

*The Energy Science and Technology Database  
on a Local Library System:  
A Case Study at the  
Los Alamos National Laboratory  
Research Library*

**Los Alamos**  
NATIONAL LABORATORY

*Los Alamos National Laboratory is operated by the University of California  
for the United States Department of Energy under contract W-7405-ENG-36.*

*An Affirmative Action/Equal Opportunity Employer*

*This report was prepared as an account of work sponsored by an agency of the United States Government. Neither The Regents of the University of California, the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by The Regents of the University of California, the United States Government, or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of The Regents of the University of California, the United States Government, or any agency thereof.*

October 1994

*The Energy Science and Technology Database  
on a Local Library System:  
A Case Study at the  
Los Alamos National Laboratory  
Research Library*

*Irma S. Holtcamp*

**The Energy Science and Technology Database  
on a Local Library System :  
A Case Study at the Los Alamos National Laboratory Research Library**

by

Irma S. Holtkamp  
(Los Alamos National Laboratory Research Library, Los Alamos, NM 87545)

**Abstract**

*This paper presents an overview of efforts at Los Alamos National Laboratory to acquire and mount the Energy Science and Technology Database (EDB) as a citation database on the Research Library's Geac Advance system. The rationale for undertaking this project and expected benefits are explained. Significant issues explored are loading non-USMARC records into a MARC-based library system, the use of EDB records to replace or supplement in-house cataloging of technical reports, the impact of different cataloging standards and database size on searching and retrieval, and how integrating an external database into the library's online catalog may affect staffing and workflow.*

**Introduction**

The Los Alamos National Laboratory Research Library's Report Collection contains over one million unclassified technical reports, 90% of which are on microfiche. The library acquires over 25,000 new unclassified reports per year, primarily from the US Department of Energy's Office of Scientific and Technical Information (OSTI). Other major providers of technical reports are the Defense Technical Information Center (DTIC) and the National Aeronautics and Space Administration (NASA). To date, use of this collection by researchers at Los Alamos National Laboratory has been seriously limited by the lack of bibliographic access to these materials in the library's online catalog. The successful integration of bibliographic data from OSTI's Energy Science and Technology Database (EDB) into the Research Library's Geac Advance online catalog is a vital step in automating the Report Collection and providing Laboratory researchers with the ability to search all of the Research Library's unclassified holdings using a single database. This paper summarizes efforts to date to improve bibliographic access to technical report literature and make the internal management of the collection more efficient.

**Technical Report Access and Management at Los Alamos**

Bibliographic access to and management of the Report Collection at Los Alamos has been less than effective due to a lack of automation. Library staff have explored various options to improve the use of the collection by providing alternative methods of

bibliographic access. The following approaches have been used with varying degrees of success and cost effectiveness:

- in-house original cataloging
- increasing the use of records in bibliographic utilities
- expanding access to external commercial and government databases

In the past, the library has attempted to provide original cataloging for an ever increasing volume of technical reports. This has been severely affected by recent staff reductions and is the least cost effective approach. Without adequate staffing to support the volume of acquisitions, timely processing and availability of materials has not been possible. Lack of bibliographic access that adheres to established, documented standards in the in-house cataloging process has resulted in uneven bibliographic access. Past cataloging practices, coupled with not having this information in a machine-readable form have provided less than effective end user results and have restricted the library's ability to share cataloging records with other institutions.

A second approach has been to rely on the availability of bibliographic records in bibliographic utilities. In the 1980s, the library cataloging community witnessed the emergence of MARC (Machine-Readable Cataloging) as the single most important factor in the growth of library automation in the United States and other countries. The MARC bibliographic standards form the basis by which the majority of automated U.S. libraries store bibliographic information in a consistent form, share that information thus reducing in-house cataloging costs, and manipulate that information by computer. Acceptance of national standards such as MARC and AACR2 (Anglo-American Cataloging Rules) have resulted in more cost effective library operations because bibliographic data that is needed for acquisitions and circulation functions and by online searchers to locate relevant primary information sources could be shared instead of duplicated in separate systems. The selection of RLIN (Research Libraries Information Network) as the chief bibliographic utility for shared cataloging also gave the library flexibility to create and contribute bibliographic records that contain data deemed essential for local use without impeding the ability of other libraries to use original records contributed by Los Alamos. In order to provide consistency of bibliographic information in our cataloging records with less in-house effort, the library accepted MARC and AACR2 as the de facto local bibliographic standards. These early decisions on standards for bibliographic data, also influenced our selection of an automated library system, since the functional requirements for an in-house library system included the ability to import, create, and export bibliographic data in the MARC formats.

