

# **DEFINITION OF TECHNICAL DATA AND ANALYSIS OF INFRASTRUCTURE WITHIN PARTICIPANT ORGANIZATIONS FOR PROVIDING LSS ACCESS**

*Prepared for*

**Nuclear Regulatory Commission**

*Prepared by*

**C. L. Acree, Jr.  
S. R. Young**

**Center for Nuclear Waste Regulatory Analyses  
San Antonio, Texas**

**February 1991**

**TABLE OF CONTENTS**

1. INTRODUCTION . . . . . 1

    1.1 HIGH-LEVEL WASTE REPOSITORY LICENSING . . . . . 1

    1.2 THE LICENSING SUPPORT SYSTEM . . . . . 1

    1.3 TASK 1: DEVELOPMENT OF ACCESS PROTOCOLS TO LSS  
    TECHNICAL DATA . . . . . 2

        1.3.1 Subtask 1 . . . . . 2

        1.3.2 Subtask 2 . . . . . 3

        1.3.3 Subtask 3 . . . . . 3

        1.3.4 Subtask 4 . . . . . 4

    1.4 TECHNICAL APPROACH . . . . . 4

2. TECHNICAL DATA DEFINITION . . . . . 5

    2.1 TECHNICAL DATA DEFINITION . . . . . 5

    2.2 CATEGORIES OF DOCUMENTARY MATERIAL . . . . . 5

        2.2.1 Classification of Categories II and III  
        Documentary Material . . . . . 8

        2.2.2 Description of Category II Documentary Material . . . . . 8

        2.2.3 Description of Category III Documentary Material . . . . . 8

    2.3 ENTRY AND RETRIEVAL OF CATEGORY II AND III  
    DOCUMENTARY MATERIAL . . . . . 12

3. ORGANIZATIONAL INFRASTRUCTURES . . . . . 17

    3.1 THE YUCCA MOUNTAIN PROJECT OFFICE (YMPO) . . . . . 17

        3.1.1 The YMPO Technical Data Management System . . . . . 18

        3.1.2 Technical Data Storage at the CRF . . . . . 19

        3.1.3 Data-Record Packages . . . . . 20

        3.1.4 Backlog Consolidation . . . . . 22

        3.1.5 Technical Data Access . . . . . 22

        3.1.6 Plans for Loading Materials into the LSS . . . . . 23

    3.2 THE DEPARTMENT OF ENERGY (DOE) . . . . . 26

    3.3 THE NUCLEAR REGULATORY COMMISSION (USNRC) . . . . . 27

    3.4 THE STATE OF NEVADA . . . . . 28

    3.5 OTHERS . . . . . 28

4. CONCLUSIONS ..... 29

5. ISSUES ..... 32

    5.1 PACKAGING ISSUES ..... 32

    5.2 STORAGE ISSUES ..... 34

6. PLANS FOR TASK 1 COMPLETION ..... 36

7. REFERENCES ..... 37

## LIST OF FIGURES

<u>Figure</u>	<u>Description</u>	<u>Page</u>
1	Categories of documentary material described by the LSS rule.	6
2	Bibliographic header fields required for description and retrieval of technical data and information from LSS documentary material.	9
3	Categories of Media Type.	10
4	Categories of Document Type.	11
5	Categories of Regulatory Basis from 10 CFR Part 60.	13
6	Categories of Regulatory Basis from USNRC Draft Regulatory Guide: TOPICAL GUIDELINES FOR THE LICENSING SUPPORT SYSTEM.	14
7	Organization of Data Record Package.	16
8	YMPO Technical Data flow.	24

# 1. INTRODUCTION

## 1.1 HIGH-LEVEL WASTE REPOSITORY LICENSING

The Nuclear Waste Policy Act (NWPAA) of 1982, as amended, addresses requirements for the licensing of a high-level nuclear waste (HLW) repository and establishes respective responsibilities of the Department of Energy (DOE - the license applicant), the Nuclear Regulatory Commission (USNRC - the license review and issuing agency), and the Environmental Protection Agency (EPA - the promulgator of standards for long-term performance).

In support of its HLW program under the NWPAA, the USNRC has established the Center for Nuclear Waste Regulatory Analyses (hereafter referred to as "the Center"). The Center's mission is to provide a sustained high quality of technical assistance and research for the HLW program, in order to help identify and resolve technical and scientific issues.

## 1.2 THE LICENSING SUPPORT SYSTEM

Under the NWPAA, Section 114(d), the USNRC is required to review the DOE license application for construction of a HLW repository within a three-year period. To assist in meeting this timeframe, the USNRC has revised the Commission Rule of Practice in 10 CFR Part 2, for the adjudicatory proceeding on the license application, to incorporate the use of an electronic information management system known as the HLW Licensing Support System (LSS).

The LSS will contain documentary material of the DOE, USNRC, and other parties to the HLW licensing proceeding that may be relevant to the review process. It is intended to facilitate that process through: centralized submission of discoverable documents before the DOE application is filed; early technical review of materials by potential parties; and electronic transmission of filings during the hearing.

Policies for the development and use of the LSS are set forth in 54 FR 14925, dated April 14, 1989. This regulation, known as "the LSS rule," tasks the DOE with the design and development of the system, while it tasks the LSS Administrator (LSSA), housed at the USNRC, with its subsequent management and operation. To discharge its oversight and operational responsibilities, the LSSA will also review DOE design and development of the system and assist the creation of standards and procedures for the submission and processing of LSS materials.

The Center's scope of work has been broadened to include support for the LSSA so that it may assist project management; evaluate technical issues related to design, development, quality assurance, and operation of the LSS; help create procedures and standards for the submission of documentary materials, including compliance evaluation and protocols for access to technical data; assist the maintenance of LSS guidance documents; and support LSS workload processing, distribution, and volume assessment.

6

### **1.3 TASK 1: DEVELOPMENT OF ACCESS PROTOCOLS TO LSS TECHNICAL DATA**

As a first task ("Task 1"), the LSSA has tasked the Center to develop recommended access protocols to LSS "technical data." The objective is to resolve ambiguities pertaining to this type of material in order to reduce their potential impact upon the design, development and operation of the LSS, as well as upon the development of LSS participants' internal records management processes.

The problem is that the LSS will, for the most part, contain documents in two electronic forms: "bit-mapped" images and ASCII representations (searchable full text). However, the LSS Rule, in Section 2.1003(c), addresses the need for access to graphic-oriented and other documentary material which is unsuitable for entry in searchable full text. Some of it can be captured in bit-mapped image form, by means of a digital scanning device, and some cannot. In Section 2.1011(d)(10), the Rule tasks the LSSA with the development of access protocols for the inclusion of this material in the LSS. If the material is to be successfully retrieved, it must be described by sufficiently informative bibliographic headers. Due to definitional and practical difficulties, procedures for its entry into the LSS and later retrieval (before and after participants are provided LSS access) have not yet been established. Moreover, no definitive requirements have been instituted for treating these materials in the bibliographic headers which must adequately describe them.

The LSSA has set forth four specific subtasks under Task 1 which are described below:

#### **1.3.1 Subtask 1 - Review and analyze the LSS Rule and other documents, and thoroughly define "technical data"**

The Center will review and analyze the LSS Rule and its Supplementary Information, the minutes of the USNRC Advisory Committee on the LSS, and other relevant documents related to technical data.

The Center will prepare a report for the LSS Administrator that:

1. Defines the categories of technical data contemplated by the LSS Rule.
2. Identifies ambiguities in the LSS Rule pertaining to technical data and recommends specific guidance that could be provided to LSS participants to remove those ambiguities.
3. Analyzes how LSS header information on technical data could best be linked to LSS documentary material employing this data.

**1.3.2 Subtask 2 - Identify and analyze existing infrastructure within participant organizations for providing access to technical data**

The Center will document the existing infrastructures of LSS participants and their contractors/subcontractors, as well as any other organizations with potentially relevant technical data, for providing access to technical data. "Infrastructure" includes related policies, procedures (e.g., quality assurance requirements, the "Morgan-Davis" agreement), organizations involved in the process, locations involved, etc. The Center will then prepare a report for the LSS Administrator which accomplishes the following:

- 1. Identifies the organizations that are generating or are expected to generate technical data.
- 2. Documents existing and/or planned infrastructures of these organizations for providing access to the technical data they produce.

**1.3.3 Subtask 3 - Propose a plan for providing access to technical data, with access protocols and submission requirements**

After LSSA review, comment and acceptance of the Center's deliverables under Subtasks 1 and 2 above, the Center will prepare a plan for implementing the requirements of the LSS Rule relative to technical data. This report will include, but is not limited to, the following:

- 1. Analysis of the information LSS users should have relative to gaining access to the various types of technical data and to understanding any restrictions placed on that access. Recommendations on the placement of this information (e.g., in the LSS header or with the owner).
- 2. Recommended content of the LSS headers to be supplied by participants concerning technical data. The header fields should provide for effective searching of the LSS database and impart useful access information about the technical data to the searcher when the image is not in the LSS database.
- 3. Guidance for LSS participants that addresses the required timing of their submission of technical data headers and, if required, technical data images.
- 4. Guidance for LSS participants that describes when images of technical data are required to be submitted and when they are not.
- 5. Identification of the potential impacts of the above recommendations on existing infrastructures for providing access to technical data, and identification of any disparities between the above recommendations and existing infrastructures.

Access protocols will be developed in the context of any existing quality assurance, recordkeeping, and other applicable rules and policies of the USNRC, DOE, and

8

other federal, state, or local agencies. These protocols will also recognize any privileges from disclosure claimed by LSS participants.

**1.3.4 Subtask 4 - The Center will provide briefings on access protocols at meetings of the LSS Advisory Review Panel and other meetings as required**

**1.4 TECHNICAL APPROACH**

The Center's Information Management Staff (IMS) initiated work on the first two subtasks in June 1990, pursuing them concurrently, as requested by the LSSA. The LSSA sent a letter to members of the LSS Advisory Review Panel (LSS/ARP), advising them that the Center would be contacting them.

Center investigators reviewed existing documents related to technical data, spoke with experts within the Center itself who are familiar with it, and visited the most significant known producers of the material, in order to define it, identify existing/potential producers, and document existing/planned infrastructures for providing access.

Fact-finding visits were made to the DOE's Yucca Mountain Project Office (YMPO) in Las Vegas, Nevada (twice), to a representative of the State of Nevada, to the DOE and USNRC, both in Washington, D. C., and to the two primary YMPO contractors producing technical data - the U. S. Geological Survey in Denver, Colorado, and Sandia National Laboratories in Albuquerque, New Mexico.

