Building the Infrastructure of Resource Sharing:
Union Catalogs, Distributed Search,
and Cross-Database Linkage

CLIFFORD A. LYNCH

ABSTRACT

Effective resource sharing presupposes an infrastructure which permits users
to locate materials of interest in both print and electronic formats. Two
approaches for providing this are union catalogs and Z39.50-based dis-
tributed search systems. The advantages and limitations of each approach
are considered, paying particular attention to a realistic assessment of
Z39.50 implementations. This article argues that the two approaches
should be considered complementary rather than competitive. Technolo-
gies to create linkage between the bibliographic apparatus of catalogs
and abstracting and indexing databases and primary content in electronic
form, such as the new Serial Item and Contribution Identifier (SICI) stan-
dard are also discussed as key elements in the infrastructure to support
resource sharing.

INTRODUCTION

Effective information access within a library and, to an even greater
extent, interlibrary resource sharing, both presuppose that library pa-
trons have the ability to effectively identify and locate materials of inter-
est. As library materials include an increasing amount of electronic con-
tent, even materials that are part of the “local” collection may not be
stored on site. With the growth of resource sharing as an explicit strate-
gic response to the inability to fund sufficiently comprehensive local col-
lections, access across multiple collections is becoming increasingly criti-
cal. Specifically, the ability to locate and identify materials in this context
implies that patrons must be able to search the holdings of multiple li-

braries and to navigate among disciplinary or citation (abstracting and indexing) databases defining logical views of a literature and primary content (both in printed form and electronic formats). Three key technologies to support these requirements are union catalogs, distributed search, and cross-database linkage systems. This article attempts to take a realistic look at these infrastructure components and examines the promises and limitations of the technological approaches available to implement them.

We have come to take the relatively mature and well-tested technology of union catalogs (both in the narrow sense of union catalogs for clusters of libraries and the broader sense of international community-wide union catalogs such as those offered by OCLC and RLG) very much for granted and, at least in our rhetoric, sometimes cast them as archaic constructs that will soon be replaced by fully distributed search approaches enabled by standards such as the Z39.50 computer-to-computer information retrieval protocol. The development of Z39.50 from an experimental protocol to a viable commercial technology has given rise to a great deal of confusion. Z39.50 is a seriously misunderstood standard. Common perceptions of the capabilities of this standard, and of systems that implement it, have shifted from skepticism to unreasonably high expectations. The limitations of Z39.50, both as a protocol and as deployed in current implementations, are discussed in some detail.

This article will make the argument that, in practical terms, the union catalog is far from obsolete—indeed, union catalogs complement the emerging distributed search models by offering substantially different functionality, quality, performance, and management characteristics. The key question for libraries and their patrons is how to use most effectively the two approaches together.

Abstracting and indexing (A&I) databases are now well established resources for library patrons that exist alongside the various types of catalogs; increasingly, the extended functionality of local integrated library systems and the availability of Z39.50 is making it possible to offer access to catalogs and abstracting and indexing databases through common user interfaces. The multiplicity of partially overlapping A&I databases available to users is beginning to raise design issues that have considerable similarity to those involved in the development of union catalogs. In addition, A&I databases need to be linked both to the catalogs and to lists of serials (representing print holdings) and to electronic primary content that is now becoming available on the network. The final sections of the article examine some of these issues.

The Functional Characteristics of a Union Catalog

Union catalogs provide a coherent view of the holdings of multiple libraries or library collections. They go beyond the normal functions of
a single-collection catalog, not only bringing together works by the same author or about the same subject in response to user queries, but also by bringing multiple instances of the same work (perhaps described differently by different institutions) together for the user searching the database. They often offer uniform (or unifying) name and subject authorities as a means of furthering the basic catalog objectives of bringing together works of common authorship or subject; this can compensate for variations in cataloging practice among the participant collections.

Union catalogs provide users with the ability to perform consistent searching of records from multiple institutions, in the sense that these records are indexed consistently. (For example, there is uniformity in the choice of fields from the records used to construct the various search indexes and also uniformity in the way in which search keys such as keywords or personal names are extracted from these fields and normalized for indexing.) In contrast to distributed search approaches, a union catalog almost trivially ensures consistent query interpretation—for example, the application of personal name algorithms and the treatment of case and punctuation in search terms in the user query.

Finally, a union catalog is presented to its users as a high-quality managed information access system. This means that the system should meet standards for reasonably rapid and predictable response time, high availability and reliability, and good communication about outages; and the user should expect its behavior to be highly repeatable from session to session.

