A Reference and Planning Model for Library Online Public Access Catalogs

TR90-043

December, 1990

Jeanne C. Sawyer

The University of North Carolina at Chapel Hill
Department of Computer Science
CB#3175, Sitterson Hall
Chapel Hill, NC 27599-3175

A Reference and Planning Model for Library Online Public Access Catalogs

by

Jeanne Clifford Sawyer

A Dissertation submitted to the faculty of The University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Departments of Computer Science.

Chapel Hill

1990

Approved by:

Stephen F. Weiss, Adviser

John B. Smith, Reader

Jerry D. Campbell, Reader

JEANNE CLIFFORD SAWYER. A Reference and Planning Model for Library Online Public Access Catalogs (Under the direction of Stephen Weiss.)

ABSTRACT

This dissertation explores the problem of how library catalogs can be connected or combined, providing library users with access to materials beyond the local library collection while maintaining local autonomy and ease of implementation. The first step in solving the problem was to identify the fundamental ways library systems can be connected, and to determine the characteristics of each.

The Reference Model for Online Public Access Catalogs (OPAC Model) does this by categorizing the basic architectures into three models: centralized, distributed and stand-alone. The reference model provides a way to classify a system according to its architecture and identifies the basic characteristics that a system must have. The distributed model (DLN Model) is explored in depth because it is the least well understood of the models and because the characteristics of distributed systems must be standardized if such systems are to be connected effectively.

Whereas the OPAC Model defines the system architectures, the Library Systems Architecture Planning Model (LSAP Model) provides a tool for choosing among them. The system characteristics defined in the reference model are included to meet real-world needs, such as providing access to another library's holdings or preserving local autonomy. The LSAP Model follows from the Reference Model by making explicit the connection between a set of system characteristics and a set of environmental characteristics.

The concepts included in the Reference Model are new and untested, especially for the distributed architecture. Therefore a case study of the Triangle Research Libraries Network's system was included in the dissertation specifically to demonstrate that:

a) the reference model can be implemented, and

b) the implementation is reasonable and is an appropriate choice for that environment.

Verifying the LSAP Model was then necessary to demonstrate that the planning model works, i.e., that the Model accurately reflects expert judgements of appropriate choice of system architecture. In addition, verification of the LSAP Model further validates the Reference Model since the LSAP Model is built from the architectures delineated in the Reference Model. If those architectures were inappropriately defined, the LSAP Model could not work properly.

ACKNOWLEDGEMENTS

Like every other Ph.D. student, I had no concept of what I was undertaking when I began this dissertation. I could never have completed it without the assistance and support of many friends and colleagues. The following were of particular assistance:

My committee, for their help in making this a coherent piece or work and also for their patience and confidence that I would finish, despite my other personal and professional obligations.

Steve Weiss, my adviser who has helped me not only through this dissertation, but also through two master's theses.

The TRLN Board, who supported me in this undertaking while I was Director of TRLN and allowed the use of the BIS transaction logs and TRLN's computer.

The librarians from the TRLN staff and the test libraries who participated in developing and verifying the LSAP Model.

Gwyneth Duncan. The transaction log analysis never would have happened without her. She provided everything from programming assistance to moral support.

Miriam and Edward Clifford. I expected moral support from my parents, but they did much more. My mother got me through the statistical analysis of the transaction logs and also served as general editor for the whole dissertation. My father suggested the Multi-Attribute Utilities Model as the basis for my LSAP Model, and helped make sure I applied the technique properly.

Henry Sawyer. My wonderful husband never complained when I spent evenings and weekends working, and always encouraged me to keep going no matter how endless the process seemed.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION . . . 1
1.1 Three Library System Models . . . 3
1.1.1 The Stand-alone Library Model . . . 4
1.1.2 The Centralized Union Catalog Model . . . 5
1.1.3 The Distributed Network Model . . . 7
1.2 Dissertation Overview . . . 8

CHAPTER 2: LITERATURE REVIEW . . . 11
2.1 Distributed Computer Systems . . . 12
2.2 Distributed Library Systems . . . 17
2.2.1 Task Distribution . . . 18
2.2.2 System Distribution . . . 19
2.2.2.1 The Triangle Research Libraries Network . . . 21
2.2.2.2 History of Cooperation and Resource Sharing . . . 22
2.3 Conclusion . . . 24

CHAPTER 3: THE REFERENCE AND PLANNING MODELS . . . 26
3.1 The Reference Model . . . 26
3.1.1 Scope of the Model . . . 27
3.1.2 Reference Model for Stand-Alone Library Systems . . . 29
3.1.3 Reference Model for Centralized Library Systems . . . 30
3.1.4 Reference Model for Distributed Library Networks . . . 31
3.1.4.1 System Functions: Switched Level . . . 35
3.1.4.1.1 Search and Data Presentation Capabilities . . . 35
3.1.4.1.2 Session Control Capabilities . . . 36
3.1.4.1.3 User Information Services . . . 37
3.1.4.2 System Functions: Linked Level . . . 37
3.1.4.2.1 Search and Data Presentation Capabilities . . . 37
3.1.4.2.2 Session Control Capabilities . . . 39
3.1.4.2.3 User Information Services . . . 39
3.1.4.3 Physical Components . . . 39
3.1.4.3.1 Workstations . . . 40
3.1.4.3.2 Telecommunications Facilities . . . 40
3.1.4.4 Discussion . . . 41
3.1.5 Conclusion . . . 43
3.2 A Planning Model for Determining Library System Architecture . . . 43
3.2.1 Determining Proper Location in the Design Space . . . 46
3.2.2 Other Considerations in Selecting System Architecture . . . 52
3.2.2.1 Catalog Policies . . . 53
3.2.2.2 Circulation and Interlibrary Loan Policies . . . 55

3.2.2.3 Collection Development Policies 56
3.2.2.4 System Operation and Management 56
3.2.3 Conclusion 57

CHAPTER 4: CASE STUDY: A LINKED LEVEL IMPLEMENTATION OF THE DLN MODEL 58
4.1 Overview 59
4.2 Underlying Principles 60
4.3 The Distributed System 63
4.3.1 Functions 63
4.3.1.1 Search capabilities 64
4.3.1.2 Session Control 66
4.3.1.3 Status Information 67
4.3.1.4 Help 70
4.3.2 Data 71
4.3.2.1 Retrieval Set Organization 71
4.3.2.2 Display Formats 72
4.3.3 Physical Components 76
4.3.3.1 Processing facilities 77
4.3.3.2 Workstations 79
4.3.3.3 Telecommunications facilities 80
4.4 Conclusion 81

CHAPTER 5: CASE STUDY: AN ANALYSIS OF USAGE PATTERNS . . . 82
5.1 Other catalog use studies 83
5.2 Methodology 86
5.2.1 Procedure 88
5.3 Findings 90
5.3.1 Main Effects 91
5.3.1.1 Target Institution 91
5.3.1.2 Repetition Style 92
5.3.1.3 Interactive Effects of Repetition Style and Target Institution 94
5.3.2 Secondary Effects 94
5.3.2.1 Effect of the Search Type 94
5.3.2.2 Effect of Terminal Type 96
5.3.2.3 Effect of Terminal Location 98
5.4 Analysis 101
5.4.1 Main Effects 101
5.4.1.1 Target Institution 101
5.4.1.2 Repetition 102
5.4.1.3 Interactive Effects of Repetition Style and Target Institution 103
5.4.2 Secondary Effects 104
5.4.2.1 Effect of Search Type 104
5.4.2.2 Effect of Terminal Type 105
5.4.2.2 Effect of Terminal Location 106
5.5 Conclusion 107

CHAPTER 6: VERIFICATION OF THE PLANNING MODEL 109
6.1 Verification Methodology 109
6.2 Findings 111
6.2.1 Centralized Systems 111
6.2.2 Distributed Linked Level Systems 113
6.2.3 Distributed Switched Level Systems 114
6.2.4 Stand-alone Systems 115
6.3 Analysis 117
6.4 Modification of the LSAP Model 120

CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS 124
7.1 Significance of the OPAC Model 125
7.1.1 Importance to System Design 125
7.1.2 Advantages of the DLN Model Over the Other Models 126
7.1.3 Disadvantages 126
7.2 Significance of the LSAP Model 127
7.3 Importance of the Case Study 128
7.3.1 The Case Study Approach to Model Development and Validation 128
7.3.2 The Transaction Log Analysis of BIS 129
7.3.3 Recommendations for TRLN Libraries 130
7.4 Recommendations for Future Research 131
7.4.1 OPAC Model Development 131
7.4.2 Catalog Use Studies 132
7.4.3 Implementations 133
7.4.4 LSAP Model Development 134
7.5 Conclusion 135

BIBLIOGRAPHY 137

APPENDIX A: LSAP MODEL RATINGS 150
A.1 Factor Ranking 151
A.2 Factor Ratings 154
A.3 Importance Rankings of Experts 157
A.4 Attribute Measures of Experts 158
A.4.1 Centralized Architecture 158
A.4.2 Distributed Linked Architecture 159
A.4.3 Distributed Switched 159
A.4.4 Stand-Alone 160

APPENDIX B: LSAP MODEL VERIFICATION 161
B.1 Example Cover Letter 162
B.2 Additional Questions Included With Rating Form 164
B.3 Importance Weights for Test Libraries 165
B.3.1 Centralized Test Libraries 165
B.3.2 Distributed Linked and Switched Level Test Libraries . 165
B.3.3 Stand-Alone Test Libraries 165

B.4 Test Library Attribute Vectors Compared to Modified LSAP Model 166
B.4.1 Centralized Test Libraries 166
B.4.2 Distributed Linked Level Test Libraries 166
B.4.3 Distributed Switched Level Test Libraries 167
B.4.4 Stand-Alone Test Libraries 167

LIST OF TABLES

Table 5.1: Numbers of Searches at each Institution 90
Table 5.2: Frequency of Target Institutions by Originating Institution ... 92
Table 5.3: Frequency of Repetition Style by Originating Institution 93
Table 5.4: Frequency of Repetition Style by Target Institution 94
Table 5.5: Frequency of Search Type by Originating Institution 95
Table 5.6: Frequency of Search Type by Target Institution 95
Table 5.7: Frequency of Search Type by Repetition Style 96
Table 5.8: Frequency of Terminal Type by Originating Institution 97
Table 5.9: Frequency of Terminal Type by Target Institutions 98
Table 5.10: Frequency of Terminal Type by Repetition Style 98
Table 5.11: Frequency of Terminal Location by Originating Institution ... 99
Table 5.12: Frequency of Target Institutions by Terminal Location 100
Table 5.13: Frequency of Repetition Style by Terminal Location 100

LIST OF FIGURES

Figure 3.1: Stand-Alone System Conceptual Structure 30
Figure 3.2: Centralized System Conceptual Structure 31
Figure 3.3: Distributed System Conceptual Structure 33
Figure 3.4: Switched System Architecture 36
Figure 3.5: Linked System Architecture 38
Figure 3.6: Library System Design Space 45
Figure 3.7: Expected Use Attribute Rating 50
Figure 3.8: LSAP Attribute Vectors 51
Figure 4.1: Examples of Single Institution and Multi-institution Searches . 65
Figure 4.2: Session Menu 67
Figure 4.3: Multi-institution Searching Help 71
Figure 4.4: Multi-institution Index Search 73
Figure 4.5: Multi-institution Basic Display 74
Figure 4.6: Multi-institution Holdings Display 75
Figure 4.7: TRLN Network Structure 76
Figure 4.8: Network Access to TRLN Databases 78
Figure 4.9: TRLN System: Installed Equipment, 4/30/90 79
Figure 4.10: TRLN Network Configuration 80
Figure 5.1: Transaction Log 88
Figure 5.2: Output of Categorizing Program 90
Figure 6.1: LSAP Attribute Vectors for Centralized Systems 112
Figure 6.2: LSAP Attribute Vectors for Distributed-Linked Level Systems . 113
Figure 6.3: LSAP Attribute Vectors for Distributed-Switched Level Systems 115
Figure 6.4: LSAP Attribute Vectors for Stand-alone Systems 116
Figure 6.5: Original and Modified LSAP Attribute Vectors 121

CHAPTER 1: INTRODUCTION

Libraries, like other organizations, have been eager to adopt computer technology to improve the quality of their services and to make their business operations more efficient. Replacing card catalogs with online systems is an obvious application where both objectives can be achieved. Online public access catalogs offer faster, more flexible search and retrieval mechanisms than card catalogs, are better able to incorporate the vastly expanded amount and variety of information sources that modern libraries manage, and do not require the user to be at the same physical location as the catalog. At the same time, the automated systems allow libraries to share cataloging data efficiently while eliminating the labor-intensive task of filing cards. Online catalogs offer the same numerous advantages that automated information systems typically have over their manual predecessors.

However, online catalogs can be much more than simple replacements for card catalogs. Libraries have adopted computer systems as integral parts of their organizations, making it increasingly important to assure that full advantage is taken of the powerful capabilities that such systems can provide.

Several different models of online public access catalogs for libraries have been and continue to be developed. The differences are based on the types of services to be provided, the particular circumstances of the libraries involved, and the technology available at the time system development began. Attention has not been focused systematically on the basic issues of how library catalogs should be connected or combined, balancing such concerns as meeting specific local requirements and ease of implementation with the need to provide library users with access to materials beyond the local collection. This dissertation addresses these gaps by:

1) developing a descriptive, or reference, model for multi-library online public access catalogs, thus defining the basic architectures possible for such catalogs;

2) creating a planning model to assist library system designers in determining the architecture of choice for a given implementation;

3) examining an example distributed library network to demonstrate that the reference model can be implemented in a reasonable way;

4) testing the planning model in libraries that have implemented systems with the architectures defined by the reference model.

1.1 Three Library System Models

As with every type of automated system, the development of automated library systems has been evolutionary, with architecture largely dictated by the currently available computer and telecommunications technologies. The first library systems were intended to be entirely self-sufficient for a single library, integrating all necessary functions. Development of such systems was usually undertaken by large academic research libraries. However, the development and operational costs were too high for most libraries to undertake alone, so libraries, which already had a strong tradition of cooperation, began to create shared, multi-function systems operating in a centralized network environment [Stevens 1980; DeGennaro 1983].

The shift to cooperative networks was driven by the expectation that this type of resource sharing would result in more effective library service, as well as by economics. As the technology has continued to develop, the existence of inexpensive, small but powerful computers makes feasible the original goal of functionally integrated systems for individual libraries. At the same time, the ability to establish connections among computers to form networks makes resource sharing possible.

This evolutionary process can be viewed as resulting in three primary models for online catalog systems: the stand-alone library model, the centralized union catalog model, and the distributed network model. Each of these has its own strengths and weaknesses.

1.1.1 The Stand-alone Library Model.

The stand-alone library model is based on a single system for one library with one bibliographic database built and maintained for that library. To search the catalog, a user first establishes a direct connection to the library, then searches the database for any relevant items. Only the holdings of the one library can be searched. This model calls for the system to include technical processing, circulation, and public inquiry capabilities in a functionally integrated manner because the same database of the library's holdings is fundamental to all. Most currently available systems, at whatever stage of completion, are based on the stand-alone library model. Systems based on this model include the Northwestern Online Total Integrated Library System (NOTIS) developed at the Northwestern University Library, GEAC Integrated Library System, developed by GEAC Computers, Ltd., and LS/2000, originally developed at the Lister Hill Center for Biomedical Communications of the National Library of Medicine but reworked and marketed by the Online Computer Library Center, Inc. (OCLC).

The stand-alone library model offers the key advantage of independence to libraries using systems based on it. Each library is able to assess its own requirements for specific system features, capacity, cost, etc., and to select the system that most nearly meets those requirements. Similarly, the library can implement the system according to its own needs and budget, establishing its schedule for which functions will be installed and when. Independence, however, is also the primary disadvantage of the stand-alone library model: resource-sharing activities are not supported.

The stand-alone library model offers the additional significant advantages that systems based on it are already available for purchase, their functional capabilities are demonstrable, and libraries have considerable experience installing and operating such systems. Thus for libraries that require online catalog systems immediately and that operate with relative independence, i.e., do not have strong requirements for close coordination with other libraries, the stand-alone library model is most appropriate.

1.1.2 The Centralized Union Catalog Model.

The stand-alone library model is readily expanded to be a centralized, single system for a group of libraries, i.e., a union catalog. As in the stand-alone model, there is only one database; to search that database the user must establish a direct connection to the system. Rather than including the holdings of only one library, however, the holdings of multiple libraries are merged into one database. Once the connection is established, the user can retrieve records belonging to any or all participating libraries. Although extending the stand-alone library model to a union catalog model is straightforward intellectually, implementing systems based on the model is very difficult. The problem is not so much for technical reasons as for political reasons related to loss of local autonomy and dependence on remote (and therefore uncontrollable) organizations [Martin 1986; DeGennaro 1983].

Although the centralized approach has proved highly successful for shared cataloging systems such as OCLC, there are few centralized public access catalogs, perhaps because it is difficult for independent organizations to set up, fund, and manage such systems. Where such systems have been developed, the participating libraries are usually already related in some way, as are the separate campuses of a state-wide university system. The University of California MELVYL system, which provides an online public access catalog for all nine campuses of the University of California, is based on the centralized model [Lynch 1983].

Similarly, the Florida Center for Library Automation operates a centralized system for the libraries in the Florida university system. Instead of a single bibliographic file, however, the database is composed of separate files for the different participating libraries with a combined index to provide access to the group of individual files [Corey 1988].

The advantages and disadvantages of the centralized model are essentially opposite to those of the stand-alone library model. The libraries cannot operate independently, but rather must coordinate their activities closely. All costs of participating, developing priorities, setting equipment requirements, etc., must be determined collectively.

However, the centralized catalog is a union catalog, so mutual access to the collections of all participants is provided. A requirement for immediate, routine access to multiple library catalogs can alone be sufficient reason to select a centralized approach rather than a stand-alone library model. Until recently, the centralized, union catalog approach was the only practicable method for providing mutual access.

Since distributed networks are only now becoming feasible, it is not yet clear under what circumstances a centralized, union catalog system will continue to be the most appropriate choice. In the past, the ability to share the acquisition and operational costs of large, expensive computers made economy of scale an important advantage of shared, centralized systems. However, there is increasing evidence that those economies no longer exist [Mendelson 1987].

1.1.3 The Distributed Network Model.

The third model, a distributed network, consists of separate systems serving each library included in a network. The separate systems are interconnected so that users with access to any system in the network can retrieve records belonging to any or all member libraries. The nodes of the network could use identical software specifically designed to operate together or, in theory, could use totally different software linked through standard communications protocols such as those under development through the Linked Systems Project [Buckland 1987; Davison 1983].

The library community sees this model as a way of preserving local autonomy and providing primary service to its local clientele, while facilitating shared cataloging and other resource-sharing activities such as inter-system searching and interlibrary loan [Aveney 1984; Boss 1985; Casey 1970; Martin

1986; Stevens 1980; Swank 1971]. However, the characteristics of the model have not been explored other than at the most superficial level.

The library community, primarily interested in networks of heterogeneous systems, recognizes that although libraries need mutual access to their collections they will often have selected different online catalog systems. Before such diverse systems can be effectively connected in a network, however, a complete model is needed to describe necessary and desirable system interworking capabilities. The same model should apply to networks of either homogeneous or heterogeneous systems, although the task of implementing the model should be much more straightforward with homogeneous systems.

1.2 Dissertation Overview

The research described in this dissertation defines a reference model for library online public access catalog systems, called the Online Public Access Catalog (OPAC) Model. The OPAC Model identifies the fundamental ways library systems can be connected or combined, and defines the basic characteristics of each way. A planning model is also developed, called the Library System Architecture Planning (LSAP) Model, which enables library system designers to determine which of the architectures defined by the reference model, distributed, centralized or stand-alone, is most appropriate for a given implementation.

First, relevant literature is reviewed. Then the OPAC Model and the LSAP Models are presented. The OPAC Model, because it is a reference model, is descriptive in nature, providing a framework within which functional specifications for a specific implementation can be developed. The primary focus is the Distributed Library Network (DLN) Model because it is least understood and because a standard reference model is necessary before such networks can be readily implemented. Although the stand-alone and centralized architectures are outlined by the OPAC Model, a detailed reference model is not needed to enable successful implementations of those architectures.

Whereas the OPAC Model defines the basic types of system architecture possible for online library catalogs, the LSAP Model identifies the critical factors in determining which of those architectures is most appropriate. In addition, it provides a structured procedure for evaluating those factors with regard to a specific pair of libraries.

The DLN Model is illustrated by examining the only reported operational distributed network of highly coordinated online catalog systems, the Bibliographic Information System (BIS) developed by the Triangle Research Libraries Network (TRLN). The functional characteristics of that network are examined to demonstrate that the DLN Model can be implemented.

The second part of the case study analyzes the use of inter-institution and multi-institution searching capabilities in an operational environment. The purpose of this analysis is to provide additional insight into what characteristics of distributed networks are important by determining how one distributed library network was actually used. The analysis also demonstrates that the implementation is reasonable and is an appropriate choice for that environment.

The LSAP Model is verified by demonstrating its application to libraries with established systems corresponding to the centralized, distributed and stand-alone library system models. Finally, the significance of this research is evaluated and recommendations are made for further investigations.

CHAPTER 2: LITERATURE REVIEW

Development of the OPAC and LSAP Models was done, as it must be, in the context of both library science and computer science. However, the specific area of interest in library science is the application of computer systems to public access online catalogs, whereas the interest in computer science is on general characteristics of computer systems in relation to application design. The following sections establish the context of this work in these two major fields.

In both cases, the primary focus is distributed computer systems since such systems are new and not yet well understood. The literature on centralized and stand-alone systems is not reviewed in detail since such systems are common and very well understood. Early studies of application system design do not even discuss network architecture since a single computer with attached terminals was the only choice. Those studies proceed directly to discussions of functional design, database or file structure, input/output devices, etc. [c.f., Brooks 1969; Kroenke 1977; Shelly 1975]. Similarly, the library literature includes numerous references to centralized and stand-alone systems [c.f., DeGennaro 1983; King 1983; Lynch 1983; Segal 1985; Weir 1979].

The literature on reference models and the Multi-Attribute Utilities Model, which provide the basis for the OPAC Reference Model and the LSAP Model respectively, is discussed in the sections on those models. Similarly, the literature related to catalog use and transaction log analysis are discussed in the intorduction to the chapter where the case study analysis is presented.

2.1 Distributed Computer Systems

The computer science community has, of course, been very interested in distributed systems; the body of literature in that area is voluminous. Most of it, however, is not particularly relevant to those aspects of concern to libraries, dealing with environments that are substantially different or focused on various aspects of implementation.

Most of the distributed systems described either operate virtually autonomously, not requiring shared databases (e.g., electronic mail systems), or they operate within the confines of a single organization (e.g., office information systems or inventory control systems). The former are too loosely structured to accomplish the goals of a library network, while the latter require a higher level of coordination than is necessary or appropriate for library systems. In both cases, research addresses implementation issues (such as file allocation problems), methods of updating distributed databases, and telecommunications issues (such as line capacity and routing) [c.f., Bucci and Streeter 1979; Chang and Liu 1978; Chen and Akoka 1980; Chu 1969; Fisher and Hochbaum 1980; Storz 1982; Wah 1984]. These issues are, of course, related to the problem explored here, but are not the principal focus of this research.

In addition to concern with specific implementation issues, research on distributed systems deals with general purpose systems based on various assumptions about the nature of a typical application. For example, although a real-time, transaction based system is generally assumed, research efforts varied widely in the amount of data manipulated by a typical transaction, the size of the database, the value of a transaction (i.e., how important it is to be able to recover without loss), and the degree to which distribution of data should be invisible to the user [Rothnie and Goodman 1977]. In any particular study, the assumptions made tend not to be typical of a library online catalog. If, for example, one assumes that transactions have high value, as would certainly be true for a distributed network of automatic teller machines each updating customer-accounts, ~~considerable~~ attention must be focused on preventing loss and recovering properly in case of failure. Simply ignoring a lost transaction would not be acceptable. In the library environment, however, such a high degree of fault tolerance is not a required characteristic. Lost searches, while annoying, are not catastrophic and expecting a user to re-enter a search if there is a failure is not unreasonable. Assumptions such as these are so fundamental to the entire system structure that studies of such generalized systems are not relevant to library systems.

Efforts to define what constitutes a distributed system and how it differs from a centralized system are more pertinent, as are investigations of various non-library information system models. Although these efforts are not directly applicable to the library environment, they do provide guidance in terms of issues to be considered.

The definition of a distributed system and what distinctions, if any, should be made between distributed processing systems and distributed database systems is subject to considerable debate. Kaunitz [1984], for example, concerned with processing activities rather than data location, defines distributed systems as those having multiple nodes, each node either a computer or an intelligent terminal, where application specific processing occurs at more than one node.

Ceri and Pelagatti [1984], on the other hand, focus on database distribution as the key feature, arguing that there are two important aspects of a distributed system: distribution, where all data are not located at the same site (in contrast to a centralized database), and logical correlation, where the data must have properties in common (in contrast to a set of unrelated files or local databases that are resident on different computers in a network). They propose the following working definition for a distributed computer system:

> A distributed database is a collection of data which are distributed over different computers of a computer network. Each site of the network has autonomous processing capability and can perform local applications. Each site also participates in the execution of at least one global application, which requires accessing data at several sites using a communication subsystem. [p. 6].

