Communication Standards for Online Interchange of Library Information

THOMAS P. BROWN

Introduction

THIS PAPER PRESENTS a new perspective on computerized information interchange within the library community. First, a simple yet revealing model for linking computer systems is explained. The model is then used as a foundation for describing a solution for connecting three major bibliographic services in a cooperative, nationwide network.

The impetus for examining automated mechanisms for exchanging library information lies within a very fundamental precept of the library world—resource sharing. Information resources are shared not only between libraries but with the patrons themselves. As recognition of the value of information spreads within the political community, there will come a reassessment of existing library policies for resource sharing. One should expect not a dampening of cooperation but, on the contrary, an increased awareness of the assets which are presently being maintained by libraries. An automated mechanism for tracking and managing these assets will further facilitate the transformation of libraries into the information age.

In order for the increasing variety of computer systems which provide library services to be able to interconnect logically, communications standards must be specified. The computing community has been highly active in this area for the past ten years; their term for it is *distributed data processing*. In recent years the computer vendor parochial attitudes have been restructured by the popular acceptance of new telecommunication standards on an international scale. This has

Thomas P. Brown is Manager of Computer Services, Washington Library Network.

been very beneficial for all computer users and has broadened their expectations for computer system interconnectability. In this light the chance for the adoption of computer communication standards for the library community will not be constrained by technical issues; the responsibility rests most heavily on the effective automation of the existing library resource-sharing principles.

Open Systems Interconnection Model

A simple building-block approach to solving communication problems is being promoted by the International Organization for Standardization (ISO), the American National Standards Institute (ANSI), and the National Bureau of Standards (NBS). This Open Systems Interconnection model serves the purpose to "provide a common basis for the coordination of standards development that will enable computer systems to interconnect."[1] The open nature of the model implies that standards development is not predicated on any single vendor of computer systems. However, the model can be applied to existing communication network architectures, such as IBM's Systems Network Architecture (SNA), Digital Equipment Corporation's DECnet, and the U.S. Department of Defense's Arpanet, as well as to new communications network architectures that are being developed based as satellite and cable systems.

There are two underlying assumptions built into the Open Systems Interconnection model. First, the communications between computer systems has bi-directional and real-time requirements; that is to say, it is online. For instance, the model does not pertain to the familiar exchange of magnetic tapes which contain bibliographic records. Secondly, the computer systems involved are treated as "peer" systems. No master/slave relationships are allowed within the model. Consequently, computer terminals to host computer system communications are not examined here. Host-computer to remote-host-computer communications is the major problem domain for the model, and that is the configuration which is discussed in this paper.

A prerequisite for computer communications is that both machines must be processing data for somewhat similar purposes. For example, it makes no sense to connect a weather-predicting computing machine to one that is solely processing income tax records. At this highest conceptual level, there must be some prior agreement that the applications being automated serve similar purposes. This can be true if there is a uniform definition of the data record, such as an existing communications standard. Such is the case with the bibliographic services in this country—OCLC, Inc., the Research Libraries Group (RLG), the Washington Library Network (WLN)—and bibliographic services in other countries. The way in which the data are processed onto a database, how data are indexed and updated, and what transactions the terminal user issues are distinctly different in these three systems. However, all three provide the basic bibliographic cataloging application.

To interconnect online bibliographic services which differ in external and internal operation, a new application-to-application communication standard must be defined. This is much more than the standard MARC format for bibliographic data. It must standardize, at least on the communications link, the functions, sequencing, and meaning of the data. A new applications protocol for message and data interchange must be put into place.

Figure 1 represents two computer systems in the building-block Open Systems Interconnection model with the database applications residing at the highest level. Proceeding downward, the lower levels provide more rudimentary communications functions, until at the lowest level the physical interconnecting wire (or wires) is represented.

At each level there must be prior negotiation and agreement by the computer system designers as to the precise message and indication protocol to which the computers are to adhere. Logically, there are seven layers of protocols spanning the computer systems, viewed as horizontal layers in figure 1. A malfunction in any one of these layers will result in an interruption of the communications. For example, a termination of the database application on one computer will have the same effect, at least to the user, as disconnecting the physical telephone wire connecting the systems. The main benefits of the scheme are: (1) communication problems, whether design or operational, can be classified by layer and scientifically resolved; (2) standard design nomenclature is adopted for each layer; and (3) a change in the design of one layer has small impact on other layers.