To this point, I have described functional characteristics of a union catalog, independent of implementation; in theory, such a union catalog could be implemented by a single centralized database, a distributed database which is centrally administered, or by a user interface to a distributed search system which accepted user queries, derived and dispatched appropriate queries to multiple autonomously managed heterogeneous databases and then post-processed the results for presentation to users. In practice, all of the systems I know about which meet these functional criteria are essentially centrally administered systems. The distinction between a centrally designed and operated system that is implemented as a centralized database and one that is implemented technically as a distributed database is increasingly meaningless; even a single large mainframe is now effectively a set of distributed machines on a very fast local area network. Thus, for the purposes of the discussion here, I will contrast the centrally designed, managed, and operated implementation with the distributed search model, which is characterized by heterogeneity and local autonomy in the design and management of the individual databases.

The next few sections will be an examination of how centralized implementations of union catalogs meet the functional characteristics—the
broad areas of consistent searching/indexing, consolidation of records, and performance/management—described above. I will then examine the technology of distributed search and consider the extent to which it can meet the same functional objectives.

**Centralized Implementations of Union Catalogs**

Online union catalogs have been around since the 1970s. They take three major forms which reflect evolutionary paths of development and, to some extent, the business and organizational models that currently support them.

*Commercial Services.* Commercial services—i.e., OCLC, RLG, WLN—are services where one pays to search (either transactionally or by subscription) and where the databases were at first a byproduct of very large-scale shared cataloging activities. These are the largest of the "union" catalogs, but they really represent multipurpose national or international resources rather than the union catalog of a specific organized community of libraries (though with appropriate search restrictions they can fill that function). These systems do not have real time links to institutional integrated library systems; they cannot, for example, indicate the circulation status of materials at holding institutions. These union catalogs include links to complex sophisticated interlibrary loan requesting and routing systems. Record consolidation approaches in these systems are strongly influenced by the design objectives of the shared cataloging activities that created them rather than the needs of patrons who want union catalog services. OCLC, for example, retains one base record, whereas RLG retains records for each contributing institution; neither approach is optimal from the union catalog standpoint.

*Union Catalogs.* Pure union catalogs, such as the University of California's MELVYL system, were developed specifically as public access union catalogs rather than as outgrowths of shared cataloging systems. These systems are only now starting to integrate with external integrated systems belonging to contributors via distributed computing technology in order to provide patrons with information such as real-time circulation status. These systems typically have at best limited links for forwarding requests to external interlibrary loan systems. In these systems, consolidation is designed specifically to address the needs of users to see multiple cataloging of the same work brought together.

*Shared Union Catalogs.* Shared union catalogs are part of an integrated library system shared by a group of libraries. Here there is very close integration between the catalog and other information about materials contained in the integrated system, such as circulation and serials receiving data. Typically these systems offer sophisticated direct borrowing or interlibrary loan among the libraries sharing the system. Because of the need to maintain individual site records for cataloging purposes,
the emphasis on consolidation is lower than in pure union catalogs. Examples include the Florida State Center for Library Automation and, to an extent, OhioLINK. Many large multibranch public libraries also use systems of this type. The vast majority of these systems still run on large mainframes, typically IBM or IBM compatibles.

Searching and Indexing Consistency

Because all records in a central union catalog are indexed in the same way, and all searches are processed through common software, searching and indexing consistency is almost axiomatic in a centrally managed implementation. Some indexing inconsistencies may appear because of varying cataloging practices used by the contributors; different systems have assigned greater or lesser emphasis on implementing software to smooth over these inconsistencies by performing source-specific record reformatting and/or indexing.

It is worth noting that searching and retrieval technologies based upon forty years of research in information retrieval are starting to appear, finally, in production systems. For example, a tremendous amount of work has been done on ranking retrievals in response to a query. The commonly used ranking schemes assign a rank to each record in the result set based on both the properties of the record and the statistical properties of the database from which it has been retrieved—the most common of these being variations on the so-called term frequency/inverse frequency distribution weighting for each term.

These technologies are all based on single database models which fit naturally with the union catalog environment. Combining ranked result sets from multiple distributed databases, or ranking the results that have been obtained from multiple distributed databases accurately without having a full characterization of the statistical properties of terms in each database from which a result has been retrieved remains an open and difficult research problem.