This method has been effectively used at Los Alamos to acquire machine-readable cataloging records for most of the materials in its book and journal collections. Using records from the two major bibliographic utilities, the level of original cataloging has been reduced from 100% to about 25% for new book and journal acquisitions. Beginning in 1981, all cataloging, including that for technical reports, adhered to these new de facto standards and enabled the library to cross-train and make more effective use of its limited cataloging staff. Unfortunately for the Report Collection, the availability of cataloging copy for technical reports in the major bibliographic utilities is extremely low since most academic institutions rarely contribute cataloging copy for technical reports in their collections. We estimate that less than 5% of our new technical report acquisitions have cataloging copy. Creating original cataloging records for these materials that adhere to the standards maintained in the bibliographic utilities benefits other libraries that acquire these publications, but it has not resulted in reduced cataloging costs or timely access to new technical reports for Laboratory researchers.

This approach did, however, provide some important information about how Los Alamos researchers make use of the library's online catalog. The ability to retrieve information about unclassified technical reports as well as books, journals, and audiovisual materials in the local library system has resulted in an increased use of technical reports from our collection. We find that if researchers can easily ascertain the existence of potentially relevant information sources using the library's online catalog, they request them from the library or use them in-house. Unfortunately, locating some technical report citations in the online catalog leads searchers to assume that not locating a given citation in the online catalog means that the item must not be available in the library. This incorrect assumption indicated a real need to expedite the retrospective conversion of the Report Collection's holdings and to fully automate library functions for this extensive research collection.

The third approach implemented at Los Alamos to improve bibliographic access to technical report literature was to expand the use of external commercial and government databases. Until most recently, few researchers have made effective personal use of these online services. Time is viewed as one of the most precious commodities of Laboratory researchers. Use of most online services still requires a considerable investment of time to learn efficient and effective querying methods and strategies, so many researchers rely on the skills of information professionals to reduce online search costs. The dissemination of commercial and government databases as CD-ROM products has eliminated telecommunications costs and expensive connect-time, but has added additional overhead costs for the library in the form of additional computer hardware and peripherals to support the CD-ROM format, additional networking requirements and associated maintenance and support for a CD-ROM network, as well as subscription costs for the CD-ROM products themselves. Subscription to a given CD-ROM product doesn't necessarily mean ownership of the bibliographic information either. In some cases, cumulations of bibliographic information acquired on a subscription basis must be returned to the vendor if the library decides to discontinue its subscription. Additionally, the CD-ROM versions of specific databases don't usually provide the same years of coverage as their online counterparts, so the library must still use online services and maintain expertise in these services in order to provide comprehensive bibliographic results to Laboratory researchers.

Expanding the in-house availability of external databases has increased the identification of potentially useful primary energy-related information sources, but a major retrieval element is still missing. Researchers must still go through a multi-step process that includes searching the library's online catalog or, in the case of technical reports, to either search the card catalogs or rely on library staff to determine the local availability of these sources. More importantly from a library perspective, these commercially developed bibliographic databases still lack the functionality essential to the internal management of the library's collections as well as the capability to tailor bibliographic information to a local needs.

The following are viewed as limitations of external databases:

- lack of local ownership information for specific citations
- lack of immediate status/availability of specific items
- lack of statistical data on local usage of materials for collection development purposes
- bibliographic information cannot be altered or augmented to meet local needs
- bibliographic information cannot be imported, created, or exported in the MARC format to facilitate exchange of cataloging records between institutions

- citations don't necessarily coincide with the library's holdings causing confusion for local users
- search/retrieval capabilities and methods can vary from database to database making it more difficult for users to become proficient searchers

In short, external databases do little to help library staff organize and maintain their institution's research collections or to enable researchers to quickly determine the availability of primary energy-related materials.