At the junction of any two levels of the model is an interface which specifies how communications service requests and indications are transferred into the adjoining level. These interfaces must be physically traversed within a system to effect the logical horizontal layer communication.

The second highest layer, the presentation layer, deals primarily with data transformations into compatible syntax and character sets without loss of meaning. The transformations performed by either system at this level can even define a new computer language to repre-

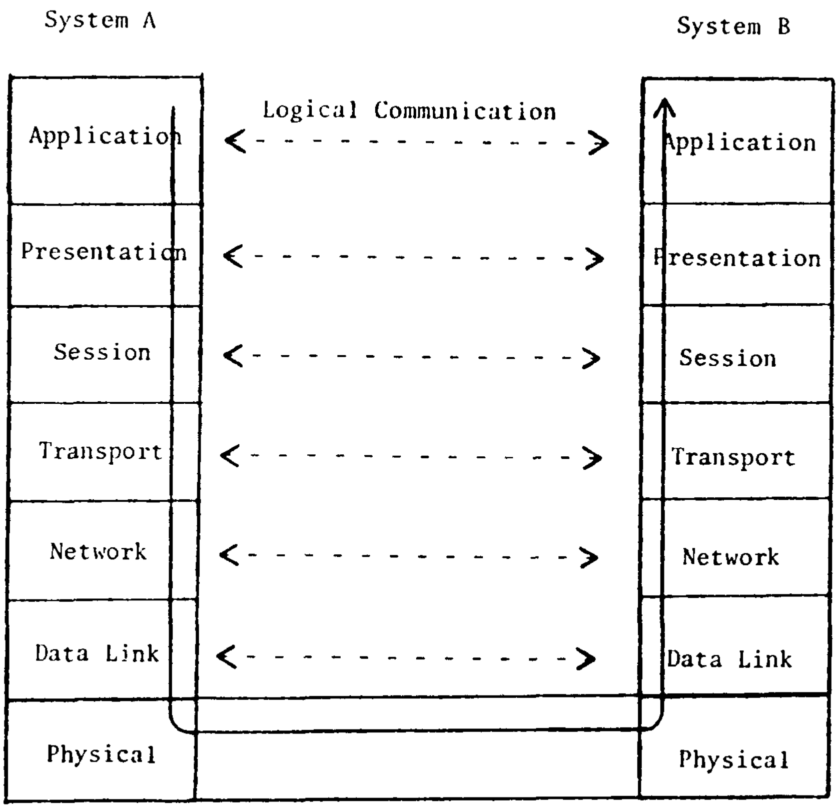

Fig. 1. Open Systems Interconnection Model

sent the application data. For example, encryption/decryption transformations may be implemented for secure communications.

Below the presentation layer is the session layer, which manages multiple concurrent dialogues between the two remote applications. These dialogues may be taking place on behalf of persons using either computer system, or may be taking place as independent application-to-application communications. In either case the session layer controls characteristics of the dialogue, such as which computer system's "turn" it is to "speak." It also detects when an abnormal dialogue condition arises, so that the upper presentation and applications levels are posted with the problem status. Two of the basic operations to be performed at this level are session establishment and session disestablishment.

The transport layer responsibility entails the end-to-end delivery of complete and correct messages. It may select among various grade network services to deliver messages to a target system, and it may compensate for less reliable network layer services.

The network layer may potentially use many separate communications processers to deliver message packets. The services of a public packet switching network such as Telenet or Tymnet may be used at this level. Breaking messages into small packets requires sequencing control, flow control (so that network processers are not overworked), routing control, and time-multiplexing packets over network circuits. With processer and circuit redundancy, it is possible to improve network reliability while at the same time reducing costs by dynamic adjustment of the network node configurations upon component failure. Obviously there are many control messages built into a network control protocol between network nodes to enable this type of sophisticated message-delivery system.