Consolidation

Just as with indexing, different union catalogs have placed differing emphasis on the importance of consolidation and the lengths to which they will go in performing record consolidation. These choices are strongly influenced by the context in which the union catalog was developed as discussed above.

One of the most striking and little-recognized characteristics of union catalogs that do attempt extensive consolidation is the amount of batch processing time that is typically spent in dealing with database quality and consistency issues. In a real sense this is off-line precomputation to support user needs to see a coherent picture of the union database. OCLC has had a large-scale program underway for some years addressing database quality through algorithmic record editing (sometimes with manual
review) and duplicate detection and elimination—which is their view of consolidation, given that they maintain a single “correct” base record for each work and do not record institutional cataloging variations for these records in their database. As another example, the MELVYL system incorporates a very expensive record consolidation process as part of its loading process; the result of this highly I/O bound process is that MELVYL can only load about 1,000 to 2,000 bibliographic records/hour/load stream on high-end IBM mainframe hardware. This consolidation process actually searches the database for candidate matches as each record is loaded, and then, for members of this candidate match set, performs a very complex field-by-field comparison and weighting to decide whether to consolidate the incoming record with one of the records in the candidate pool; in cases where consolidation occurs, individual contributor site cataloging variations are recorded and maintained on a field by field basis. In some cases, these variant fields are all used for indexing and display purposes; in other situations, a “best” version of a field is selected for display or indexing purposes.

Certainly, high-quality consolidation is possible in centralized union catalogs that have made it a priority, though this is often achieved at the cost of considerable background processing and software development effort.

Performance and Management

The management of large centralized database systems supporting high query volumes is now a relatively well understood process. Except for the network connections out to the end-user, the managers of such a system can typically control all of the variables and add capacity (disks, I/O channels, CPU cycles, etc.) as needed. There are sophisticated tools for measuring response time and system utilization and for performing capacity planning. Typically, there are extensive quality assurance and release management procedures in place for moving new operating system and applications software into production.

Union catalogs scale well. Search response time is typically primarily a function of the number of unique records. To the extent that the union catalog performs extensive consolidation, the number of unique records will grow slowly at the margin. Because indexes typically use B-tree type data structures, the number of additional I/O operations necessary to service a query will grow logarithmically in the number of unique records (or index terms derived from these records); there is also a linear component due to the increased size of the hit lists. When a new institution joins an existing union catalog, the additional load will be determined by the increase in query volume that the new institution generates and the additional per query cost, which is likely to be relatively low. It is only in situations where the union catalog is retrieving circulation status for each
holding institution on each record that the additional load for a new institution will be really significant. Put another way, adding a new institution typically exacts only a small cost in terms of increased resources per query.

**Distributed Search and Z39.50**

In the past few years, the concept of distributed search using Z39.50 has been proposed as a substitute for creating a “static” union catalog. Basically, the idea here is that some method is used to identify a set of online catalogs which logically represent a union catalog, or which are to be viewed as a union catalog for the purposes of a given user query. The remote systems which will contribute to this temporary virtual union catalog might be relatively fixed or highly dynamic and variable. The user then submits a query to a distributed search interface, which might be provided by the consortium offering access to the virtual union catalog or might be provided directly to the end-user by some third party. This distributed search interface translates the user’s query into an appropriate query for each of the constituent databases, submits it via Z39.50 to each of the remote systems comprising the virtual union catalog, and retrieves and consolidates the results, which are then presented to the user. The results coming back from the remote systems (which will typically be full-scale locally integrated library systems) may well include information such as circulation status.

**Searching and Indexing Consistency**

Theoretically, in functional terms, the distributed search model should be able to produce results that are equivalent to what can be obtained from a centralized union catalog. In practice, there are two problems. The first is that the query language that can be supported will be, effectively, the lowest common denominator of all of the query languages supported by the systems servicing the distributed search. If even one of these participant systems cannot support a given index or search option against that index (such as truncation), then the search option cannot be correctly supported by the distributed search system or will produce potentially inconsistent results. If the distributed search interface is sufficiently intelligent to recognize the limitations of constituent systems, it may be able to compensate for at least some of these shortcomings by doing additional query processing prior to presenting a result to the user at a significant performance penalty. For example, if the user asks for a personal author query and one of the constituent systems only supports an undifferentiated (personal and corporate combined) author query, then the search interface could in theory filter the records returned by that particular participating system. The more sophisticated the query language supported by the distributed search system interface, the less
likely that, in the general case, all participating systems will be able to support these queries correctly. It is worth noting that to the extent that the participating systems are relatively homogeneous—for example, supplied by the same vendor—the limitations of a lowest common denominator search language are minimized.