It is not clear that a single definition of a distributed system is either necessary or desirable. However, because the term is used so broadly, simply describing a system as distributed does not provide sufficient information for the reader to understand the environment in which the system operates.

There have also been efforts to classify different types of distributed systems by the structure of the database (e.g., replicated data, partitioned data, central data) or by the nature of the processors involved (e.g., network of mini- or micro-processors connected to a central mainframe, network of co-equal mainframes) [Kaunitz 1984]. However, little effort has been made thus far to develop implementation independent descriptive models for various types of processing environments. Kaunitz [1982; 1984] suggests that the first step in developing such models is to identify and define the major components of the system such as workstations, processing facilities, and communication facilities. He has begun that task, viewing the model, in part, as the first step in the system design process. Kaunitz's model focuses on these components as basic building blocks.

Gibbs and Tsichritzis [1983] have done some work with data models for an office environment that is somewhat relevant to the library environment. Specifically, they find standard types of data models inadequate for the office environment in three ways:

1. the conceptual structure of the model fails to correspond directly to entities in the real world in terms of both data and operations,

2. the models do not include the capability of handling unformatted data such as text and video images, and

3. the models do not distinguish between internal processing formats and external representations of data.

Lyngbaek and McLeod [1984] have also considered data models for office systems with specific regard to distributed systems. Their model is based on objects (e.g., text, video images, formatted information such as names and phone numbers) and the relationships among objects. The model defines various types of objects and provides a set of operators for manipulating them.

Like Gibbs and Tsichritzis, Lyngbaek and McLeod find problems in applying most of the research on distributed databases to office systems. Specifically, they find inadequate handling of location transparency, inadequate accommodation of information sharing among otherwise autonomous databases, and inadequate facilities for managing individual databases without central control. They suggest an object oriented approach because it allows flexibility in granularity: all objects do not need to be at the same level.

There are few studies found in the computer science literature of research dealing with user interface aspects of end-user information systems, such as library systems, or with how such features are influenced by system architecture. However, computer scientists are paying increasing attention to general user interface issues such as dialog management and screen displays. Principles of user interface design identified by such research can be applied to designing the user interface in library systems [cf. Galitz 1981]. Recent efforts to develop user interface reference models for general purpose systems are also of some interest [Lantz 1986; Lynch and Meads 1986;

Sisson 1986]. Although these efforts are not specifically concerned with distributed systems or with library systems, they help identify issues to be considered, such as the need to characterize the external environment in which the model operates. These studies also provide an example of the analytical process necessary to develop a specific reference model.

2.2 Distributed Library Systems

The development of distributed library systems is a natural extension of the existing cooperative tradition of American libraries. Such long-standing activities include shared cataloging (i.e., creating bibliographic descriptions) as well as resource sharing, not only in the sense of interlibrary loan but also for cooperative acquisitions and storage of materials [Stevens 1980]. The national interlibrary loan code, for example, was first adopted in 1917. This cooperative tradition arose both from the service orientation of librarians and from economic necessity. Providing efficient bibliographic access to collections beyond those of the home library by using modern technology is an obvious extension of these activities.

Automated library systems have developed in an evolutionary manner, following the available technology rather than causing new technological developments. The relationship of available technology to the services to be provided is critical to librarians who would not be as intensely interested in distributed networks if the technology did not exist to create them. The primary reason for interest, however, is that distributed networks seem to offer the possibility of meeting resource-sharing goals, i.e., contributing to the common good, without sacrificing local service priorities as would probably be necessary in a centralized system. It remains to be seen whether these expectations are met as operational distributed library systems are developed and installed.

The National Library of Canada (NLC) and the Research Libraries Information Network (RLIN) are two organizations that have expressed particular interest in distributed library systems. NLC co-sponsored the Bibliographic and Communications Network Pilot Project with several other institutions to investigate options for a nationwide, decentralized library and information network for promoting resource sharing [Linking 1984]. The Research Libraries Group (RLG) has also conducted a study, in part to develop a distributed architecture for RLG [Schroeder et al. 1984]. Neither of these studies was directly concerned with communications between online public access catalogs, but they did begin the process of identifying required functions and possible mechanisms for providing those functions.

Since library systems could be distributed in various ways, it is necessary to understand distinctions among various systems, taking them into account when considering a particular system. Systems could be distributed by allocating various tasks among separate systems or by connecting separate complete systems into a network.

2.2.1 Task Distribution.

The Online Multiple User System (TOMUS), a single institution library system developed and marketed by Carlyle Systems, Inc., uses multiple processors each designed to accomplish a specific task. Designated processors handle terminals and the user interface, which includes parsing the command language and providing error messages, prompts and help screens. Separate database processors manage the bibliographic database, handling disk allocation, index creation, index searching, and record retrieval [Salmon 1986]. Barnholdt and Nybroe [1983] describe a circulation system that is distributed in a somewhat similar fashion, using a separate computer as a terminal handler.

The possibility for distribution by function within a single institution system is being explored by other librarians. Buxton [1984], for example, describes an approach to distributed processing where reserve processing is handled on its own computer, which then communicates with a host circulation system. Acquisitions and serials control are also functions that are sometimes handled by systems separate from but communicating with an online public catalog.

2.2.2 System Distribution.

The Irving project is a particularly interesting attempt to connect disparate online catalog systems into a network [Hooker 1985]. Irving was originally a consortium of four public libraries in the Denver metropolitan area, each using one of several different stand-alone, single institution, commercially available online catalogs.

The project links the disparate systems using a Common Network Language in which users enter their queries, specifying which system they wished to search. The network processor, a separate hardware and software component, translates queries from the common language to the language native to the target system. The results are then transmitted back to the user via the network processor, which displays the results in a common format.

An important restriction in this particular project is the requirement that no modifications could be made to any of the individual library systems [Hooker 1985; Luce 1984; Irving 1986]. Major problems occur for the user because the native languages of these systems include quite different capabilities. In addition, users get results that are not only unpredictable but also are apparently inconsistent from one system to another.

The term "user friendly" is perhaps overused, but the concept that it represents is crucial to the success of any system. It is in this area that the IRVING project faces especially serious problems, causing it to be described as a "horrendous kludge" [Aveney 1984]. A network environment is inherently more complex than a single, centralized system, making proper design of the user interface considerably more difficult. Substantial additional work would be needed to define the user interface for a network before the Irving project would be an effective distributed library system. Despite the limitations of the implementation, however, the Irving project represents a basis for considering the interconnection of heterogeneous systems.

The general approach of providing access to multiple online catalog systems by allowing users to select among available catalogs is also being explored by increasing numbers of libraries [Zimmerman 1990; Sweeney 1990]. One of the best examples is the Colorado Alliance of Research Libraries (CARL) system, which enables separate systems using the CARL software and Tandem hardware to interact in a distributed network. After selecting the desired catalog from a menu, users interact with the remote system via the host system on a pass-through basis. Since the network is homogeneous, command language differences and other problems of heterogeneous networks are avoided. The CARL system, however, allows only one catalog to be searched at a time, with no capability for creating merged, multi-institution retrieval sets [Shaw 1988].

2.2.2.1 The Triangle Research Libraries Network. The Bibliographic Information System (BIS) developed by the Triangle Research Libraries Network (TRLN) is the first operational system specifically based on the distributed network model. TRLN has been committed to implementing a highly interactive distributed system since its inception in 1979 [Bregzis and Knapp 1978; Cooperative Development 1979]. BIS is conceived as a single system that, besides providing local online catalog services for each node of the network, also provides inter-institution and multi-institution searching capabilities as an integral part of the library catalog system. That is, it provides a system with the ability to search any one of the catalogs in the network as well as the ability to search more than one catalog simultaneously.

The network consists of three computer systems, one located on each of the three member university campuses (The University of North Carolina at Chapel Hill (UNC-CH), Chapel Hill, NC; Duke University in Durham, NC; and North Carolina State University (NCSU), in Raleigh, NC). Each computer supports the online catalog for that campus, and can also access the other nodes of the network via telecommunications facilities.

2.2.2.2 History of Cooperation and Resource Sharing. The member libraries of TRLN have a long history of cooperation in several areas of library service. As early as 1933, Duke and UNC-CH formed a Joint Committee on Intellectual Cooperation, which included libraries as one aspect of its charge. In 1934, the libraries began duplicating main entry cards from their card catalogs to form a joint library catalog. Duke and UNC-CH have formal written collection development agreements dating to 1935, when responsibility for acquiring materials in various subject areas was divided according to institutional strengths. The Duke-UNC Inter-University Committee on Cooperation, established in 1953, urged that cooperative activities be expanded to include what was then North Carolina State College (now North Carolina State University). These early efforts have been continued and expanded, encouraging scholars to use the three collections as a single resource. The joint collection development agreements are supported by extensive interlibrary lending agreements that include direct borrowing privileges, special interlibrary loan and photocopy arrangements, and a daily truck service that moves materials among the libraries [Bregzis and Knapp 1979; Cooperative Development 1979; Byrd et al. 1985].

As the collection development agreements became more extensive and more detailed, the need for mutual access to bibliographic information about the collections became acute. In addition to their primary clienteles, the libraries provide services to many researchers affiliated with organizations in Research Triangle Park and elsewhere in the Triangle area. Since these non-university-affiliated library users are not physically located on any campus, they cannot use the combined resources of the libraries as a single collection unless they can quickly and easily determine which of the three institutions holds a particular item.

In 1978, Ritvars Bregzis and John Knapp were retained as consultants by the three libraries to recommend an approach for cooperative development of an automated bibliographic access system. They recommended that the three institutions undertake joint development of a distributed system as the most cost effective means of providing mutual bibliographic access while still allowing a suitable measure of local autonomy [Bregzis and Knapp 1979]. The libraries accepted the consultants' recommendations, applied for and received funding under Title II-C of the Higher Education Act, Strengthening Research Library Resources Program; in October 1979, development of the system began.

TRLN is, in many ways, an ideal setting to test the characteristics and appropriateness of a distributed architecture for library systems:

> the Triangle university setting, the resource strength, their national role in research, and their currently existing high level of cooperation in library resource building and usage argue strongly for the establishing of an effective solution to the bibliographic access problem as a high priority requirement. The existing wealth of resources and the resource delivery mechanism

> constitute an almost ideal situation for developing a solution to the bibliographic access problem, while the methodology of development could be of assistance in similar situations elsewhere. [Bregzis and Knapp 1979, p. 13]

A single institution, stand-alone online catalog would not meet the mutual access requirements dictated by the long standing collection development agreements and the diverse clienteles of the three institutions. Indeed, TRLN is a good example setting for a distributed network in part because the requirement for mutual access is so strong and so well defined. A centralized union catalog would have met this requirement, but the consultants recommended, and the libraries concurred, that the distributed system would be more appropriate for many of the same reasons that other types of organizations have elected to implement them [King 1983; Ceri and Pelagatti 1984]. The organizational advantages of considerable site autonomy, both operationally and in terms of cataloging policies, and the expected performance benefits resulting from parallelism, led the libraries to implement a distributed system [Sawyer and Duncan 1981].

2.3 Conclusion

The literature review demonstrates considerable interest in distributed systems and how they differ from other system architectures. However, there is little evidence of research on architecture-dependent functional characteristics necessary to make a system effective or on how to determine when an environment makes a particular choice of architecture appropriate.

The computer science literature addresses general characteristics of distributed systems, with primary emphasis on implementation issues. Although the generalized models provide some guidance on the kinds of characteristics that must be considered, they do not fit specific application environments. The computer science literature does provide definitions of distributed systems, which were helpful. More important, however, was the basic concept that an agreed upon definition in a specific context is necessary to provide a common frame of reference. The reference model developed here expands on this concept.

The library science literature emphasizes the long-standing tradition of cooperation among libraries, suggesting that libraries are an excellent environment to test the viability of distributed networks among independent organizations. There are existing examples of library systems that meet the basic requirements of each architecture described by the OPAC Model, but the number of examples is very limited, especially for distributed networks. Although these systems provide evidence as to what characteristics must be included in the reference model, they do not comprise the model.

There is considerable interest shown in distributed networks of online catalogs, with particular recognition that standards, such as provided by the reference model, are needed to make it generally possible to build such networks. However, even with the appropriate standards, networks are complicated and expensive to implement. There is little guidance in the literature on how to determine when the environment makes that choice of architecture appropriate.

CHAPTER 3: THE REFERENCE AND PLANNING MODELS

3.1 The Reference Model

The first step in exploring the issues of connecting or combining library catalogs is to develop a general, descriptive model for library online public access catalogs. Such a model identifies the fundamental ways such systems can be constructed and defines the basic characteristics of each. First, however, it is necessary to clarify what is meant by the term "model" since the term has been used in the literature to mean anything from a simulation of a particular environment to an example of how a given function might be accomplished [Buxton 1984; Korfage 1971; Montgomery 1983; Rouse 1979; Weir 1979].

A type of model that has attracted considerable attention recently, particularly within standards-making efforts, is called a "reference model". A well known example of this type of model is the Reference Model for Open Systems Interconnection (ISO 7498) of the International Standards Organization [International 1984]. A reference model is not an implementation specification, but rather the basis for such a specification. It represents a conceptual framework for identifying logically related components of a complex system and, at a general level, defines how these components relate to each other [Fong 1986].

The concept of a reference model is quite new, and existing reference models show considerable variation in structure. For example, the Open

Systems Interconnection Model is expressed in prose while the Dexter Hypertext Reference Model is expressed using a formal specification language [Halasz 1989]. Although the computer science community may establish specific criteria for what constitutes a reference model, at present the definition is quite broad. The type of model developed in this dissertation is a reference model that is expressed in prose.

The connection of reference models with standardization is particularly appropriate in this case. In part, the justification for developing this particular model is to provide a common understanding of the functions and data required to support effective interconnections of library systems. As has been frequently observed, most library networks will be mixtures of various commercially available systems, and it is only through standardization that the individual systems can be modified to communicate effectively with each other [Avram 1980; Boss 1985; Pope 1981; Segal 1985; Sugnet 1986]. Development of this Reference Model for Online Public Access Catalogs (OPAC Model) is a step towards creating the needed standards.

3.1.1 Scope of the Model.

The purpose of the OPAC Model is to provide an abstract structure that describes the relationships among independent libraries' online public access catalog systems. The Model categorizes systems into three basic architectures, each defined by its own reference model. Thus the OPAC Model comprises three parts: the Reference Model for Stand-Alone Library Systems (SALS Model), the Reference Model for Centralized Library Systems

(CLS Model), and the Reference Model for Distributed Library Networks (DLN Model).

The OPAC Model provides a way to identify the factors that must be considered in designing systems that comply with one of the three architecture models. It characterizes the basic functions and services required to provide effective public access to bibliographic and holdings information for each architecture. The OPAC Model also provides a way to classify a system according to its architecture, which can help reduce confusion when describing various systems. Only the DLN Model is explored in depth. Because the SALS and CLS Models are based on well-understood single system architectures [c.f., DeGennaro 1983; King 1983; Lynch 1983; Segal 1985; Weir 1979], the reference model is not needed before systems based on them can be implemented successfully. Standard system design techniques are sufficient, as evidenced by the relatively large number of operational systems in these architectures. In contrast, the DLN Model is not well understood and is needed to enable wide-spread implementation of distributed networks, especially of heterogeneous systems.

The model presented here does not include functions beyond the online public access catalog itself. Although system architecture has important implications for such additional functions as circulation, acquisitions, and serials control as well as such corollary functions as database creation and maintenance, such considerations are beyond the scope of this dissertation. The focus here is on the functions specifically supporting public retrieval of bibliographic and holdings information from multiple libraries.

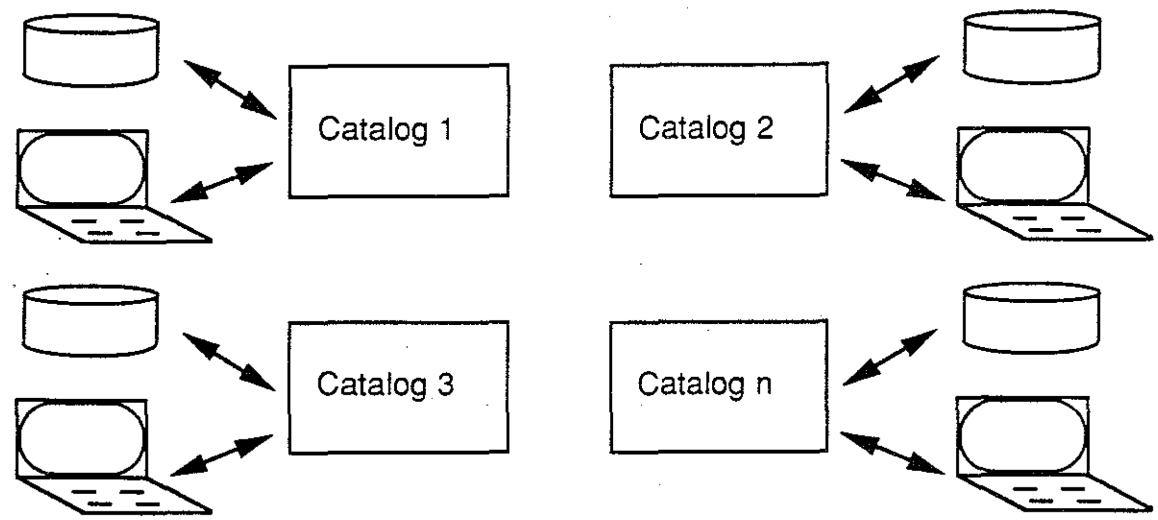
The OPAC Model has the general attribute of independence. The model does not require specific hardware, telecommunications facilities, operating systems, or applications software. Systems that comply with the OPAC Model have the following characteristics:

- Resilience: The model requires robust systems with the ability to manage hardware or software failure to cause minimal disruption to users.
- Responsiveness: The model requires an online transaction processing environment, with all search requests and retrievals handled in real time.

3.1.2 Reference Model for Stand-Alone Library Systems.

The SALS Model defines a system that provides online catalog services for only one library. Like the individual nodes in a distributed network, a single computer system provides search and retrieval of bibliographic and holdings information for its library. Unlike a node in a distributed network, however, stand-alone systems do not have the capability to allow users to search the holdings of other libraries. The conceptual structure of the stand-alone system is shown in figure 3.1.

Figure 3.1: Stand-Alone System Conceptual Structure.



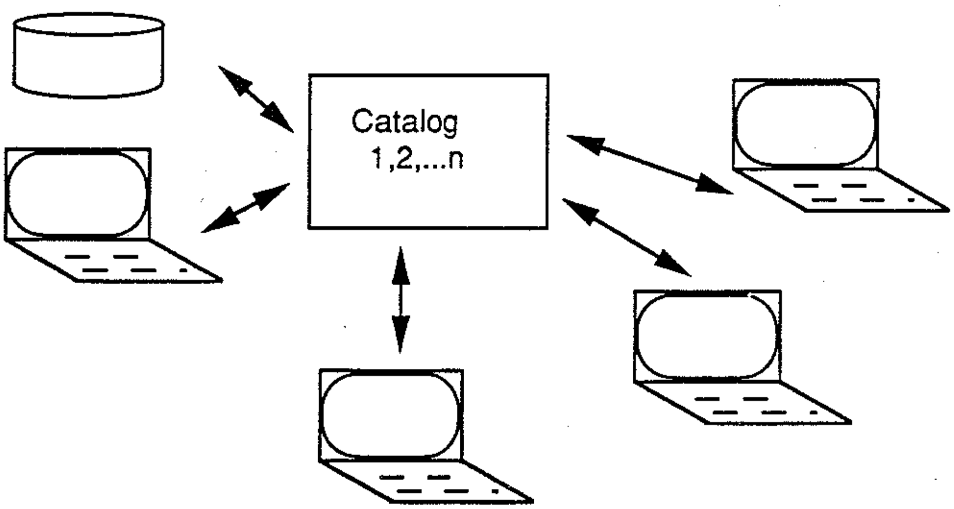
Each stand-alone catalog provides its own searching facilities, with no capability for searching the holdings of another library. Command language, data presentation, session control, etc., are all handled by that system with no requirement for coordination with any other system.

The SALS Model imposes no requirements on the physical components of the system. The type of computer used, the nature of the workstations, and the structure of the telecommunications facilities are all implementation decisions that can be made by the one library without need for coordination among multiple institutions.

3.1.3 Reference Model for Centralized Library Systems.

The CLS Model defines a system that uses one computer to provide online catalog services for multiple libraries. The holdings of the participating libraries are represented in a single database, enabling system users to search and retrieve bibliographic and holdings information for all of the libraries. A centralized system is much like a stand-alone system in basic architecture: a single computer system provides access to a single database. However, the database in the centralized system represents the holdings of more than one library. The conceptual structure of a centralized system is shown in figure 3.2.

Figure 3.2: Centralized System Conceptual Structure.



Like the stand-alone system, the centralized catalog provides its own capabilities for searching, data presentation, and session control without requiring coordination with any other system. However, the requirements for coordination are even stronger than for a linked-level distributed network. Not only must the librarians and system designers coordinate their activities, but they must agree on a single implementation design that meets the needs of all the participating institutions.

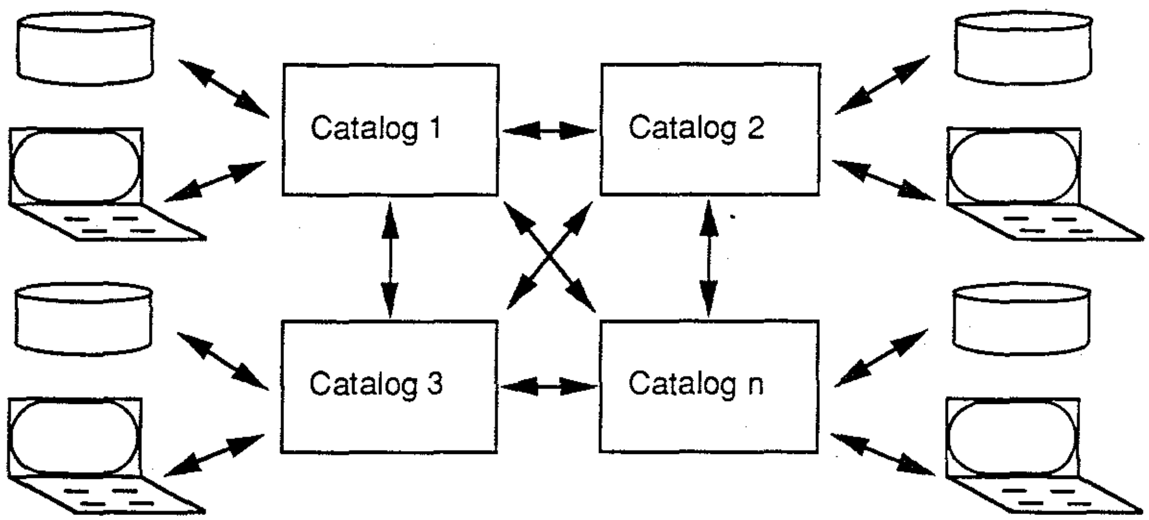
3.1.4 Reference Model for Distributed Library Networks.

The DLN Model, in accordance with the Ceri and Pelagatti definition given earlier, defines the distributed library network as an interconnected group of different computers each with data storage capability and each performing autonomous processing as well as participating in at least one global application. In the DLN Model, each node,or computer in the network, performs autonomous processing including all activities necessary to support the local online catalog as well as other strictly local applications, which are not required to be library applications.

The global application of each node allows search and retrieval of bibliographic and holdings information from other catalogs in the network. The nodes are not precluded from participating in additional global applications such as electronic mail or an interlibrary loan facility. However, such additional applications are not part of the DLN Model.

Each node provides online catalog services for one or more libraries, i.e., a centralized union catalog may be a node in a distributed network at the same time. The database is partitioned over all nodes in the network, with each node supporting the records representing materials held by the libraries using that system as their local catalog. The systems in the network interact as peer level systems, each with direct virtual access to all other participating systems. Physical connections depend on the structure of the supporting telecommunications network. The conceptual structure of the distributed system is shown in figure 3.3.

Figure 3.3: Distributed System Conceptual Structure.



The DLN Model provides for systems at two levels of interaction, the switched level and the linked level, based on the degree of connectivity and coordination of the network nodes. Systems interacting at the switched level are loosely connected, providing only basic system selection and pass-through capabilities to each other. This requires the user to switch from one catalog to another using the native language and capabilities provided by each catalog. Linked level systems operate in an integrated manner, providing general intersystem search and retrieval capabilities. This approach provides a consistent view of the network as a whole, letting the user search any or all catalogs in the network using the same user interface. Both levels are online transaction processing systems, operating in a real-time environment.

The two levels defined by the DLN Model specify the endpoints of a range of closeness of interaction for distributed library networks. The switched level provides complete access to remote system capabilities, but the host system provides little assistance to the user in dealing with remote systems beyond simple establishment of a terminal session on the requested target. The linked level provides a seamless environment that incorporates the catalogs of all participating libraries, thus removing from the user the burden of dealing with a variety of systems.

A system could be implemented to function at either level, or at some intermediate level of interaction. In addition, although performance would suffer, a switched system could simulate a linked system if the host had enough information about the target. This approach has been tried but has not worked well, possibly because of insufficient knowledge of the target system.

In a switched level connection, the host system is responsible for all aspects of the interaction, including requesting a terminal session and passing all communications to and from the target system. Since the interaction appears to the target like any other terminal session, no special facilities are required on the target system to handle requests from other catalog systems.