The widely used Telenet and Tymnet packet switching systems both specify the CCITT X.25 standard network interface.[2] They both meet the packet handling requirements of the model's network layer.

The data link layer specifies how streams of bits are delineated into manageable segments called frames. Certain bit patterns are specified to indicate the start of a frame and end of a frame. Most data link protocols such as BISYNC (Binary Synchronous) protocol and the ISO standard HDLC (High-Level Data Link Control) protocol provide for error-free transmission by including a bit pattern at the end of each frame, called a checksum, which is precisely derived from each preceding bit in the frame. If a communication line transient condition alters one or more bits within the frame, then the error is detected when the re-derived checksum is found to differ from the transmitted checksum. Errors are reported to the network level so that proper retransmission takes place without involving higher level layers.

Finally, the lowest level, the physical level, specifies how electrical signaling is used to transmit bit streams. The Electronics Industries Association RS232-C connection standard is the most popularly used physical-level interface standard.[3] Newer standards are being developed for use with large band-width cable and satellite systems.

There is certain flexibility, at least conceptually, in applying the seven-level model as a new architecture for intersystem communications. Alternative implementations of the levels may be considered in either a top-down or bottom-up direction. Evaluation and negotiation of communications protocols may proceed in a structured fashion. And finally, adoption of existing standards or development of new standards can be done stepwise by level.

This discussion has given only a cursory introduction to the Open Systems Interconnection model. There is substantially more detail to be found regarding the functioning of each of the seven layers of the model within the formal definition and in a recent paper entitled "Network Protocols" by Andrew S. Tanenbaum.[4] Since the model is a relatively new concept in telecommunications, much more refinement in the model is expected as it is exercised in practical situations.

Linked System Project

Background

An ambitious computer linking project has been undertaken by the Library of Congress (LC), the RLG, and the WLN through funding from the Council on Library Resources (CLR) to begin the formulation of a national bibliographic network. New communications standards are being developed using Open Systems Interconnection methodology so that in the future other computer systems may be readily interfaced to the network on a peer basis. The first application to use the linkage will be a nationwide authority service drawing on the authority databases being maintained by each of the three organizations.

Experiments have been conducted on the interchange of bibliographic data between computer systems and have proved that a thorough and complete specification of application, presentation and session protocol must be done before reliable communication can take place. In 1976 an experimental online link was developed and tested between LC's computer and RLG's computer at the New York Public Library. Technically, the linkage worked on a simple basis; it was unidirectional. The message transport mechanism was implemented by having the New York Public Library computer put on the guise of an LC computer terminal. Although this terminal emulation-mode linkage was achieved in a speedy fashion, it was not reliable, efficient nor easy to manage. Furthermore, the emulation mode severely constrained the possibilities for expanding its usage.

Concurrent with this experiment, several efforts were proceeding to lay a better foundation for coordinated communications. Foremost was a task force cosponsored by the National Commission on Libraries and Information Science (NCLIS) and NBS to establish an applications-level protocol for library information.[5] Indeed, this was conceived partly on the early work on the Open Systems Interconnection model.

Unfortunately, the NCLIS/NBS task force did not adequately address protocol functions in a database environment, leaving important protocol negotiation to a national registry which was never formed. However, the draft protocol did point out two very important problems to be resolved: (1) the need for a common presentation of a query language, and (2) the need to establish a session-level layer in the protocol model. Additionally, the NCLIS/NBS protocol was judged to be difficult to off-load onto minicomputer systems.[6]

A second effort was initiated by the Library of Congress and the Network Technical Advisory Group to propose a Message Delivery System to be developed in a well-organized and thorough manner.[7] The CLR-funded Linked Systems Project has taken over the pathway set out for the Message Delivery System.

Development

The need for a consistent national union catalog is obvious. Repeated use of data stored in a national union catalog would result in very large savings of librarian staff work. However, a consistent national union catalog presupposes a consistent and coordinated authority.[8] Approximately 18 percent of bibliographic record data is authoritative in nature and requires more staff work to assemble and validate than the other 72 percent of the descriptive cataloging. This makes a shared nationwide authority file a principal objective of a national union catalog.

