system. This includes information such as pages missing, portions illegible, and marginalia. It is always assumed that the unit is the "best copy" available to the submitter, but that the "best copy available" may not be a perfect copy. In some cases, the difference between two identical documents may be that one document contains marginalia; this indication makes the distinction between "duplicate" documents.

Edition/Version: the edition number, version number, revision number, or draft status of all units (including computer codes) that have multiple iterations.

Event Date/Code: the date(s) a particular event, such as an inspection, audit, meeting, or hearing, which is not the date of the document and a code indicating the type of event, e.g., inspection (IN) or meeting (MT). This enables the user to retrieve all documents concerning a particular inspection or meeting.

Protected Status: a coded field indicating the type(s) of privileges or exceptions claimed for the underlying document upon which the header is based.

Related Documents: units within the LSS can have relationships among them which are important to retain. There are several types of relationships, such as: parent/child (a document and its attachments); original/subsequent (a document and later versions, comments, corrections or errata); and whole/part (a book and its chapters, a journal and its articles, an information package and the cataloging units it contains). This field is intended to be used by a participant to store these relationships by identifying the type of relationship and the units involved. The LSSA will translate these references to a standardized form of pointers for navigating between the units.

Special Class: further classifies units in a manner that would assist a user in locating all units belonging to a special class of units, regardless of the type of unit. These classifications are not necessarily the subjects of the document, but rather are another way of grouping certain kinds of materials in order to facilitate retrieval or inform the user of some unusual aspect of the group. Examples of the use of this field, respectively, would be for units that are part of the Site Characterization Plan administrative record or for units that have only a header and image (no full-text of the unit is available).

Abstract/Summary (for non-documents and image-only materials): description of physical samples, raw data, hand-written notes, and other units that are not available for full-text searching in the LSS.

DESCRIPTIONS OF REMAINING RECOMMENDED FIELDS (FULL HEADER)

Document Type: the type of unit, i.e., the format or physical form of the document. This field is a two-part field consisting of both a major document type and a subset of the major document type. For example, a major document type might be correspondence and the subset would include letters, memos, etc.

Sponsoring Organization: the name of the agency or agencies responsible for funding or otherwise sponsoring the work reported in the unit is stored in this field. Generally, it is assumed the work has a sponsor if there is a contract number, if it is stated that the work was "Prepared for", or if a conference or workshop was presented or organized by a society or agency.

Copyee: used for correspondence only and contains the names of all persons to whom a copy of the unit was sent (as listed on the unit). This field is linked to the Copyee Organization for search and reports purposes.

Copyee Organization: the affiliation of each copyee, if indicated.

Publication Data: bibliographic information that is not

covered in other fields but is important in identifying or citing the unit. Examples of such information are journal name, conference title, conference location. This field in combination with author and title fields provides the user with a standard, consistent bibliographic citation for use in creating bibliographies and references for reports.

Descriptors: terms selected from the LSS Thesaurus that represent the subject content of the unit. The descriptor may or may not be a word or phrase contained in the text of the document. The use of a descriptor obviates the need for synonyms in a search statement. The number of descriptors assigned will vary from unit to unit, depending upon how many are needed to fully cover the content of the unit.

Identifiers: those terms that are not contained in the thesaurus, but that the submitter or cataloger believe will assist a user in retrieving the unit. These may be "buzz words" or words representing new concepts that have not yet appeared in the LSS Thesaurus. The terms in this field will provide a candidate list of terms for inclusion in the LSS Thesaurus.

Comments: any information not covered in the bibliographic fields which the submitter or indexer believes will be necessary to identify or retrieve the unit is stored in this field. This field should tell the user what language the unit is written if it is not English. It is also important because foreign language documents will not have any ASCII text. This will assist the user in determining whether the document is in a language which he will be able to understand.

Abstract/Summary (for documents only): description of the content of the document, generally written by the author but may be prepared by the submitter or cataloger.

DESCRIPTIONS OF SYSTEM FIELDS

LSS System Accession Number: a unique internal identification code assigned to each cataloging unit entering the system. The capture station at which the unit enters the LSS processing system is also identified as part of this number. The LSS Accession No. will also be used in the LSS Pointer field for units which have relationships to other units in the data base.

Number of Images: the exact count of the number of images that will be created from the pages of the unit. This informs the user as to how many pages will be printed if he executes the print command, as well as how many images he will need to view for "image only" documents. Since it is anticipated that an image represents an 8-1/2 X 11" page, there will be more images than the submitter page count indicates to allow for oversize pages (foldouts, maps, etc.) that will need to be tiled.

Pointers: references to related documents after they have been standardized by the LSSA. See Related Documents description.

ADMINISTRATIVE AND PROCESS TRACKING FIELDS

Additional elements of data are required to track the processing of the documents, their headers and their ASCII files and images for statistics and quality control. Such information might be captured in fields of the LSS Full Header either by the LSS system automatically or by the Process Tracking Data Bases. Typically, this type of information is available for use and viewing only to the LSS Operations and Management staff and is not displayed to the users of the search system.

The exact form and content of such fields will be determined by the future LSS design and development integrator. The following are typical examples of tracking information:

- a. Information about the dates that submissions were received, accepted, returned, resubmitted, finally accepted
- b. Initials of Indexer and Station ID
- c. Initials of QC staff
- d. Initials of subject and abstract cataloger
- e. Initials of cataloging QC staff
- f. Status field indicating the current process stage
- g. Date loaded into the LSS
- h. Date and Initials of Submitter Center personnel reviewing and verifying the loaded information.
- i. Change Tracking - a log of who, when, and what changes, additions, and/or corrections are made to the header record, if any, after the header is loaded into the search database.