The limitations of each of the approaches cited above is challenging technical services professionals at Los Alamos to find new ways to integrate bibliographic records from multiple sources into the library's online catalog. While this may appear to be relatively straightforward, an in-depth examination of the standards by which bibliographic records are created in a given database quickly reveals that this process can be very complex.

### **Expected Benefits of Providing Access to EDB in the Local Online Catalog**

We expect that providing local access to bibliographic data for energy-related information sources via the in-house library system will result in more cost effective dissemination of that information to Laboratory researchers, will increase use of the library's research collections, and will facilitate the management of those collections. In 1993, the library sent out a survey to 1400 laboratory employees to get feedback on services from the Report Collection. An overwhelming number of respondents indicated that the collection was essential for their research. They also indicated strong support for automation of the Report Collection. The library was at that time migrating its in-house database to a new system, Geac Advance. Contract negotiations with the vendor also called for a second in-house library system for classified documents. A decision was made to include records for library collections in these two systems based on classification level and public releasability of the materials. This meant that all unclassified, unlimited access materials (books, journals, technical reports, audiovisual materials) would be in one system and all materials with use restrictions would be in a separate system. The implications of this decision are still being studied in light of significant differences between records developed using MARC and AACR2 standards and those developed according to COSATI or other cataloging guidelines.

As with the automation of the book and journal collections, library staff has identified potential sources of machine-readable bibliographic records that could be acquired to facilitate conversion of the technical report collection. A preliminary analysis of existing holdings indicates two major sources of bibliographic data: EDB for technical reports which have resulted from energy-related research, and DTIC (Defense Technical Information Center) records for technical reports which have resulted from defense-related research. We expect to match 65-70% of all technical report holdings to EDB records, making this database the primary starting point for retrospective conversion. For new OSTI microfiche acquisitions, we expect a 100% match, and at least a 95% hit ratio for technical reports acquired in paper format that report results for DOE-funded research.

In negotiating with OSTI for the purchase of EDB, we had to determine whether it was more cost effective to acquire and mount the entire database or to selectively acquire only records for technical reports. It is estimated that over half of all citations in EDB are for journal articles. For this reason, we have opted to acquire the entire database because it provides bibliographic access to "analytic" level citations for conference papers and

journal articles which facilitates identification of relevant sources of information in our book and journal collections as well as in the Report Collection.

### **Citation Database Concept in A Local Library System**

The concept of providing bibliographic access to external databases through a single system is not a new one. Academic libraries have been using this approach for many years to reduce the cost of online database searching in their locally developed systems. Until recently, few library automation vendors provided this capability. The Geac Advance system, recently implemented at Los Alamos has optional capabilities that enable the library to mount external databases. In Geac, citation databases are separate from the master bibliographic database which contains bibliographic records for materials in the library's collections. These citation databases are loaded at the system administration level and support read-only capability. This feature was initially designed to provide bibliographic access to commercial journal citation databases. Traditionally, the library has not provided bibliographic access to journal articles and other analytic level citations. The ability to load these citation databases on the local library system and to link these citations to library holdings expands the usefulness of the online catalog. The search/retrieval capabilities of the online catalog apply to the citation databases eliminating the need to learn new a search syntax. The citation database software in Advance assumes the mutual presence of certain types of data in the citation record and the master database record to generate information about the local availability of a given citation. Generally, the data elements used for matching is the ISBN (International Standard Bibliographic Number) or ISSN (International Standard Serial Number). Since technical reports rarely have an ISBN or ISSN associated with them, modifications to the matching algorithm will be needed to match report citations to the library's technical report holdings. The first implementation of the citation database capability in Geac Advance will be to link citations in EDB to our unclassified library holdings thus providing bibliographic access that cannot be created locally and which is not readily available in bibliographic utilities normally used for shared cataloging in a cost effective manner.

### **Mapping EDB Data Elements into a MARC-based System**

The bibliographic format used to create EDB records and the MARC formats represent different implementations of ANSI standard Z39.2, American National Standard for Information Sciences--Bibliographic Information Interchange. This standard specifies the requirements for a generalized interchange format that accommodates the description of all types of bibliographic materials. The standard does not, however, specify the content of the bibliographic record, nor does it assign meanings to tags, indicators, or data element identifiers. Such specifications are provided by particular implementations of the standard.