One of the key elements of the CLR Bibliographic Services Development Program's five-year plan is the development of an integrated consistent authority file service for nationwide use.[9] To this end a task force was formed in 1979 to address the organizational issues and make recommendations for a Name Authority File Service (NAFS).[10]

The Linked Systems Project will develop the technical mechanisms by which the NAFS will be distributed in an online computerized mode. While the "master" NAFS file will reside at LC, the maintenance responsibilities will be distributed among many participating libraries at RLG and WLN. In addition, LC will fill the critical overseer role to ensure quality control as well as supply a large portion of the authority data.[11] Future participation in the NAFS will be encouraged via standard linkage mechanisms defined by the Linked Systems Project.

There has been substantial work on revising the MARC Authority record format standard in anticipation of its use in linked, online communications. Simplifications in control subfield data and in updating images will reduce the implementation time for interfacing to the standard. With the implementation of AACR2, the data content of authority records has been more closely aligned throughout the library world, enabling a wider base of participation in the creation of data. The revised MARC Authority format has been approved as a standard through the MARBI (Machine Readable Form of Bibliographic Information) American Library Association committee which controls all changes in MARC formats.[12]

Use of the Model

In order to negotiate decisions on the implementation of the link among the three organizations (LC, RLG and WLN), a classic cost/benefit analysis mechanism was attempted. Alternatives for the system components were developed, described and then weighed against a set of criteria to judge benefits. The vendor community was surveyed to determine what off-the-shelf software existed and what developmental services were available. Standards activities, especially at NBS, were studied to determine whether these activities were far enough advanced to be of use. Finally, the capabilities of the three organizations were examined to determine the level of development each could contribute to the project.

The following criteria were used for the evaluation of the telecommunication linking alternatives.

1. The reference model of Open Systems Interconnection should be followed in the design of the telecommunications link.
2. The link must support multiple simultaneous application-to-application interactions among host computers (i.e., it does not use a terminal-emulation protocol).
3. The software involved in the link must have available a sufficient level of support to ensure its operational reliability. The link should not require excessive operational personnel. Network management functions must not be dependent on any single node.
4. The software used must belong to a clear evolutionary software architecture to ensure longevity for software support.
5. The link should be cost-effective to operate.
6. The telecommunications system should be extendable to other computer systems (mainframe, minicomputer and perhaps microcomputer), following the "open" nature of Open Systems Interconnection.

7. The telecommunications system should support substantially higher traffic volume (such as would be generated by bibliographic exchange, interlibrary loan, etc.) without major redesign.

Based on these criteria, the vendor-dependent solutions, such as Systems Network Architecture (SNA), DECnet and Arpanet, were rejected. A minicomputer Network Front-End Processer (NFEP) approach was selected as the best solution for establishing the initial network, and local system development was selected to define the application. In terms of the Open Systems Interconnection model, the layers will be defined as follows:

Authority Services:
Application Layer—Develop new standards
Presentation Layer—Develop new standards and use MARC communications formats.

Telecommunication Services:
Session Layer—Use NBS proposed standard
Transport Layer—Use NBS proposed standard
Network Layer—Telenet X.25 level 3 service
Data Link Layer—Telenet X.25 level 2 service (LAP-B)
Physical Layer—Telenet X.25 level 1 service (RS232-C)

The authority services will define new application and presentation communication protocols for searching remote authority files and displaying results of the searches, and for intersystem maintenance of authority files. These standard protocols will be set in general terms to be independent of implementation, but will be usable as a specification for development at WLN, RLG and LC.

The telecommunications services will define a standard network interconnection for delivery of messages from one computer to another.[13] It will accommodate authority messages, as well as messages from other applications that may eventually be created.

Figure 2 gives a schematic representation of the system components for two of the three computer systems. Since all three organizations make use of large IBM mainframe processers and front-end processers, the same logical system component diagram is applicable to each. However, nothing in the selection of the communications protocol standards requires a similar system component arrangement for future interfacing sites. In fact, the large effort that has been applied to standards development in the context of the Open Systems Interconnection model has assured independence of vendor and system component configurations.