The second issue is the extent to which the various systems participating in the distributed search implement common semantics for Z39.50 search attributes and are consistent about how they process these attributes. Z39.50 implementations vary widely, and it is difficult to make any general statement other than to observe that Z39.50 is not a database indexing standard, and current Z39.50 attribute sets are not defined in terms of database indexing. Ultimately, inconsistencies in query processing have their roots in varying choices about indexing of databases on Z39.50 servers. For example, many systems accept and respond to queries that specify the AUTHOR or TITLE use attributes in the Z39.50 query; they do not necessarily use the same fields in their database records to build author or title indexes. There are also problems with extraction and normalization algorithms for search keys with stopword handling and with a host of other messy implementation details.

Consolidation

The second problem in distributed search systems is consolidation. Z39.50 clients are only now moving out of their initial implementation environment, which usually allowed a local interface to be used with a single remote database at a time and thus did not require the client to deal with consolidation issues. Typically, if a Z39.50 interface does anything in the area of consolidation today, it is duplicate elimination based on some sort of unique key like the ISBN or LCCN. Most often consolidation functions are still completely omitted.

It is worth recognizing that to duplicate the level of consolidation quality that is found in a central union catalog like the MELVYL system, for example, it is not sufficient to just process the records retrieved from each of the searches sent to the participating Z39.50 servers. Records retrieved from one participating server might potentially consolidate with records that were not retrieved as part of the search result from another system. To fully emulate the consolidation performed by a union catalog like the MELVYL system, it would be necessary to search for matching records in each participating Z39.50 server using the records retrieved from the other participating Z39.50 servers iteratively until convergence occurred. While there are undoubtedly heuristics that could be used to prune the possibilities and speed convergence, and to trade off the overlooking of remotely possible but unlikely consolidations against extra searches, this is still, as far as I know, an unexplored research area.

Even without seeking this level of consolidation, doing any type of consolidation merge of records from multiple servers will require either
that all result records from all participating servers be brought back to
the client for merging or that all participating servers be able to sort
their results in a consistent fashion. If the user is to be provided with a
query result count for his or her search against the virtual union catalog
database, then all records from all servers will have to be examined first—
clearly a sizable performance penalty. Yet system design experience sug-
gests that the ability to provide such a consolidated result size report is
very important for users.

Performance and Management

The performance of a distributed search system is critically depen-
dent on the performance of the network links between the client and
the participating servers; if these links are over the commodity internet,
then network performance may be a major problem at times, particu-
larly if consolidation of records is being done at the client, which im-
plies the transfer of potentially large amounts of data rather than just
the interchange of queries and search execution reports.

The performance of a distributed search system will be paced by
the performance of the slowest participating system in each distributed
search, since it will be necessary for all of the distributed queries to
complete before consolidation processing can begin and the aggregate
results can be reported back to the user.

The scaling properties of a distributed search system can be quite
unattractive when compared to a centralized union catalog. Each par-
ticipating system must be capable of handling the query load that all
users of the union system represent, since each search will be sent to
each participating system. A local system joining such a distributed
search constellation might have to be able to handle a magnitude of
more queries in support of distributed search than it needs to be able
to support its local patron base. Relatively small institutions joining
large virtual union catalogs implemented through distributed search
are at a particularly notable disadvantage. And, to the extent that any
one constituent system falters under the distributed search load, it will
degraded search response time for all searches run through the distrib-
uted search interface.

Reliability is also a problem. In any sufficiently large constellation
of systems supporting distributed search, there is a high probability that
at least one participating system will be out of service at any given time.
Some protocol to poll all constituent systems to determine which ones
have failed at any given time—a function outside the scope of the Z39.50
protocol—will be needed in order to avoid resorting to timing out Z39.50
connections and searches, which would have a serious impact on re-
sponse time in such a distributed search environment.
APPROPRIATE ROLES FOR THE CENTRALIZED CATALOG AND DISTRIBUTED SEARCH APPROACHES