In contrast, the linked level requires both target and host systems to support appropriate software to manage the interactions. Target systems in a linked level interaction will return complete retrieval sets rather than screen images. Especially if the environment is heterogeneous, both target and host will need to support a standard command language, such as the intersite searching language specified for the Information Retrieval Application of the

Linked Systems Protocol, which in turn requires the OSI communication protocols.

To develop the model for each level, it is necessary to consider two aspects of the interaction. Function is the first consideration, defining what operations a user can perform as well as what support capabilities are necessary. The second consideration is the operating environment, including the different physical components necessary to support intersite interactions.

3.1.4.1 System Functions: Switched Level.

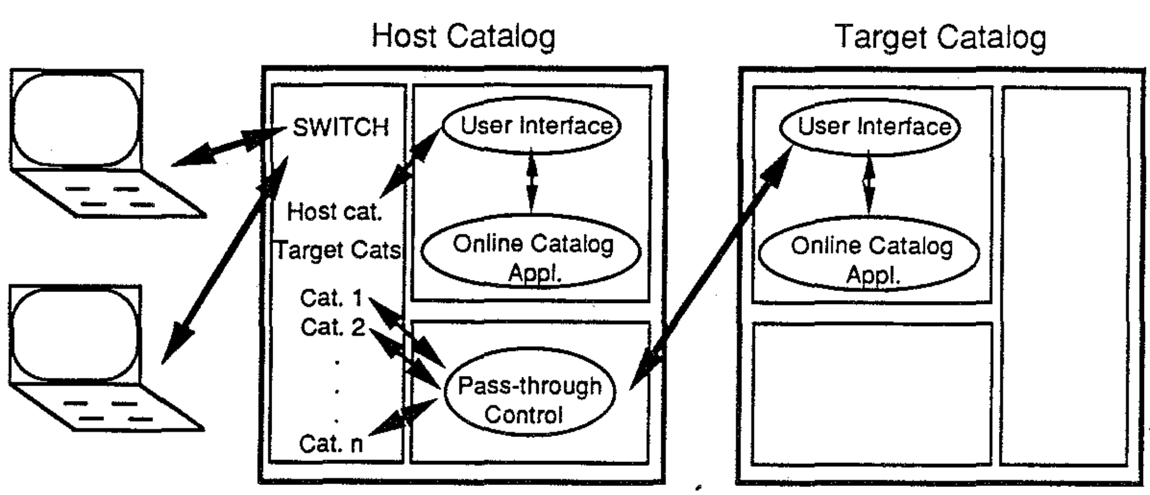
3.1.4.1.1 Search and Data Presentation Capabilities. Each participating catalog in a switched network provides its own searching capabilities: switched level systems do not provide coordinated searching capabilities. Inter-institution searching is provided by allowing users to switch from one system to another. Switched level systems do not support multi-institution searching or search forwarding.

Once the user selects an institution, the host system establishes communication with the system supporting that institution. Communication with the target system at the switched level is handled on a pass-through basis, with the user entering commands in the language of the target system as though using a terminal directly connected to that system. The host also transmits target system responses without interpretation. Data presentation, including display content and format as well as retrieval set order, is the responsibility of the target system. The user has access to any capabilities provided by the target system such as selecting alternate display formats.

The functions performed by the target are the same as it provides for all its terminal sessions, so no special capabilities are required for the target to support switched level connections.

The software that handles the system connections and pass-through can either be resident on the host system or can be further distributed to a separate processor. The diagram in figure 3.4 shows the basic structure of a switched system with the switching software included in the host.

Figure 3.4: Switched System Architecture.



<u>3.1.4.1.2 Session Control Capabilities</u>. Participating catalogs in a switched network establish terminal sessions on remote systems upon explicit user request. Once the remote session is established, all user input is transmitted to the designated system without interpretation except to remove or translate workstation-specific control information. The remote session is ended when the user requests that control be returned to the host system or transferred to another target system. If the user ends the session on the host

system, any remote session is also terminated. The host system is responsible for recovering from remote system crashes and telecommunications failures by returning the user to the host terminal session and providing appropriate information regarding what happened.

3.1.4.1.3 User Information Services. The host system in a switched level network is responsible for providing directory information regarding systems available within the network and for providing help information and system messages regarding establishing a session on a target system. Once such a session is established, all help facilities, messages, etc. are provided by the target system on a pass-through basis until the session is terminated.

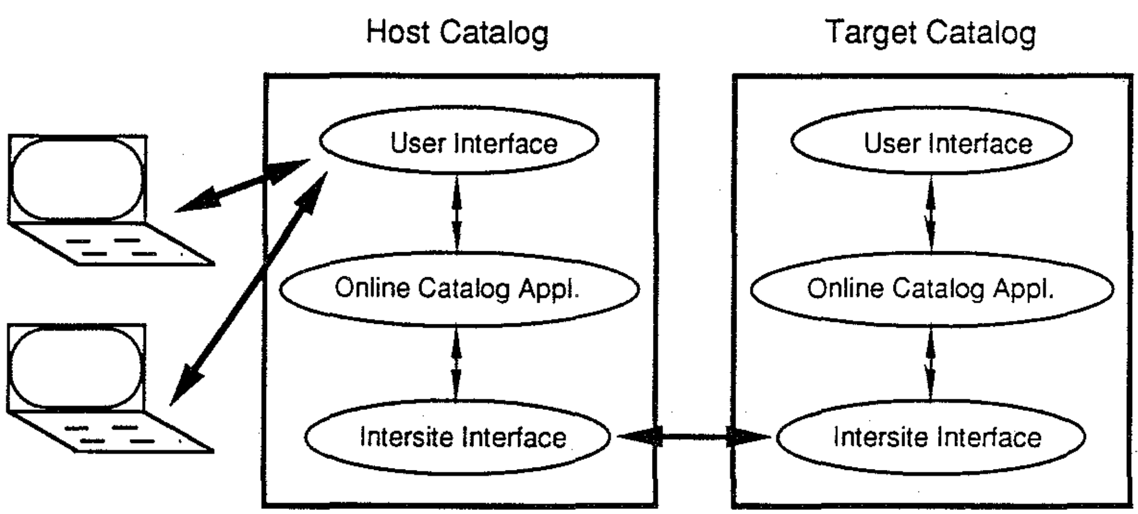
3.1.4.2 System Functions: Linked Level.

3.1.4.2.1 Search and Data Presentation Capabilities. Inter-institution searching among systems interacting at the linked level is handled in a standard way, regardless of the specific systems involved in the interaction. Users enter search commands in the language of the host system, which translates them if necessary into an agreed upon intersite searching language. Translation is not necessary if the interacting systems use the same native command language and both host and target systems agree to use the native language instead of a standard intersite language. All terminal interactions, including display formats, retrieval set order, and message wording are handled by the host system.

Retrieval sets and error messages are returned by the target system using agreed upon protocols such as those being developed by the Linked

Systems Project [Buckland 1987; Linked 1988]. The target system returns entire retrieval sets, including status information, rather than just screen images as occurs in switched level systems. Figure 3.5 shows the basic structure of a linked network. In this structure, the Linked Systems Protocol (LSP) could be used to provide the Intersite Interface. LSP uses standard OSI protocols for the lower six layers, and specifies the application layer functions for information retrieval and record transfer of bibliographic and authority data. Although the LSP protocols provide the required function, the choice of protocol for the Intersite Interface is an implementation decision and therefore is not specified by the DLN Model.

Figure 3.5: Linked System Architecture.



Multi-institution searching is supported by the DLN Model, and is implemented at the option of the host system. The Model does not limit the number of institutions that may be searched simultaneously. However, limits may be required as a practical matter in implementing such systems. It is the host system's responsibility to collect and merge retrieval sets from the target catalogs, and to display the results in a meaningful way.

3.1.4.2.2 Session Control Capabilities. The host system in a linked network is responsible for all aspects of supporting terminal sessions. Remote searching is handled by the host system establishing a communications connection with the appropriate target system(s) for transmitting search requests and receiving search results. It is the host system's responsibility to assure that results are returned to the correct user: the connections between host and target systems do not necessarily correspond one-to-one to terminal sessions. Connections are established and terminated under both normal and error conditions according to the requirements of the communications protocol adopted by the participating libraries. The host system is responsible for providing the user with appropriate information regarding success or failure in executing requested searches.

3.1.4.2.3 User Information Services. The host system in a linked network is responsible for the content and format of all user displays, including information services. Therefore, the host system provides directory and status information, all kinds of error messages, and general help facilities for the user. Host and remote systems provide status and error information to each other in accordance with the specifications of the adopted communications protocol.

3.1.4.3 Physical Components.

Although the general architecture of the distributed network is defined by the model, specific physical components of the network are implementation choices that are made individually by each participating node. Both the linked and switched levels of the model support heterogeneous networks.

3.1.4.3.1 Workstations. Workstations are provided according to the capabilities of the host system directly supporting each workstation. Linked level systems need no special accommodation for workstations since all presentation services are handled entirely by the host system.

Workstation support in switched level systems can be more complicated because of the pass-through communication method. Any system hosting remote sessions must be able to emulate some terminal supported by the target system. Since most online catalog systems support "dumb" asynchronous terminals, at least for dial-in services, this requirement should usually require minimal changes to the host system software.

3.1.4.3.2 Telecommunications Facilities. The model does not require specific telecommunications facilities or the use of a specific communications protocol. The participating libraries may select whatever equipment, capacities and protocols are most appropriate for their purposes. For example, switched systems may establish connections on demand, perhaps using the target system's regular dial-in capabilities or some type of data-switch.

Linked level systems, however, may require connections that are maintained whenever the participating systems are up. In a heterogeneous network, a standard communications protocol such as that provided by the Linked Systems Protocol is needed. Homogeneous networks may use either a standard protocol or system-specific facilities.

3.1.4.4 Discussion.

The general characteristics of the distributed network specified by the DLN Model were selected to provide maximum local autonomy while still meeting interworking requirements. The database and processing distribution allows each library to own and operate its online catalog system independently, with coordination required only for those functions that are truly part of network capabilities.

The DLN Model was established with two levels to allow libraries flexibility in determining the appropriate closeness of coordination for their particular situations. Since the two levels offer different advantages and disadvantages, a library must carefully evaluate the trade-offs to determine what blend of approaches will be most effective.

Choice of level is made separately for each target library since the same approach may not be appropriate for all libraries. A given library can support both levels of interaction, each with a different partner library. For example, a library could interact with a close neighbor at the linked level while interacting with more distant libraries on a switched basis. Thus, each pair of libraries determines whether mutual access is needed, and if so, whether the linked or switched approach is more appropriate.

However, the connection between two libraries in a pair may be asymmetric, i.e., when one of the libraries is host, the interaction is linked, but when the other is host the interaction is switched. Additionally, it is possible for a library to have a switched connection to a target library, where the target library not only has no inverse connection but also where the target is not even aware that the switched connection to it exists. An asymmetric connection could be useful when libraries have substantially different requirements for access to each other's collections. For example, a large library may not need any connection to the many smaller libraries located near it, but each of the small libraries may want to support a linked level connection with the larger library to make the larger collection readily available to its users.

The fundamental trade-off in the two levels is ease of use versus cost. The linked level system puts the burden of coordination on system designers and librarians to implement and operate the systems in the network and establish appropriate supporting policies. A linked level network requires substantial software development for all participating systems, higher capacity computing and telecommunications facilities, and staff trained to operate systems in a complicated network environment. Telecommunications and remote system failures can have serious effects since interaction with remote systems in the network is integral to each system. However, the linked level system allows users to interact with all participating catalogs as easily as their own, without having to understand different cataloging procedures or learn different system command languages.

In contrast, a switched network forces the user to cope with institutional differences. Systems in a switched network operate almost as stand-alone systems, with remote system interactions handled as a front-end to the online catalog. Network management is much simpler, and therefore less costly, than in the linked environment. Hardware and software requirements are also simpler, and little additional capacity is required to support network capabilities.

3.1.5 Conclusion.

The OPAC Model provides the conceptual framework for understanding the architecture choices when designing online public access catalogs. It further provides the designers of distributed library networks the basis from which they can prepare functional specifications for particular implementations.

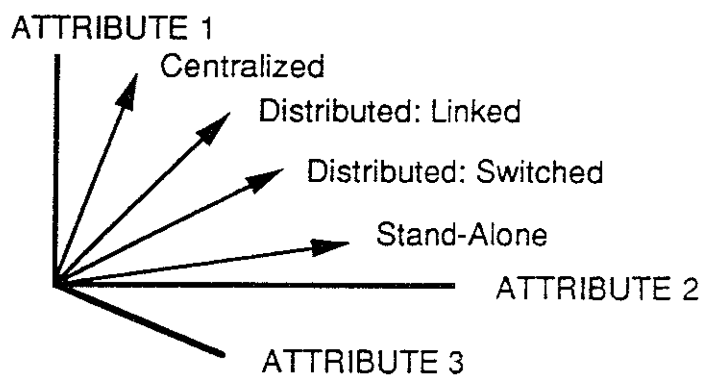
However, although the OPAC Reference Model defines the basic architectures available, it does not provide guidance as to when to use each one. Therefore, a tool is needed to help choose among them. This tool, called the Library System Architecture Planning Model (LSAP), is described in the next section.

3.2 A Planning Model for Determining Library System Architecture

Systems in distributed library networks, regardless of level, present different management issues than systems based on either the stand-alone or centralized model. There are differences not only in implementing and operating the systems themselves, but also in developing appropriate policies in other areas of library services. Libraries must consider carefully how closely coordinated their systems will be. First, it must be determined whether a distributed system is appropriate at all and, if so, which level will provide the best balance of costs and services for the particular libraries in the network. High levels of coordination make it easier for users to take advantage of the full network, but are more costly in administrative overhead and cause greater loss of local autonomy.

The DLN Model, together with the centralized and stand-alone models, can be viewed as fitting into a multi-dimensional design space within which all online catalogs fit. Each model can be represented by a vector of descriptive attributes, with the position of the vector defining where the model fits in the design space. Each axis of the design space represents an attribute; the stronger the attribute for a model, the closer the model's descriptive vector is to that axis. A concept diagram is shown in figure 3.6, illustrating the design space as though it had only three attributes. In this example, attribute 2 is much stronger in stand-alone systems than in centralized systems, so the vector for stand-alone systems is closer to the attribute 2 axis.

Figure 3.6: Library System Design Space.



The design space with its descriptive vectors for the models can be used as a decision-making tool for determining where in the design space a given implementation should be. The tool, called the Library Systems Architecture Planning Model (LSAP), and how it was developed are described in the following sections.

The LSAP Model is based on the outcome evaluation process known as a Multi-Attribute Utilities Model, which provides a general "framework and methodology that links inferences about states of the world, the values of decision makers, and decisions" [Edwards 1975, p. 140]. In this type of analysis, each outcome is located on a scale by measuring each significant attribute. The measurements are determined by the decision maker, and may be based on experimentation, judgement, observation or any combination of those.

Although the fundamental techniques for constructing the tool for a specific decision using the Multi-Attribute Utilities Model are well understood [for example, Clifford 1989, Edwards 1975], significant research is necessary to build and test the planning tool for making a decision in any specific domain. The LSAP Model is the result of a research effort to build a planning tool for determining library system architecture.

3.2.1 Determining Proper Location in the Design Space.

The steps necessary to build a planning tool based on the techniques of the Multi-Attribute Utilities Model are:

1) identify the possible outcomes of the decision making process;
2) identify the attributes, or key factors, that discriminate among the outcomes;
3) establish the relative importance of the attributes in discriminating among the outcomes;
4) establish the relationship between the possible outcomes and the attributes.

The result of these steps is a design space comprising all possible combinations of the attributes. Each possible outcome is represented by a vector whose component values are the attribute measurements that correspond to that outcome adjusted by the importance ratings. Thus the position of each vector in the design space is determined by the relationship between that outcome and the attributes.

The LSAP Model treats the two levels of the distributed architecture as distinct models. Thus, the possible outcomes in the LSAP Model are the four models of library system architecture defined by the OPAC Model: centralized, distributed--linked level, distributed--switched level, and stand-alone. The position of the models in the design space is determined by measuring eight attributes, or key factors, that observation suggested were discriminators of system architecture.

These attributes include service issues as well as political factors. The eight attributes measured by the LSAP model are as follows:

INTERNAL PURPOSE describes the strength of the library's purpose in considering a connection with another library. Internal purpose is a measure of how strongly a library wants an interconnection with another library, without regard for why such a connection might be wanted. For example, if the library is large with a strong collection, that library staff may see little reason to support a connection with a much smaller library. In this case, internal purpose would be rated as very weak.

USE describes the expected degree of use of access to the other library's catalog. For example, libraries with long-standing cooperative collection development agreements might expect use of their partner libraries to be very heavy.

GEOGRAPHIC PROXIMITY describes the physical closeness of the libraries. For example, libraries that are within walking distance of each other would rate geographic proximity as very near.

SCOPE describes the similarity in scope of the library collections and services, considering such factors as library type and size. For example, the scopes of two libraries from similar size liberal arts colleges would be much more similar than for an engineering school with a liberal arts college.

COOPERATIVE ACTIVITIES describes the existence of cooperative endeavors between libraries and their parent institutions that generate a need for access to the other library collection, such as joint teaching or research activities.

SUPPORT SERVICES describes the strength of existing or planned supporting facilities such as direct borrowing privileges, expedited inter-library loan, inter-library bus services for users, or special document delivery services.

LOCAL AUTONOMY describes the extent to which the library is able and willing to limit its independence to assure the coordination necessary to support interaction with another library catalog. It may be important to maintain local autonomy because of issues such as whether diverse automated systems are already installed, differences in governance (e.g., public versus private institutions), and differences in funding levels or control of funds.

EXTERNAL INFLUENCES are factors outside the library that affect the decision of whether to establish the connection, and if a connection is to be established, what the nature of that connection should be. Such factors could include political requirements, as might occur if interconnections were mandated by a library's governing body, or economic expedients, such as availability of funding for cooperative projects. The possibility that the external influence is so strong and specific that it determines the entire decision is excluded: if that circumstance should arise, the decision is already made and there is no use for a planning tool.

Cost is not considered an attribute because it does not discriminate among the architectures, but rather depends on implementation decisions beyond choice of architecture. For example, a centralized system could be more or less expensive to implement than a distributed system, depending on other implementation choices.

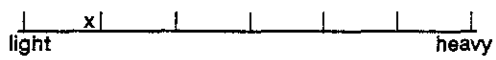
The locations of the four outcome vectors in the design space of the LSAP Model were determined by agreement of a group of four professional automated library system designers. The group received a brief explanation of the four architectures and definitions of the eight attributes. They were asked first to rate the relative importance of the attributes and then, for each architecture, to measure each attribute on a seven-point scale. For example, the experts thought expected use of another library catalog would be very light in an environment where the stand-alone architecture was most appropriate, so they rated the Expected Use attribute for the stand-alone architecture at 1.7, as shown in figure 3.7.

Figure 3.7: Expected Use Attribute Rating for Stand-Alone Systems.

USE

USE describes the expected use of access to the other library's catalog, if the two catalogs were interconnected in some way. The more heavily users are expected to use multi- or inter-institution searching capabilities, the closer the mark should be placed to the "heavy" end of the line.

Expected use of other library catalogs



The importance rankings were determined through discussion and negotiation until agreement was reached. The attribute measures were obtained through a similar process, although in this case the experts first indicated their independent judgements by simultaneously placing their fingers on a flipchart with the scale drawn on it. They then discussed differences and moved their fingers until everyone agreed on the proper rating. The ratings were then transferred to the printed form for easier manipulation. The negotiating process was adopted because allowing the experts to discuss the issues and modify their opinions to reflect a collective judgement would result in a more effective model than averaging their independent judgements.

The attributes were then measured as the percent of the distance from the "weak" end of the scale. For example, the Expected Use rating for stand-alone systems was marked at 17mm from the "light" end of the scale. When divided by 123mm, the total length of the scale, and multiplied by 100 to express the number as a percent, the rating for Expected Use was 13.8.

The adjusted ratings were then multiplied by the importance weighting factor for each attribute to yield the final descriptive vectors for each architecture. Thus for the Expected Use Example, the adjusted rating of 13.8 was multiplied by the importance weight of .210, giving an attribute measure of 2.90. Figure 3.8 shows the vectors of attribute measures for all four architectures. The forms used by the experts, together with the details of their importance rankings and attribute measures, are given in Appendix A.

Figure 3.8: LSAP Attribute Vectors.

Attribute	Centralized	Distributed Linked	Distributed Switched	Stand-Alone
Internal Purpose	9.39	8.28	5.21	2.05
Use	16.04	13.31	7.85	2.90
Geographic Proximity	9.53	7.89	6.35	5.01
Scope	1.66	1.42	1.14	0.96
Cooperative Activities	10.33	9.58	8.04	1.66
Support Services	4.31	3.93	2.79	1.23
Local Autonomy	2.05	7.22	13.49	15.05
External Influences	10.95	8.40	5.30	3.21

To use the Model, a library system planner evaluates each of the attributes with regard to the circumstances of his library's need for interconnection with some other specific library. The planner ranks the relative importance of the factors and also evaluates the strength of each attribute using the same scale as was used to establish the position of the outcomes in the design space.

The importance weights and rankings are combined to create a descriptive vector, which represents the position in the design space most suitable for the library. This vector, describing the library's requirements, is then compared to the vectors that describe the possible outcomes of the model, i.e., the architectures, using a standard cosine matching function. The cosine measure of similarity is used because experimentation has shown that it has the necessary properties for comparing vectors of this type.

Specifically, it is a normalized measure, it behaves properly even with missing attributes (as might occur if a library considered one of the attributes irrelevant for its circumstances), and it is easy to calculate [Salton 1983].

The most closely matching vector in the design space determines which architecture is most appropriate for that pair of libraries. This procedure is demonstrated in Chapter 6, where the model is verified by creating the descriptive vectors for a number of libraries with already existing systems of the different architectures.

3.2.2 Other Considerations in Selecting System Architecture.

The LSAP Model is general, intended to apply to all academic libraries. It suggests the system model that is most desirable based on general service and political requirements. However, the architecture indicated by the LSAP Model is only a starting point for more detailed systems planning since unique circumstances affecting a particular implementation could override the decision indicated by the model. For example, a library with an existing stand-alone system may find implementation of a switched distributed architecture to be the only practical choice in providing access to another library catalog, even if the service requirements indicated that a closer interaction was warranted.

Similarly, although the DLN Model allows for heterogeneous networks, a given implementation may not. In that case, to participate in the network, a library would have to adopt the specified online catalog system, possibly sacrificing other system capabilities. Such restrictions may merely be inconvenient or may have serious financial implications, as would be the case if the library already owned a system and the network implementation required a different one. In some cases, libraries may be prevented by law or policies of their parent institutions from making some types of agreements such as to purchase specific types of equipment.

In addition, there are other issues beyond the system itself that libraries must consider in planning a specific implementation. These issues are not discriminators that help determine which architecture is most appropriate and therefore are not part of LSAP itself. Rather, they are general supporting policies necessary to assure that the chosen system functions as an integral part of the library as a whole and that full advantage can be taken of the entire set of system capabilities.

Thus the issues for additional consideration discussed here have two possible uses, depending on the circumstances of the particular library. First, although typically they are not, in certain cases these issues may be architecture discriminators. If so, evaluation of the importance of the factor and its strength in comparison with the attributes of the LSAP Model will suggest whether an architecture other than that suggested by the LSAP Model is a better choice for the particular library. Second, once the library has chosen an architecture, the policies discussed in this section should be analyzed and changed appropriately as part of the system design process to provide the best possible environment for the catalog system.

3.2.2.1 Catalog Policies. To make inter-institution and multi-institution search capabilities as effective as possible, the bibliographic record for a given item should be indexed the same way for all participating catalogs.

Otherwise, it would be possible for an item to be represented in multiple catalogs, but not be retrieved by the same search. Since centralized systems normally use a single database, uniform indexing policies are readily enforced by the system itself. In contrast, coordination of cataloging policy is relatively unimportant for stand-alone systems since users do not routinely use them together. For distributed systems, however, especially those implemented at the linked level, cataloging policies are very important and require a conscious decision to enforce. Two aspects to this problem must be considered: the contents of the catalog record and the way system indexes are built.

For linked level systems, uniform indexing is critical to creating accurate merged retrieval sets, which by their existence imply a degree of uniformity to users. Since indexing is controlled through the systems, once the agreement that records will be indexed the same way at all sites is established and the decisions are made as to what that way will be, the systems in a distributed network, or even stand-alone systems, can guarantee uniform indexing. Although uniform indexing is desirable in switched level systems, the effects of differences are not as severe since users are interacting with individual systems in the network.

Except in centralized systems, controlling the content of the catalog record is more difficult, and in certain cases, would not be desirable despite the possible confusion in multi-institution searches. One of the advantages of the distributed approach is the flexibility for librarians to establish policies that best serve their own primary clientele. In addition, requiring closely coordinated cataloging increases the cost of cataloging. However, users are unlikely to recognize when cataloging differences have caused items to appear differently in various catalogs or to understand why there are differences. The more heavily multi-institution or inter-institution search functions are used, the more frequently differences in cataloging will create confusion, causing users to make incorrect judgements about the location of material. Therefore, librarians must consider carefully to what extent cataloging policies should be coordinated.

