FINAL REPORT
March 17, 2006

The Changing Nature of the Catalog and its Integration with Other Discovery Tools

Prepared for the Library of Congress

By

Karen Calhoun
Cornell University Library
Ithaca, NY 14853
ksc10@cornell.edu

Table of Contents

The Changing Nature of the Catalog and its Integration with Other Discovery Tools

Executive Summary

Highlights

The destabilizing influences of the Web, widespread ownership of personal computers, and rising computer literacy have created an era of discontinuous change in research libraries—a time when the cumulated assets of the past do not guarantee future success. The library catalog is such an asset. Today, a large and growing number of students and scholars routinely bypass library catalogs in favor of other discovery tools, and the catalog represents a shrinking proportion of the universe of scholarly information. The catalog is in decline, its processes and structures are unsustainable, and change needs to be swift. At the same time, books and serials are not dead, and they are not yet digital. Notwithstanding widespread expansion of digitization projects, ubiquitous e-journals, and a market that seems poised to move to e-books, the role of catalog records in discovery and retrieval of the world's library collections seems likely to continue for at least a couple of decades and probably longer.

This report, commissioned by the Library of Congress (LC), offers an analysis of the current situation, options for revitalizing research library catalogs, a feasibility assessment, a vision for change, and a blueprint for action. Library decision makers are the primary audience for this report, whose aim is to elicit support, dialogue, collaboration, and movement toward solutions. Readers from the business community, particularly those that directly serve libraries, may find the report helpful for defining research and development efforts. The same is true for readers from membership organizations such as OCLC Online Computer Library Center, the Research Libraries Group, the Association for Research Libraries, the Council on Library and Information Resources, the Coalition for Networked Information, and the Digital Library Federation. Library managers and practitioners from all functional groups are likely to take an interest in the interview findings and in specific actions laid out in the blueprint.

Notes to the Reader

The report has two chapters and three appendixes. Chapter 1 describes the background, methodology, and objectives of the investigation. Chapter 2 describes what to do about the catalog. Readers on a fast track should look quickly at chapter 1, which is short, then peruse chapter 2. Appendixes A to C, which provide context and evidence for chapter 2, offer a detailed analysis of the current situation and key findings from the literature and a set of structured interviews. Readers who prefer a full analysis and elucidation of the topic will want to glance through chapter 1, examine the appendixes, then return to chapter 2.

The Changing Nature of the Catalog and its Integration with Other Discovery Tools

Chapter 1: Introduction

Background

The library catalog has long been the keystone supporting the mission of libraries—to save readers' time and advance the state of knowledge within the library's community. By directly connecting users and information, the library catalog (not Google or one of the other popular search engines) was the first service to support unmediated information discovery and access, and it has been doing so for over a hundred years.

Today, the catalog operates against a backdrop of flat or declining use of library collections, flashy and powerful alternatives for information discovery, rapid changes in information technology, rising expectations of library patrons, a rapid increase in new kinds of digital assets, mass digitization projects, and an incipient revolution in scholarly information exchange. If one accepts the premise that library collections have value, then library leaders must move swiftly to establish the catalog within the framework of online information discovery systems of all kinds. Because it is catalog data that has made collections accessible over time, to fail to define a strategic future for library catalogs places in jeopardy the legacy of the world's library collections themselves. For this reason, the option of rejecting library catalogs is not considered in this report.

As part of the celebration of its bicentennial year, from November 15–17, 2000 the Library of Congress (LC) hosted a conference of approximately 125 invited participants [1]. The theme of the conference was bibliographic control in the 21st century. One objective of the conference was to begin defining strategic issues related to the future of catalogs. Conference speakers and participants posed many questions, among them:

- What is the current state of standards and technologies to support unified access to multiple repositories, including catalogs?
- What are the future roles of MARC and cataloging content rules?
- What are the challenges to the economic sustainability of the current model of the catalog?
- What do 21st century information seekers need from catalogs?
- In what ways might libraries leverage catalog data for new uses?
- What partnerships are worthy of pursuing with the publishing, systems, scholarly, and information technology communities?

Since the bicentennial conference, library leaders at LC and elsewhere have pushed beyond the questions posed at that conference to challenge the value of the catalog itself. At a seminar held at the American Library Association 2005 Midwinter Conference, Deanna Marcum, LC's Associate Librarian for Library Services, asked, in light of the

high cost of cataloging and swift changes in information seeking behavior and tools, "just how much do we need to continue to spend on carefully constructed catalogs?" [2]

Objectives

In spring 2005, the director for Acquisitions and Bibliographic Access at the Library of Congress authorized a research project to carry out LC's Bicentennial Conference on Bibliographic Control Action Item 6.4, "support research and development on the changing nature of the catalog to include consideration of a framework for its integration with other discovery tools." [3]   Action item 6.4 was one of many action items that LC leaders defined following the bicentennial conference. For action item 6.4, LC engaged the author as principal investigator.