The YMPO provided briefing charts, management plans and administrative procedures relating to its existing and projected technical data management system. These were analyzed by the Center in connection with the briefings that the YMPO and its contractors provided.

A status report on the first two subtasks was prepared on September 26, 1990, so that the Center could provide a briefing on its initial work to the LSS/ARP on October 11, 1990, in Reno, Nevada. That briefing and discussion provided an opportunity for organizations represented on the LSS/ARP to characterize their respective needs for the materials in question and to suggest ways of providing convenient access to it.

The two draft reports on (1) technical data definition and (2) organizational infrastructure follow in the next two sections of this combined report. The final three sections deal respectively with conclusions, identified issues and the Center's plans for Task 1 completion.

## **2. TECHNICAL DATA DEFINITION**

### **2.1 TECHNICAL DATA DEFINITION**

The Licensing Support System described by 10 CFR Part 2 Subpart J (54 FR 14944) will provide for the submission and retrieval of documentary material relevant to the licensing of a geologic repository for high-level radioactive waste. The LSS is intended to facilitate timely document discovery and early technical review of this material. Technical review by both the USNRC and other participants of the licensing proceedings will require access to the massive volume of technical data held by the LSS.

Technical data, as considered in this report, is that data or information that is acquired during, or generated by, scientific and engineering observation, measurement and analysis. The LSS is not intended to be a technical data base, offering online access to digital listings of numerical parameter values, but rather will implement procedures which will allow technical data to be retrieved from the documentary material archived within the system and from technical data bases referenced within the system. Bibliographic headers can be designed with an array of header fields that will allow the technical data content of documentary material to be examined and retrieved. Header fields must be selected to allow all documentary material that contains technical data to be indexed and retrieved. Header fields should also provide sufficient description of the technical data content to allow efficient search and retrieval.

### **2.2 CATEGORIES OF DOCUMENTARY MATERIAL**

The LSS Rule (Subpart J) does not specifically address comprehensive management of technical data, but rather describes categories of documentary material to be submitted to the LSS Administrator (Figure 1). Ten categories are suggested with respect to LSS physical submission requirements, and are distinguished on the basis of these requirements. Depending on the category, the LSS requires some combination of ASCII text code, a physical image and a bibliographic header. It is generally anticipated that almost all of the material of interest will fall into Categories I - IV. Technical data occurs in material now being considered for submission as categories I through IV, and is likely to occur in categories V through X as well. However, as requested by the LSSA, this report specifically addresses the characteristics, management procedures and bibliographic header requirements of Category II, III and IV material.

CATEGORIES OF DOCUMENTARY MATERIAL		LSS SUBMISSION REQUIREMENTS		
		ASCII	IMAGE	HEADER
<b>I</b>	<b>DOCUMENTARY MATERIAL SEC 2.1003(a)(1)(2)(3) 2.1003(b)(1)(2)</b>	<b>YES</b>	<b>YES</b>	<b>YES</b>
<b>II</b>	<b>GRAPHIC-ORIENTED DOCUMENTARY MATERIAL SEC 2.1003(c)(1)</b>	<b>NO</b>	<b>YES</b>	<b>YES</b>
<b>III</b>	<b>DOCUMENTARY MATERIAL NOT SUITABLE FOR ENTRY SEC 2.1003(c)(2)</b>	<b>NO</b>	<b>NO</b>	<b>YES</b>
<b>IV</b>	<b>PACKAGE OF INFORMATION SEC 2.1003(c)(3)</b>	<b>NO</b>	<b>NO</b>	<b>YES</b>
<b>V</b>	<b>PRIVILEGED MATERIAL SEC 2.1003(d)(1)</b>	<b>NO</b>	<b>NO</b>	<b>YES</b>
<b>VI</b>	<b>CONFIDENTIAL MATERIAL SEC 2.1003(d)(2)</b>	<b>NO</b>	<b>NO</b>	<b>YES</b>
<b>VII</b>	<b>SAFEGUARDED MATERIAL SEC 2.1003(d)(3)</b>	<b>NO</b>	<b>NO</b>	<b>YES</b>
<b>VIII</b>	<b>REQUESTED MATERIAL SEC 2.1003(e)</b>	<b>IF APPLICABLE</b>	<b>YES</b>	<b>NOT STATED</b>
<b>IX</b>	<b>AMENDED DOCUMENTS SEC 2.1004(b)(1)(2)(3) 2.1004(c)(1)(2)</b>	<b>YES</b>	<b>YES</b>	<b>YES</b>
<b>X</b>	<b>INCORRECTLY EXCLUDED SEC 2.1004(d)</b>	<b>NOT STATED</b>	<b>YES</b>	<b>NOT STATED</b>

Figure 1. Categories of documentary material described by the LSS rule.

//

The boundaries between categories of documentary material are not exact and some categories of material are, in some cases, subsets of other categories. Category I documentary material is comprised primarily of text with some subsidiary graphs, figures, tables and included or attached photographic or graphic plates. In general, material will be judged to be Category I if it is composed of a sufficient amount of text to be scanned and converted to full text-searchable form. Many reports and manuscripts which are discrete documents have graphic-oriented, non-text-searchable attachments. The U.S. Geological Survey Open File Reports, SAND reports from Sandia National Laboratory, and LA reports from Los Alamos National Laboratory are examples of these. These reports are properly Category I documentary material, but have Category II material attachments, and are being placed within Data-Record Packages which are Category IV.

Category II material is graphic-oriented documentary material that is displayed in any hard copy form and for which the embedded text is not suitable for capture in searchable full text form. This is generally not published material. It exists as supporting documentation, basic and computed data and material composed primarily of graphic elements rather than text. Category II material is currently being compiled into packages (Category IV material) that usually include the final report (Category I material). On a page by page basis, considerable Category II material exists within documents that are currently judged by LSS criteria as Category I material. For example, any given page of a NUREG document may be completely devoted to a figure, table, printout or source code listing. This page considered alone is Category II material, but is contained in a document that belongs in Category I. The technical data on this page is retrievable within a bibliographic archive to the extent that its content is reflected in the text and header of the document. This same figure may also exist as a piece of stand-alone Category II material and may therefore be placed in the appropriate Data-Record Package with the final report. In this instance, it will be listed on the Table of Contents of the package and will be held in the LSS as an image.

Category III documentary material is material that is not suitable for entry into LSS in image or searchable full text form. Current consensus on LSS operations holds that this material must be physically archived in some location where it can be retrieved upon request by an LSS user. The LSS Administrator (LSSA) requires the submission of a bibliographic header that describes the material. Category III material is primarily magnetic tape, bore hole geophysical logs and oversize geologic maps, surveys and facilities designs.

Category IV documentary material is a package of information comprised primarily of Categories II and III material items. The LSS Rule requires the submission of a bibliographic header for a Table of Contents describing the package of information. YMPO Data-Record Packages commonly contain a final report (Category I material) in addition to the supporting data. Submission requirements for the material within the package is determined by the type of material and not by the submission requirements for the package itself. If the package contains a final report, that report (Category I documentary material) will have been scanned, converted to searchable full text and have a header prepared as per the requirements of Sec. 2.1003(a)(1)(2)(3) and 2.1003(b)(1)(2). The graphic-oriented material (Category II documentary

material) that can be scanned, will have been scanned and a header prepared according to the requirements of Sec. 2.1003(c)(1). The documentary material not suitable for entry into the system (Category III documentary material) will have had a header prepared as per the requirements of Sec. 2.1003(c)(2). The contents of the entire package, then (Category IV documentary material) are listed on a Table of Contents and a header is prepared as per the requirements in Sec. 2.1003(c)(3).

Some redundancy exists, but the rule is not ambiguous with respect to technical data management or the handling of Categories II and III documentary material in general. However, ambiguities are likely to arise during implementation and operation of the system. An obvious caution at this point is that the Tables of Contents of the PACKAGES OF INFORMATION must be very carefully and completely filled out. By far, most of the material in a package will be Category II. Likewise most Category II material will be in packages. Any item not listed on the Table of Contents may be, for practical purposes, lost in the system. This problem exists because the only search path for Category II material in packages is through the Table of Contents of the package.

### **2.2.1 Classification of Categories II and III Documentary Material**

As with technical data in general, Categories II and III documentary material are best characterized by technical subject area. The LSS Thesaurus is developing into an adequate source of technical subject categories under which all the Category II and III material can be classified. Effective categorization, for system entry and retrieval, must be achieved by proper use of header fields. The bibliographic header (Figure 2), then, properly determines the working description of Category II and III material.

### **2.2.2 Description of Category II Documentary Material**

Category II documentary material is best described using a combination of MEDIA TYPE (Figure 3) and DATA type (Figure 4). The (TECHNICAL SUBJECT) DESCRIPTORS, derived mainly from the TITLE/DESCRIPTION of the material and based on the LSS Thesaurus will offer the most useful categorization of the material.