3.2.2.2 Circulation and Interlibrary Loan Policies. If they have not already done so, the libraries involved in a centralized system or distributed network must determine whether they will establish reciprocal borrowing agreements that will allow users to check out materials from other participant libraries, and if so, what the specific policies will be. Of course, it may be necessary to establish different policies for different categories of users, just as the library has different policies internally.

In addition to reciprocal borrowing agreements, libraries should consider whether adjustments in their interlibrary loan policies are needed and whether any special arrangements for delivering materials should be made. In general, interlibrary loan policies among libraries are well standardized, but with a network or centralized system, it may be appropriate to establish an expedited procedure for requests from other participants to get material delivered more quickly. If arrangements are not made to get materials efficiently to and from users at all participating libraries, the justification for providing a system that provides mutual bibliographic access comes into question.

3.2.2.3 Collection Development Policies. The availability of complete bibliographic access to other library collections makes possible very closely coordinated collection development policies. Even without such systems, many libraries have established cooperative collection development programs. However, such agreements have usually been limited to defining general areas of responsibility and have primarily been concerned with expensive research materials. Only where users can locate materials in the other library collections as easily as in their own and where librarians can determine quickly and accurately whether a given item is already in one of the other collections, can libraries implement more refined collection development programs that significantly extend the range of materials mutually available. Thus the availability of complete, immediate bibliographic access to other library catalogs creates an opportunity for stretching a library's collection development resources in a way that otherwise would not be possible.

3.2.2.4 System Operation and Management. Centralized systems require a central operations and system management organization, which must be jointly supported by the participating libraries, whereas stand-alone systems are operated independently by the owning library. Although the nodes in a distributed network can largely be administered separately, there are also many issues that must be resolved jointly, particularly among linked level systems. The libraries participating in a network must establish a governance structure and assure that appropriate staff will be available to operate the network. Although the systems themselves may be configured and operated by the owning institution, coordination of some activities is necessary. For example, the telecommunications facilities that connect the nodes in the system must be planned cooperatively, communications protocols must be selected, and all nodes must be equipped to use those facilities. Similarly, responsibility for resolving any problems that affect more than one node in the network must be clear.

3.2.3 Conclusion.

The LSAP Model is a tool to help determine which system architecture, as defined by the OPAC Model, is most appropriate for a specific situation. It provides a structured, quantitative approach to evaluating those factors that are critical to system architecture, making it easier for system designers to decide which approach is best for their circumstances.

This chapter has presented the two models. However, both are new and untested. The next two chapters present a case study of the TRLN system to demonstrate that the OPAC Model can be implemented in a reasonable way. Chapter 6 describes how the LSAP Model was verified.

CHAPTER 4: CASE STUDY
A LINKED LEVEL IMPLEMENTATION OF THE DLN MODEL

The OPAC Model provides the framework within which online catalogs of the four basic architectures can be implemented. As discussed earlier, there are numerous examples of stand-alone and centralized systems, and the techniques for implementing such systems are well understood. Distributed networks, especially highly interactive ones such as defined by the linked level of the DLN Model, are rare and it is not obvious that effective implementations of the DLN Model can be built. Therefore, to explore the linked level of the DLN Model further and to demonstrate that the Model can be implemented, this chapter examines one implementation of the Model, TRLN's Bibliographic Information System (BIS).

Of interest are those functional characteristics that are fundamental parts of supporting the distributed network. The focus is on network-based facilities, primarily multi-institution and inter-institution searching functions. General capabilities of the system are included only to the extent necessary to understand the functions that are relevant to the distributed nature of the system. As the only operational network of this type, it provides an important example of how the DLN Model at the linked level can be implemented.

4.1 Overview

The functions that support the distributed aspects of BIS are a fully integrated part of the system, which allows library users to search one or more bibliographic databases by author, title, subject, call number or various control numbers (such as ISBN) [Triangle 1988; Triangle 1989a; Triangle 1989b]. At the time of writing, additional searching capabilities were being designed, including keyword access, the capability to combine search terms using Boolean logic, and the capability to restrict search results by "limiting factors" such as language and date of publication. The system is primarily command driven although prompts are displayed in many circumstances to remind the user of the most probable next step, and menus are used in several cases. The multi-institution and inter-institution search commands are specified by adding a parameter (the institution's name) to the search command normally used for searching the home institution.

The results of a BIS search can be displayed in any of a variety of formats: index entries only (called the INDEX display), a brief bibliographic display with location information (the BASIC display), a complete bibliographic display (FULL), or a complete holdings display giving detailed information on specific volumes held by the library (HOLDINGS). Simple extensions of these same display formats allow for merged retrieval sets resulting from multi-institution searches.

Although many users gain considerable experience with the system, there are always many more who have little or no experience. Users generally receive no formal instruction in system use, although aids such as brochures and flip charts are available at many terminals. The system includes online help, which allows users to request explanations of system capabilities. Contextual help is also available to help the user understand the system response to a particular command and determine what the appropriate next command might be.

4.2 Underlying Principles

Besides the general principles of software design that are expected in any modern system, BIS is based on several principles specifically addressing the issues that arise in the distributed environment. These principles were developed specifically for TRLN where, as shown by the background information above, the very close coordination of various library activities was long established. The principles enumerated below derive from the general principle that the system should be easy to use, thus making consistency, predictability of results and user control of terminal sessions of particular importance.

1. All three systems should behave identically in all aspects.

The three systems were to behave identically in all aspects, not just those specifically concerned with inter-institution functions. Although restricting the independence technically possible in a distributed network and not required by the DLN Model, the principle was adopted for two key reasons. First, BIS was to be homogeneous to simplify initial implementation as well as ongoing software maintenance. Second, identical system function would promote ease of use in an environment where users not only searched the holdings of the other libraries from their home system, but also often travelled to the other campuses and used the catalogs directly.

This principle resulted in identical display formats and identical message and help screen wording, so that users find the same information in the same place on the screen regardless of which system is used. The principle also requires that records be indexed in an identical manner, i.e., the individual libraries could not choose different mappings of the MARC fields into the indexes.

Because the complexity of the MARC record is such that there are multiple logical options in selecting which fields should even be indexed, and if indexed, how the indexing should be done, stand-alone online catalogs usually allow libraries to determine how their records will be indexed. This has resulted in considerable variety from one system to another.

Such variability is disruptive in a distributed network, however, since identical records may not be retrieved in all catalogs in the network in response to the same search. Not only is such behavior inconsistent and confusing to users, it also results in users incorrectly believing materials to be unavailable at a library that does actually hold them.

2. The command for specifying a multi-institution or inter-institution search should be as much like the command for specifying a search of the home institution's catalog as possible.

Ease of use was again the motivation for this principle. If searching other catalogs in the network could be accomplished by an intuitive extension of the command language for searching the home institution catalog, users would be more likely to take advantage of the distributed network's capabilities than if they were required to learn a special set of commands. Since the whole idea of searching other catalogs would be new to library users, system designers were particularly concerned that the process be as straightforward as possible.

3. Minimal rekeying should be necessary to forward searches from one institution to another.

It was considered unnecessarily burdensome for users to rekey searches when the identical search was to be carried out serially in the different catalogs.

4. The display format for search results should be basically the same for multi-institution searches as for single institution searches.

As with the command language, ease of use argued for keeping display formats as similar as possible regardless of whether one, two, or all three institutions were being searched.

5. In multi-institution searches the bibliographic record should be displayed only once with location information attached for all holding libraries.

Although the distributed network would maintain separate bibliographic records for each online catalog, and those records might contain somewhat different information, displaying multiple copies of records for the same item would be confusing and make navigation through retrieval sets more difficult. An exception to this principle was made for displays specifically designed for library staff who needed an easy way to examine each library's version of the record.

4.3 The Distributed System

4.3.1 Functions.

As indicated earlier, BIS is essentially command driven. Users may enter searches, manipulate the displays of resulting retrieval sets, establish session parameters for controlling the way the commands are interpreted, request information about any of the systems in the network, and request explanations of system functions. This section describes the functions available in BIS that are directly affected by the distributed system environment.

4.3.1.1 Search capabilities. BIS was designed to support two basic types of searches called items searches and index searches, for both single institution and multi-institution retrievals. Both kinds of searches are typical of many online catalog systems, although a variety of terms are used to describe and to invoke them. An items search is a search that retrieves bibliographic records directly. The result of an items search is the creation of a retrieval set. The way specific systems handle these types of searches varies considerably.

In BIS, when an items search is successful, the retrieval set can be displayed in any of several formats; the user can page back and forth through the set, changing display formats as desired. When no items are retrieved, a message stating that no items were retrieved is displayed and the search is treated as an index search.

Index searching, often called browsing, results in the alphabetical display of index terms by which records can be retrieved rather than the direct creation of a retrieval set. The BIS display was designed to begin at the point in the index most nearly matching the search argument entered by the user. From there, the user can scan forward and backward through the index until the desired term is found. The user can select one or more terms from the index display, causing the system to create a retrieval set of all bibliographic items indexed by the selected term(s).

In agreement with the principles described earlier, the commands for specifying multi-institution searches are exactly like the commands for specifying single institution searches except that the user must specify which catalogs are to be searched. Unless session parameters are changed or the search statement specifies otherwise, searches are executed as items searches of the home institution. The user specifies which index (e.g., author, title or subject) is to be searched.

Figure 4.1 shows some examples of search commands as they would be entered for a single institution as compared to multiple institutions. In the examples, command terms are shown in full form although abbreviations are allowed as shown by the underlining. The system assumes search arguments are right truncated unless the user enters an explicit non-truncation symbol, so it is not necessary to enter complete names, titles or subject headings.

Figure 4.1: Examples of Single Institution and Multi-institution Searches.

Search Type	Single Institution	Multi-institution
Items	<u>au</u>thor asimov	all <u>au</u>thor asimov ncsu <u>du</u>ke <u>t</u>itle foundations edge
Index	<u>in</u>dex <u>au</u>thor asimov <u>in</u>dex <u>du</u>ke <u>subj</u>ect expert systems	<u>in</u>dex all <u>au</u>thor asimov <u>in</u>dex <u>du</u>ke ncsu <u>subj</u>ect expert systems

Once a search has been executed in any catalog, it can be forwarded for execution in one or both other catalog(s). This allows users to search their home institution first, repeating the search at another catalog only if the search of the home institution was unsuccessful. This capability promised to be quite convenient since, if the materials were available there, it was assumed users would generally prefer to use their home institution where it would usually be quicker and easier to get the material. A search can be forwarded by re-keying the search command with the appropriate change in institution parameter, but, if the user has done no other search meanwhile, a search can also be forwarded by keying just the institution name or abbreviation. Thus, for example, if a user at UNC-CH enters "subject expert systems" and does not find what is needed, the search can be repeated at the other two institutions by keying only "duke ncsu". Both items and index searches were structured to be forwarded in this way.

4.3.1.2 Session Control. In addition to entering institution parameters explicitly in search commands, it is convenient for users to be able to request multi-institution and inter-institution searches by setting a session parameter. A session parameter controls how all searches are executed from the time the parameter is set until either the parameter is reset or a start or end command is entered, indicating a new session. Thus for example, if a user changes the institution session parameter to all three institutions, all subsequent searches would be conducted as multi-institution searches of all three catalogs (unless an overriding institution parameter is then entered explicitly in the search). Start and end commands automatically reset all session parameters to the default, which, for the institution parameter, is the home institution. An end command is automatically executed after ten and a half minutes of inactivity.

Session parameters can be set directly using the session command, or can be set from a menu. For example, to set the session parameter to conduct all searches as Duke searches, the user enters "session duke". To get the session menu, the user enters "session menu", then follows the instructions on the menu to set not only the institution parameter but also any other session parameters. The session menu for general (i.e., non-staff) users at UNC-CH is shown in figure 4.2. The only differences for other institutions and for staff users is in the defaults, which are indicated by the asterisks.

Figure 4.2: Session Menu.

```
ses menu

                    SETTING SESSION PARAMETERS
Current Settings: Display Type: BASIC          Review Method: ITEMS
                  Search Type:  (has not been set)Institution: UNC

To leave current settings as they are, press "ENTER" or enter a new search.

To set parameters for the entire session, enter a number from each group below
for the parameter you wish to set, leaving spaces between numbers, then press
"ENTER". (The next screen will let you set a parameter for Search Type.)

    Display Type: *1=basic  2=full   3=holdings
                  4=tagged  5=technical (holdings)

    Review Method:          *6=items  7=index

    Institution:    8=Duke   9=NCSU  *10=UNC  11=Duke+NCSU  12=UNC+Duke
                   13=UNC+NCSU          14=all three

To use the automatic settings, or defaults (marked with asterisks), type "ses",
then press "ENTER". No session parameter will be set for Search Type.

To set Search Type without changing other settings, press "NEXT".
```

4.3.1.3 Status Information. As with any computer system, nodes in the TRLN network could be unavailable for a variety of reasons (software failure, hardware failure, routine scheduled maintenance, new equipment installation, communication line failure, flood, fire, etc.), and the duration could be minimal or extended. One of the complications that occurs in the distributed environment, however, is the possibility that nodes in the network may be unavailable at different times. The unavailability of one node does not prevent the continuing operation of the others, which is one of the advantages of the distributed approach, but questions are raised as to what information about the unavailable nodes should be provided to users of the still operational ones.

It is necessary to distinguish between two cases when no items are retrieved because the user's response would be different. The first case is when the catalog has no entries that match the search. In this case, that no items were retrieved may indicate that the library does not own the material the user is seeking, or it may indicate that the user should reformulate the search. The second case is when no items are retrieved because the catalog is unavailable, as would occur when equipment or telecommunications fail. In addition, it is important that users be able to determine when an unavailable system is expected to be back up (to the extent that it can be predicted) to avoid unnecessary frustration. For example, a message saying that users should try again in a few minutes would be entirely appropriate sometimes, but not when system managers were well aware that the system would be unavailable for several hours or more.

These requirements are addressed in BIS in several different ways. If a user enters a multi- or inter-institution search, and one or more of the catalogs requested is unavailable, a system message to that effect is displayed. However, if the home institution is unavailable, no communication with the user is possible, so such messages cannot be displayed.

The message wording is varied according to which catalogs are included in the search and which catalogs are available. For example, an attempted inter-institution search of the Duke catalog while it is down results in the message, "The Duke online catalog is temporarily unavailable; try again in a few minutes." Several alternative wordings are available for cases where the system is expected to be down for longer.

However, if the search is multi-institutional and one of the non-home institutions is unavailable, the situation becomes more complicated because there is a choice of whether to continue the search of whichever catalogs are still available or stop the search entirely. BIS leaves the decision to the user by displaying a message and asking for additional instructions. For example, a multi-institution search of Duke and NCSU results in the following message if the Duke catalog is unavailable: "Duke is unavailable; type "y" and press "ENTER" to continue with NCSU." If the user does not want to continue, any system command can be entered at that point.

BIS makes no attempt to notify the user when a system comes back up. Since the system cannot tell accurately which users previously attempted searches requiring the unavailable system, it was considered more disruptive than helpful to interrupt users with this type of message.

When the catalogs involved in a search are either available or down when a search is first entered, explanations are straightforward, but if a system goes down while a search and its retrieval set manipulations are in progress, the situation gets quite confusing. The system identifies the retrieval set contents by a list of record identification numbers, but does not retrieve specific records until they are needed for display. Therefore, for example, a request for the next screen of items in a multi-institution retrieval set would require records to be retrieved on the remote system, which would be impossible if the remote system had crashed meanwhile. It is difficult, if not impossible, to handle this gracefully: BIS displays a message showing which catalog(s) have just become unavailable, leaving it to the user to re-enter the search for available institutions if that is desired. The user cannot continue to work with a retrieval set that includes records from an unavailable catalog.

A user can also ask explicitly for status information by entering the status command. The screen displayed in response to this command includes a statement about the availability of each system in the network. The specific text is determined by the system manager, thus allowing the flexibility to provide different messages about when an unavailable system was expected to be back up. Because a full screen is available, in contrast with the single line available for system messages such as those described above, considerably more detail can be provided. The news command is an alternative means for users to obtain information about the systems. Besides current status information, users can request information about database coverage, scheduled downtime, new features and future plans.

4.3.1.4 Help. Three types of online help are available in BIS. Specific help allows the user to request explanations of various topics, contextual help provides information about what the user has just done and possible options for next steps, and location help furnishes descriptions of the various libraries on each campus such as their hours, types of collections and locations. No changes to the help system itself were necessary to accommodate the network environment, but many additional help screens were provided to help the user learn how to use the multi-institution and inter-institution search features. Figure 4.3 shows one of the specific help screens that explains multi-institution searching.

Figure 4.3: Multi-institution Searching Help.

```
help institution

            SEARCHING THE DIFFERENT INSTITUTIONS' CATALOGS

BIS can search the catalogs at Duke (du), NC State (ncsu), UNC-CH (unc) or all
three at once (all). There are a number of ways to request this:

WHEN YOU ENTER A SEARCH:
    Begin your search with "du", "ncsu", and/or "unc", or "all".

TO "FORWARD" YOUR LAST SEARCH TO A DIFFERENT INSTITUTION:
    Type the institution name(s) or "all" and press "ENTER" (without
    retyping the rest of the search).

EXAMPLES:
    To search Isaac Asimov at Duke and UNC: du unc au asimov isaac (ENTER)
    To repeat the same search at NCSU:       ncsu (ENTER)

To learn how to do a series of searches at another institution(s),
press "NEXT".

To see the other Help screens available, press "HELP".
To return to where you were before requesting Help, press "ENTER".
If you need further assistance, consult a reference librarian or the manual.
```

4.3.2 Data.

4.3.2.1 Retrieval Set Organization. BIS always organizes retrieval sets as indicated by the type of search, generally in accordance with the American Library Association filing rules [American 1980]. Thus, for example, the retrieval set for an author search is sorted first by author's name (retrieval sets frequently include works by more than one author), then all works by the same author are sorted by title, and all works with the same title are sorted by publication date. The distributed environment created no special circumstances that affected retrieval set organization, given the underlying principle that bibliographic records were to be displayed only once with holdings statements attached.

4.3.2.2 Display Formats. Users can display the results of BIS searches in a variety of different formats. The system responds initially to a search with a display determined by the search type (items or index) and the current display parameter (which could be set by default or be explicitly set).

An index search results in the display of a list of index terms beginning with the term matching (or preceding, if there was no exact match) the search argument entered by the user. The number of items matching each index entry is then listed so that the user can begin to determine whether the search should be broadened or narrowed.

In a multi-institution search, a merged display is created, but index terms are listed separately for each institution that has items matching the term. This is somewhat in conflict with the principle of "collapsing" bibliographic records to display only once, but was considered necessary to avoid a worse predicament. Specifically, the only feasible option other than the one chosen was to display the term once, but the number of items would include duplicates. Response time considerations precluded determining the size of the merged retrieval set when the index display was constructed. Library staff believed that the option with duplicates counted would be more confusing than separate listings because they felt most users were not conscious of the duplicate elimination process and thus would not understand why the system listed a term as having a certain number of items, yet when that term was selected, the actual size of the retrieval set was smaller. An example of the results of a multi-institution search involving all three institutions is shown in figure 4.4.

Figure 4.4: Multi-institution Index Search.

in all sub information retrieval

SUBJECT HEADING INDEX

INDEX#		INST	#ITEMS
1	Information retrieval.	UNC	47
2	Information retrieval.	Duke	16
3	Information retrieval.	NCSU	12
4	Information retrieval--Addresses, essays, lectures.	UNC	1
5	Information retrieval--Addresses, essays, lectures.	Duke	1
6	Information retrieval--Addresses, essays, lectures.	NCSU	1
7	Information retrieval--Bibliography.	UNC	1
8	Information retrieval--Bibliography.	NCSU	2
9	Information retrieval--Case studies.	UNC	1
10	Information retrieval--Congresses.	UNC	7
11	Information retrieval--Congresses.	Duke	3
12	Information retrieval--Congresses.	NCSU	4
13	Information retrieval--Periodicals.	UNC	1
14	Information retrieval--Periodicals.	Duke	2
15	Information retrieval--Periodicals.	NCSU	1
16	Information retrieval--Popular works.	UNC	1

Index#'s marked with an * exceed the system limit of 1000 items.
To retrieve records, enter the Index# or Index#'s, then press "ENTER".
For INDEX screens in sequence, "NEXT" or "PREVIOUS".

Once the retrieval set is created, either by selecting terms from an index display or by entering an items search directly, the system permits the user to display the records in any of several formats. The summary format is a very brief (usually two line) display of bibliographic information designed for quick scanning, i.e., its primary purpose is to aid in retrieval set navigation. The basic display is the general purpose bibliographic display. It includes the information usually needed to identify an item and determine its location. An example of a basic display resulting from a multi-institution search where all three institutions hold the item is shown in figure 4.5. A full bibliographic display is also available, including more complete bibliographic information. Holdings information in the full display is shown in the same way as in the basic display. There is also a detailed holdings display available, showing information about specific copies, volumes, etc. An example of the holdings display is shown in figure 4.6. The user can switch at will from one format to another or from one record to another.

Figure 4.5: Multi-institution Basic Display.

```
all sub expert systems (50 items retrieved)

        BASIC DISPLAY

Item 3 of 50                                    SYSTEM NO.: #13215722
SUBJ   : EXPERT SYSTEMS (COMPUTER SCIENCE)
TITLE  : Applying expert systems in business /
AUTHOR: Chorafas, Dimitris N.
IMPR   : New York : McGraw-Hill, c1987.
LANG   : English
 UNC-CH LIBRARIES
LOC    : Davis                        CALL#: HD30. 2.C475 1987
 DUKE UNIVERSITY LIBRARIES
LOC    : Fuqua School                 CALL#: HD30. 2.C475 1987
 NCSU LIBRARIES
LOC    : D.H. Hill Lib.               CALL#: HD30. 2.C475 1987
SUBJECT HEADINGS BY WHICH THIS RECORD MAY BE SEARCHED:
LCSH   : Management--Data processing.
This record is continued on the next screen.

For BASIC screens in sequence, press "NEXT" or "PREVIOUS".
For HOLDINGS of an item, type "ho" with the item no., then press "ENTER".
For FULL display of an item, type "fu" with the item no., then press "ENTER".
To go to the SUMMARY, type "su", then press "ENTER".
```

Figure 4.6: Multi-institution Holdings Display.

all sub expert systems (50 items retrieved)

HOLDINGS INFORMATION
Item 3 of 50 Screen 1 of 1 for this item

AUTHOR: Chorafas, Dimitris N.
TITLE : Applying expert systems in business /

UNC-CH LIBRARIES
Location : Davis
Call no. : HD30.2 .C475 1987
Holdings: copy 1

DUKE UNIVERSITY LIBRARIES
Location : Fuqua School
Call no. : HD30.2 .C475 1987
Holdings: copy 1

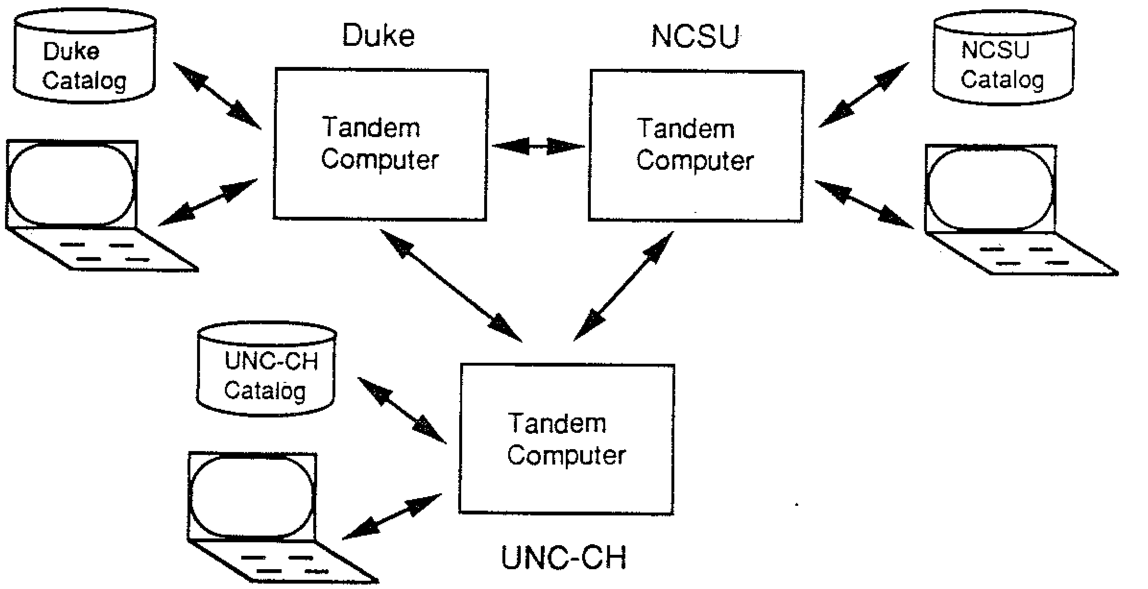
NCSU LIBRARIES
Location : D.H. Hill Lib.
Call no. : HD30.2 .C475 1987
Holdings: copy 1

To go to the BASIC or FULL display, type "ba" or "fu", then press "ENTER".