In environments where a fixed-scope union catalog needs to be presented to a large patron community as a basic high-quality highly available resource, it seems clear that, with current technology, centralized union catalogs have major advantages both in function and in performance. Yet the power to permit a user to build ad hoc virtual union catalogs for specific searches, and to delegate to a Z39.50 client the tedium of at least first-pass consolidation and duplicate record elimination, is unquestionably attractive, and it seems likely that users who need such capabilities will be willing to pay some performance penalty for them. The ability to create such dynamically defined virtual union catalogs will be used relatively rarely and by fairly serious and sophisticated searchers; these searchers will also likely weigh the pros and cons of using international-scope centralized union catalogs such as OCLC and RLG to satisfy their searching requirements as an alternative and will be prepared to pay the costs of using these services where appropriate (or will be able to have their relatively infrequent searching of these resources subsidized by their host institutions). This kind of searching will complement, rather than supplant, high volume searching of predefined union catalogs that represent the holdings of consortia that offer explicit resource-sharing agreements for obtaining the materials cataloged in these union catalogs.

It should also be noted that while Z39.50 is limited in its ability to support the dynamic federation of databases for distributed searching today, it has been highly successful in the more limited role of extending familiar local user interfaces to remote databases outside a local system, particularly if this is done as a crafted implementation rather than just the ad hoc incorporation of random external databases.

CROSS-DATABASE LINKAGES AND ABSTRACTING AND INDEXING DATABASES

Article citation (abstracting and indexing or A&I) databases and other secondary information resources (such as reviews) are now commonplace services offered to library patrons alongside access to catalogs; they are available from a wide range of sources scattered across the network, from local mounts and from CD-ROM based systems. Increasingly, technologies like Z39.50 are enabling consistent user interfaces to wide ranges of A&I databases accessible through the network. Many of these A&I databases have coverage that overlaps with other competing or complementary A&I databases in complex ways. While most library patrons today search A&I databases sequentially, one at a time, there is a growing need for interfaces that will consolidate records retrieved from multiple A&I databases into a logical "union" A&I database. The characteristics of such
a consolidation process are highly dynamic and are likely to be based more on distributed search approaches than on traditional union catalog style consolidation—although specific popular clusters of A&I databases may be predefined into logical union A&I databases (using centralized union catalog approaches) for performance and functional quality reasons. It is interesting to note that some of the commercial search services, such as DIALOG, have offered such capabilities for consolidation of records from multiple A&I databases for some time, though the implementations of these services seem to be based more on static database architectures typical of union catalogs.

The primary content described by these abstracting and indexing databases is now springing up everywhere in electronic formats: locally mounted databases, publisher-provided servers, and intermediary (third party) aggregation and access services. Of course, not all available primary content is described by the available abstracting and indexing databases (much less the subset of these databases available to a given patron) due to limitations in chronological coverage and editorial policy scoping the A&I databases. This means that while abstracting and indexing databases will be an important path to identifying and locating primary content, they cannot serve as the only path. Similarly, not all of the primary content described by the A&I databases is even available currently in electronic formats, nor is comprehensive availability likely to occur for the foreseeable future due to the wide variation in publisher strategies for electronic dissemination of their materials. Retrospective coverage for many journals may also be slow in coming in electronic form, even after the publisher has made the decision to offer electronic access prospectively. Further, even if the materials are available electronically, many libraries may choose not to pay the price to make them available to patrons. Business relationships and models among patrons, libraries, and publishers in the electronic environment are far from clear; one can readily envision situations where a library will offer a patron a choice between paying directly for immediate access to an electronic copy of an article or having the library obtain it in printed form through interlibrary loan for free or at a lower price. For all of these reasons, abstracting and indexing databases must be linked to databases of print serials holdings as well as to electronic primary content.

While key standards (such as the revised 1996 version of Z39.56, the Serial Item and Contribution Identifier or SICI) to support linkages from abstracting and indexing databases to primary content are now coming into place, actual implementation of such linkages is relatively new, and considerable work needs to be done on appropriate matching algorithms. The revised (1996) SICI code incorporates a number of partially redundant data elements which may or may not be present (and explicitly tagged) in specific database records from which a SICI code is computed.
When using SICI codes to make interfile linkages, one cannot simply do an exact match; rather, one needs to perform a matching computation that is sensitive to these optional data elements in the SICI code. There are numerous other technical issues as well. For example, many of the publishers and third party primary content aggregators are mounting articles on Web sites rather than in Z39.50 databases, so the actual linkage mechanism is a Uniform Resource Locater (URL); one needs URLs that can include SICI codes and invoke CGI scripts or other services in the server that map the SICI code to the appropriate article file, or external published algorithms for computing appropriate URLs from SICIs that can be implemented in the client dynamically or used in programs that build linkages in the A&I files as a batch process.