### **2.2.3 Description of Category III Documentary Material**

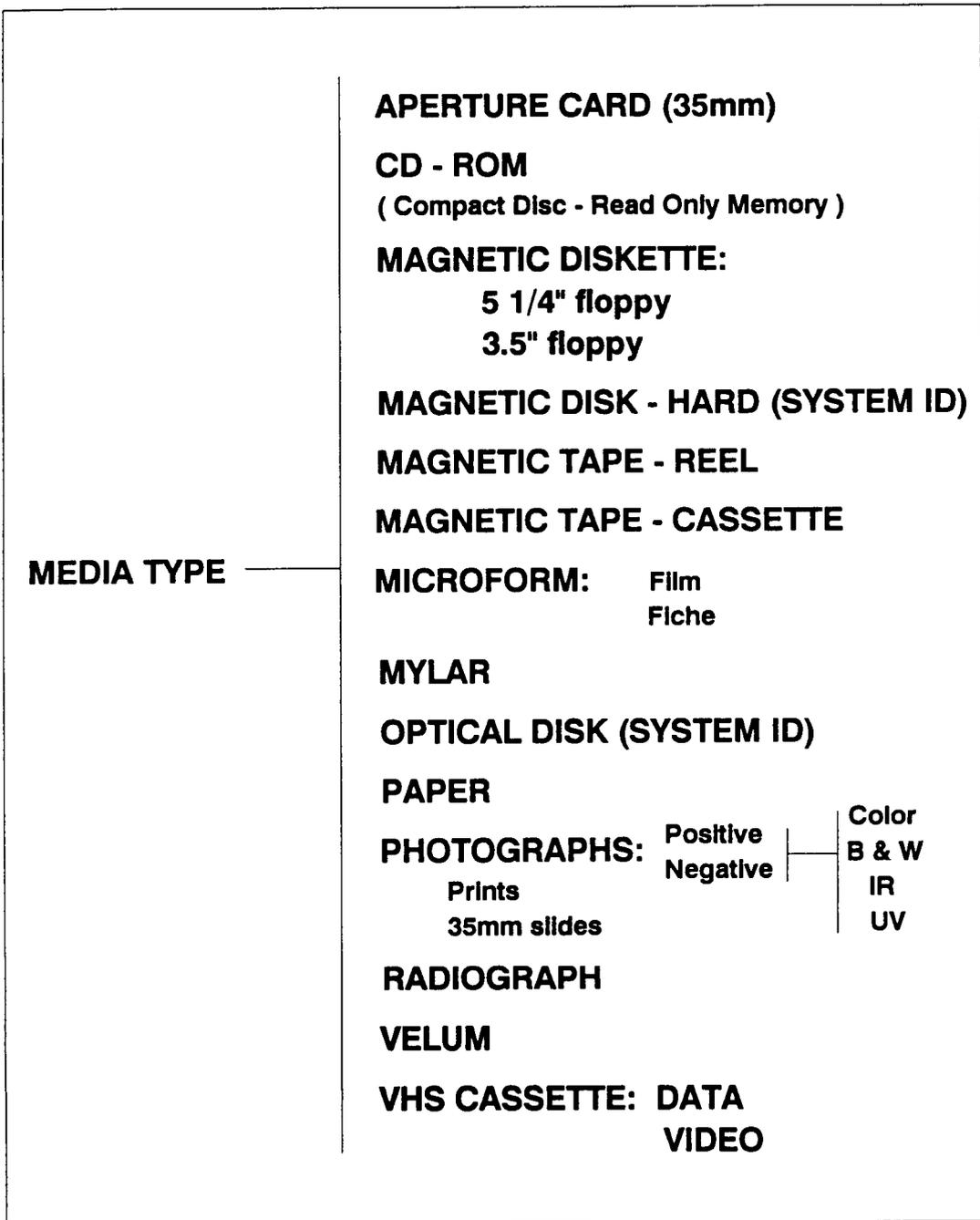
Category III documentary material is effectively described mainly on the basis of MEDIA TYPE (Figure 3), and (TECHNICAL SUBJECT) DESCRIPTORS from the LSS Thesaurus.

**HEADER FIELDS IN ORDER OF RELATIVE IMPORTANCE TO RETRIEVAL OF CATEGORY II AND III DOCUMENTARY MATERIAL**

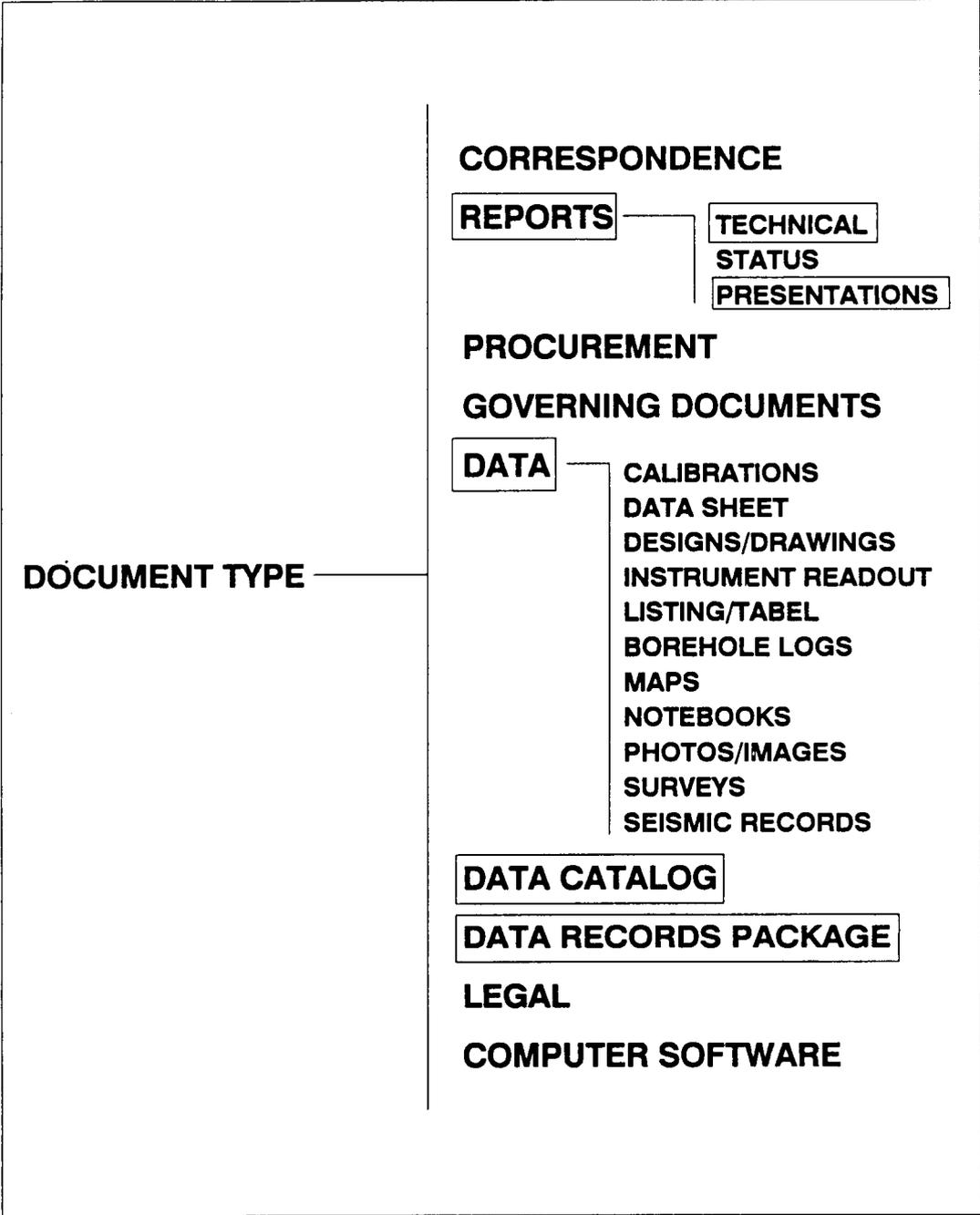
- 1. TITLE / DESCRIPTION**
- 2. DESCRIPTORS (TECHNICAL SUBJECT)**
- 3. AUTHOR**
- 4. AUTHOR ORGANIZATION**
- 5. REGULATORY BASIS \***
- 6. DOCUMENT TYPE**
- 7. MEDIA TYPE \***
- 8. IDENTIFIERS**
- 9. COMMENTS**
- 10. POINTERS**
- 11. SUBMITTER CENTER**
- 12. EVENT DATE / CODE**

**\* FIELDS NOT CURRENTLY IN THE LSS PROTOTYPE HEADER DESIGN**

**Figure 2. Bibliographic header fields required for description and retrieval of technical data and information from LSS documentary material.**



**Figure 3. Categories of Media Type.**



**Figure 4. Categories of Document Type.**

### 2.3 ENTRY AND RETRIEVAL OF CATEGORY II AND III DOCUMENTARY MATERIAL

Category II and III material can be entered, indexed and retrieved using bibliographic header fields selected to optimize management of technical data in general (Figure 2). The most common categorization of technical data is on the basis of technical subject. This is the way that most people think about and search for this type of information. Commercial and National technical data bases are normally accessed in this manner (e.g. Dialog, GeoRef, NTIS and STN International). Ideally, technical data can be categorized on the basis of any one of the header fields, but that normally is not very effective. The header fields listed on Figure 2 are intended to be an optimum array of fields for entry and retrieval of technical data. All Category II and III material reviewed by the Center can be entered and retrieved using this set of fields. All but two of the fields are included in the LSS header fields that have been accepted by the LSS/ARP.

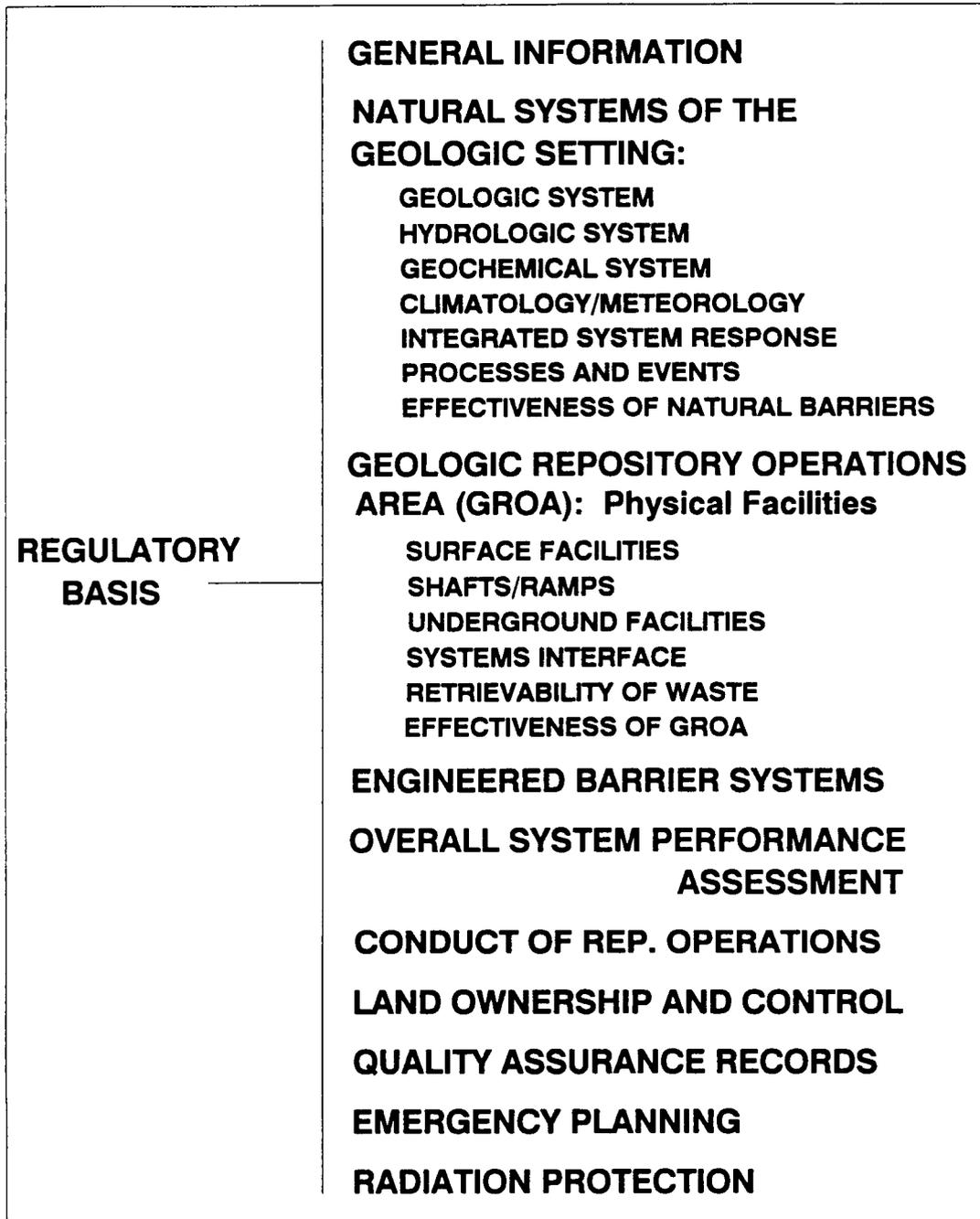
A MEDIA TYPE (Figure 3) field is a suggested addition because it is particularly descriptive with respect to the Category II and III material. Most of the Category III material will be one of these media types and should be identified as such on the header for individual items. Alternatively, a user could search the LSS on the basis of MEDIA TYPE if that user were particularly interested in finding data for a specific computer system (e.g. retrieve all 1/2", 9-track tapes in LIS format). This is a common geophysical log data format and that person may have a tape reader for it. Technical users can have a tendency to search for media that they may have a special capability to read. CD-ROM players are not common, so a particular user may not be able to utilize this data.