4.3.3 Physical Components.

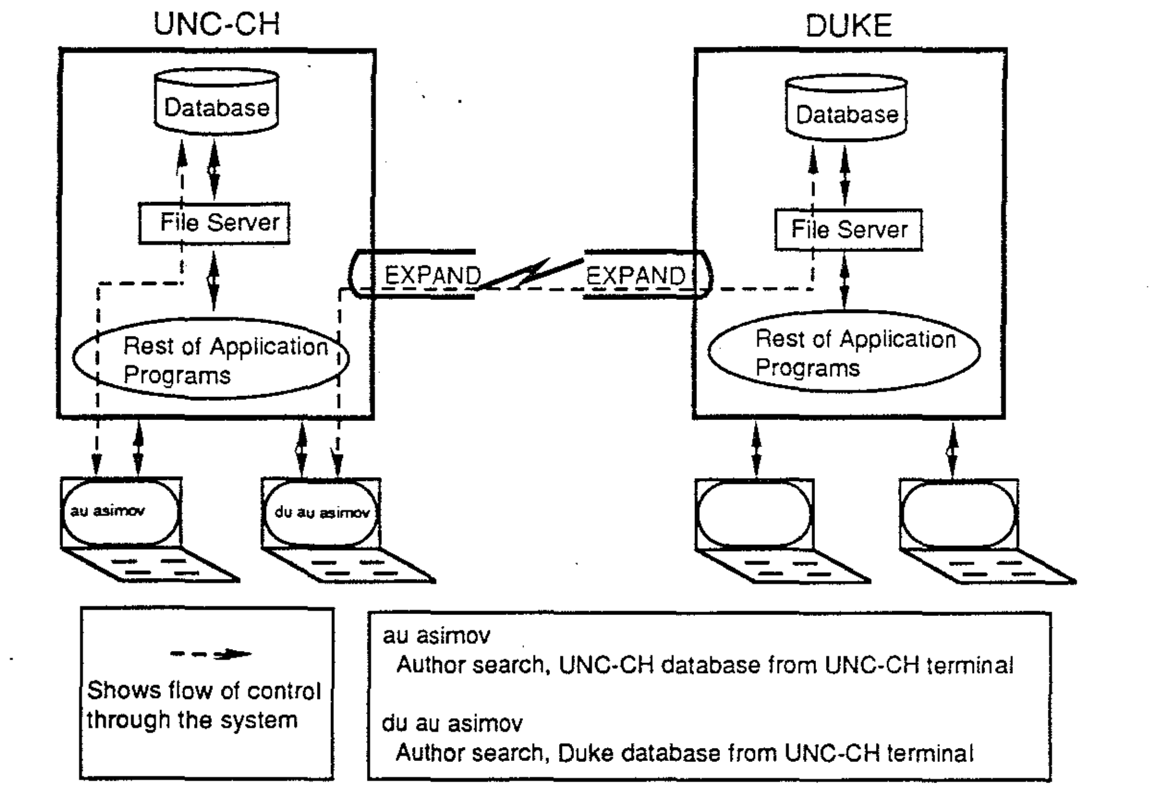
BIS is set up as a distributed network consisting of three nodes, one located on each campus. The computer at each node supports the bibliographic database for the institution at which it is located, all supporting files and indexes, and a complete copy of the application software. The network was designed and implemented using the same hardware and identical applications software to provide inter-institution and multi-institution searching capabilities as an integral part of the library catalog. The three computers are linked via telecommunications facilities that allow queries to be directed to whichever database(s) is needed and the retrieval set(s) to be transmitted back to the computer that is host to the user's terminal session for merging and display. The basic structure of the network is shown in figure 4.7.

Figure 4.7: TRLN Network Structure.



4.3.3.1 Processing facilities. Specifically, the computers used are a mixture of Tandem NonStop II and TXP processors. The particular configuration is determined by the individual requirements of each campus, such as the number of terminals to be supported and the size of the database. Each system acts as host for all terminals connected directly to it, i.e., is responsible for command parsing, maintaining session context information, formatting results for display, etc. for its terminals. When a query requires access to a remote database, the application software on the host system simply requests the necessary service from the database server on the remote system. Figure 4.8 shows that this is quite straightforward since the Tandem proprietary operating system, GUARDIAN, and networking software, EXPAND, allow application processes on one system to address application processes on a remote system directly, simply by naming the required process.

Figure 4.8: Network Access to TRLN Databases.



In addition to the production nodes supporting the online catalogs on each campus, there is a fourth system in the network used exclusively for development and testing: a two processor Tandem EXT system. Figure 4.9 shows the primary equipment in the network as of April 30, 1990.

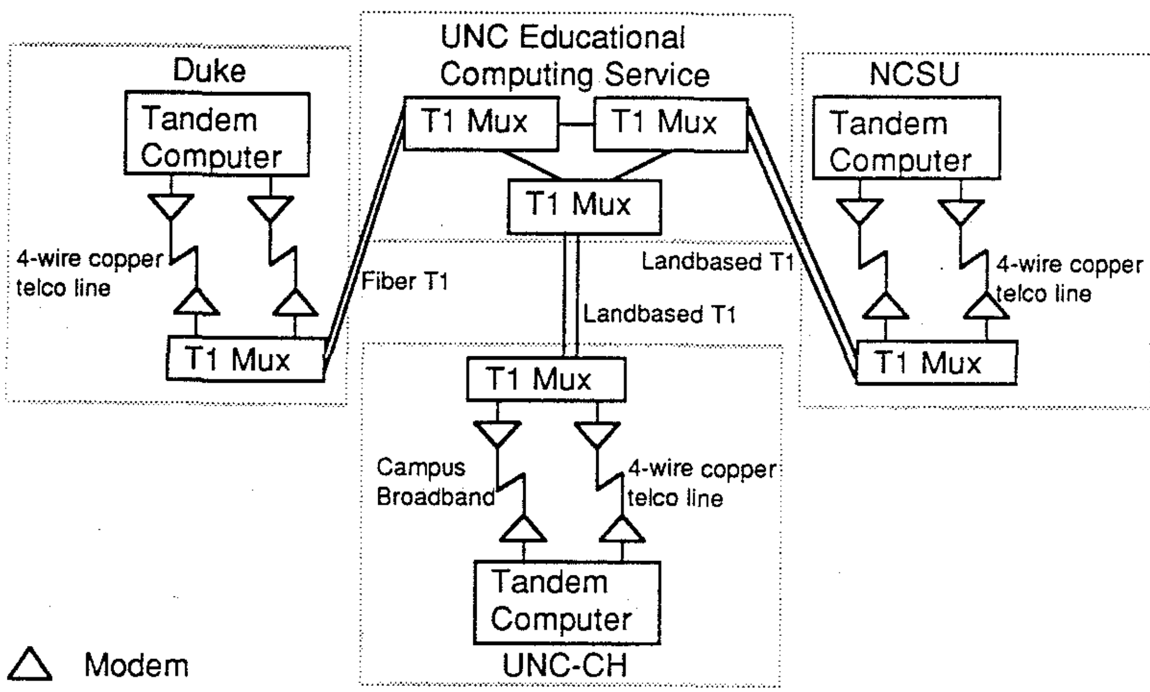
Figure 4.9: TRLN System: Installed Equipment, 4/30/90.

	Duke	NCSU	UNC-CH	TRLN	Total
Terminals					
Public	28	41	64	2	135
Technical (editing)	33	9	23	2	67
Dial access ports	11	14	14	0	39
Circulation	-	6	-	-	6
Processors					
NonStop II	0	2	0	0	2
TXP	3	2	4	2	11
Memory (total MB)					
NonStop II	-	20	-	-	20
TXP	36	24	64	16	140
Communications					
Synchronous ports	28	8	28	4	68
Asynchronous ports	64	64	64	32	224
Disc (total GB)	3.0	2.8	6.2	0.9	12.9

4.3.3.2 Workstations. At the time of writing, the primary terminal supported by the TRLN system was the Telex 476L, a 3270-type terminal providing the extended library character set that occurs in MARC records. This character set provides diacritical marks and special characters (such as the Scandinavian "O" with a slash) that occur in Roman alphabet languages. BIS was also designed to support dumb ASCII terminals or any microcomputer emulating such terminals, as would be used by users of the dial-in service. Investigations continue to identify other full character set terminals, and using intelligent workstations certainly represents a future possibility.

4.3.3.3 Telecommunications facilities. The initial telecommunications configuration uses a combination of telephone leased lines, campus broadband, and T1 subchannels to form a triangle by connecting each node to the other two. Tandem's proprietary networking software, EXPAND, is used to drive the telecommunications facilities. As additional capacity is needed, additional subchannels may be added, possibly using a different physical network to provide greater fault tolerance. The EXPAND software handles multiple channels, selecting optimal routes and rerouting traffic as necessary to avoid channels that are either very busy or out of service. The TRLN telecommunications network, as of April 30, 1990, is shown in figure 4.10.

Figure 4.10: TRLN Network Configuration.



4.4 Conclusion

Although the DLN Model had not been defined at the time BIS was implemented, the system meets the requirements for a linked-level distributed network. As the only operational network of this type, it provides an important example of how the DLN Model can be implemented, demonstrating one possible solution to the set of key implementation issues in building a linked-level network. More importantly, however, the case study verifies that the DLN Model can function properly as a reference model, i.e., that the DLN Model does, in fact, provide a framework within which a specific implementation can be developed.

CHAPTER 5: CASE STUDY
AN ANALYSIS OF USAGE PATTERNS

Although examination of the TRLN system demonstrates that the DLN Model can be implemented, it is also necessary to demonstrate that the implementation is reasonable, i.e., that not only can the Model be implemented but also that it can result in a system that is actually used as intended. Analyzing the use patterns of BIS, since it is an implementation of the DLN Reference Model, serves that purpose. In addition, however, although not directly related to characteristics included in the reference model, the use patterns provide general insight into how linked level distributed networks are actually used. Patterns according to criteria such as terminal location were examined to determine whether multi-institution and inter-institution searching capabilities were used differently by different parts of the university community.

The contents of the complete network transaction logs for a full semester were examined. Total usage was investigated, measuring the extent to which library users actually used inter-institution and multi-institution searching capabilities. The choice of institutions as the search target was studied, determining the numbers of searches that extended beyond the home institution. Use of the forwarding capabilities of the system was also studied, determining the numbers of searches that were repeated by rekeying and by command.

Beyond these general usage patterns, the study investigated whether there were statistically significant differences in usage based on terminal location. Institutional differences were examined, as were differences by type of library (i.e., medical, law, or general academic).

Since users do not identify themselves to the system in any way, it was not possible to study the variations in usage according to individual characteristics (e.g., to compare student versus faculty usage). However, the complete content of each search, exactly as entered by the user, was available, so it was possible to investigate whether the inter-institution and multi-institution search capabilities were used significantly more for certain types of searches, i.e., known-item versus subject searches.

5.1 Other catalog use studies

Although the opportunity to study a distributed network of library catalogs is new, catalog use studies have a long tradition. William M. Randall is generally acknowledged as the first to call for systematic catalog use studies as a means of improving the library catalog [Randall 1930]. Randall emphasized that studies were needed if the catalog was to be truly useful to anyone other than librarians, and urged that the focus of such studies should be catalog users, not the catalog itself.

Since that time, many catalog use studies have been done. Until recently, of course, the focus of these studies was use of the card catalog. Obtaining large, unbiased samples in the usual library environment was a difficult problem because of the manual methods required. Most studies used labor intensive and obtrusive survey methods to study catalog users. These studies were limited in scope, generally involving a single type of library or single type of library user, and were designed to supply data on who was using the catalog and why, how they were approaching it, what information included in the catalog was found useful, and how successful users were [Krikelas 1972].

The realm of catalog use studies has been expanded to include the online catalog. Often the purpose of the study and the techniques employed were the same as those used in previous card catalog use studies. Questionnaires and surveys continue to be the most common methods, usually being used to study user attitudes and to determine the need for various catalog features [Markey 1984].

There is increasing interest in automatic monitoring of information systems, a method that was not a possibility until the systems themselves were automated. Such methods have the important advantage of being unobtrusive since user interactions with the catalog are simply logged for later analysis. In contrast with other methodologies, transaction log studies record what users do rather than what they say they do. The thoroughness of logging has raised concern regarding users' rights to privacy and the need for appropriate caution in the design of unobtrusive monitoring stems [Penniman 1980, Rice 1983].

Automatic monitoring produces much larger volumes of data than was previously possible. However, the volume of data can also be a disadvantage, and may explain why this method of study has not become more widely used: additional computing resources are required to transform the vast quantity of raw data into analyzable form and appropriate methods of analysis must be understood and used.

A recent survey of significant findings from online catalog use studies shows that, for the most part, these studies were investigating user satisfaction, how users learned about the online catalog, types of problems encountered by users, and the amount of subject searching done in online catalogs [Lewis 1987]. The finding that online catalog users do much more subject searching than do card catalog users is of particular interest here.

A summary, using transaction logs from eight online catalogs, showed 24 to 65 percent of searches in online catalogs were subject searches as compared with 10 to 62 percent reported for card catalogs [Markey 1984]. Because it is a system for university users, the results of the University of California study of the MELVYL system showing 50 to 60 percent subject searches may be a better indication than the Markey study of what the TRLN libraries should expect [Larson 1983a, Larson 1983b]. Kaske [1988b] found 47 percent of all searches were subject searches at the University of Alabama, but it ranged from 35 to 52 percent over the weeks of a fall semester. He also found [1988a] that the amount of subject searching varies considerably among the different branch libraries.

The demand for subject searching in online catalogs, which was also borne out by questionnaire studies, caught librarians by surprise since card catalog studies had repeatedly shown that known item searching was used much more heavily than subject searching. These findings emphasize the importance of such studies in helping designers develop systems that meet user requirements [Besant 1982, Ferguson 1982, Lewis 1987, Markey 1984].

There have been several unpublished transaction log studies of BIS. David Bennett [1987] studied transaction logs from the UNC-CH catalog to determine what kinds of errors users made and how they used the online help system. A small-scale study of multi-institution searching activity was carried out by Stockton [1988]. Although the study included the transactions for public and dial-in terminals at UNC-CH for only two weeks (the fourth full week of the Fall 1987 semester and the third full week of the Spring 1988 semester), and did not include all the search characteristics studied here, it provides an interesting basis for comparison with the results obtained in this analysis. Specific findings of interest are discussed in direct comparison with the results of this study.

5.2 Methodology

The complete transaction logs for the Fall 1987 semester for all three systems were examined. The analysis was based on individual search commands as actually executed. A search was considered to be any command that first creates an index or items display for a given search argument. Subsequent manipulations such as choosing items from an index display or requesting the corresponding index display from a retrieval set were considered display commands, and therefore were not counted as searches.

However, each repetition of a search was counted as a separate search. Thus doing a search and then forwarding it counted as two searches. Invalid searches, such as search commands with no argument or index control numbers searches, were not counted as searches. Correctly formed searches that resulted in empty retrieval sets were counted as searches.

Searches were not context-free since prior commands within a session could redefine the way a specific search command was interpreted. In BIS, a new session begins whenever an explicit start or end command is executed, or when the system time limit of ten and a half minutes of inactivity is exceeded. Any of these conditions causes all session parameters to be reset to the system defaults. Different users can share a session; users also can, and often do, begin new sessions unnecessarily. Therefore, different system sessions do not necessarily correspond to different user sessions. This study used the system definition of session boundaries because it is that definition that controls how commands are in fact interpreted and executed.

There are two particular cases where session boundaries affect search interpretation. First, session parameters cause searches to be directed either to an institution other than the home institution or to multiple institutions without requiring an explicit indication in the search command itself. Second, a previous command affects the proper categorization of a search command in a repeated search, i.e., where the user requests repetition of the same search, presumably at another institution, by rekeying the search rather than using the explicit forwarding capability of the system. Therefore, it was necessary to compare each search with the previous search to determine whether the search was repeated, i.e., forwarded by rekeying.

5.2.1 Procedure.

Several different steps were taken to transform the data in the transaction logs into analyzable form. The logs consisted of a strictly chronological trace of every command line entered at every terminal in the system, whether valid or not. Specifically, whenever the user pressed any key resulting in a transmission to the system, an entry was made in the log that included the terminal identifier, date and time stamp, and information on what command was entered. The specific format of the log and an example are shown in figure 5.1.

Figure 5.1: Transaction Log.

Format:

Position	Content
1 - 16	Terminal identifier
17 - 22	Date (YYMMDD)
23 - 30	Time (HHMMSSSS: e.g, 14261588 = 2:26pm, 15.88 seconds)
31 - 32	Screen type ("M" if the command was entered from a menu)
33	Function key number
34	Terminal type ("E" for editing, "N" for public)
35 - 36	Command length in characters (binary)
37 - end	Text of typed command (what the user typed, max. length 480 characters)

Example:

```
$TERM24            87081015064407   N..au clabby john f
$TERM24            87081015070573   N..au munter mary
$BSCTR42#LS1       8708101507473301 N..
$TERM24            87081015075843   N..ti progress in sensory physiology
$BSCTR42#LS1       87081015081719   N..au osborn c
$BSCTR42#LS1       8708101508287412 N..
$TERM24            87081015082880   N..au morris ian
$BSCTR42#LS1       8708101508359912 N..
```

The first step in the analysis was to reparse the commands to obtain a shorter, encoded version of the transaction log. For example, a two-character command code allowed the categorization program described below to distinguish search and session commands important to this study from all the other commands, which were not relevant here. Similarly, the reparsing identified explicitly which institutions' catalogs were being searched. The parser from the BIS system itself was used to insure that the commands were interpreted for analysis in exactly the same way as the BIS system interpreted the commands for execution.

The parser output was then sorted chronologically by terminal so that all commands of a terminal session were adjacent. The next step resulted in a file that categorized the search commands according to information about session parameter changes, terminal locations and types, and information contained in the search commands themselves. Each record in the categorized file represented one search, but, unlike the search represented in the original transaction log, included all the context information necessary to count the search in the proper categories. The output of the categorizing step is described in figure 5.2.

Figure 5.2: Output of Categorizing Program.

Terminal type: Editing terminals, public terminals in libraries, or terminals using dial-in services.

Institution searched: The institution(s) explicitly specified, either as part of the search command itself or as a session parameter.

Home institution: The institution at which the terminal is located.

Separately administered library: The separately administered library at which the terminal is located (e.g., UNC-CH Health Sciences Library, Duke School of Law).

Search type: Subject (subject headings) or known-item (author, title, control number) search.

Forwarded search indicator and count: Indication of whether the search was forwarded explicitly or rekeyed, and how many times the same search was forwarded.

5.3 Findings

The data analyzed included all searches on the BIS network from August 24 through December 18, 1987, which included the entire fall semester for UNC-CH, Duke, and NCSU. Frequency distributions were compared using a Chi-Square test of difference among proportions. This section presents the results of this analysis. The interpretation is discussed in section 5.4.

During the period studied, a total of 921,256 searches were executed on the network. The general distribution of searches among the institutions is shown below.

Table 5.1: Numbers of Searches at each Institution.

Institution	Number of Searches	% of Total
Duke	294,492	31.97
NCSU	291,179	31.61
UNC	335,585	36.43
Total	921,256	100.01

In the studies that follow, the Chi-Square test was used to determine whether there were statistically significant differences in the frequency distributions as a whole. Although this test does not indicate whether any two particular frequency values are significantly different, such detailed analysis was unnecessary for the purpose of this study, which was to study overall searching patterns. More detailed analysis of the variable interactions is left for future research.

5.3.1 Main Effects.

5.3.1.1 Target Institution. The institution searched (target institution) had five possible values:

1. The home institution only was searched (HomeOnly),
2. One institution other than the home institution was searched (OneOther),
3. Two institutions, including the home institution, were searched simultaneously (Home+1),
4. Two institutions other than the home institution were searched simultaneously (BothOther),
5. All three institutions were searched simultaneously (All).

Analysis showed similar trends in which target institutions were searched among the three institutions. As might be expected, by far the majority of searches originating at all three institutions were directed to the home institution only (table 5.2). However, searches originating at UNC-CH were most likely to include target institutions other than the home institution, while searches at originating at NCSU were least likely to have multi-institution targets.

Table 5.2: Frequency of Target Institutions by Originating Institution.

Target Institution	Originating Institution: DUKE Frequency	DUKE %	NCSU Frequency	NCSU %	UNC-CH Frequency	UNC-CH %	TOTAL Frequency	TOTAL %
HomeOnly	281,322	95.53	280,304	96.27	318,267	94.84	879,893	95.51
OneOther	6,386	2.17	5,659	1.94	8,320	2.48	20,365	2.21
All	6,386	2.17	4,895	1.68	8,451	2.52	19,732	2.14
Home+1	228	0.08	150	0.05	472	0.14	850	0.09
BothOther	170	0.06	171	0.06	75	0.02	416	0.05
TOTAL	294,492	100.01	291,179	100.00	335,585	100.00	921,256	100.00

Chi-Square = 956.15, DF = 8, $p < 0.0001$

When searches did include a target institution other than the home institution, users at all three institutions showed a strong preference for searching all three institutions together, or one other institution, over searching any combination of two institutions. Searches of all three institutions represented 49.23 percent of non-HomeOnly searches, while OneOther searches represented 47.70 percent of non-HomeOnly searches. However, searches originating at UNC-CH were more likely to include all three institutions simultaneously while searches originating at NCSU were more likely to include only one of the non-home institutions. Searches originating at Duke were equally likely to include one other institution or all three institutions.

5.3.1.2 Repetition Style. The second primary effect studied was the pattern of search repetition, which also had five possible values:

1. The search was not repeated (NONE),

2. The search was repeated once or twice, by explicit forwarding command (COMMAND-1,2),
3. The search was repeated once or twice, by rekeying it (REKEY-1,2),
4. The search was repeated three times or more, by forwarding explicit command (COMMAND-3),
5. The search was repeated three times or more, by rekeying it (REKEY-3).

There were similar trends in repetition patterns among the three institutions. As expected, most searches (95.39 percent) were not repeated. When searches were repeated, however, they were generally repeated by rekeying the entire search: 95.09 percent of repeated searches were rekeyed rather than forwarded by command. Searches originating at Duke were repeated more often than those originating at either of the other two institutions, and the command style of repetition was used proportionally more at Duke.

Table 5.3: Frequency of Repetition Style by Originating Institution.

Repetition Style	Originating Institution							
	DUKE		NCSU		UNC-CH		TOTAL	
	Frequency	%	Frequency	%	Frequency	%	Frequency	%
NONE	279,517	94.91	275,707	94.69	323,529	96.41	878,753	95.39
REKEY-1,2	12,291	4.17	13,708	4.71	10,984	3.27	36,983	4.01
REKEY-3	1,686	0.57	1,135	0.39	613	0.18	3,434	0.37
COMMAND-1,2	960	0.33	577	0.20	431	0.13	1,968	0.21
COMMAND-3	38	0.01	52	0.02	28	0.01	118	0.01
TOTAL	294,492	99.99	291,179	100.01	335,585	100.00	921,256	100.00

Chi-Square = 1830.35, DF = 8, $p < 0.0001$

5.3.1.3 Interactive Effects of Repetition Style and Target Institution. Analysis showed significant interactions between the repetition pattern and the target institution(s). As shown in table 5.4, repeated searches were usually directed to the home institution: 76.62 percent of repeated searches were of the home institution only. When institutions other than the home institution were searched without forwarding, the searches were usually of the three institutions together. When multi- or inter-institution searches were repeated, rekeying was the more common style, and the search was typically of one institution other than the home institution.

Table 5.4: Frequency of Repetition Style by Target Institution.

Repetition Style	Target Institution							
	HomeOnly		OneOther		All		TOTAL	
	Frequency	%	Frequency	%	Frequency	%	Frequency	%
NONE	847,327	96.30	13,145	64.55	17,238	87.36	878,753	95.39
REKEY-1,2	29,481	3.35	5,391	26.47	2,008	10.18	36,983	4.01
REKEY-3	2,888	0.33	468	2.30	72	0.36	3,434	0.37
COMMAND-1,2	171	0.02	1,292	6.34	394	2.00	1,968	0.21
COMMAND-3	26	0.00	69	0.34	20	0.10	118	0.01
TOTAL	879,893	100.00	20,365	100.00	19,732	100.00	921,256	99.99

Data are not shown for Home+1 and BothOther searches, which together represent 0.14% of all searches.

Chi-Square = 88027.84, DF = 16, $p < 0.0001$

5.3.2 Secondary Effects.

5.3.2.1 Effect of the Search Type. Search type had two possible values:

1. Known-item searches were those where the user appeared to be looking for a specific piece of material. Author, title and control number searches were considered known-item searches (KNOWN),

2. Subject searches were those where the user was searching for items on a particular topic. All subject headings searches were considered subject searches (SUBJ).

During the period studied, 33.19 percent of all searches of the TRLN catalogs were subject searches. Searches originating at NCSU were more likely to be subject searches than searches from either of the other institutions.

Table 5.5: Frequency of Search Type by Originating Institution.

Search Type	Originating Institution: DUKE Frequency	%	NCSU Frequency	%	UNC-CH Frequency	%	TOTAL Frequency	%
KNOWN	202,907	68.90	179,121	61.52	233,490	69.58	615,518	66.81
SUBJ	91,585	31.10	112,058	38.48	102,095	30.42	305,738	33.19
TOTAL	294,492	100.00	291,179	100.00	335,585	100.00	921,256	100.00

Chi-Square = 5420.03, DF = 2, p < 0.0001

Interactions between search type and institution target were significant. Table 5.6 shows a higher proportion of multi- and inter-institution searches than of HomeOnly searches were for known items.