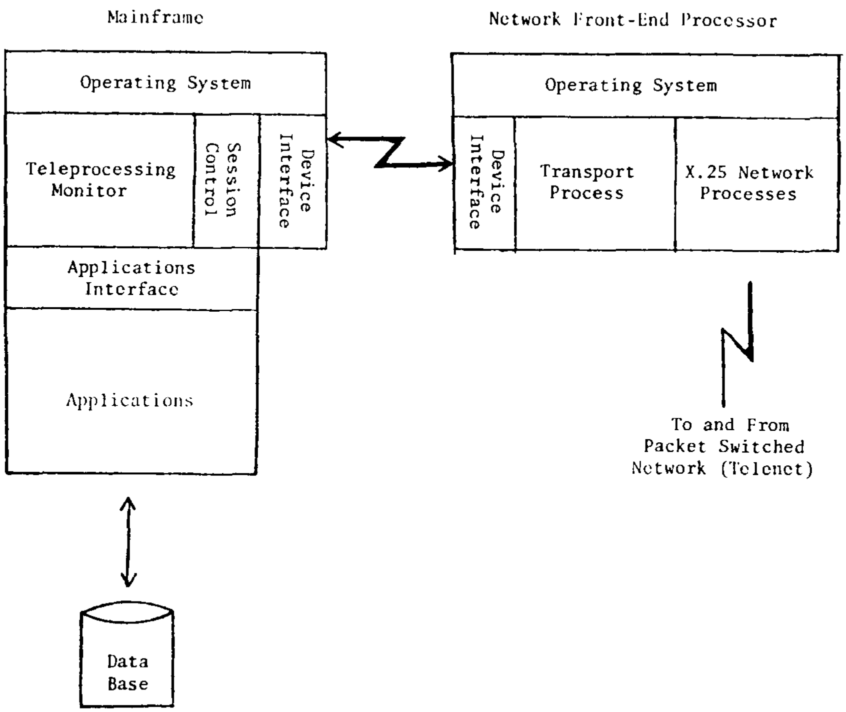

Fig. 2. Linked Systems Project—Software Components of System A or B

Crucial to the success of the layer implementations are perfectly fitting system component interfaces. One technique that can be used is to focus on an interface and reapply a transport/session-level submodel which will help in analyzing how the interface is to be used. However, caution is due here in using this technique, since the interface is between two specific devices and does not follow the general model's requirement "openness."

In the broad picture, it is clear that the importance of the model has been to help establish a common vocabulary to negotiate the characteristics of the computer-to-computer protocol. The terminology defined by the model is growing in popularity as further agreement on the conceptual functioning of each layer is reached.

Search/Response Application

At the application level are two new application protocols being defined for interlibrary system communications; the first of these is the search/response protocol. Each of the three current participants in the Linked Systems Project operates its own authority database and query subsystems. A search is formulated by the computer system user according to the syntactic and semantic rules of the local system and the resulting display is structured in a manner that the user can understand. Of course, to link these three systems together in a fashion whereby a user on one system may obtain database services from the other systems, a translation function must be performed to reconcile the differing search and display syntaxes. It was decided that this translation function should not be performed by the user, but should be automated. This avoids the confusion that can be seen arising with the diversity of online reference services, and avoids the burden it places on the user to remember the proper computer command syntax for each system. A generalized search/response application-level protocol is being defined through funding from CLR to be used in the Linked Systems Project and for proposal as a national library standard.[14] Initially this search/response protocol will be applied to authority database searching, but it will be formulated in a general way so that it can also be applied to bibliographic and other database searching.

There are two roles for a host system to play in the search/response protocol—the originator system for a search, and the target system for a search. As originator, the host system must translate the user's search into the communications search standard and direct the search to the specified target system. Records received in answer to the search will be in MARC communications format and will be used by the host system to format and present to the user a conventional display screen. On the other hand, a host functioning as the target system must receive the incoming search request, translate it into internal format, search its database according to the search criteria, extract the records satisfying the search, and send them in communications format to the originating system. The target system must retain search results during a search session so that further qualification or restoration of previous search results can be requested by the user. The originating system, on behalf of the user, establishes the start and end of the search session using services of the session layer in the Open Systems Interconnection architecture.