DATA CATALOG should be included as a DOCUMENT TYPE (Figure 4) to accommodate searches for catalogs of the major technical data bases [EQ3/EQ46, Site Engineering Properties Data Base (SEPDB), Reference Information Base (RIB) and the developing ARC/INFO Geographic Information System (GIS)]. DOCUMENT TYPE codes applicable to technical data management are highlighted by boxes on Figure 4.

REGULATORY BASIS is a very important categorization scheme for technical data in the high-level waste project. Indeed, most of the data will be obtained or developed under direct provision of 10 CFR Part 60. Regulatory analysts will certainly have opportunity to search the LSS on the basis of the portion of 10 CFR Part 60 to which the desired information pertains. For example, certain technical data pertains strongly to important performance objectives such as groundwater travel time, and it is conceivable that certain users would want to search on this basis. Figure 5 lists a potential set of REGULATORY BASIS field codes taken selectively from 10 CFR Part 60. The USNRC Draft Regulatory Guide (TOPICAL GUIDELINES FOR THE LICENSING SUPPORT SYSTEM) (Figure 6) closely mimics the topical headings of 10 CFR Part 60 and could also be used as the basis for REGULATORY BASIS codes. Practical implementation of this field, however, could be hindered by a lack of personnel trained to properly categorize documentary material in this area.

<p><b>REGULATORY BASIS</b></p>	<p><b>GENERAL PROVISIONS</b></p>
	<p><b>LICENSES</b></p>
	<p>PREAPPLICATION REVIEW</p>
	<p>LICENSE APPLICATION</p>
	<p>CONSTRUCTION AUTHORIZATION</p>
	<p>LICENSE ISSUANCE AND AMENDMENT</p>
	<p>PERMANENT CLOSURE</p>
	<p><b>PARTICIPATION BY STATE GOVTS. AND AFFECTED INDIAN TRIBES</b></p>
	<p><b>RECORDS, REPORTS, TESTS AND INSPECTIONS</b></p>
	<p><b>TECHNICAL CRITERIA</b></p>
<p>PERFORMANCE OBJECTIVES</p>	
<p>LAND OWNERSHIP AND CONTROL</p>	
<p>SITING CRITERIA</p>	
<p>DESIGN CRITERIA FOR GROA</p>	
<p>DESIGN CRITERIA FOR WASTE PACKAGE</p>	
<p>PERFORMANCE CONFIRMATION REQUIREMENTS</p>	
<p><b>PERFORMANCE CONFIRMATION</b></p>	
<p><b>QUALITY ASSURANCE</b></p>	
<p><b>TRAINING AND CERTIFICATION OF PERSONNEL</b></p>	

**Figure 5. Categories of Regulatory Basis from 10 CFR Part 60.**



**Figure 6. Categories of Regulatory Basis from USNRC  
Draft Regulatory Guide: TOPICAL GUIDELINES FOR THE  
LICENSING SUPPORT SYSTEM.**

Most Category II and III material will occur in packages (Figure 7). The same header fields can be effectively applied to these packages for entry and retrieval. Packages will be entered into the system and retrieved mainly by TITLE/DESCRIPTION and (TECHNICAL SUBJECT) DESCRIPTORS. Category III material will have its own header, so it can be found either through the package or directly. This is probably a useful redundancy. Retrieval of Category II material, that occurs only in a package, depends strongly on the Table of Contents (TOC) of the package. If an item is not listed on the TOC, a user would have to look at every item in the package to find it. Unlisted items would be, for practical purposes, lost in the system. The TOC of the package is extremely important in this respect. The Rule stipulates that Category II material will have a header, however, the Rule can be interpreted to preclude the development of a header for each and every Category II item in a package.

At this time, knowledge of the physical STORAGE LOCATION of the majority of YMPO Category III material is not considered essential to successfully search for and retrieve this material. The Yucca Mountain Project Office has repeatedly stated that the only access to this material will be through direct request to the YMPO. In general, suggested header fields will allow identification of the material and a request through YMPO will complete the retrieval. However, situations may arise in which Category III material that is not in the direct control of the YMPO is requested. In these cases, some procedures will be necessary to allow the LSSA to accomplish the retrieval. An example of a situation like this would be a USNRC request for data on tape in the possession of a State of Nevada contractor. YMPO could not directly control access to this material, so the LSSA would be responsible. For these reasons, a STORAGE LOCATION field may ultimately be required.

QUALITY ASSURANCE status is not currently considered to be necessary to effective search and retrieval of Category II and III material. Quality assurance records could be indexed in a REGULATORY BASIS field (Figure 6). Even though QA status does not significantly impact the search and retrieval process, a QA STATUS field may be desirable in the absence of the REGULATORY BASIS field.

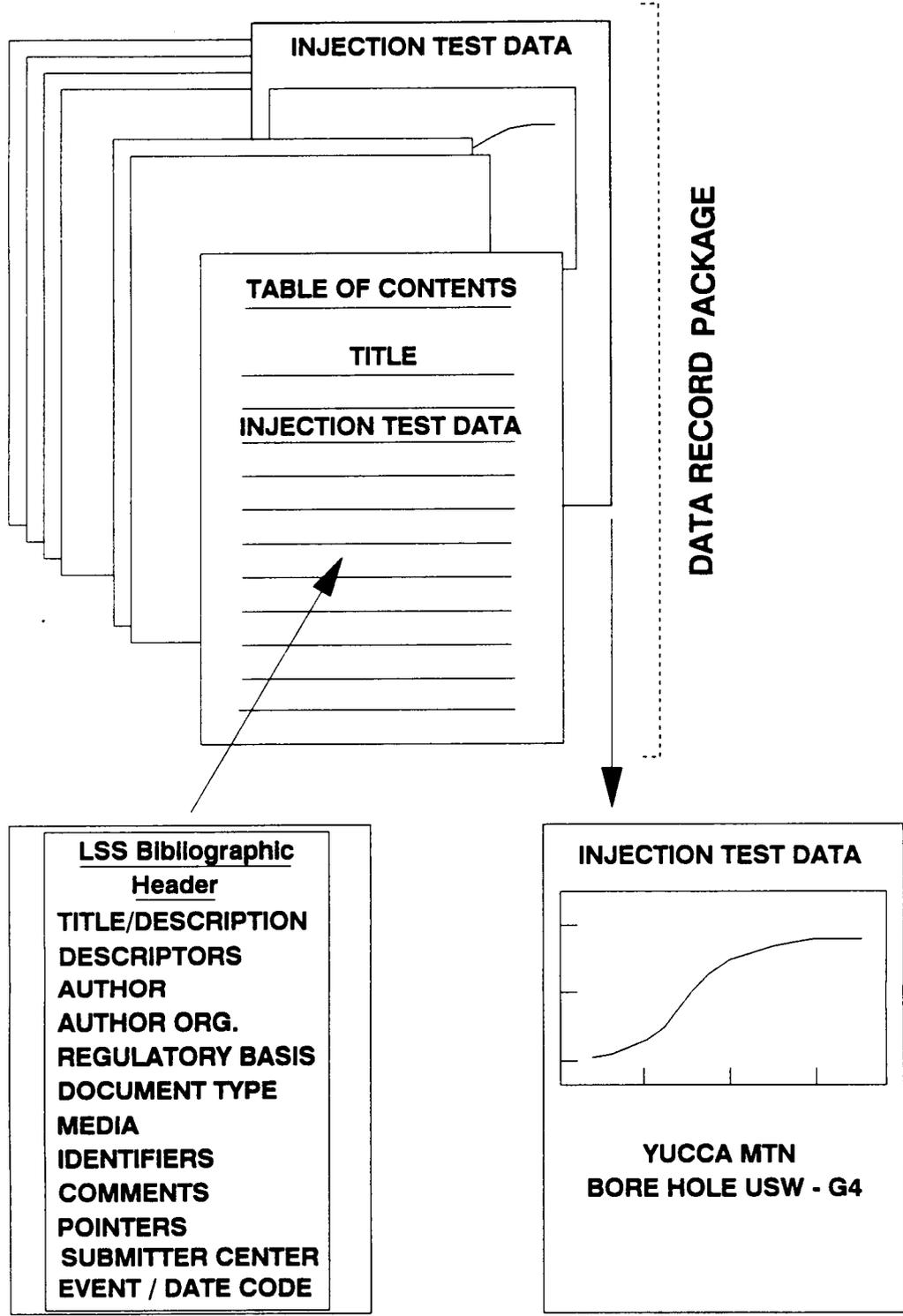


Figure 7. Organization of Data Record Package.

### 3. ORGANIZATIONAL INFRASTRUCTURES

Everyone interviewed by the Center agreed that the vast bulk of documentary material relating to the licensing of an HLW repository, including technical data and the subset of it with which this Task is concerned, will continue to be produced by the DOE's Yucca Mountain Project Office (YMPO) and by its supporting contractors.

As pointed out above, the term "technical data" was employed in the initial phase of this Task as a convenient way to describe documentary material which must become accessible to LSS users but is not suitable for entry into the LSS in text-searchable form, because it is graphic, handwritten, or contained on media other than paper - the material, in other words, that will be successfully found only with the help of a sufficiently descriptive LSS bibliographic header. Some of it can be scanned to create LSS images (e.g., photographs, numeric tables, handwritten notes) and some cannot (e.g., magnetic media).