Successful integration records created according to different implementations of ANSI Z39.2 depends largely on the ability of the library to map each bibliographic field into the equivalent data field(s) of the library's local system. This has proven to be more difficult than initially expected with EDB records because of many differences between EDB descriptive cataloging and AACR2 descriptive cataloging rules as implemented in the MARC formats. The remainder of this paper presents the methodology used by Los Alamos to create loader specifications for the library's automation vendor.

An important aspect of successfully mapping data elements from EDB records to MARC was soliciting input from library staff with extensive experience in searching EDB through OSTI's ITIS system as well as via Dialog. Discussions with these experienced users focused on identification of specific data elements in EDB that provide unique retrieval capabilities which librarians routinely use in searching the ITIS and Dialog versions of EDB. Through an iterative process, each EDB data element was identified and a decision was made whether it would be indexed, and if so, in what index. In general, the indexing of specific EDB data elements was defined so that it followed the same indexing conventions used for similar data in the library's master bibliographic database. This effort resulted in the conceptualization of the online catalog search menu for the EDB citation database in the Advance system (see Figure 1). This visual representation will be used by the system administrator to set up the initial search screen for the EDB citation database.

<b>International Energy Citation Database</b>	
Use arrow keys to select search option.	
Enter first words of search and press <Enter>:	
<b>Browse citations by:</b> Subjects Broader terms Descriptors Major descriptors EDB subject categories International Patent Classification codes	<b>Keyword searches by:</b> Subject keywords Broader terms Descriptors Major descriptors EDB subject categories
<b>Titles (includes Conferences)</b>	<b>Title keywords (includes Conferences)</b>
<b>Authors</b> Corporate names Personal authors Conferences	<b>Author keywords</b> Corporate names Personal authors Conferences
	<b>Abstract and other notes</b>
<b>Number searches:</b> CODEN ISBN/ISSN Report numbers (includes accession, order, SuDocs, contract, patent, and technical report numbers)	

Fig. 1. Online Catalog Search Menu for EDB Citation Database in Geac Advance

Once the project coordinator had a good understanding of what experienced EDB searchers expected in search/retrieval capabilities, a careful analysis of the Advance system's search/retrieval capabilities was done. It is important to stress that loading EDB records into the local library system does not mean that users will have the same retrieval capabilities that are available in the online or CD-ROM versions of the database. The advantage to searchers is that they can use the same search syntax for the EDB citation database that they use in searching the library's online catalog. In cases where the Advance system did not provide an equivalent search capability, the library carefully considered whether lack of a specific capability warranted negotiating with the vendor to develop that capability for the local system instead of waiting for the development of that capability in a future release of the generic Advance product. The Advance system currently lacks the capability to do adjacency and proximity searching. Although, these search/retrieval capabilities are deemed essential by advanced users, it was felt that most users would not make use of such a capabilities thus eliminating the need to pay for the vendor for special programming at this time.

The next step in mapping EDB data elements to MARC involved acquiring a more in-depth understanding of the content of each EDB data element in order to identify its nearest MARC equivalent. It is important to stress nearest because few EDB data elements were identified as having an exact MARC equivalent.

Because we intend to load the entire EDB database, it was important to acquire as much documented information about the content of EDB records from the inception of the database to the present. A study of OSTI's past cataloging practices as identified in earlier compilations of OSTI's Energy Data Base: Computer Media Description for Data Exchange and in information gleaned from drafts of the OSTI's internal descriptive cataloging manual provided additional information about the development and use of specific EDB data elements which would have otherwise been overlooked. Additional clarification on OSTI cataloging practices was solicited from experts at OSTI. A representative sample of EDB records was studied and provided information on the presence of bibliographic data elements not previously documented by OSTI.

The loader specifications for the vendor include detailed instructions on what the incoming data would look like, whether the data were of a fixed or variable length, whether to parse the data into separate MARC tags or to take separate EDB data elements and map them to distinct subfields within a given MARC tag. Some EDB data elements contain information that is defined as system-generated display constants for specific MARC tags. In these instances, the loader specifications indicate how EDB data that is defined as a print constant in MARC will appear in incoming records so that it can be appropriately stripped in the load process. MARC records employ the use of indicators to convey additional information about the data in variable data fields. Each MARC variable field contains two indicators. The EDB format does not make use of indicators, so for each EDB data element, default indicator values were defined to eliminate the possibility that EDB records would fail the system's edit checks in the batch load process solely because of the lack of indicator values. Table-defined edit checks are an important method used in Advance to determine the data integrity of records loaded into the system. It provides some measure of control over the mapping of each piece of bibliographic data and will alert the system administrator to bibliographic data elements that are not recognized by the loader software.