Table 5.6: Frequency of Search Type by Target Institution.

Search Type	Target Institution: HomeOnly Frequency	%	OneOther Frequency	%	All Frequency	%	TOTAL Frequency	%
KNOWN	582,251	66.17	16,437	80.71	15,890	80.53	615,518	66.81
SUBJ	297,642	33.83	3,928	19.29	3,842	19.47	305,738	33.19
TOTAL	879,893	100.00	20,365	100.00	19,732	100.00	921,256	100.00

Data are not shown for Home+1 and BothOther searches, which together represent 0.14% of all searches.

Chi-Square = 3643.37, DF = 4, p < 0.0001

Interactions between search type and repetition style were also significant. As shown in table 5.7, rekeyed searches were more likely than non-repeated searches to be subject searches. However, searches forwarded by command were more likely than non-repeated searches to be known-item searches.

Table 5.7: Frequency of Search Type by Repetition Style.

Search Type	Repetition Style: NONE Frequency	%	KEY-1,2 Frequency	%	KEY-3 Frequency	%	COMMAND-1,2 Frequency	%	TOTAL Frequency	%
KNOWN	594,415	67.64	18,513	50.06	1,018	29.64	1,493	75.86	615,518	66.81
SUBJ	284,338	32.36	18,470	49.94	2,416	70.36	475	24.14	305,738	33.19
TOTAL	878,753	100.00	36,983	100.00	3,434	100.00	1,968	100.00	921,256	100.00

Data are not shown for COMMAND-3 searches, which represent 0.01% of all searches.

Chi-Square = 7167.50, DF = 4, p < 0.0001

5.3.2.2 Effect of Terminal Type. Terminal type was defined by the following possible values:

1. Public terminals, i.e., terminals located mainly in public areas of the libraries, and used primarily by library users and staff not involved in building and maintaining the databases (PUB),
2. Terminals or microcomputers that gain access to the catalog via the campus computation centers, either through campus telecommunications networks or by dialing in (DIAL),
3. Terminals for technical services staff, primarily catalogers responsible for creating and modifying bibliographic and holdings records (TECH).

Most searches were from public terminals (74.91%), with the remaining 18.60 percent from technical terminals and 6.49 percent from dial-in terminals. As with previous analyses, the general trend among the three institutions was the same. However, Duke showed a higher proportion of searches from technical terminals than the other two institutions, while NCSU showed somewhat higher use of dial-in terminals.

Table 5.8: Frequency of Terminal Type by Originating Institution.

Terminal Type	Originating Institution							
	Duke		NCSU		UNC-CH		TOTAL	
	Frequency	%	Frequency	%	Frequency	%	Frequency	%
PUB	215,591	73.21	217,987	74.86	256,506	76.44	690,084	74.91
TECH	62,260	21.14	50,602	17.38	58,525	17.44	171,387	18.60
DIAL	16,641	5.65	22,590	7.76	20,554	6.12	59,785	6.49
TOTAL	294,492	100.00	291,179	100.00	335,585	100.00	921,256	100.00

Chi-Square = 2827.32, DF = 4, $p < 0.0001$

Interactions between terminal type and target institution were significant (see table 5.9). A higher proportion of multi- or inter-institution searches were done from dial-in terminals than from public terminals. Technical terminal users rarely searched beyond the home institution catalog, with 99.1 percent of their searches of the home institution only. When searches from public or technical terminals did include another institution, the tendency was to search the other institutions one at a time. Multi-institution searches from dial-in terminals, however, were more likely to include all three institutions together.

Table 5.9: Frequency of Terminal Type by Target Institutions.

Terminal Type	HomeOnly Frequency	HomeOnly %	OneOther Frequency	OneOther %	All Frequency	All %	TOTAL Frequency	TOTAL %
PUB	658,998	74.90	15,822	77.69	14,232	72.13	690,084	74.91
TECH	169,839	19.30	1,139	5.59	374	1.90	171,387	18.60
DIAL	51,056	5.80	3,404	16.71	5,126	25.98	59,785	6.49
TOTAL	879,893	100.00	20,365	99.99	19,732	100.01	921,256	100.00

Target Institution

Data are not shown for Home+1 and BothOther searches, which together represent 0.14% of all searches.

Chi-Square = 20944.80, DF = 8, p < 0.0001

Table 5.10, however, shows searches from all terminal types displayed the general pattern of searching the desired institution(s) first rather than using forwarding. When searches were repeated, they were usually rekeyed rather than forwarded by explicit command, regardless of terminal type. However, a higher proportion of repeated searches were from dial-in terminals than from other terminal types. Of searches forwarded by command, considerably more were from dial-in terminals and fewer from technical terminals than for non-repeated or rekeyed searches.

Table 5.10: Frequency of Terminal Type by Repetition Style.

Terminal Type	NONE Frequency	NONE %	KEY-1,2 Frequency	KEY-1,2 %	KEY-3 Frequency	KEY-3 %	COMMAND-1,2 Frequency	COMMAND-1,2 %	TOTAL Frequency	TOTAL %
PUB	654,836	74.52	30,687	82.98	2,995	87.22	1,495	75.97	690,084	74.91
TECH	167,825	19.10	3,326	8.99	178	5.18	55	2.79	171,387	18.60
DIAL	56,092	6.38	2,970	8.03	261	7.60	418	21.24	59,785	6.49
TOTAL	878,753	100.00	36,983	100.00	3,434	100.00	1,968	100.00	921,256	100.00

Repetition Style

Data are not shown for COMMAND-3 searches, which represent 0.01% of all searches.

Chi-Square = 3966.51, DF = 8, p < 0.0001

5.3.2.3 Effect of Terminal Location. Terminal location by separately administered library was defined by the following values:

1. The main group of academic libraries, i.e., all libraries on campus except the law and medical libraries (ACADEMIC),
2. The law libraries (LAW),
3. The medical libraries (MEDICAL).

Since NCSU does not have either law or medical libraries, data were available only for UNC-CH and Duke. For those two institutions, the distribution of searches was as shown in table 5.11. Differences among the separately administered libraries were significant. As expected, the academic libraries were responsible for the majority of the searches, but the separately administered libraries at Duke represented a higher proportion of searches originating at Duke. The Duke law library was a heavier user than the Duke medical library, whereas at UNC-CH the medical library was responsible for more searches than the law library, which was a very light system user.

Table 5.11: Frequency of Terminal Location by Originating Institution.

Terminal Location	Originating Institution					
	Duke		UNC-CH		TOTAL	
	Frequency	%	Frequency	%	Frequency	%
ACADEMIC	260,424	88.43	307,684	91.69	568,108	90.16
MEDICAL	11,756	3.99	25,875	7.71	37,631	5.97
LAW	22,312	7.58	2,026	0.60	24,338	3.86
TOTAL	294,492	100.00	291,179	100.00	630,077	99.99

Chi-Square = 23557.65, DF = 2, $p < 0.0001$

Analysis showed significant interactions between the terminal location and the target institution (see table 5.12). Searches originating at all three types of separately administered libraries were most likely to be of the home institution alone. However, for searches involving an institution other than home, the academic libraries tended to search all three institutions together while the medical and law libraries tended to search the other institutions one at a time.

Table 5.12: Frequency of Target Institutions by Terminal Location.

Target Institution	Terminal Location ACADEMIC Frequency	%	MEDICAL Frequency	%	LAW Frequency	%	TOTAL Frequency	%
HomeOnly	540,927	95.22	35,391	94.05	23,271	95.62	599,589	95.16
OneOther	12,906	2.27	1,164	3.09	636	2.61	14,837	2.35
All	13,359	2.35	1,050	2.79	428	1.76	14,706	2.33
Home+1	674	0.12	24	0.06	2	0.01	700	0.11
BothOther	242	0.04	2	0.01	1	0.00	245	0.04
TOTAL	568,108	100.00	37,631	100.00	24,338	100.00	630,077	99.99

Chi-Square = 237.34, DF = 8, p < 0.0001

Table 5.13 shows that when searches were repeated they were most likely to have been rekeyed, regardless of separately administered library. The medical libraries repeated searches proportionally more than the other types of libraries, and were more likely than the others to use forwarding by command.

Table 5.13: Frequency of Repetition Style by Terminal Location.

Repetition Style	Terminal Location ACADEMIC Frequency	%	MEDICAL Frequency	%	LAW Frequency	%	TOTAL Frequency	%
NONE	543,605	95.69	35,895	95.39	23,546	96.75	603,046	95.71
REKEY-1,2	21,079	3.71	1,465	3.89	731	3.00	23,275	3.69
REKEY-3	2,184	0.38	83	0.22	32	0.13	2,299	0.36
COMMAND-1,2	1,179	0.21	183	0.49	29	0.12	1,391	0.22
COMMAND-3	61	0.01	5	0.01	0	0.00	66	0.01
TOTAL	568,108	100.00	37,631	100.00	24,338	100.00	630,077	99.99

Chi-Square = 241.75, DF = 8, p < 0.0001

5.4 Analysis

While there are statistically significant differences among the three institutions on the search characteristics studied, clearly network utilization has not been dominated by any one or two of the institutions. Despite differences in number of terminals installed, database size (proportion of library materials represented in the online catalog), or student body and faculty size, the three universities are using the network similarly. The significant effects that were found seem not attributable to differences among the institutions. Rather they seem more likely to be due to the dramatic differences in choice of target institution (95.51 percent HomeOnly searches versus 0.05 percent BothOther searches) and repetition style (95.39 percent not repeated versus 0.01 percent repeated by command more than twice) than to differences among the originating institutions.

5.4.1 Main Effects.

5.4.1.1 Target Institution. That 95.51 percent of searches were of the home university reflects the expectation that most users would satisfy their searching requirements without searching the catalogs of the other institutions. The system default for target institution, which is the home institution, was established to support this expected common behavior. However, one of every twenty searches involved at least one institution other than the home library, indicating that multi-institution and inter-institution searching capabilities of the network were heavily used.

The 5.16 percent rate of multi- and inter-institution searches for UNC-CH measured in this study is somewhat less than the 7.35 percent rate reported by Stockton [1988]. Since this study included a full semester, including holidays and low use periods at the beginning and end of the semester, while Stockton's study covered only two weeks, each selected from the middle of a semester, one explanation for the discrepancy is that multi- and inter-institution searching activity was not evenly distributed throughout the semester. In addition, since one of the two weeks of the Stockton study was during the Spring, 1988 semester, his results may show increasing user awareness of these features over time.

5.4.1.2 Repetition. Users tended not to use the system's forwarding capabilities, with only 4.61 percent of searches repeated. This agrees with Stockton's finding that users typically seem to decide first which catalog(s) to search rather than searching the home institution first and then forwarding the search to the other institutions if necessary.

Many users repeated searches many more times than necessary to search the catalogs of the three institutions, as indicated by the 3,552 searches that were repeated three or more times. This pattern occurred more at Duke than the other two institutions. The most extreme example of repeated searches was a user who rekeyed the same search 42 times in a row during one terminal session. Although there were intermittent problems with the telecommunications at various times during the period of the study, which probably account for some extra search repetitions, this seems an unlikely explanation for their large numbers. Informal observation of the logs showed that many of the searches done repeatedly were of the home institution where telecommunications problems were infrequent. In addition, users often made changes in punctuation or capitalization that would be eliminated when the search argument was normalized. These findings suggest a lack of understanding of how the catalog works: perhaps when users do not get the desired response from the system, they try the same search again, or what they incorrectly think is an alternative form, hoping for a different response.

The data also showed more searches were repeated by the laborious method of rekeying the search argument than by command. This result may also show a lack of understanding of how to use the catalog effectively. Another reason for the difficulty experienced by some users may be that the forward by command feature was added to the system after many of them had already learned to forward searches by rekeying. Whatever the reason, the result suggests a need for user education.

5.4.1.3 Interactive Effects of Repetition Style and Target Institution. As indicated earlier, users tended to direct their original search to multiple institutions rather than searching their home institution first and then forwarding the search. When original searches were multi-institution, they were most likely to have been directed to all three institutions simultaneously. Apparently users who searched beyond the home institution either wanted everything together, letting the system do the work of merging the retrieval sets, or had prior expectations about which institution(s) would have the material they required.

When searches were repeated, however, the most popular strategy, after searching the home institution alone, was to search the other institutions one at a time. This pattern reflects the type of behavior the capability was designed to support. However, searches were also often forwarded to all three institutions together. It is not clear why this was done, since this strategy repeated a search that was already done in at least one of the catalogs. User confusion is one possible explanation. However, this pattern also might result when users initially search one catalog, probably their home institution and, finding some material but not everything desired, decide to expand the search to the entire network, using the system to create a merged retrieval set. Another explanation may be simply that it is easier to type "all" than two institution names.

The interactive effects further demonstrated user confusion as shown by inappropriate repetition of searches. Not only were users rekeying searches unnecessarily, but also most of the repeated searches were of the home institution.

5.4.2 Secondary Effects.

5.4.2.1 Effect of Search Type. The division of searches by type into known-item and subject searches showed a considerably different pattern than expected from other online catalog use studies, which showed a predominance of subject searching. NCSU showed a higher rate of subject searches than the other two institutions, but its rate of 38.48 percent subject searches was still lower than the 50 to 60 percent rate expected from other studies. Reasons for the discrepancy with other studies are not clear. However, the result of only 30.42 percent subject searches for UNC-CH in this study is quite comparable to Stockton's result of 30.7 percent subject searches in the UNC-CH catalog.

As expected, multi-institution and inter-institution searches were more likely to have been known-item searches than were searches of the home institution alone: 66.17 percent of home only searches were known-item whereas approximately 81 percent of multi- and inter-institution searches were known-item. This result confirms Stockton's finding that subject searches were less likely to include other institutions.

Rekeyed searches were more likely to be subject searches than were searches that were not forwarded. When searches were rekeyed more than three times, the proportion of subject searches rose dramatically to 70.36%. In contrast, searches forwarded by command were more likely to be known-item searches (75.86%). This pattern suggests that users looking for specific items were more sophisticated in their use of the system than users searching by subject.

5.4.2.2 Effect of Terminal Type. The general distribution of searches among terminal types was as expected, based on the numbers of terminals installed (or ports, for dial-in users) and the expected type of use, i.e., cataloging staff versus general searching. However, it is not clear why Duke showed proportionally more activity at technical terminals. NCSU showed heavier use of dial-in services, which may reflect the scientific orientation of that campus.

As expected, dial-in users, who were searching the catalog from campus locations other than the libraries or from off-campus, showed the greatest tendency to search the three institutions simultaneously and to use forwarding. Quite a high proportion of searches forwarded by command were from dial-in terminals, possibly showing a greater willingness of dial-in users to learn some of the system's more sophisticated features.

The result that technical users made least use of multi-and inter-institution searching capabilities was also expected since the primary use of technical terminals is for database maintenance activities on the home institution catalog. Although technical users might be expected to be better trained than public users, technical users also rekeyed searches rather than forwarding by command and repeated some searches unnecessarily.

5.4.2.2 Effect of Terminal Location. There were some differences in the use patterns of the separately administered libraries, which are probably due to such factors as differences in database size or number of users. The Duke medical library was responsible for 3.99 percent of Duke's searches while the UNC-CH medical library, which has both a larger database and larger number of clients, was responsible for 7.71 percent of searches at UNC-CH. The few searches (0.61 percent) originating at the UNC-CH law library may be due to the fact that, although their records are represented in database and terminals are located in the library for both editing and public use, that library is not an active member of TRLN. In addition, the UNC-CH law library database is small, representing only approximately one percent of the UNC-CH database. In contrast, the Duke law library database represents approximately ten percent of the Duke database.

When searching libraries other than the home institution, the academic libraries were more likely to search all three institutions together. In contrast, the medical and law libraries were more likely to search one other institution. Although this result is probably because only one other institution also included medical and law libraries, there was no increased number of Home+1 searches that would result in merged retrieval sets of the home institution and the one other institution.

The medical libraries made somewhat more sophisticated use of forwarding capabilities, using the command style of forwarding proportionally more than the other types of libraries. However, all libraries used rekeying more than forwarding by command, indicating that the problems with user confusion occur at all types of libraries.

5.5 Conclusion

The transaction logs of the BIS system were analyzed partly as a general exploration of how such a system is used. In addition, however, the analysis had the specific goal of determining whether inter-institution and multi-institution searching capabilities were used, thus demonstrating the reasonableness of the implementation. The finding that such usage was quite heavy, with one of every twenty searches involving an institution other than the home library, accomplishes this goal. The finding also suggests that, to the extent that usage is the key factor in determining system architecture, the TRLN libraries appropriately chose the distributed architecture at the linked level.

However, the usage was low enough to indicate that, although some libraries may require the highly interconnected approach provided by the linked level of the DLN Model, it is likely that many libraries will find a switched level network will meet their users' needs and be far less costly to implement. The analysis performed here confirms the concept that the DLN Model must allow for both highly and loosely connected networks of library catalogs.

CHAPTER 6: VERIFICATION OF THE PLANNING MODEL

The case study indicates that the distributed model at the linked level is a viable architecture for the TRLN libraries, but does not indicate other than by example when that architecture should be used. However, the experience of those libraries and others provides the basis for the LSAP Model, which offers a formal structure for the decision making process. Future library system planners who use the LSAP Model lessen the risk that they will overlook factors critical to the choice of appropriate system architecture. However, it is necessary to verify that the Model works before it can be effectively used. In addition, since the outcomes used in the LSAP Model are defined in the OPAC Reference Model, verification of the LSAP Model further validates the OPAC Model.

6.1 Verification Methodology

The LSAP Model was verified by using it to select an appropriate system architecture in situations where libraries have operational centralized, distributed or stand-alone systems that are considered appropriate for their circumstances. Systems librarians from the selected libraries were asked to rate the importance of the LSAP attributes and to establish the rankings for their library with regard to an appropriate partner library, i.e., to use the Model to characterize their relationship to the chosen partner as it would be used if the choice of system architecture were still open.

Systems librarians were used as raters although a library actually using the LSAP Model is more likely to use a committee to recommend system architecture. Since such committees demand a substantial commitment of high level staff, establishing a committee at each test library was not feasible.

Each systems librarian was also asked to complete a short questionnaire about the test library's current system. The methodology for verifying the LSAP Model is based on the assumption that the choice of architecture is appropriate, so if the Model suggests a different architecture, there may be a problem with the Model. The questionnaire was designed to cross-check that assumption by providing supplemental information that would verify that the installed system architecture was, in fact, considered by the systems librarian to be appropriate.

The systems librarian at each test library was given brief instructions, definitions of the attributes, a form for ranking the relative importance of the attributes in the test library environment, a form for rating the test library's relationship to another library with respect to those attributes, and the questionnaire about the test library's current system. The importance rankings and library rating forms, identical to those used initially to establish the Model, are shown in Appendix A. The remaining forms are shown in Appendix B.

6.2 Findings

6.2.1 Centralized Systems.

Three libraries with centralized systems were chosen to test how well the LSAP Model selects centralized systems. The systems librarians in those libraries were asked to select one of the other libraries participating in their system as their partner for the purposes of testing the Model.

The three libraries that participated in this study were Florida International University (FIU), Florida State University (FSU), and the University of North Florida (UNF). FIU used Florida Atlantic University as its partner; FSU and UNF both used the University of Florida.

The attribute vector was calculated for each library using the same formula described earlier for developing the attribute vectors in the Model. Figure 6.1 shows the attribute vectors for the three libraries with centralized systems as well as the vector for the centralized architecture of the Model. The cosine measure of similarity for each library in comparison to the attribute vectors for all four architectures in the LSAP Model is also given.

Figure 6.1: LSAP Attribute Vectors for Centralized Systems.

Attribute	Model	FIU	FSU	UNF
Internal Purpose	9.39	0.53	7.14	3.16
Use	16.04	3.30	13.68	7.60
Geographic Proximity	9.53	16.46	4.36	3.80
Scope	1.66	4.86	5.29	2.26
Cooperative Activities	10.33	24.48	3.60	0.63
Support Services	4.31	7.01	5.16	2.10
Local Autonomy	2.05	0.45	2.08	22.70
External Influences	10.95	6.28	18.38	24.60
Similarity to Model				
COS (Centralized)		.70	.89	.57
COS (Distributed-Linked)		.69	.84	.67
COS (Distributed-Switched)		.62	.66	.78
COS (Stand-Alone)		.33	.43	.82

The LSAP Model selects the centralized architecture correctly for two of the three test libraries, FIU and FSU. However, for FIU, the attribute vector is almost as close to the distributed linked architecture, which indicates that architecture would also have been a reasonable option to consider.

For UNF, the Model indicates the stand-alone architecture would be more appropriate than the centralized architecture that the library actually has. However, examining the attribute vector for that library indicates conflicting priorities that may have caused the Model to indicate a different architecture than the one actually chosen. Specifically, the attribute vector places very high priority on local autonomy, which is a strong characteristic of stand-alone systems, but notably absent in centralized systems. In addition, cooperative activities, which the Model rates as an important indicator of environments suitable for centralized systems are rated as very weak for UNF. However, the attribute vector for the library also shows extremely heavy external influences, which may indicate that UNF participates in the centralized network primarily because of the external influences.

6.2.2 Distributed Linked Level Systems.

The three TRLN libraries were chosen to test how well the LSAP Model selects distributed systems at the linked level. The systems librarian from each library was asked to perform the evaluation with regard to each of the other two TRLN participants.

As with the centralized libraries, the attribute vector was calculated for each library with each partner using the same formula described earlier for developing the attribute vectors in the Model. Thus, for example, the systems librarian at Duke first rated the attributes with regard to Duke's closeness to UNC-CH and then with regard to NCSU. The choice of architecture is evaluated separately for each partner library. Figure 6.2 shows the attribute vectors for each of the three libraries with its two partners as well as the vector for the distributed architecture at the linked level in the Model. The cosine measure of similarity for each library in comparison to the attribute vectors for all four architectures in the LSAP Model is also given.

Figure 6.2: LSAP Attribute Vectors for Distributed-Linked Level Systems.

		Duke with		NCSU with		UNC-CH with	
Attribute	Model	UNC-CH	NCSU	Duke	UNC-CH	Duke	NCSU
Internal Purpose	8.28	18.29	18.29	8.10	8.37	19.65	19.47
Use	13.31	17.09	17.09	4.94	8.24	17.95	17.95
Geographic Proximity	7.89	4.06	4.06	7.99	2.52	2.73	3.12
Scope	1.42	12.04	12.04	4.89	2.40	1.02	0.94
Cooperative Activities	9.58	8.37	8.37	9.20	9.59	14.35	14.10
Support Services	3.93	6.26	6.26	6.55	6.31	13.24	10.78
Local Autonomy	7.22	12.30	12.30	3.77	3.69	0.53	0.64
External Influences	8.40	2.18	2.18	9.30	9.30	2.71	3.28
Similarity to Model							
COS (Centralized)		.78	.78	.89	.93	.84	.87
COS (Distr.--Linked)		.84	.84	.90	.93	.83	.85
COS (Distr.--Switched)		.80	.80	.81	.81	.64	.66
COS (Stand-Alone)		.62	.62	.55	.51	.30	.31

The LSAP Model correctly selects three of the connections as distributed at the linked level, two of the connections as centralized, and the remaining connection as equally close to the centralized and distributed linked level architectures. However, the three connections that the Model selects as centralized are almost as close to the distributed linked level architecture. This may indicate that either architecture would be a reasonable choice for the libraries.

6.2.3 Distributed Switched Level Systems.

Two of the three TRLN libraries, NCSU and UNC-CH, have a switched level connection through a state-wide network to other libraries in the state university system. These two were chosen to test how well the LSAP Model selects distributed systems at the switched level. The systems librarian from each of those libraries was asked to perform the evaluation with regard to both North Carolina Central University (NCCU) and the University of North Carolina at Greensboro (UNC-G). (However, the systems librarian from NCSU did not complete the ratings for NCCU).

These relationships were especially useful tests of the LSAP Model for distributed systems at the switched level since NCSU and UNC-CH made a different architecture choice for NCCU and UNC-G than they made for the TRLN libraries. In addition, NCCU and UNC-G were once considered as possible members of TRLN, but it was decided that the need for interaction was insufficient to justify the linked level. Figure 6.3 shows the attribute vectors for the two libraries with their partners as well as the vector for the distributed architecture at the switched level in the Model. The cosine measure of similarity for each library in comparison to the attribute vectors for all four architectures in the LSAP Model is also given.

Figure 6.3: LSAP Attribute Vectors for Distributed-Switched Level Systems.

Attribute	Model	NCSU with UNC-G	UNC-CH with NCCU	UNC-CH with UNC-G
Internal Purpose	5.21	2.82	5.09	5.26
Use	7.85	1.65	5.85	9.76
Geographic Proximity	6.35	0.95	9.07	0.78
Scope	1.14	2.48	0.61	0.57
Cooperative Activities	8.04	1.15	3.94	3.43
Support Services	2.79	3.71	3.52	3.52
Local Autonomy	13.49	3.89	0.53	0.64
External Influences	5.30	1.13	0.94	1.45
Similarity to Model				
COS (Centralized)		.62	.86	.88
COS (Distr.--Linked)		.74	.84	.86
COS (Distr.--Switched)		.81	.70	.66
COS (Stand-Alone)		.74	.44	.34

The LSAP Model selects only one of the three connections correctly as distributed at the switched level, whereas the other two are selected as centralized. For those two, the key difference between the library ratings and the Model appears to be local autonomy: the Model expects the need for local autonomy to be very strong, but the attribute vector for UNC-CH to both partner libraries rates the need for local autonomy as insignificant.