Use of the term "technical data," however, proved unsatisfactory. That term is employed far more broadly by the YMPO and by others to include not only all of the raw data that is acquired from technical investigative activities, but also the data that is developed through analysis of raw data and data that is used during the process of interpretation (e.g., socioeconomic data). The YMPO, in fact, calls the totality of the data that it includes within its non-administrative information storage and retrieval systems "technical data" - the "qualitative" portions, as well as the graphic and quantitative (numeric) portions. While computer tapes are included within the YMPO's definition, physical samples themselves are not - being properly a source of data, not data itself.

Henceforth, in this report, the term "technical data" will be used in the way that the YMPO uses it - to include all data related to technical activities involving the HLW repository. The subset of technical data which this Task addresses, namely the data which cannot be entered into the LSS in text-searchable form, will, in accordance with the previous section, be called either Category II (imageable) material or Category III (non-imageable) material. Often, the context will call for them to be lumped together as Category II/III material.

#### 3.1 THE YUCCA MOUNTAIN PROJECT OFFICE (YMPO)

In order to understand how the YMPO intends to provide access to Category II/III material, it is necessary to describe the YMPO technical data management system as a whole, because Category II/III material is not separately processed or stored within the YMPO's system. When the time comes for the YMPO to enter materials relevant to HLW licensing into the LSS, it will enter Category II/III material together with Category I (text-searchable) material. The three types will usually be combined within Data-Record Packages.

### 3.1.1 The YMPO Technical Data Management System

Detailed YMPO management plans and procedures pertaining to technical data have been forwarded earlier to the LSSA, attached to References. In sum, the YMPO's technical data management system has the following primary components:

- Nine Local Records Centers (LRC's) - at the YMPO itself and at eight prime contractor sites: Science Applications International Corporation (SAIC), Sandia National Laboratories (SNL), Los Alamos National Laboratory (LANL), Lawrence Livermore National Laboratory (LLNL), U.S. Geological Survey (USGS), Holmes & Narver, Inc. (H&N), Fenix & Scisson (FSN), and Reynolds Electrical & Engineering Company (REECo). These LRC's are libraries to which project organizations submit their acquired (raw) and developed (analyzed) data. Each of them employs its own filing and computer systems, as needed, to assist data-consolidating activity.
- The Central Records Facility (CRF), the project-level library at the YMPO to which the nine LRC's will eventually submit all of their collected data in the form of Data-Record Packages.
- The Sample Management Facility (SMF), which stores and tracks access to physical samples. Information derived from those samples is eventually submitted to the CRF as technical data. A Core Sample Inventory Tracking System (CSITS) is used to catalog core samples. There may, in the future, be an LRC at this facility.
- The Technical Data Base (TDB) - an "umbrella" structure of Project-level "controlled data bases," which includes, most importantly, the Site and Engineering Properties Data Base (SEPDB) kept at SNL, but also a Geographic Information System and a Thermodynamic database. These computer systems are used by project investigators to manipulate raw numeric data through the use of modeling and graphic-display techniques, thereby assisting analysis and interpretation. The data are available on-line only at SNL itself. A SEPDB catalog of digital data holdings is updated and distributed quarterly so that others may request copies on computer tape, or in hard copy if they prefer.
- The Reference Information Base (RIB), produced at SNL - a printed folder that contains a controlled, distilled, summary description of fully interpreted technical data, on which consensus has been achieved on a Project-wide basis, with references to more detailed data. This summation of the "current state of knowledge" is updated quarterly, or as investigations warrant, and is revised annually.

- The Data Catalog, also produced at SNL - a manual index, published quarterly, which describes both acquired and developed technical data stored by the SNL, stating its origins, methods of acquisition, and whereabouts. It will eventually become a comprehensive document, describing all YMPO-produced technical data.
- The Records Information System (RIS), the CRF's computer-based general records index.
- The Project Automated Technical Data Tracking System (variously called the ADTS, ATDT, or TDTS), the CRF's supplemental, computer-based technical-data index now under development, which will trace analyzed data to its raw-data origins through a hierarchy of references to previously-submitted material. When it becomes fully operational, the ADTS will be the basis for a Project-level Data Catalog (describing data obtained from all LRC's) and will record RIB/TDB submissions and requests.

**3.1.2 Technical Data Storage at the CRF**

The CRF is receiving quarterly and indexing on a timely basis Data-Record Packages that are generated by the Project, its contractors and subcontractors in connection with HLW repository licensing.

Technical data storage at the CRF is governed by several management plans and implementing procedures, which were formally adopted in early 1990. The LRC's have developed internal practices of their own which are intended to conform to the centralized procedures that the CRF has promulgated.

During the past year, the LRC's have been forwarding Data-Record Packages to the CRF, which they have been receiving from their associated project organizations within forty-five days after acquisition or analysis. The YMPO is working to incorporate a large pre-1990 backlog of these packages into its CRF holdings and computer indices.

The CRF stores both qualified and non-qualified technical data - identified accordingly in its indices. However, submissions to the Project-level TDB data bases and RIB summary must satisfy all quality-assurance requirements.

Unlike the projected LSS, the YMPO does not store any documents on-line within its computer records system (RIS), in either text-searchable or image ("bit-mapped") form. Its permanent holdings are recorded on microfilm rolls, in accordance with National Archives directives. Document retrieval is therefore entirely dependent upon the use of bibliographic headers, which lead document requestors to microfilmed copy.

### 3.1.3 Data-Record Packages

The YMPO is employing Data-Record Packages to store all of its technical data (reduced to microfilm), and is using a single header and Table of Contents to describe each package.

A package of technical data amounts to a case file on a particular YMPO investigative activity, which is usually undertaken by a contractor. In most cases it is gradually compiled at an LRC as materials arrive there separately from the responsible local investigator. All relevant material to the investigation must be included in the package, so that it will be available in one place. When the package is complete, an appropriate title is assigned. Usually, the investigator originates the title - particularly when a published report is included, in which case that title becomes applicable to the package as a whole. Otherwise, an appropriate title is assigned by the LRC.

A package's Table of Contents is generally created by the responsible LRC, once it has collected all of the associated material. The LRC alone determines how much detail is adequate to describe the accumulated items. YMPO procedures dictate only basic requirements ("concise" descriptions and page counts for each). A table-of-contents form has been made available by the CRF, but its use is not mandatory. Formats, therefore, differ considerably among LRC's.

A package must be forwarded to the CRF bearing an authenticating signature. Ideally, this is done when it is complete. Backlogged material is being forwarded in this manner. However, to assure that data submissions to the CRF are not delayed if an investigative activity takes more than three months to finish (which is usually the case), LRC's are told to submit authenticated package "segments" to the CRF quarterly, unless they are specifically granted extensions by the YMPO. This is a particularly important point with regard to future investigations, which are commencing once again.

When an investigator submits data to an LRC (in a package or a segment), he must attach a Technical Data Information Form (TDIF) describing the data, its collection process, and its relationship to other package segments. Sometimes these forms are created by the LRC (particularly the case of backlogged materials). The forms are intended to accompany Data-Record Packages, or the segments to which they relate, when they are forwarded on a quarterly basis to the CRF. They are used primarily for input into the CRF's new computer-based data-tracking system (ADTS).

The bibliographic header that describes each package in the RIS, for retrieval purposes, is written entirely by the CRF upon receipt of the package or segment. The header will include an abstract when one has already been written by the investigator (usually for the benefit of an associated published report). Otherwise, it will not. Header creation is limited also by the fact that it will usually include keywords only from words encountered in the given title. CRF employees do not possess substantive expertise to write abstracts or scan packaged

material for additional keywords that might assist retrieval.

While every package will have an associated RIS header (derived from its Table of Contents - making reference to all accessioned records), certain materials within a package may be assigned headers of their own. Items within a package that cannot be microfilmed or would lose meaning if microfilmed, including what the YMPO calls "special-process" (duplicable) items such as tapes and films and what it calls "one-of-a-kind" items such as radiographs and multi-colored maps (Category III items), will always be assigned individual RIS headers. Other "stand-alone" items, including (Category I) correspondence, published reports, and materials that an LRC wants for some reason to be separately retrievable, will also be assigned individual headers, at the discretion of the CRF. Any incoming group of records which received a CRF accession number will be assigned at least a minimal header, according to current procedures.

After indexing, every package is microfilmed for the CRF at the Project Microfilming Center, which is managed by a contractor (Holmes & Narver) in Las Vegas. The microfilm master copies are then stored at an off-site, strictly-controlled, environmentally-secure, permanent storage facility in that city, where the paper originals are also kept. Backup copies of the microfilm and the "special-process" items are sent to the DOE in Washington, D.C., as official records.

"One-of-a-kind" and "special-process" items (always kept in the field while still in use) are eventually separated from their packages and stored in this same Las Vegas facility. There, they are readily available via accession number. A descriptive, handwritten "slip sheet" (formally called a "Special Instruction Form"), displaying that number, is inserted within a package in place of the missing item and is, of course, included in the microfilm copy.

The YMPO intends to incorporate all technical data, including Category II/III material, ultimately, within Data-Record Packages. Some materials, however, notably some magnetic tapes, are currently being received individually by the CRF from some LRC's, with no relation to a package, and are being assigned their own headers.

All magnetic tapes are carefully stored, with adequate written specifications and external labeling to assure future readability by the YMPO and other users having requisite equipment. External labels identify the originator, file title, date of creation, recording density, tracks, record length, block size, and sequence number. Accompanying specifications include narrative description, physical file characteristics, coding structure, recording system, and record layout describing fields of information. Management plans call for all machine-readable records to be submitted to an LRC in magnetic tape form. Tapes are required to be transferred from an LRC to the CRF when they become inactive or when the responsible investigator or LRC cannot provide proper care and handling to guarantee preservation of their data.

Each package has a permanent LRC-assigned alphanumeric identification code (typed on the upper right-hand corner of the Table of Contents), as well as a CRF-assigned

accession number (placed in the same location, but affixed vertically).

### 3.1.4 Backlog Consolidation

The YMPO and its contractors are still getting used to this relatively new and evolving technical data management system, for which approved plans and procedures are generally less than a year old. Much work will be required on the part of investigators and LRC's alike to compile, itemize, duplicate (as needed), and transmit the huge mass of backlogged data - estimated to amount to four million pages of pre-1990 material, in acceptable form, so that the CRF can incorporate it within its holdings. The YMPO hopes to complete this task by 1995 if projected funding is received.

There are local problems. SNL, for example, has not yet forwarded any Data-Record Packages permanently to the CRF because the standard format of its completed packages does not conform with currently enforced CRF requirements. If its requested exceptions to those requirements, now under discussion, cannot be granted, SNL will have to perform a significant amount of conversion work at considerable cost.