It is important to note that the mapping process to convert EDB data elements to MARC-like fields does not provide MARC compatibility and adherence to AACR2 descriptive cataloging guidelines for most bibliographic data. It has been impossible to convert the content of personal, corporate and conference data elements in EDB to their MARC/AACR2 equivalents. The order and level of fullness for these data elements is substantially different. The structure and content of subject terms in EDB also exhibit major differences from Library of Congress subject headings which are utilized for subject retrieval in the library's master bibliographic database. These differences are best handled at the present time by loading the EDB database as a separate, distinct citation database with its own associated name and subject indexes and related authority records. This avoids the creation of multiple authority records in the master bibliographic for specific entities and subject headings because of different cataloging standards in the incoming source records. The integrity of the EDB database is preserved without having an adverse effect on the master bibliographic database.

Once written loader specifications were developed they were forwarded to our system vendor for comments. The vendor has worked closely with library to ensure that data from EDB records is mapped to MARC according to the library's specifications. This iterative process with the vendor has been extremely beneficial. It has provided a better

understanding of the Geac Advance system capabilities. In specific instances, discussions with the vendor have led to alternative ways to achieve the same end result without specialized programming. In the past, loader programs developed by the vendor according to the library's specifications, have been written in such a manner that only the vendor can make subsequent modifications. In the development of the EDB loader, the library vendor has implemented a table-driven environment that will allow the library to modify the field mapping locally. This will minimize costly vendor programming in the future and is expected to give the library more control over the loader's instructions. This capability will be invaluable during the current testing phase. It will also allow the library to easily modify the loader when and if OSTI implements new EDB data elements. Los Alamos and its vendor, Geac, are presently involved in testing the EDB loader software. We expect a minimum of two to three months to fully test the loader before making the EDB citation database available to Laboratory researchers.

## Summary

This paper has provided one library's approach to loading EDB records into its local library system. It is important that any library considering a similar project carefully evaluate how to undertake such an effort based on its local cataloging standards, the expected benefits to its online catalog users, and the capabilities of its local library system. The following points are general in nature, but bear reviewing throughout the planning and implementation of such a complex project:

- Involve as many of the library staff in the planning process to ensure that the majority of their expectations are identified and addressed.
- Study existing documentation to ensure that all data elements to be mapped are identified.
- Supplement information gleaned from document by soliciting clarification of data element contents from those who create them.
- Study existing documentation on the bibliographic data fields to which external data elements are being mapped carefully.
- Identify local cataloging practices that may influence the mapping of external data elements into the library's online catalog.
- Be flexible. Provide optional ways to map specific external data elements. Carefully evaluate the merits and limits of each option in terms of ease of mapping and value to the end user.
- Work closely with the library vendor to understand the local library system's search/retrieval capabilities and to ensure that the loader specifications are clear to the vendor.
- Ensure that data integrity mechanisms are implemented as part of the system or loader functionality.
- Carefully test each of the mapping specifications during the testing phase with the vendor to avoid having to reload major portions of the database
- Be patient, this is a long, iterative process.

---

Irma S. Holtkamp is the Special Projects Librarian for the Los Alamos National Laboratory Library's Report Collection. Most of her professional efforts are in cataloging where she has been instrumental in the development of local cataloging policies and procedures that compliment national bibliographic standards. She is currently the project coordinator for the implementation of the EDB citation database on the local library system and also coordinates library projects to automate both classified and unclassified technical reports in the Report Collection. She is also engaged in the library's efforts to create optical images of technical reports, to link those images to bibliographic information in the library's online catalog, and to deliver both to the Laboratory researcher's desktop.

This report has been reproduced directly from the best available copy.

It is available to DOE and DOE contractors from the Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831. Prices are available from (615) 576-8401.

It is available to the public from the National Technical Information Service, US Department of Commerce, 5285 Port Royal Rd. Springfield, VA 22161.