6.2.4 Stand-alone Systems.

Three libraries with stand-alone systems were chosen to test how well the LSAP Model selects stand-alone systems. The systems librarians were asked to select a library that it might seem reasonable for their library to consider an interconnection or union catalog with, but with whom they had no intention of establishing one.

The three libraries that participated in this study were Harvard University, the University of Texas at Austin, and the University of Wisconsin--Madison. Harvard used the Massachusetts Institute of Technology as its partner, Texas used Austin Public Library, and Wisconsin used Marquette University.

The attribute vector was calculated for each library using the same formula described earlier for developing the attribute vectors in the Model. Figure 6.4 shows the attribute vectors for the three libraries with stand-alone systems as well as the vector for the stand-alone architecture in the Model. The cosine measure of similarity for each library in comparison to the attribute vectors for all four architectures in the LSAP Model is also given.

Figure 6.4: LSAP Attribute Vectors for Stand-alone Systems.

Attribute	Model	Harvard	Texas	Wisconsin
Internal Purpose	2.05	1.88	1.49	20.27
Use	2.90	1.06	0.73	2.30
Geographic Proximity	5.01	4.41	10.06	2.48
Scope	0.96	1.05	0.48	8.00
Cooperative Activities	1.66	0.42	0.53	3.10
Support Services	1.23	0.80	0.76	4.46
Local Autonomy	15.05	29.76	15.79	16.34
External Influences	3.21	0.35	1.79	13.34
Similarity to Model				
COS (Centralized)		.19	.37	.61
COS (Distr.–Linked)		.42	.55	.70
COS (Distr.–Switched)		.75	.80	.76
COS (Stand-Alone)		.95	.96	.71

The Model selects the stand-alone architecture correctly for two of the test libraries, Harvard and Texas, whereas for the third library, Wisconsin, it selected distributed at the switched level. The attribute ratings for internal purpose and external influences both are much stronger for Wisconsin than for either of the other two test libraries or for the Model, indicating that the distributed switched level architecture would be appropriate to consider.

6.3 Analysis

Results of the verification testing show that the LSAP Model selected the actual system architecture in nine of the test cases, but not in the other six cases. However, in three of those six cases, examination of the attribute vectors and knowledge of the library circumstances suggest that the Model's choice might be as appropriate a choice of system architecture, at least for initial consideration. One of the choices that matched the actual architecture was equally close to another architecture, so either would have been appropriate. Although statistical significance is not a meaningful measure for this type of data, the accurate performance of the Model indicates that the vector positions are not occurring by chance.

A planning model of this sort is intended to assist the decision maker in the thought process that leads to making a particular decision. Such a model does not consider the many external factors that enter such a process and, of course, is not expected to determine the decision totally. The verification testing shows that the LSAP Model is a successful planning model. It is sufficiently accurate to assist system designers in evaluating the factors critical to choosing the appropriate system architecture. It can also successfully be used as a tool to suggest which architecture(s) should be most seriously considered.

Nevertheless, a model that more accurately reflects expert judgements as evidenced by choices actually made in libraries would be a more useful tool. There are several possible reasons other than inaccuracies in the Model itself that may explain why the Model did not work more accurately in the verification tests.

As explained earlier, an unavoidable weakness in the verification methodology was dependence on the judgement of the systems librarian alone. The consensus process of determining the library ratings by committee would probably resolve some of the inconsistencies in the ratings as well as increase the probability that the ratings would actually reflect the organizational position of the library. In addition, some of the systems librarians who participated in the test had assumed their positions relatively recently. Thus they did not participate in the original decision making process and may have been unaware of some of the circumstances, especially regarding external influences, that led to the original choice of architecture.

Another factor that may explain why the Model is not more accurate may be that the chosen architecture is no longer considered most appropriate, at least by the systems librarian doing the ratings for this test. For example, at the time systems in the test libraries were installed, systems of a more desirable architecture may not have been available. Indeed, distributed systems at the linked level are still not available commercially, and most libraries are unable to undertake in-house development. The possibility that the installed system is not actually considered the optimal architecture was anticipated in designing the verification tests, but the questionnaire intended to gather information that would detect this situation was misunderstood by most respondents and thus did not serve its purpose.

Nevertheless, informal discussions with the systems librarians indicate that they are satisfied with the architecture of their systems and have no plans to change to a different architecture.

Insufficient instruction to the systems librarians in the test libraries may have been another factor affecting the results of the verification tests. The systems librarians were given only brief instructions on how to fill out the rating sheets, and were provided no general context information on how the Model works. This probably resulted in wider variation of the ratings than might have occurred if the librarians had been given a better overview of the total framework. More detailed instructions were not given out of concern that this would bias the results. However, questions from participants and inconsistencies within individual ratings seem to indicate that more complete instructions would have been helpful.

Finally, the Model was based on the consensus judgement of four experts on automated library systems. These four individuals had been working together for some time, hence could have developed a similarity in their judgements that does not reflect the diversity that occurs in the broader community of library system designers. Although this group represented a reasonable point for a first attempt at developing the Model, it was probably too small and too uniform for best results.

This analysis suggests some approaches to modifying the Model and to using it most effectively:

1) the library ratings should be developed by a committee of librarians familiar with all aspects of the library's internal and external circumstances rather than by an individual;

2) the process of developing the ratings should be directed by a facilitator who can explain what the purpose of the Model is, what the planning group is supposed to accomplish, how the procedures work, and assure that the members of the group work with a common framework;

3) the Model itself should be based on the collective judgements of a larger group than the original team of four experts.

6.4 Modification of the LSAP Model

Analysis of the original LSAP Model suggests that accuracy of the Model could be improved by incorporating the judgements of additional library system planners. To test this theory, the LSAP Model was modified by averaging the judgements of attribute vectors for the test libraries together with the attribute vector for the appropriate architecture in the original LSAP Model. The verification comparison was then redone using the modified LSAP

Model to determine whether the accuracy improved. Figure 6.5 shows the modified LSAP attribute vectors together with the original vectors.

Figure 6.5: Original and Modified LSAP Attribute Vectors.

Attribute	Centralized		Distributed Linked		Distributed Switched		Stand-Alone	
	Orig.	Modif.	Orig.	Modif.	Orig.	Modif.	Orig.	Modif.
Internal Purpose	9.39	5.06	8.28	14.35	5.21	4.60	2.05	9.82
Use	16.04	10.16	13.31	13.80	7.85	6.28	2.90	4.29
Geographic Proximity	9.53	8.54	7.89	4.62	6.35	4.29	5.01	5.04
Scope	1.66	3.51	1.42	4.96	1.14	1.20	0.96	4.03
Cooperative Activities	10.33	9.77	9.58	10.51	8.04	4.14	1.66	2.54
Support Services	4.31	4.64	3.93	7.62	2.79	3.38	1.23	2.39
Local Autonomy	2.05	6.82	7.22	5.78	13.49	4.64	15.05	16.98
External Influences	10.95	15.05	8.40	5.33	5.30	2.20	3.21	3.52

The ratings for each attribute in the original Model showed a relatively smooth transition in strength from one architecture to the next, in order of closeness of interaction supported by the architecture. For example, Internal Purpose was expected to be strongest for environments where the centralized architecture was most appropriate, slightly weaker for distributed linked level environments, and weakest for stand-alone environments.

The Modified Model does not show such consistency. Rather, a specific factor or two pushes the choice, with different factors being key in each architecture. For example, it appears that the External Influences factor determines when libraries are likely to participate in centralized systems, but are relatively unimportant in determining any of the other architectures. Local Autonomy seems critical to choosing a stand-alone architecture, but if it is rated as a relatively weak factor, and a connection of some sort is otherwise indicated as appropriate, Local Autonomy does not discriminate well among the types of connections. Strong ratings for Expected Use and Cooperative Activities seem to push toward either the centralized or distributed linked level architecture, but not to distinguish clearly between the two.

Examination of the original Model as compared to the modified Model helps explain why the original Model did not perform as accurately as expected. However, the next step in testing the modification is to determine whether it performed more accurately than the original in indicating system architecture. The attribute vectors for each test library were compared to the modified Model, using the cosine measure of similarity, as in the tests of the original Model. Tables showing the results of these comparisons are included in Appendix B.

The results with the modified LSAP Model show considerably improved accuracy: the Model selects the actual system architecture in eleven of the test cases, and a different architecture in only four cases. Two of the cases where a different architecture is selected are for Duke connecting with the other two TRLN libraries, cases where the original Model selected the actual architecture. The modified Model suggests that the distributed architecture at the switched level would be more appropriate than the linked level, but the attribute vector is almost as close to the linked level vector. The modified Model suggested that the NCSU to Duke connection should be centralized rather than distributed linked, and that the connection between UNC-CH and UNC-G should be at the linked level rather than the switched level. Observation does not suggest that any attribute in particular is responsible for the discrepancies, and in all four of these cases, the architecture selected by the Model seems reasonable for consideration.

The results of the verification testing indicate that the LSAP Model is a viable planning tool for determining most appropriate library system architecture. Modifying the Model by using library ratings from test libraries where system architecture decisions have been made improved the accuracy of the Model, but without causing dramatic changes in any of the attribute vectors. This suggests that the sensitivity of the Model is appropriate and that additional modifications should be made to reflect the judgement of more decision makers as systems are selected and installed in libraries.

CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS

This dissertation explores the basic problem of how library catalogs can be connected or combined, balancing such requirements as the need for autonomy and ease of implementation with the need to provide library users with access to materials beyond the local library collection. The first step in solving the problem was to identify the fundamental ways library systems can be connected, and to determine the characteristics of each.

The Reference Model for Online Public Access Catalogs (OPAC Model) does this by categorizing the basic architectures into three models: centralized, distributed and stand-alone. The reference model provides a way to classify a system according to its architecture and identifies the basic characteristics that a system must have. The distributed model (DLN Model) is explored in depth because it is the least well understood of the models and because the characteristics of distributed systems must be standardized if such systems are to be connected effectively.

Whereas the OPAC Model defines the system architectures, the Library Systems Architecture Planning Model (LSAP Model) provides a tool for choosing among them. The system characteristics defined in the reference model are included to meet real world needs, such as providing access to another library's holdings or preserving local autonomy. The LSAP Model follows from the OPAC Model by making explicit the connection between the possible system architectures and a set of environmental characteristics.

The concepts included in the reference model are new and untested, especially for the distributed architecture. Therefore a case study of the TRLN system was included in the dissertation specifically to demonstrate:

a) the reference model can be implemented, and

b) the implementation is reasonable and is an appropriate choice for that environment.

Verifying the LSAP Model was then necessary to demonstrate that the planning model works, i.e., that the Model accurately reflects expert judgements of appropriate choice of system architecture. In addition, verification of the planning model further validates the reference model: since the LSAP Model outcomes are the architectures delineated in the OPAC Model, if those architectures were inappropriately defined, the LSAP Model could not work properly.

7.1 Significance of the OPAC Model

7.1.1 Importance to System Design.

The OPAC Model, specifically the DLN Model, is of particular importance to system designers developing distributed library networks since it provides the conceptual framework for such systems. It is the basis from which designers can prepare functional specifications for particular implementations, whether new systems or modifications to existing systems. To be most effective, however, the DLN Model must be accepted universally, i.e., developed into a standard. The DLN Model presented in this dissertation, therefore, could be the first step in the standardization process.

7.1.2 Advantages of the DLN Model Over the Other Models.

The distributed model has a number of advantages over the stand-alone (SALS) or centralized union catalog (CLS) models for online library catalogs. The DLN Model, as developed here, provides a balance between the independence inherent in a single library catalog and the extremely close cooperation required for a centralized, shared system. As with the SALS Model, the DLN Model allows an individual library to implement an online catalog system according to its needs, i.e., selecting or designing a system with function and capacity appropriate to its user population and budget. Like the CLS Model, the DLN Model provides bibliographic access to multiple library catalogs. Unlike these other models, however, systems based on the DLN Model can provide both local autonomy and union catalog access capabilities.

7.1.3 Disadvantages of the DLN Model.

However, some constraints are imposed if a library system is to participate in a network with other library catalogs. Specifically, it must conform to the requirements of the DLN Model for function and physical components, and must establish the appropriate corollary policies to enable users to make effective use of those functions. Like the centralized model, coordination is required, although not to the degree required for a centralized configuration.

7.2 Significance of the LSAP Model

Although the DLN Model represents an important part of the library system architecture design space, distributed systems are not appropriate in all circumstances. The LSAP Model provides a planning tool to help determine, for a specific situation, whether the distributed architecture or some other approach is more appropriate. The LSAP Model provides a systematic method for evaluating those factors critical to choosing system architecture.

However, beyond providing a planning tool specifically for designing library systems, the LSAP Model demonstrates how the Multi-Attribute Utilities Model, a structured approach to decision making, can be used to improve the computer system design process. Software engineers have developed a variety of techniques to help structure the thought process that results in a system design, with a primary focus being to reduce complex decisions to groups of simpler ones that can be dealt with one at a time. Thus, for example, top-down design can be viewed as a series of decisions. Each decision in the process has a set of possible outcomes; once an outcome is selected, that choice determines what other decisions must be made and constrains the possible outcomes of those decisions. The technique demonstrated by the LSAP Model extends this general approach by adding a method for systematically evaluating the possible outcomes of a specific decision.

7.3 Importance of the Case Study

The functional and transaction log analyses not only help validate the DLN Model by demonstrating that it can be implemented effectively, but also provide an important example of how such analyses can be used. The transaction log analysis type of catalog use study is still quite new, and each additional study adds to the collective understanding of how transaction log analyses can be carried out and the types of information that can be obtained from them.

7.3.1 The Case Study Approach to Model Development and Validation.

This analysis demonstrated that using the case study as a springboard to develop a reference model is an effective methodology. Developing the DLN Model without an example would have been more difficult. Rather than attempting to identify in the abstract the factors to be considered, it was possible to consider functions available in the implemented system to determine whether they were of general importance or only significant to the specific implementation. A model developed without a supporting implementation would be more likely to omit key features. The existence of a fully operational network that complies with it demonstrates that the DLN Model can be implemented, thus at least partially validating the model.

However, the possibility exists that a model derived from a single case study is too narrowly based. Every attempt was made to consider the generalized environment, but until other implementations of the model are developed, it will not be possible to demonstrate the DLN Model is complete.

7.3.2 The Transaction Log Analysis of BIS.

In addition to being an important step in developing the DLN Model, the transaction log analysis of BIS represents a significant contribution towards understanding how online catalog systems in a distributed network environment are used. Librarians have assumed that distributed networks would efficiently support the expected usage patterns, but until now there has been no evidence that the underlying assumptions about use were correct.

The transaction log analysis of BIS demonstrated that the multi-institution and inter-institution searching capabilities were used as intended, thus indicating that the interest in a distributed system was not misplaced. However, usage was low enough to indicate that, while the highly interconnected linkages demonstrated by BIS are important, a less sophisticated method of providing mutual access for libraries might also have been provided. The finding that the three TRLN libraries use the multi- and inter-institution searching capabilities in similar ways is an important indication that the patterns identified in the TRLN libraries may apply to other libraries as well.

7.3.3 Recommendations for TRLN Libraries.

The evidence that the choice of model is supported and that the specific functions included are, in fact, frequently used is significant, but indications of problems are perhaps more important. In particular, the transaction log analysis shows that many users do not understand basic searching strategies, as indicated by the number of unnecessarily repeated searches.

The TRLN libraries should consider establishing training programs or expanding existing programs to help users understand how to use the catalog more effectively. Changes to the user interface may also help, such as making the online help system more aggressive in offering unsolicited assistance. For example, since repeated searches can be detected automatically, the system could display an informative message whenever repeated searches to the same target institution(s) are detected.

The functional comparison of BIS to the DLN Model shows several areas where the implementation approach of BIS provides less general capabilities than specified by the Model. If TRLN plans to market BIS or to expand TRLN itself to include other systems, especially in a heterogeneous environment, it will be necessary to adopt the more general approach specified by the model. However, such changes should not be made until an intersite searching language and communications protocol such as defined by LSP are adopted as standards.

7.4 Recommendations for Future Research

The work completed here suggests a number of areas for future research, including extensions to the OPAC Model itself, additional catalog use studies, additional work with implementations of the DLN Model, and further investigations of the LSAP Model and how it and decision making tools like it can be used in systems design.

7.4.1 OPAC Model Development.

An important area of future research is to verify the accuracy of the OPAC Model, especially the DLN Model, or identify necessary changes by building additional implementations based on it and studying the effectiveness of those implementations. First, as distributed systems are implemented, the DLN Model should be reviewed to verify its accuracy and completeness. In addition, it should be extended to include added functions beyond the online catalog, including circulation and interlibrary loan, acquisition/serials control, database maintenance functions, and possibly full text retrieval and access to other types of bibliographic databases.

Explicit study of the role in the system development process played by models such as this would also be of interest. Questions to be considered include what effect using reference models has on development time and what methods for developing functional specifications are most effectively used in conjunction with such models.

7.4.2 Catalog Use Studies.

There are a number of additional catalog use studies that would both provide information about the specific systems involved and add to the general knowledge of how library catalogs are used. For example, the same analysis performed here should be repeated to obtain a profile over time of how the multi- and inter-institution capabilities are used. The analysis reported here was done when the features studied were quite new, and before any substantive changes to the prior collection development agreements were made. As the library databases grow, when more specific collection development agreements have been established, and as users become more familiar with the network features, how the usage changes should be studied. In addition, there is some evidence that usage varies considerably during the semester. More detailed study to determine when the peak periods of use occur may assist the libraries in optimizing their systems to provide better service.

As discussed earlier, the transaction log analysis indicated that many users do not understand adequately how the catalog searching functions work. Once training programs and changes in the user interface are implemented, the logs should be analyzed to verify that effectiveness is actually improved.

A variety of related studies could be performed to learn more about user errors with not only the multi- and inter-institution searching capabilities, but also other catalog features such as display manipulation, session control, searching capabilities in general, and the online help facilities. The transaction log analyses could be accompanied by questionnaires, perhaps administered online at the beginning or end of terminal sessions, to obtain user profiles. A better understanding of which segments of the user community use specific catalog features could suggest improvements to the catalog design as well assist libraries in more effectively focussing training programs.

Transaction log studies should be carried out for various distributed networks once they are implemented. Other linked systems should be studied, comparing results with these findings, to determine whether important differences occur in other contexts. Transaction log analyses are also needed to determine how distributed networks implemented at the switched level are used.

7.4.3 Implementations.

Although the DLN Model provides a framework, considerable additional research is needed to determine the characteristics of effective implementations. As other implementations are developed, they should be compared to each other and to BIS. Linked and switched implementations should be compared to provide additional information on effectiveness for users, costs and operational considerations. The user interface is of particular importance, and controlled studies of alternative forms of user interfaces for both levels would be helpful to system designers.

7.4.4 LSAP Model Development.

The LSAP Model developed here demonstrates the soundness of the basic technique. However, even the modified version still represents the judgement of a very small group of experts. In addition, systems that conform to the DLN Model, especially at the linked level, are not commercially available, which makes that architecture impractical for most libraries to implement. The LSAP Model should be modified as additional libraries actually choose systems based on a particular architecture, distributed systems become more readily available, technology evolves, and as the political and economic environment changes. The Model, if modified over time, could become a tool for studying changes in system design as reflected by expert judgement.

Additional study of the attributes of the LSAP Model would also be of value. This study shows that some of the attributes seem to play a key role in determining architecture whereas others seem relatively unimportant. Further research should be aimed at determining specifically which of the attributes are key to choosing system architecture, and to achieving better understanding of how the attributes interact.

Experiments should be conducted using the LSAP Model as part of the design process to determine how the Model can most effectively be used. In addition, models specific to other applications should be developed and tested.

Further investigation is also necessary to determine whether the methods demonstrated by the LSAP Model can be applied to computer systems design in general, becoming part of the structured design process. For example, if attributes key to determining system architecture can be shown not to be application specific, then one LSAP type model could become a standard system design tool, regardless of the particular application.

7.5 Conclusion

The LSAP Model provides a planning tool that assists decision makers in determining what system architecture is most appropriate in their library environment. Although the distributed architecture is not always the best choice in today's libraries, it is expected to be increasingly important as distributed software systems are developed, as the technology to support distributed systems becomes more established, and as libraries increasingly endeavor to share their resources.

The DLN Model provides a basis for libraries to begin moving from the world of isolated catalogs to the interconnected world of a variety of information systems. Library collections represent a wealth of material that can be used only if adequate finding tools are available. Information is disseminated in an increasing variety of formats, but the fundamental library service of providing access to that information is unchanged. The library catalog is a key tool in providing that function, and although its form will evolve, the catalog will continue to serve as a finding tool in future information systems.

BIBLIOGRAPHY

American Library Association. 1980. *ALA Filing Rules*. Chicago: American Library Association.

Aveney, Brian, ed. 1984. *Online Catalog Design Issues: A Series of Discussion. Report of a Conference sponsored by the Council on Library Resources*. Washington, D.C.: Council on Library Resources.

Avram, Henriette D., and McCallum, Sally H. 1980. Directions in Library Networking. *Journal of the American Society for Information Science* 31(6): 438-444.

Barnholdt, B., and Nybroe, J. 1983. ALIS - The Online Literature Localization and Document Ordering System of the National Technological Library of Denmark - an Example of a Distributed System. *Computer Networks* 7(6): 407-411.

Bennett, David B. 1987. *BIS Online Catalog Use: A Transaction Log Analysis*. Master's paper, School of Library Science, University of North Carolina at Chapel Hill.

Besant, Larry. 1982. Early Survey Findings: Users of Public Online Catalogs want Sophisticated Subject Access. *American Libraries* 13(3): 160.

Boss, Richard W. 1984. Technology and the Modern Library. *Library Journal* 109(11): 1183-1189.

Boss, Richard W. 1985. Integrating and Interfacing Library Systems. *Electronic Library* 3(2): 124-31.

Bregzis, Ritvars, and Knapp, John F. 1979. Library Systems Development for the Triangle Universities. Consulting report.

Brooks, Frederick P., Jr., and Iverson, Kenneth E. 1969. *Automated Data Processing*. New York: John Wiley and Sons, Inc.

Bucci, G., and Streeter, D. N. 1979. A Methodology for the Design of Distributed Information Systems. *Communications of the ACM* 22(4): 233-45.

Buckland, Michael K., and Lynch, Clifford A. 1987. The Linked Systems Protocol and the Future of Bibliographic Networks and Systems. *Information Technology and Libraries* 6(2): 83-88.

Buxton, D. T. 1984. The Clemons Library Reserve Circulation System: A Model for Distributed Processing in Library Networks. *Information Technology and Libraries* 3(2): 178-182.

Byrd, Gary D., et al. 1985. The Evolution of a Cooperative Online Network. *Library Journal* 110(2): 71-77.

Casey, Genevieve M. 1970. Emerging State and Regional Library Networks. In *Proceedings of the Conference on Interlibrary Communications and Information Networks*. Chicago: American Library Association.

Ceri, Stefano, and Pelagatti, Giuseppe. 1984. *Distributed Databases: Principles and Systems*. New York: McGraw-Hill Book Company.

Chang, S. K., and Liu, C. N. 1978. Modeling and Design of Distributed Information Systems. In *Advances in Information Systems Science*. New York: Plenum Press.

Chen, P. P. S., and Akoka, J. 1980. Optimal Design of Distributed Information Systems. *IEEE Transactions on Computers* C-29(12): 1068-80.

Chu, W. W. 1969. Optimal File Allocation in a Multiple Computer System. *IEEE Transactions on Computers* C-18(10): 885-889.

Clifford, Edward. 1988. Several conversations with the author, Nov.-Dec.

Cooperative Development of a Distributed Processing System to Provide Improved Bibliographic Access for Users at Duke University, The University of North Carolina at Chapel Hill, and North Carolina State University. 1979. Application for Grant under Strengthening Research Library Resources Program.

Corey, Jim. 1988. Telephone conversation with the author, 23 August.

Davison, Wayne E. 1983. The WLN/RLG/LC Linked Systems Project. *Information Technology and Libraries* 2(1): 34-46.

DeGennaro, Richard. 1983. Library Automation and Networking: Perspectives on Three Decades. *Library Journal* 108(7): 629-635.

Edwards, Ward, et al. 1975. A Decision-Theoretic Approach to Evaluation Research. In *Handbook of Evaluation Research, Vol. I*, edited by Elmer L. Struening and Marcia Guttentag, 139-181. Beverly Hills, CA: Sage Publications.

Ferguson, Douglas, et al. 1982. The CLR Public Online Catalog Study: An Overview. *Information Technology and Libraries* 1(2): 84-97.

Fisher, M. L., and Hochbaum, D. S. Database Location in Computer Networks. 1980. *Journal of the ACM* 27(4).

Fong, Elizabeth N., and Jefferson, David K. 1986. Reference Models for Standardization. In *Proceedings of the Computer Standards Conference, 1986: Striking a Balance Between Technology, Economics, and Reality--For Substance, Not Form*, 86-90. Washington, D.C.: IEEE Computer Society Press.

Galitz, Wilbert O. 1981. *Handbook of Screen Format Design*. Wellesley, MA: Q.E.D. Information Sciences, Inc.