The LRC's are having some difficulty deciding how best to forward certain magnetic media, data sheets and other materials that have been in storage for a long time and are not readily identifiable. Some materials are not being sent to the CRF due to lack of specific guidance, lack of funds, or non-duplicability (e.g., developocorder film, well-log tapes, analog-recharge tapes, water-level-monitoring tapes and barometric tapes). TDIF forms describing the data on these items will be sent to the CRF, so that their existence will be known centrally, but people interested in retrieving them might have to inspect some of them at the outlying LRC's - indefinitely.

### 3.1.5 Technical Data Access

DOE policy dictates that all technical data in its possession will be made available to the public if no special privileges (permitted by the Rule) apply to the requested items. However, this is a complex matter which must yet be resolved in detail.

Under current arrangements, the YMPO expects all external (non-DOE) requests for its technical data to be channelled centrally, in written, descriptive form, through the YMPO Manager to the CRF. It intends to respond to all requests in a timely manner using microfilm-produced hard-copy or stored hard-copy. If a request involves materials which have not yet been forwarded, or will not be forwarded, by an LRC to the CRF, the CRF will obtain them from the LRC or authorize access at the LRC location itself. This arrangement means that there are to be no external requests addressed directly to LRC's, no requests submitted in electronic form, and no on-line access permitted to CRF or other Project-level computer indices, such as the SEPDB. The YMPO will furnish numeric electronic data in the form of magnetic tapes, which it will duplicate for requestors.

Non-DOE requests for technical data are made using written publications for reference, namely the RIB technical data summary, the SNL Data Catalog, and the SEPDB report - all of which are updated quarterly.

The DOE's goal, on a national scale (aside from LSS planning), is to create a networked, distributed data-base environment for HLW materials - providing non-DOE access, as appropriate, through standardized protocols, insofar as possible.

**3.1.6 Plans for Loading Materials into the LSS**

The CRF, with its centralized holdings and computer indices, will be the foundation for the YMPO's support of the LSS (See Figure 8).

When LSS loading begins, the CRF will be provided an LSS capture station. It will employ that device to scan all of its Data-Record Packages pertinent to licensing activity - copying the paper originals it has been saving for that very purpose (unless the capture process should unexpectedly provide for direct microfilm input). The CRF will submit the resulting images, plus the ASCII text that is recognized within those images, to the LSS by means of optical disk (according to LSS design documents).

In addition, the CRF will submit bibliographic headers for its packages and for its individually-stored items to the LSS, in accordance with the approved LSS header format, also on optical disk. To that end, the CRF is seeking to assure that the header fields which it includes in its RIS (computer records index) are compatible with the header fields which the LSS is in the final stages of adopting. This will allow the RIS headers to be converted expeditiously to LSS headers, with as little intervention as possible, perhaps in combination with the emerging ADTS, the supplemental technical-data tracking system which is updated using the TDIF forms that accompany materials forwarded by the LRC's to the CRF.

In the case of items that the CRF cannot usefully scan, from either a practical or a cost-effective point of view (Category III materials), the CRF will send only bibliographic headers to the LSS.

Each updated version of the RIB (summary description of analyzed technical data) will be entered into the LSS as a text-searchable document. (Merely entering updated pages of the RIB on each update occasion would probably not be satisfactory.)

The SEPDB quarterly report (catalog) will also be available to LSS users on-line - providing references to numeric data sets that may be ordered from the YMPO either in hard copy or on tape. An alternative arrangement, which has arisen during the course of our discussions, would provide for the down-loading of desired data from the SEPDB directly to a requestor, who could process it without delay using an associated computer system.

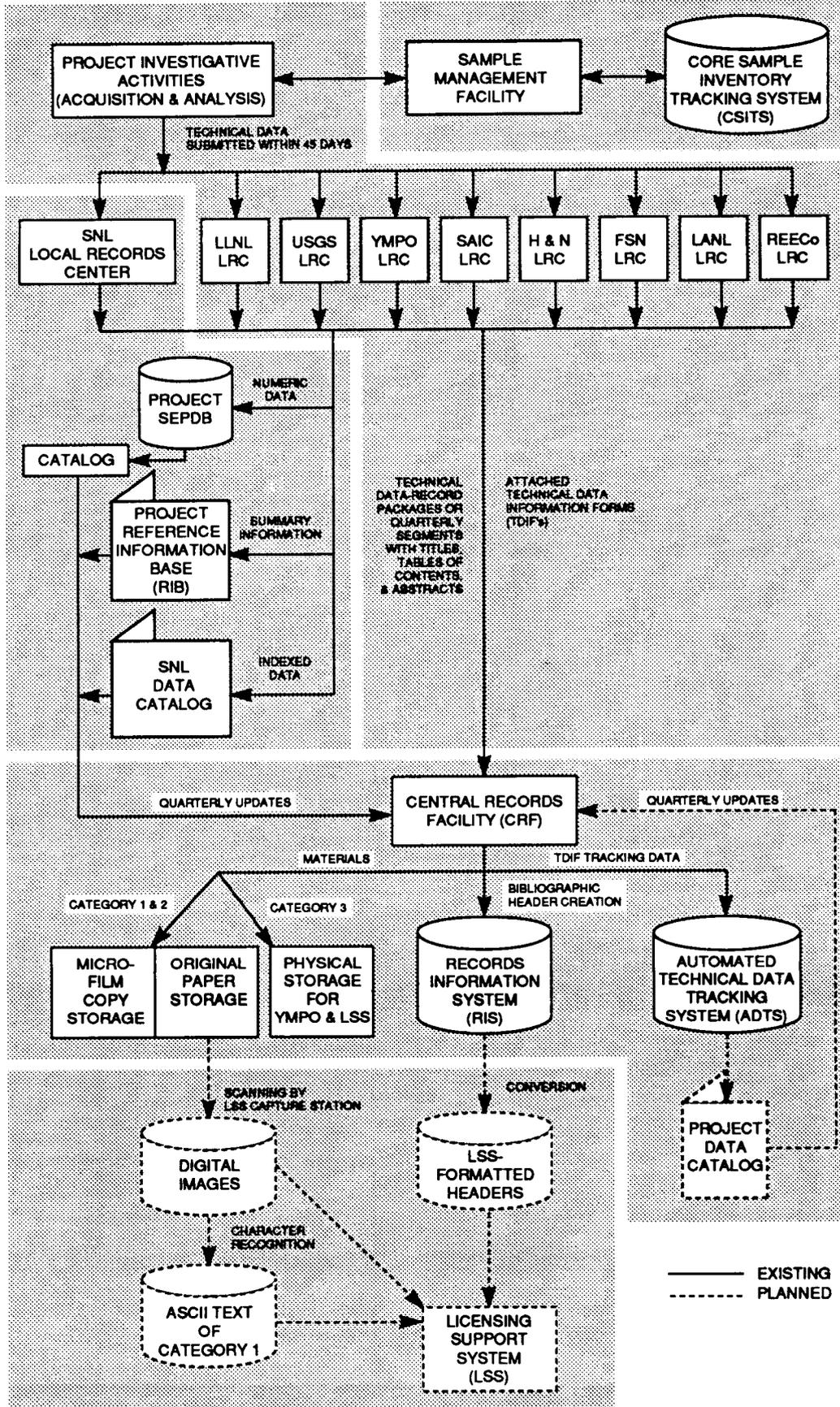


Figure 8. YMPO Technical Data flow.

By the time the LSS is operational, the Data Catalog, a reference to technical data that is currently published exclusively by SNL (not yet by other LRC's to catalog their own materials, as the YMPO intends) should be fully automated by the ADTS (technical-data tracking system). This will then be a Project-level, on-line reference and printed report, indexing all CRF-held technical data. Any direct relationship of the ADTS to the LSS, as mentioned above, would occur in combination with the RIS (the CRF's general records index) - providing additional fields of information to the LSS header in a way that has not yet been defined.

The CSITS (core sample tracking system) is supposed to interface with the LSS in a way that has not yet been defined.

So far as we can determine, YMPO planning for the advent of the LSS has not progressed beyond these few generalizations. No projected YMPO procedures (written or unwritten) yet exist for the input of technical data into the LSS - including the loading of the Category II/III data with which this report is specifically concerned. That circumstance is not at all surprising, given the fact that LSS procedural requirements have not yet been established. It is, of course, the purpose of Task 1 to help define those requirements.

In particular, the CRF has not planned in any detail how it will handle the scanning of Data-Record Packages, which usually contain many interspersed pages of graphs, photos, numeric arrays, and handwritten notes (that is to say, Category II materials). Presumably, the images of the packages will be converted to ASCII representation (for retrieval through text-search) as thoroughly as possible, by means of the capture station; but such a policy has not yet been set, much less the intricate procedures necessary to put it into effect.

Similarly, the CRF has not defined which items cannot or should not be scanned (and therefore constitute Category III material), although it has tentatively included such things as computer tapes, video cassettes, radiographs, extra-large maps and drawings, color-significant maps, and color photographs within this category.

Logically, the materials which the CRF considers inappropriate for microfilming should be identical to those that are inappropriate for scanning (to create LSS digital images). It was noted above that a handwritten "slip sheet" (substitution page) is being inserted into a package in place of items which are not microfilmed and that these identifying sheets are included in the microfilm copy. Presumably, this same practice will apply to materials that are not scanned (the Category III materials).

It was noted above that when packages do not already contain abstracts, the CRF will not create abstracts especially for them, lacking the requisite expertise among its staff to do so. This presents a potential conflict with the LSS/ARP decision that "non-documents" (meaning Category III material) must invariably contain abstracts within their associated LSS bibliographic headers, so that they may be adequately described for retrieval purposes.

It was also noted above that the CRF can only be expected to incorporate keywords in the RIS bibliographic header to the extent that they are contained in the package title itself, which was devised by a principal investigator or by an LRC. As matters stand, CRF employees will not be required to scan the packaged material to find additional keywords which might assist retrieval. Since RIS headers will form the basis for LSS headers, it follows that these limited keywords will become the only keywords available to LSS users.

What is more, the LSS has chosen to distinguish between descriptors (controlled keywords in a thesaurus) and identifiers (other keywords that may be applicable). If such distinction is to be made with regard to YMPO materials, special procedures will have to be devised to separate RIS keywords into two distinct groups.