Even in the presumably simpler case of linking abstracting and indexing databases to serials records in catalogs or union lists of serials, while many of the necessary linkage data elements (such as ISSNs) nominally exist in the relevant files, experience in practice has shown that the data are often inaccurate or incomplete; this problem will gradually fade as more use is made of such linking elements and errors are reported and corrected. Some vendors will improve the quality of the linking elements in their A&I files; others will become known for offering "linkage hostile" files and will consequently face a competitive disadvantage in the marketplace.

The implementation and maintenance of high quality linkages on a large scale presents major challenges; I believe that this will become perhaps the central problem for the next generation of information access systems. These are hard problems even in the relatively controlled environment of a union catalog, where many linkages can be precomputed and validity checked off-line rather than on demand, and where the results of the linkage calculations can be reflected in indexing. For example, users often find it useful to be able to restrict a search rapidly on an A&I database to only those citations that are available in electronic format, or that represent materials held in printed form at a specific library; this would involve the use of an index rather than simply trying to compute and display a linkage to primary content as each record from the A&I database is displayed. Reliably, accurately, and quickly computing linkages in the more anarchic framework of distributed search appears to be quite difficult; using the presence of linking elements as a search restriction is likely to lead to unacceptable performance since the entire result set has to be examined record by record prior to reporting on the results of a search.

Today the problem of creating linkages to primary content is focused on A&I databases, in large part because current networked information access technology can support access to articles in electronic formats reasonably well, while access to digital format books, manuscripts, maps,
sound recordings, films, and similar materials is still problematic. The files representing these kinds of materials are enormous; they are awkward and time-consuming to transfer and difficult to navigate once retrieved. Materials other than journal articles by and large are not practical today in electronic formats; the publishers recognize this reality and have made little of this material available, so there is not much demand to create links to it. Such material is starting to appear slowly, however, in part due to library-based programs to digitize special collections and to employ digitization for preservation purposes. Over time, the set of necessary linkages will expand to include not only A&I databases to primary content and serials holdings and serials holdings to primary content (or, more precisely, to navigational systems for cover-to-cover content of journals, including material not in scope for the A&I databases), but also from (monographic) catalog bibliographic records to primary content (or to finding aids that assist in the navigation of large collections of primary content) and to secondary materials such as book reviews.

**CONCLUSION**

This has been a primarily technical analysis of the comparative benefits and drawbacks of distributed search and traditional centralized union catalogs, and of how some of these issues extend to the integration of abstracting and indexing databases and electronic primary content within the bibliographic apparatus that is needed to support resource sharing. From a technical point of view, it seems clear that both centralized union catalogs and systems that can support intelligent distributed search offer important benefits to users, and that they can be used together in a complementary fashion to great advantage. Centralized catalogs are still the best way to support high volume searching against fixed collections that reflect explicit consortia or other resource-sharing arrangements, and which users will want to search regularly with high precision and performance. Indeed, centralized union catalogs can stand as visible symbols of such resource sharing agreements. Distributed search can be used to provide a way of delivering on the promise that the networked information environment offers for enabling users to define arbitrary virtual information collections that span organizational and geographical boundaries. Both approaches continue to be relevant as we consider the broader environment of catalogs, abstracting and indexing databases, and primary content proliferating in a distributed network environment.

But, as with much of the discussion of interlibrary loan and document delivery, it is essential to recognize that the issues here are not purely technical. They have significant organizational, economic, and political components; the economics are particularly treacherous because the environment mixes explicit costs (for example, the costs of searching an international union catalog like OCLC or RLG, or of actually creating
and hosting a centralized union catalog somewhere for a group of cooperating institutions) with implicit costs (such as provisioning a set of local systems to participate effectively in a distributed search constellation). There are issues of local autonomy and control; these are given the greatest latitude in distributed search architectures, while to some extent they are sacrificed or submerged in centralized union catalog systems. In some cases, distaste for centralized organizations or distrust of centralized control may be the determining factor. The convenience of the user community, particularly when this community is as broad and poorly defined as is typically the case in a resource sharing consortium, may be less important to decision makers than retention of local control and autonomy. The emergence of distributed search as an alternative (albeit a sometimes impoverished one) to centralized union catalogs means that it is now at least possible to permit nontechnical considerations increasingly to dominate design choices. It is hoped that this article will at least provide some insights into what may be sacrificed in such choices.