Intersystem File Maintenance Application

The second application protocol being developed is for record transfer and synchronization of databases between the host systems. Provisions are made for creation of new records, and for changes and deletions of existing records to be communicated between computer systems. Two types of online links have been defined to support two separate concepts of participation in the building of a national "logical" database. A record contribution link provides the ability to create new records and change existing records in the Name Authority File (NAF), while a record distribution link between the master NAF and a remote site ensures that all changes made on the NAF database are precisely communicated to the remote database. For each distribution link there is a duplication of disc storage at the remote site for each record in the NAF. However, the cost of extra storage is offset by the lower search/response communications traffic. The intersystem search/response protocol will be minimally used at a remote site which maintains a nearly synchronized database with the NAF. With the contribution link, editing and validation of records to be added to the NAF is done before acceptance of the records.

The Linked Systems Project will implement both types of online links. LC will support the Name Authority File. WLN and RLG will contribute records to the NAF via a contribution link, will subscribe to online distribution, and will extract selected records through the search/response protocol.

Software Standards

The final result of the Linked Systems Project will be an operating link between three dissimilar mainframe computer systems over which search requests and data records will be communicated. Minicomputer hardware will be installed, software will be written for the minicomputer, and software will be changed and added to the mainframe computers. As much as is feasible, high-level computer languages such as PL/I and Pascal will be used so that portions of the software, especially the minicomputer transport and network software, and the mainframe session-layer software may be available to future NAF participants and will be supported through periodic maintenance releases. In addition, WLN will include software developed in the Linked Systems Project into its standard licensed software package.

References

1. International Organization for Standardization. Technical Committee 97. Subcommittee 16. *Data Processing-Open Systems Interconnection-Basic Reference Model* (ISO/TC97/SC16 Draft Proposal ISO/DP 7498), 3 Dec. 1980, p. 3.

2. *Recommendation X.25: Interface Between Data Terminal Equipment (DTE) and Data Circuit Terminating Equipment (DCE) for Terminals Operating in the Packet Mode on Public Data Networks CCITT*. Geneva: ISO, 1980.

3. Electronics Industries Association. *EIA Standard RS232-C: Interface Between Data Processing Terminal Equipment and Data Communications Equipment, Electronics Industries Association*. Washington, D.C.: EIA, p. 19.

4. Tanenbaum, Andrew S. "Network Protocols." *Computing Surveys* 13(Dec. 1981):453-89.

5. NCLIS/NBS Task Force on Computer Network Protocols. *A Computer Network Protocol for Library and Information Science Applications*. Washington, D.C.: National Commission on Libraries and Information Science, 1977, p. 8.

6. Protocol Implementation Planners. *Final Report to the Library of Congress Network Development Office on a Computer-to-Computer Communications Protocol Implementation Plan for the National Library and Information Service Network*. New York: New York Public Library, 1978, pp. 28-34.

7. Network Technical Advisory Group. *Message Delivery System for the National Library and Information Service Network: General Requirements*. Washington, D.C.: Library of Congress, 1978.

8. Buchinski, Edwin J. *Initial Considerations for a Nationwide Data Base*. Washington, D.C.: Library of Congress, 1978, pp. 15-17, 32-43.

9. Council on Library Resources. Bibliographic Services Development Program. "An Integrated Consistent Authority File Service for Nationwide Use." *Library of Congress Information Bulletin* 39(11 July 1980):244-48.

10. Task Force on Name Authority File Service. *Requirements Statement for the Name Authority File Service*. Washington, D.C.: Council on Library Resources, 1981.

11. Washington Library Network/Research Libraries Group/Library of Congress. "Proposal to the Council on Library Resources for an Authorities Implementation and Bibliographic Analysis," 5 March 1982, pp. 8-9.

12. American Library Association. *Handbook of Organization 1981/1982 and Membership Directory*. Chicago: ALA, 1981, p. 108.

13. Washington Library Network/Research Libraries Group/Library of Congress. "Proposal to the Council on Library Resources for a Standard Network Interconnection," Dec. 1981.

14. Council on Library Resources. "Application Level Protocol Development for Library and Information Science Applications—Features Description and Analysis" (draft), 21 April 1982.

This Page Intentionally Left Blank