Another problem which must be addressed is the timing for LRC submission of Data-Record Package segments to the CRF. The time-driven (quarterly) segmentation currently imposed by the YMPO is expected to pose an obstacle for external LSS users, as well as for internal users of the YMPO records system. Requestors will presumably want to retrieve all available segments of a given package and must therefore bear the burden of piecing them together as best they can through package identification and cross-reference numbers. If, alternatively, data submission is delayed until a complete package can be forwarded to the CRF, desired raw data could be unavailable centrally for well over a year. While the granting of extensions by the CRF may effectively alleviate this problem in practice, it would seem that a studied, logical segmentation of completed investigative activity would be a more satisfactory approach than the current arbitrary, time-driven segmentation.

Another area of potential difficulty (discussed briefly at the October 1990 LSS/ARP meeting) concerns the questionable usability of aging magnetic media. Existing YMPO procedures for LRC submittal of machine-readable data to the CRF on tape are reasonably thorough with respect to the inclusion of proper specifications. However, as mentioned above, some of these items are not being sent to the CRF as the result of non-duplicability. Others can be read only with special equipment that is becoming increasingly obsolescent. The CRF is committed to maintain sufficient equipment to read all of its accumulated holdings, as well as to ensure that information is not lost due to deteriorating magnetic media. However, as things stand, the unique equipment required to read some of these materials may be independently discarded at any time by sub-contractors, without the knowledge or approval of the accountable LRC.

### **3.2 THE DEPARTMENT OF ENERGY (DOE)**

The DOE maintains a Records Information System (RIS) at its headquarters in Washington, D. C., that is identical to the RIS employed by the YMPO - in terms of records management software. Its bibliographic header will, therefore, have the same compatibility with the emerging LSS header that the YMPO's RIS header will have.

The DOE does not expect to produce any significant amount of Category II/III documentary material at its headquarters location, but may produce some at its Chicago contractor's site. Any such material that may be produced at these locations in the future or that has been produced there in the past, when encountered, will be entered into the LSS using a convenient capture station.

Procedures have not yet been developed for the LSS document capture process (including header conversion) in general, let alone for the more complicated capture of Category II/III material. These procedures, however, are not expected to differ significantly from those that will be employed by the YMPO's Central Records Facility.

**3.3 THE NUCLEAR REGULATORY COMMISSION (USNRC)**

The USNRC and its contractors (including the Center) have produced Category II/III documentary material in the past and will continue to produce it in the future.

Currently, the USNRC employs a computer index, called NUDOCS, which provides reference to the USNRC's document holdings.

When LSS loading begins, the USNRC will submit required documentary material to the LSS centrally, through its existing Document Control Center. While the USNRC will retain its microfilm records and index, as federally mandated, NUDOCS itself will cease to exist as a separate system for HLW purposes when the LSS arrives - thereby eliminating a potential duplication of data bases between it and the LSS.

The USNRC has not budgeted within its five-year plan for an LSS capture station of its own. It, therefore, intends to submit both its backlogged and new material, including Category II/III items, to the LSSA for entry into the LSS (presumably using the Washington-area capture station envisioned in the LSS design documents) - in accordance with procedures to be established by the LSSA (following Task 1 recommendations to be made by the Center). Nothing more definite has been planned.

The USNRC will also provide LSS bibliographic headers for the Category II/III material it submits, in accordance with LSSA guidance (following Task 1 recommendations by the Center), in a manner that has not been defined specifically.

The USNRC is interested in the data-packaging concept, which would use a single bibliographic header to identify a package of associated material (in contrast to multiple headers referring to the individual items included). However, it has reserved judgment on the applicability of this practice to USNRC technical data management methods.

The USNRC intends to provide a central USNRC point of contact to provide Category III (non-imageable material) to requestors, in generally the same, centralized way that the YMPO intends to service external requests.

### 3.4 THE STATE OF NEVADA

The State of Nevada has, to date, produced only limited amounts of Category II/III material through the conduct of technical investigative activities of its own. Its Nuclear Waste Project Office (NWPO) in Carson City and that Office's contractors (primarily the University of Nevada and the Nevada Bureau of Mines and Geology) have, for the most part, analyzed DOE-produced Category II/III material (called "backup data" by the NWPO) to originate the State's formal technical reports. This backup material has been kept carefully stored, within associated packages, for eventual entry into the LSS. Much of it is now dispersed at NWPO contractors' sites.

When the LSSA makes a capture station available to the State of Nevada to begin loading documentary material into the LSS, the NWPO will scan its technical reports (Category I material) and also its dispersed backup data (mostly DOE-originated Category II material) for LSS entry. The NWPO will create a bibliographic header for each technical report and for each package of backup data.

### 3.5 OTHERS

It has been suggested that other organizations represented on the LSS/ARP - Nevada counties and the Nuclear industry in particular - might generate Category II/III material in the future, as a consequence of launching independent technical investigative activities. Currently, however, these organizations appear to have no firm plans in this regard and no fixed procedural infrastructures to provide access to such material by LSS participants. It is evident that very little Category II/III material will be forthcoming from organizations other than the DOE, the USNRC and, possibly, the State of Nevada.

## 4. CONCLUSIONS

The Center has found that the term "technical data" is not the best one to use in referring to the type of documentary material that we have been tasked to define. That term is widely used to encompass all data involved in HLW technical investigative activities and is, therefore, too broad for the intended purposes. The material which the LSS Rule and the LSSA have identified as problematic for LSS entry - the material that is in need of special identification and access protocols - may be considered a subset of technical data. It is the subset which cannot be entered into the LSS in text-searchable form and therefore can be found only with the help of sufficiently descriptive bibliographic headers.

Rather than apply a new, abbreviated name to this problematic, non-text-searchable material, which might be inadequate or confusing, the Center has chosen for purposes of discussion to sub-divide it, as the Rule itself sub-divides it, and to employ two distinct terms. The Center has arbitrarily called the portion of it that is imageable "Category II" material and the portion that is not imageable "Category III" material. These two categories differ markedly from the more commonplace text-searchable material, which the Center has called "Category I."

Beyond that dichotomy, it was considered worthwhile to sub-divide Category II/III material further, for two important reasons: 1) to be certain that no worthwhile portions will escape appropriate LSS entry, and 2) to help LSS users retrieve the items they desire by promoting the use of consistent entry/search terminology within applicable LSS header fields.

Category II and III documentary material are best classified on the basis of (TECHNICAL SUBJECT) DESCRIPTOR. Category II material is described best by MEDIA TYPE (primarily paper) and DATA TYPE, while Category III is most dependent on MEDIA TYPE.

Everyone with whom the Center has talked agreed that there are no ambiguities in the LSS Rule pertaining to Category II/III material which would require an amendment. However, if the Center is to recommend workable protocols for the submission and retrieval of this material (the focus of Subtask 3), it will be necessary to resolve some key issues, stated in Section 5.

The Center expects that its recommended plan for the development of access protocols to Category II/III technical data (Subtask 3) will include proposals for relatively minor alterations of existing plans and procedures within the organizations that produce this material, so that it may be effectively retrieved in accordance with the LSS Rule.

The most important producers of Category II/III material, and of technical data in general, are clearly the DOE, USNRC and those organizations' contractors. The State of Nevada, however, may produce more of this material in the future. Concerned Nevada counties and the nuclear industry may also produce it. The Center has learned of no other likely producers, though it is possible that any LSS participant organization or other interested party

may, in the future, produce it as a consequence of initiating HLW technical investigative activity.

It has not been part of this Task to measure or estimate respective volumes of production among these organizations. However, it is generally agreed that the largest past, current and prospective producer, by far, is the DOE's YMPO - more particularly two of its contractors: the Sandia National Laboratories (SNL) and the U. S. Geological Survey (USGS).

As the most significant producer of Category II/III technical data and the producer which has given most consideration to the LSS input process, the YMPO will unavoidably set the pattern for the storage and retrieval of this data within the LSS framework. The YMPO, however, is still in the process of refining its own requirements with respect to its extensive holdings of the material. It intends to create a centralized environment and has revised its detailed administrative plans and procedures accordingly. However, this has been accomplished during the past year and its contractors are, understandably, still adjusting to various complexities and complications. Areas of uncertainty, confusion and disagreement exist. The Center has identified problems within the YMPO infrastructure that will impact the LSS (all recognized by the YMPO and included in the summary of issues in Section 5). Many of them may be expected to have future applicability beyond the YMPO environment.

Presuming that the key issues identified below can be satisfactorily resolved, the retrieval of desired Category II/III technical data through the LSS should be a reasonably straightforward matter. Initial search for the data will depend entirely upon the adequacy of bibliographic headers that refer to it, which must contain sufficiently descriptive titles, keywords, and abstracts.

In the case of Category II material, which requestors will view in the form of on-line images, successful retrieval will depend also on the effective construction of technical data packages, which must contain helpful and thorough tables of contents.

In the case of Category III material, which requestors will receive in the form of hard-copy or media such as tape or film, successful retrieval will depend also on clearly identified, convenient, and responsive storage facilities, which will lend items, reproduce them for requestors, or permit them to be examined on-site.

Retrieval of all Category II/III material will be greatly assisted by the placement of successive versions of the YMPO's RIB (technical-data summary), with its important references (leading to further references) on the LSS in text-searchable form. It will also be greatly assisted by the inclusion of quarterly updates of the SEPDB catalog (listing of digital data available on tape from the YMPO/SNL).

Finally, the Center has concluded that additional LSS bibliographic header fields should be considered for the effective retrieval of Category II/III material, which is so thoroughly dependent upon header content:

1. Media Type - a field supplementing the Document-Type field.
2. Regulatory Basis - if expertise to assign it is available.
3. Storage Location - identifying the specific physical location of Category III material.
4. Quality Assurance Status - to label data one way or the other.

## 5. ISSUES

There are several key issues that must be resolved in the course of recommending access protocols for the retrieval of Category II/III technical data. These issues are intertwined, but may be usefully grouped under two general headings: packaging issues, which relate to both categories of material, and storage issues, which relate exclusively to Category III material.

### 5.1 PACKAGING ISSUES

The practice of collecting within one package all of the technical data pertaining to a particular investigative activity, in whatever form (in any or all of the first three categories), has the obvious merit of convenience - for facilitated information retrieval, as well as convenient storage.

But is it sufficient for a package originator, notably the YMPO, to create just one LSS bibliographic header to describe the package as a whole, rather than multiple headers to describe every one of the included items individually? This question is crucial to the packaging of Category II material because, while packaged YMPO Category I documents will sometimes be assigned headers of their own (as in the case of finished reports) and Category III items will always have them, packaged Category II items will seldom have individual headers. Buried within their package and irretrievable (by definition) through text-search techniques, their retrieval will be completely dependent upon the adequacy of package description - provided by its header on the one hand and by its Table of Contents on the other.

A further concern: Can packaging methods be left entirely to the discretion of individual producers of technical data, who are very likely to employ differing criteria for package assembly, segmentation, and scanning, which may be disadvantageous to LSS users?