Gibbs, Simon, and Tsichritzis, Dionysios. 1983. A Data Modeling Approach for Office Information Systems. *ACM Transactions on Office Information Systems* 1(4): 299-319.

Grosch, Audrey N. 1985. *Distributed Computing and the Electronic Library, Micros to Superminis*. New York: Knowledge Industry Publications.

Halasz, Frank, and Schwartz, Mayer. 1989. The Dexter Hypertext Model. In *Proceedings of the Hypertext Standardization Workshop*. Gaithersburg, MD: National Institute of Standards and Technology.

Hooker, Fran. 1985. Local Area Networking: Colorado's IRVING Project. *Wilson Library Bulletin* 59(1): 38-42.

International Organization for Standardization. 1984. *Information Processing Systems--Open Systems Interconnection--Basic Reference Model*. International Standard ISO 7498.

Irving Library Network. 1986. *Technical Information*. Boulder, CO: Minicomputer Systems Incorporated.

Kaske, Neal K. 1988a. A Comparative Study of Subject Searching in an OPAC among Branch Libraries of a University Library System. *Information Technology and Libraries* 7(4): 359-372.

Kaske, Neal K. 1988b. The Variability and Intensity over Time of Subject Searching in an Online Public Access Catalog. *Information Technology and Libraries* 7(3): 273-287.

Kaunitz, J. 1982. A Descriptive Model for the Architectural Design of Information Systems. *Australian Computer Journal* 14(3): 91-98.

Kaunitz, J. 1984. Distributed Information Systems: Motivation, Concepts and Issues. In *ACC '84: Proceedings of the Australian Computer Conference*, 329-341. Sydney, Australia: Australian Comput. Soc.

King, John Leslie. 1983. Centralized versus Decentralized Computing--Organizational Considerations and Management Options. *Computing Surveys* 15(4): 319-349.

Korfhage, Robert R. 1971. *Graph Models for Library Information Networks*, NTIS Report ED-060-861.

Krikelas, James. 1972. Catalog Use Studies and Their Implications. In *Advances in Librarianship*, 3: 195-220. Edited by Melvin J. Voigt. New York: Seminar Press.

Kroenke, David. 1977. *Database Processing: Fundamentals, Modeling, Applications*. Chicago: Science Research Associates.

Lantz, Keith A. 1986. On User Interface Reference Models. *SIGCHI Bulletin* 18(2): 36-42.

Larson, Ray R., and Graham, Vicki. 1983a. Monitoring and Evaluating MELVYL. *Information Technology and Libraries* 2(1): 93-104.

Larson, Ray R. 1983b. *Users Look at Online Catalogs: Part 2: Interacting with Online Catalogs*. Final Report to the Council on Library Resources. Berkeley, CA: Division of Library Automation and Library Studies and Research Division.

Lewis, David W. 1987. Research on the Use of Online Catalogs and its Implications for Library Practice. *Journal of Academic Librarianship* 13(3): 152-157.

Linked Systems Project. 1988. Bibliographic External Design: Intersystem Functional Requirements. Unpublished report.

Linking Today's Libraries, Tomorrow's Technologies. Report of the Bibliographic and Communications Network Pilot Project. 1984. Ottawa: National Library of Canada.

Luce, Richard E. 1984. IRVING: Interfacing Dissimilar Systems at the Local Level. *Library Hi Tech* 2(3): 55-61.

Lynch, Clifford A., et al. 1983. Computing Resources for an Online Catalog. *Information Technology and Libraries* 2(1): 58-72.

Lynch, Gene, and Meads, Jon. 1986. In Search of a User Interface Reference Model. *SIGCHI Bulletin* 18(2): 25-33.

Lyngbaek, Peter, and McLeod, Dennis. 1984. Object Management in Distributed Information Systems. *ACM Transactions on Office Information Systems* 2(2): 96-122.

Markey, Karen. 1984. *Subject Searching in Library Catalogs: Before and After the Introduction of Online Catalogs*. Dublin, Ohio: OCLC.

Martin, Susan K. 1986. *Library Networks, 1986-87: Libraries in Partnership*. White Plains, New York: Knowledge Industry Publications.

Montgomery, K. Leon, and Flynn, R. 1983. A Library Network Simulation Model: A New Approach. In *Proceedings of the American Society for Information Science, 1983*. White Plains, New York: Knowledge Industry Publications, Inc.

Penniman, W. D., and Dominick, W. D. 1980. Monitoring and Evaluation of On-line Information System Usage. *Information Processing and Management* 16(1): 17-35.

Pope, Nolan F. 1981. Library Planning for Future Networks. *Resource Sharing and Library Networks* 1(1): 11-18.

Randall, William M. 1931. The Uses of Library Catalogs; a Research Project. In *Catalogers' and Classifiers' Yearbook, Number Two, 1930*, 24-32. Chicago: American Library Association.

Rice, Ronald E., and Borgman, Christine L. 1983. The Use of Computer-Monitored Data in Information Science and Communication Research. *Journal of the American Society for Information Science* 34(4): 247-256.

Rothnie, James B., and Goodman, Nathan. 1977. A Survey of Research and Development in Distributed Database Management Systems. In *Very Large Data Bases: Proceedings, Third International Conference on Very Large Data Bases*. New York: IEEE.

Rouse, William B., and Rouse, Sandra H. 1979. *A Mathematical Model of the Illinois Interlibrary Loan Network. Final Report Submitted to the Illinois State Library*. Urbana, IL: Univ. of Illinois at Urbana-Champaign.

Salmon, Stephen R. 1986. TOMUS: The System's Design and Features. *Library Hi Tech* 4(3): 51-64.

Salton, Gerard, and McGill, Michael J. 1983. *Introduction to Modern Information Retrieval*. New York: McGraw-Hill Book Co.

Sawyer, Jeanne, and Duncan, Gwyneth. 1981. TRLN Online Catalog: Centralized vs. Distributed. Unpublished report.

Schroeder, J. R., et al. 1984. *Processing and Data Distribution within the Research Libraries Information Network, final report to the Carnegie Corporation of New York: on a Study of Distributed Processing*. Stanford, CA: Research Libraries Group.

Segal, Jo An S. 1985. Networking and Decentralization. *Annual Review of Information Science and Technology*, ed. by Martha E. Williams. White Plains, NY: Knowledge Industry Publications.

Shaw, Ward. 1988. Telephone conversation with the author, 24 August.

Shelly, Gary B. and Cashman, Thomas J. 1975. *Business Systems Analysis and Design*. Fullerton, CA: Anaheim Publishing Co.

Sisson, Norwood. 1986. Dialogue Management Reference Model. *SIGCHI Bulletin* 18(2): 34-35.

Stevens, Norman D. 1980. Library Networks and Resource Sharing in the United States: An Historical and Philosophical Overview. *Journal of the American Society for Information Science* 31(6): 405-412.

Stockton, David A. 1988. *Linking the Members of the Triangle Research Libraries Network: Multi-Institutional Searching.* Master's paper, School of Library and Information Science, University of North Carolina at Chapel Hill.

Storz, W. Process Model and Resource Management in a Distributed Database. 1982. *Information Systems* 7(3): 229-236.

Sugnet, Chris. 1986. Standards: Where are We Headed? *Library Hi Tech* 4(2): 95-103.

Swank, R. C. 1971. Interlibrary Cooperation, Interlibrary Communications and Information Networks--Explanation and Definition. In *Proceedings of the Conference on Interlibrary Communications and Information Networks*, 18-26. Chicago: American Library Association.

Sweeny, Suzanne. 1990. The Evolving Catalog. *LITA Newsletter* 11(4): 15-17.

Triangle Research Libraries Network. 1988. *TRLN Bibliographic Information System: BIS Manual*. Functional Description FD-userg-1. Chapel Hill, NC: Triangle Research Libraries Network.

Triangle Research Libraries Network. 1989a. *TRLN Bibliographic Information System: Online Catalog: Command Language.* Functional Description FD-comm-1. Chapel Hill, NC: Triangle Research Libraries Network.

Triangle Research Libraries Network. 1989b. *TRLN Bibliographic Information System: Online Catalog: Display Formats.* Functional Description FD-disp-1. Chapel Hill, NC: Triangle Research Libraries Network.

Wah, B. W. 1984. File Placement on Distributed Computer Systems. *Computer* 17(1): 23-32.

Weir, Birdie O. 1979. *A Comparative Analysis of Three Major Models Most Frequently Used in Applying Computer Technology in Academic Libraries in the Southeast.* Ph.D. dissertation, George Peabody College for Teachers, Vanderbilt University.

Zimmerman, Julia. 1990. Intrepid Interfaces. *LITA Newsletter* 11(4): 13-15.

APPENDIX A: LSAP MODEL RATINGS

The forms included in this appendix are those used by the team of four experts whose judgements constitute the original version of the LSAP Model. Also shown are the results of their importance rankings and attribute measures for situations where each architecture was considered most appropriate. Specifically, the appendix includes the following parts:

A.1 Factor Ranking

Includes definitions of the attributes and the factor ranking form used to rank their relative importance.

A.2 Factor Ratings

The form used to indicate the strength of the attributes for a particular library in relation to another specific library.

A.3 Importance Rankings of Experts

The rank, value, and weight assigned to each attribute by the group of experts. These results determined the relative importance of the attributes for the original version of the LSAP Model.

A.4 Attribute Measures of Experts

The attribute measures determined by the experts for the four system architectures of the LSAP Model. These results constitute the attribute vectors of the original LSAP Model.

A.1 Factor Ranking

The following factors contribute to the decision of what system architecture (e.g., centralized, stand-alone, distributed) is most appropriate for a given library's interaction with another library, but the factors are not equally important in making that decision. Your factor rankings indicate your judgement of the relative importance of these factors in determining system architecture.

INTERNAL PURPOSE

INTERNAL PURPOSE describes the strength of the library's purpose in considering a connection with another library. Internal purpose is a measure of how strongly a library wants an interconnection with another library, without regard for why such a connection might be wanted.

USE

USE describes the expected use of access to the other library's catalog.

GEOGRAPHIC PROXIMITY

GEOGRAPHIC PROXIMITY describes the physical closeness of the libraries.

SCOPE

SCOPE describes the similarity in scope of the library collections and services, considering such factors as library type and size.

COOPERATIVE ACTIVITIES

COOPERATIVE ACTIVITIES describes the existence of cooperative endeavors between libraries and their parent institutions that generate a need for access to the other library collection, such as joint teaching or research activities.

SUPPORT SERVICES

SUPPORT SERVICES describes the strength of existing or planned supporting facilities such as direct borrowing privileges, expedited inter-library loan, inter-library bus services for users, or special document delivery services.

LOCAL AUTONOMY

LOCAL AUTONOMY describes the extent to which the library is able and willing to limit its independence to assure the coordination necessary to support interaction with another library catalog. It may be important to maintain local autonomy because of issues such as whether diverse automated systems are already installed, differences in governance (e.g., public versus private institutions), and differences in funding levels or control of funds.

EXTERNAL INFLUENCES

EXTERNAL INFLUENCES are factors outside the library that affect the decision of whether to establish the connection, and if a connection is to be established, what the nature of that connection should be. Such factors could include political requirements, as might occur if interconnections were mandated by a library's governing body, or economic expedients, such as availability of funding for cooperative projects. Exclude the possibility that the external influence is so strong and specific that it determines the entire decision.

Use pencil so you can adjust as you go!

Please rank the factors in order of importance to determining the type of system architecture (1=least important). It's OK to assign the same rank if the importance is the same.

Then rate each factor by giving the least important factor a value of 10. Go to the next highest and assign an importance value to it.

> For example, if the factor you scored as 2 in order of importance is only a little more important, you might assign it a value of 12 or 13. If it is three times more important than the least important factor, assign it a value of 30 (3x10).

Continue assigning values until a value is given to the most important factor. You can go back and forth adjusting values until the values assigned to all factors reflect your judgement of their relative importance.

The values assigned can be as high as necessary to indicate the relative importance of the factors. Factors may be considered of equal importance, and should be assigned the same importance value.

FACTOR	RANK	VALUE
Internal Purpose	____	____
Use	____	____
Geographic Proximity	____	____
Scope	____	____
Cooperative Activities	____	____
Support Services	____	____
Local Autonomy	____	____
External Influences	____	____

A.2 Factor Ratings

The ratings requested here serve as input to a planning model to determine what system architecture (e.g., centralized, distributed, stand-alone) is most appropriate for a given library's interaction with another library.

Please rate the strength of each factor in describing the relationship of libraries for which the ______________ architecture is most appropriate.

INTERNAL PURPOSE

INTERNAL PURPOSE describes the strength of the library's purpose in considering a connection with another library. Internal purpose is a measure of how strongly a library wants an interconnection with another library, without regard for why such a connection might be wanted.

If the library has very strong reasons for providing interactions with the other catalog, mark the scale at the "strong" end. If access to the other library catalog is relatively unimportant, INTERNAL PURPOSE should be marked near the "weak" end.

Desire for connection

strong |------|------|------|------|------|------| weak

USE

USE describes the expected use of access to the other library's catalog, if the two catalogs were interconnected in some way. The more heavily users are expected to use multi- or inter-institution searching capabilities, the closer the mark should be placed to the "heavy" end of the line.

Expected use of other library catalogs

light |------|------|------|------|------|------| heavy

GEOGRAPHIC PROXIMITY

Geographic proximity describes physical closeness of the libraries. For example, if the libraries are in walking distance, mark the scale as "near". The more difficult it is for users to visit the other library, the closer the mark should be towards "far".

Distance between the two libraries

near |_____|_____|_____|_____|_____|_____| far

SCOPE

SCOPE describes the similarity in scope of the library collections and services, considering such factors as library type and size.

Similarity of libraries

similar |_____|_____|_____|_____|_____|_____| diff.

COOPERATIVE ACTIVITIES

COOPERATIVE ACTIVITIES describes the existence of cooperative endeavors between libraries and their parent institutions that generate a need for access to the other library collection, such as joint teaching or research activities. The more prevalent and important such activities, the closer to "lots" the scale should be marked.

Number and importance of cooperative activities

few |_____|_____|_____|_____|_____|_____| lots

SUPPORT SERVICES

SUPPORT SERVICES describes the strength of existing or planned supporting facilities such as direct borrowing privileges, expedited inter-library loan, inter-library bus services for users, or special document delivery services.

Availability of support services

strong |——|——|——|——|——|——| weak

LOCAL AUTONOMY

LOCAL AUTONOMY describes the extent to which the library is able and willing to limit its independence to assure the coordination necessary to support interaction with another library catalog. For example, it may be important to maintain local autonomy because of issues such as whether diverse automated systems are already installed, differences in governance (e.g., public versus private institutions), and differences in funding levels or control of funds.

Importance of independence of action

not important |——|——|——|——|——|——| important

EXTERNAL INFLUENCES

EXTERNAL INFLUENCES are factors outside the library that affect the decision of whether to establish the connection, and if a connection is to be established, what the nature of that connection should be. Such factors could include political requirements, as might occur if interconnections were mandated by a library's governing body, or economic expedients, such as availability of funding for cooperative projects. Exclude the possibility that the external influence is so strong and specific that it determines the entire decision.

External push to establish connection

strong |——|——|——|——|——|——| weak

A.3 Importance Rankings of Experts

The importance ranks and values were determined by a group of four experts who reached agreement through a negotiating process. The lowest ranked factor was required to receive a value of 10, as described in the instructions accompanying the form.

After all values were assigned, the weight for each factor was calculated as a proportion of the total value points for all factors. The rank was used to assist the experts in systematically deciding what importance values to assign. The rank itself was not used to calculate the weight. Thus, for example, Internal Purpose, which was assigned 50 of the total value points, received a weight of 50/477, or .105.

FACTOR	RANK	VALUE	WEIGHT
Internal Purpose	3	50	.105
Use	8	100	.210
Geographic Proximity	4	60	.126
Scope	1	10	.021
Cooperative Activities	6	75	.157
Support Services	2	25	.052
Local Autonomy	7	92	.193
External Influences	5	65	.136
TOTAL		477	1.000

A.4 Attribute Measures of Experts

The attribute ratings were calculated by measuring the position of the marks on the rating forms and determining the percentage of the distance from the end of the line expected to most nearly represent stand-alone systems. The experts used a negotiating process to determine where to position the marks. These ratings were multiplied by the importance weights to give the weighted attribute measures.

A.4.1 Centralized Architecture.

FACTOR	WEIGHT	RATING	MEASURE
Internal Purpose	.105	89.4	9.39
Use	.210	76.4	16.04
Geographic Proximity	.126	75.6	9.53
Scope	.021	78.9	1.66
Cooperative Activities	.157	65.8	10.33
Support Services	.052	82.9	4.31
Local Autonomy	.193	10.6	2.05
External Influences	.136	80.5	10.95

A.4.2 Distributed Linked Architecture.

FACTOR	WEIGHT	RATING	MEASURE
Internal Purpose	.105	78.9	8.28
Use	.210	63.4	13.31
Geographic Proximity	.126	62.6	7.89
Scope	.021	67.5	1.42
Cooperative Activities	.157	61.0	9.58
Support Services	.052	75.6	3.93
Local Autonomy	.193	37.4	7.22
External Influences	.136	61.8	8.40

A.4.3 Distributed Switched Architecture.

FACTOR	WEIGHT	RATING	MEASURE
Internal Purpose	.105	49.6	5.21
Use	.210	37.4	7.85
Geographic Proximity	.126	50.4	6.35
Scope	.021	54.5	1.14
Cooperative Activities	.157	51.2	8.04
Support Services	.052	53.7	2.79
Local Autonomy	.193	69.9	13.49
External Influences	.136	39.0	5.30

A.4.4 Stand-Alone Architecture.

FACTOR	WEIGHT	RATING	MEASURE
Internal Purpose	.105	19.5	2.05
Use	.210	13.8	2.90
Geographic Proximity	.126	39.8	5.01
Scope	.021	45.5	.96
Cooperative Activities	.157	10.6	1.66
Support Services	.052	23.6	1.23
Local Autonomy	.193	78.0	15.05
External Influences	.136	23.6	3.21

APPENDIX B: LSAP MODEL VERIFICATION

The forms included in this appendix are the additional forms (beyond those shown in Appendix A) sent to the systems librarian at each test library. Also shown are the results of the test library importance rankings and attribute measures as compared to the modified LSAP Model.

B.1 Example Cover Letter

3262 Indus Ct.
San Jose, CA 95127
April 4, 1989

Dale Flecker
Assoc. Director for Planning and Systems
Harvard University Library
88 Widener Library
Cambridge, MA 02138

Dear Dale:

I am finally ready to make good on my promise (threat?) to ask for your assistance with my dissertation. The basic idea is as follows.

I have developed a planning model at is based on the hypotheses that:

1. Certain factors, which I have identified, are key in determining which system architecture (centralized, stand-alone, or distributed) is most appropriate for a given environment, and

2. When library system planners rate the strength of those factors with regard to the need for their library catalog to interact with another library catalog, the factors can discriminate appropriately among the system architectures.

I am requesting your help in verifying that my model works. To do that, I need you to rank these determining factors with regard to your library's need for a cooperative catalog of some sort with another library. Using the enclosed forms, please do the following:

a. Rank the factors in order of importance in determining the best system architecture and assign importance weights. This establishes the relative importance of the factors (e.g., how much more or less important geographic proximity is than expected use). Definitions of the factors and more specific instructions are included on the "Factor Ranking" form.

b. Determine the strength of each factor as it describes your relationship with another specific library. Select a library that it might seem reasonable for Harvard to consider an interconnection or union catalog with, but with whom you have no intention of establishing one. Write the name of the library in the space provided on the "Library Rating" form, and complete the rest of the form with regard to that specific library.

I hope this makes sense, but if not please give me a call. My phone number at home is 408-929-5602 and at work is 408-725-5857. I would appreciate it if you could return these by April 30. Thanks for your help!

Sincerely,

Jeanne Sawyer

B.2 Additional Questions Included With Rating Form

How would you describe the system architecture of the system your library currently has operational?

___Centralized

___Distributed, highly interconnected

___Distributed, minimally interconnected

___Stand-alone, single institution

Considering only those capabilities concerning access to the other library's holdings:

What features of your current system are you satisfied with?

What features would you add to the system if you could?

Of the features of your current system that are not used and if you were designing a new system, which would you:

Keep the basic capability but change how it works? (please indicate the general nature of the needed change)

Omit entirely?

Additional comments?

B.3 Importance Weights for Test Libraries

B.3.1 Centralized Test Libraries.

Attribute	Model	FIU	FSU	UNF
Internal Purpose	.105	.054	.090	.189
Use	.210	.080	.180	.114
Geographic Proximity	.126	.214	.060	.076
Scope	.021	.064	.072	.068
Cooperative Activities	.157	.267	.150	.038
Support Services	.052	.160	.120	.042
Local Autonomy	.193	.054	.090	.227
External Influences	.136	.107	.239	.246

B.3.2 Distributed Linked and Switched Level Test Libraries.

Attribute	Model	Duke	NCSU	UNC-CH
Internal Purpose	.105	.194	.111	.214
Use	.210	.222	.201	.238
Geographic Proximity	.126	.055	.106	.119
Scope	.021	.167	.101	.024
Cooperative Activities	.157	.111	.156	.155
Support Services	.052	.083	.151	.143
Local Autonomy	.193	.139	.050	.071
External Influences	.136	.028	.126	.036

B.3.3 Stand-Alone Test Libraries.

Attribute	Model	Harvard	Texas	Wisconsin
Internal Purpose	.105	.229	.166	.217
Use	.210	.163	.100	.046
Geographic Proximity	.126	.046	.132	.027
Scope	.021	.039	.066	.160
Cooperative Activities	.157	.065	.073	.115
Support Services	.052	.098	.093	.069
Local Autonomy	.193	.327	.172	.183
External Influences	.136	.033	.199	.183

B.4 Test Library Attribute Vectors Compared to Modified LSAP Model

B.4.1 Centralized Test Libraries.

Attribute	Model	FIU	FSU	UNF
Internal Purpose	5.06	0.53	7.14	3.16
Use	10.16	3.30	13.68	7.60
Geographic Proximity	8.54	16.46	4.36	3.80
Scope	3.51	4.86	5.29	2.26
Cooperative Activities	9.77	24.48	3.60	0.63
Support Services	4.64	7.01	5.16	2.10
Local Autonomy	6.82	0.45	2.08	22.70
External Influences	15.05	6.28	18.38	24.60
Similarity to Model				
COS (Centralized)		.73	.92	.79
COS (Distributed–Linked)		.59	.76	.51
COS (Distributed–Switched)		.66	.75	.62
COS (Stand-Alone)		.32	.50	.75

B.4.2 Distributed Linked Level Test Libraries.

		Duke with		NCSU with		UNC-CH with	
Attribute	Model	UNC-CH	NCSU	Duke	UNC-CH	Duke	NCSU
Internal Purpose	14.35	18.29	18.29	8.10	8.37	19.65	19.47
Use	13.80	17.09	17.09	4.94	8.24	17.95	17.95
Geographic Proximity	4.62	4.06	4.06	7.99	2.52	2.73	3.12
Scope	4.96	12.04	12.04	4.89	2.40	1.02	0.94
Cooperative Activities	10.51	8.37	8.37	9.20	9.59	14.35	14.10
Support Services	7.62	6.26	6.26	6.55	6.31	13.24	10.78
Local Autonomy	5.78	12.30	12.30	3.77	3.69	0.53	0.64
External Influences	5.33	2.18	2.18	9.30	9.30	2.71	3.28
Similarity to Model							
COS (Centralized)		.72	.72	.93	.92	.68	.70
COS (Distr.–Linked)		.89	.89	.86	.93	.95	.96
COS (Distr.–Switched)		.90	.90	.88	.89	.85	.86
COS (Stand-Alone)		.77	.77	.65	.64	.52	.53

B.4.3 Distributed Switched Level Test Libraries.

Attribute	Model	NCSU with UNC-G	UNC-CH with NCCU	UNC-G
Internal Purpose	4.60	2.82	5.09	5.26
Use	6.28	1.65	5.85	9.76
Geographic Proximity	4.29	0.95	9.07	0.78
Scope	1.20	2.48	0.61	0.57
Cooperative Activities	4.14	1.15	3.94	3.43
Support Services	3.38	3.71	3.52	3.52
Local Autonomy	4.64	3.89	0.53	0.64
External Influences	2.20	1.13	0.94	1.45
Similarity to Model				
COS (Centralized)		.70	.74	.70
COS (Distr.--Linked)		.79	.82	.92
COS (Distr.--Switched)		.84	.88	.86
COS (Stand-Alone)		.84	.55	.50

B.4.4 Stand-Alone Test Libraries.

Attribute	Model	Harvard	Texas	Wisconsin
Internal Purpose	9.82	1.88	1.49	20.27
Use	4.29	1.06	0.73	2.30
Geographic Proximity	5.04	4.41	10.06	2.48
Scope	4.03	1.05	0.48	8.00
Cooperative Activities	2.54	0.42	0.53	3.10
Support Services	2.39	0.80	0.76	4.46
Local Autonomy	16.98	29.76	15.79	16.34
External Influences	3.52	0.35	1.79	13.34
Similarity to Model				
COS (Centralized)		.37	.53	.71
COS (Distr.--Linked)		.32	.40	.72
COS (Distr.--Switched)		.51	.63	.72
COS (Stand-Alone)		.86	.86	.85