These questions raise issues which are essentially ones of consistency. To what extent should packaging procedures be uniform across organizational and contractor lines to ensure the satisfaction of LSS needs? If package assembly and description could be governed by consistent, mandatory practices, the retrieval of Category II/III material would be facilitated. (Consistent practices always assist retrieval within a document storage and retrieval system.) The packaging process, however, would, as a consequence, be encumbered and complicated, especially when it is considered that packaging will be accomplished by multiple contractors in scattered locations. Hard trade-offs such as this between ease of input and ease of retrieval are inevitable in the procedural design of any computer-based system.

The related packaging issues are these:

1. Can LSS Rule stipulations regarding packaging [Section 2.1003(c)(3)] be interpreted to permit the omission of individual headers for each and every Category II item that a package may contain?

2. What types of Category II/III items should be excluded from package association, if any (and therefore definitely gain their own headers)? To date, the YMPO is not purposefully excluding any special types, but some (tapes, for instance) have simply arrived apart from packages and have not been subsequently grouped with others. From the LSS viewpoint, is there anything wrong with a policy of convenience in relation to the creation of packages?
3. What Category II/III items associated with packages should be provided individual headers? The YMPO creates headers for all Category III (non-microfilmed) items, packaged or not. (It seems to be generally agreed that multiple references to a given package are not a matter of concern.)
4. What Category II/III items associated with packages should be stored apart from their packages, with substituting slip sheets inserted to signal their separation? At the YMPO these items are the ones that are not microfilmed, and will, by extension, be the ones separately stored for the LSS.
5. What general classes of Category II/III items should definitely not be scanned, from either a practical or cost-effective standpoint? Put another way, what materials constitute Category III? At the YMPO these items will, by default, be the ones not microfilmed.
6. What procedures should be used during the capture process to achieve the maximum number of imaged pages during scanning, presuming this is an objective?
7. What procedures should be used during the capture process to achieve the maximum number of (interspersed) text-searchable pages during optical character recognition, presuming that this, too, is an objective? Should it be required, for example, that a package's Table of Contents, at least, always be text-searchable? Should the identifiers of graphs, figures, tables, etc. be made text-searchable, in view of the LSS Rule's concession that "text embedded within these (graphic-oriented) documents need not be separately entered in searchable full text?"
8. Are abstracts absolutely necessary for all Category II/III material, or at least for Category III material (the true "non-documents")? The YMPO includes abstracts in its RIS headers only when they have been written already by investigators. Are stylistic guidelines for their creation needed? How brief can these summaries be? Can tables of contents substitute for specially-written abstracts or supplement them?
9. Are minimal guidelines needed for the origination of descriptive package titles? These are individually styled within the YMPO context and greatly impact keyword creation.
10. Is it sufficient for package indexers to glean keywords mainly, or perhaps exclusively, from package titles, as the YMPO does? If not, where will the capable personnel come from to do a more thorough job? And, who will separate descriptors from identifiers at

the YMPO, unless this can somehow be done automatically (using the capture station's built-in thesaurus)? Will the LSSA perform this latter task, in view of the fact that the descriptor/identifier fields are optional for participant organizations?

- 11. Aside from the additional LSS bibliographic header fields that we have recommended for consideration, are further guidelines needed for making entries into established LSS fields to accommodate packages (e.g., the document-condition, related-documents, special-class, comments, and pointer fields)?
- 12. Will it be necessary for the LSSA to edit package headers (the fields completed by participant organizations) when they are submitted for LSS input? Or, would it be more realistic for the LSSA to focus its efforts upon the establishment of sound procedures to confine its policing of submissions to spot-checks for signs of non-compliance?
- 13. How consistent should tables of contents be among packages so that they will be readily comprehended? YMPO formats differ markedly among its Local Records Centers, rendering them less dependable.
- 14. Are special guidelines needed to make Table of Contents adequately inclusive and descriptive? Those of the YMPO are of varying quality and none show starting page numbers which would help a requestor find listed items.
- 15. How should packages be segmented when they must be submitted in parts to avoid delay in making the materials they contain available? The YMPO, which currently calls upon its Local Records Centers to submit their segments quarterly, is struggling with the complications of this issue. Ease of package reconstruction at retrieval time is essential.
- 16. How will the CSITS (core sample tracking system) "interface" with the LSS in relation to Category II/III material?

**5.2 STORAGE ISSUES**

There are several issues related to non-packaged items and to items which are separated from packages for storage. The concern here is exclusively with Category III material, because Category II material (by definition) will be available to the LSS in image form; it is inconsequential where the originals are stored. A decision must be made as to where the Category III items will be stored, who will control them, how they will be made available to requestors expeditiously through "protocols," and how to ensure that none will become irretrievable through the passage of years.

The related storage issues are these:

- 17. Must the currently geographically dispersed material be brought together into a central location - at the University of Nevada, for instance? For its part, the YMPO is

attempting to gather all of its Category III (as well as Category II) material within the CRF; but it may not be entirely successful in the case of exceptional materials, materials that cannot be duplicated easily, and materials that require obsolescent reading equipment - which will probably continue to be kept locally. In this connection, it is worth pointing out that the YMPO's SEPDB tapes will presumably remain at SNL.

- 18. Must stored material be eventually controlled by the LSSA and, if so, moved to a central site? The State of Nevada has taken the position that Category III material (excepting only physical samples) should be transferred to such a facility several months before the HLW repository licensing proceeding begins, to ensure that all parties will have timely access to it. The YMPO has no such plans. An added advantage of taking this course of action would be to eliminate continued need for special access protocols to Category III items once they are consolidated under LSSA control.
- 19. Wherever it is stored, will Category III material be loaned to requestors or viewed at its storage location - much as reference books are read at a library? This material is difficult, if not impossible, to copy (except, of course, for commonly-used magnetic media).
- 20. How will the problem of obsolescent equipment be handled? Is it enough to provide a requestor data on magnetic media with adequate specifications for use if the requestor is able to find requisite equipment? Or must equipment capable of reading it also be provided? Or must all data resident on magnetic media be routinely and repeatedly converted by originators (or by the LSSA), so that it may be interpreted by newer equipment as outdated machines are continually replaced?
- 21. Should YMPO (and any other) digital, numeric data sets, now available to requestors only through delivered copies of magnetic tapes, be made alternatively available by down-loading the data from its store (SEPDB in the case of the YMPO) directly to the requestor, who could process it immediately using an associated computer system? The YMPO has no current plans in this regard, but such a step would be welcomed particularly by technical requestors.

## 6. PLANS FOR TASK 1 COMPLETION

At the request of the LSSA, the Center will, by March 29, 1991, draft preliminary descriptions of the additional fields that have been suggested (above) for inclusion in the LSS bibliographic header, so that the header may be more fully descriptive of Category II/III material and thereby facilitate its retrieval.

After LSSA review, comment, and acceptance of this report (the two deliverables under Subtasks 1 and 2), the Center will pursue Subtask 3 - the drafting of a proposed plan for providing access to technical data, implementing related requirements of the LSS Rule. As in this report, the focus will be on Category II/III documentary material required by the specifics of the Task, not on the broader range of technical data that includes Category I material. The plan is due on June 30, 1991, and will necessarily include proposals for solutions to the issues highlighted in Section 5 above.

Before submitting recommendations formally in the June report, the intent is to discuss them with both the LSSA and the primary producers of Category II/III material, particularly the YMPO. This will be done not only to identify potential impacts of the plan and potential disparities with existing infrastructures, as tasked, but also to be certain that proposals are realistic and practicable, and that they would upset internal plans, procedures, and systems of the producing organizations as little as possible.

The Center, of course, can only make recommendations on this matter, which the LSSA, in turn, will take to the LSS/ARP, as it deems appropriate, for further advice. Such recommendations are constrained to some extent by existing circumstances, since it is necessary to accommodate, insofar as possible, not only established infrastructures, but also the approved LSS Rule and the detailed design work that has already been established in relation to the LSS. However, the intent is to offer for consideration whatever recommendations are perceived as needed to promote the effective retrieval of Category II/III material within the LSS. The YMPO appears to be amenable to sensible suggestions for procedural improvement, as it continues to fine-tune its technical data management system.

The next meeting of the LSS/ARP, scheduled to be held in Washington, D. C., in mid-June, 1991, would provide another good opportunity to discuss related issues and propose tentative solutions, if the LSSA should decide to place this topic on the LSS/ARP schedule at that time.

A final report, incorporating an amended version of this report, as well as an amended version of the next one (in June), will be submitted on September 30, 1991, in accordance with the Center's Operations Plan. Key staff will be available for briefings and meetings, as the LSSA may direct.

## 7. REFERENCES

1. Statement of Work: Center for Nuclear Waste Regulatory Analyses (CNWRA) Program Element Plan for the Office of the LSS Administrator, May 10, 1990.
2. "The LSS Rule": 54 FR 14925, April 14, 1989.
3. Trip Report (C. Acree): CNWRA/LSSA Meeting, July 30, 1990.
4. Trip Report (C. Acree, S. Young): USNRC/DOE Meeting on Technical Data, August 23-24, 1990 (attached YMPO management plans and procedures).
5. Trip Report (C. Acree, R. Johnson): CNWRA Discussion with Representative of the State of Nevada regarding LSS Technical Data, September 18, 1990.
6. Task 1 Status Report on Development of Access Protocols to LSS Technical Data (Intermediate Milestone #137), September 26, 1990.
7. Trip Report (C. Acree, R. Johnson, S. Young): CNWRA Discussions with USNRC regarding LSS Technical Data, October 1-2, 1990.
8. Trip Report (C. Acree, R. Johnson, S. Young): CNWRA Discussions with DOE Regarding LSS Technical Data, October 2, 1990.
9. Trip Report (C. Acree, R. Johnson, S. Young): CNWRA Discussions with LSSA Regarding Projected CNWRA Briefing of the LSS/ARP on Technical Data within the LSS, October 3, 1990.
10. Trip Report (C. Acree, R. Johnson, S. Young): Meeting of the LSS/ARP, October 10 - 11, 1990.
11. LSS Advisory Review Panel Meeting: Official Transcript of Proceedings, October 10 - 11, 1990.
12. Trip Report (C. Acree, S. Young): CNWRA Discussion with DOE/YMPO Regarding its Intended Use of Data-Record Packages as the Retrieval Vehicle for Technical Data within the LSS, November 1-2, 1990 (attached other relevant YMPO procedures).
13. Trip Report (C. Acree, R. Johnson, S. Young): CNWRA Discussion with USGS and SNL (DOE/YMPO contractors) on Packaging Technical Data for Entry into the LSS, January 24-25, 1991.