Acting on the advice of LC's research project sponsors, the author defined the research objectives broadly, from the perspective of major research libraries in general, rather than focusing on the issues as they relate to LC specifically. This report contains recommendations for the future of research library catalogs, a preliminary assessment of the technical and organizational feasibility of next steps, and a vision and blueprint for change. The content is intended to elicit support, dialogue, collaboration, and movement toward solutions and a phased approach to change at LC and in the library community at large.

Methodology: Research Process and Starting Points

The investigation began with gathering and reviewing a range of literature, mainly from the last five years (2000-2005), on the nature of the scholarly information universe; information seeking behavior and the role of search engines, particularly Google; markets, market positions, and competitive strategy; the economics of information services; federated access methods and tools for digital libraries; and of course the future of catalogs and cataloging.

Structured Interviews

The literature review served as the basis for selecting six questions to use in structured interviews. Each question was intended to elicit the kind of information that an investor might want to know about any product or service (in this case, the catalog) whose market position is eroding. Ample evidence documents the declining market position of the library catalog (see Appendix A). The next step, choosing interviewees, was done in collaboration with John Byrum and Judy Mansfield of LC plus members of the ALCTS (Association of Library Collections and Technical Services) Task Force on the LC Action Plan, which served in an advisory capacity for all aspects of this study. The desire to gather a range of perspectives guided the choice of interviewees, who are listed in Appendix B. Martin Kurth, a colleague at Cornell and the head of the library's metadata services group, assisted the author with the interviews. Appendix C lays out the key findings of the structured interviews and literature review.

Chapter 2: The Catalog's Future

Where We Are

This investigation has taken place in a time when it is more important than ever to position the research library catalog successfully within a rapidly evolving information universe for scholarly research, teaching, and learning, and to adapt to sea changes in information seeking behavior. As discussed in Appendix A, the legacy of the world's library collections is for the time being tied to the future of catalogs. At the same time, a large and growing number of students and scholars routinely bypass library catalogs in favor of other discovery tools, and the catalog represents a shrinking proportion of the scholarly information universe.

Findings from the structured interviews and literature analysis (Appendix C) suggest that today's library catalogs are long on problems and short on unique benefits for users. The cost-effectiveness of cataloging tradition and practice is under fire. The typical research library catalog's strongest suit is its support for inventory control and as "last mile" technology to enable delivery of the library's assets into the hands of local users. A new technology for expanding the service model of the catalog to cover more of the scholarly information universe—metasearch—has generated much hope but is not meeting early expectations for tying together the fragmented landscape of scholarly information resources.

There are a number of prevailing strategies for integrating the catalog with other discovery tools and one huge opportunity. First the strategies. All feature some degree of data consolidation (mainly for discovery) paired with distribution of the functions of discovering, requesting, and getting information among multiple services. Initiatives like Google Book Search, Open WorldCat, and RedLightGreen hold promise, but so far finding and obtaining items from library collections on the open Web is not a practical alternative for students and scholars. Nevertheless there is an expectation that such initiatives will eventually make research library collections more visible to a worldwide audience. Some influential library and information science professionals are beginning to suggest relying more on state-wide, national or global aggregations of catalog data for discovery, and using library ILSes as a middle "switching" layer to enable delivery. The huge opportunity of integrating catalogs with open Web discovery tools is the long tail—surfacing research libraries' rich collections in ways that will substantially enhance scholarly productivity worldwide.

Taking advantage of research libraries' opportunities for leveraging their investments in their catalogs and collections requires overcoming some daunting obstacles (Appendix C). Many research library leaders, most staff members, and some university faculty are not ready for change of this magnitude. Progress toward reliable and easy interoperability is painfully slow. Precedents for large scale collaboration among research libraries are few. Copyright law has not caught up with the digital world. It's unclear whether research libraries and library service firms are sufficiently capitalized to build the necessary technical infrastructure.

What to Do About It

The online library catalog has been a successful product. Like other products, it has passed through a life cycle. In the late 1970s, online catalogs were a rarity and adoption crept along. By the late 1980s, online catalogs had taken off, and any self-respecting research library had to have one. By the mid 1990s, online catalogs were mature. A new cycle of replacement had begun and is now nearly complete. Product replacement took the form of migrations from character-based catalogs to systems based on client-server technologies and relational databases. Today, the online catalog is losing appeal for students and many scholars. Catalog usage, drifting downward compared to other discovery tools, may soon plummet. Fortunately, there are ways to use the knowledge that today's catalog has reached the end of its life cycle. Theodore Levitt [4] and his successors in the business world offer a number of strategies for revitalizing products:

1. Promote more frequent use among existing users
2. Develop new uses among existing users
3. Find new users for the existing product
4. Find new uses and new users

Figure 1 applies these concepts to extending the life cycle of the research library catalog. The examples in the figure are merely aids to understanding how to use the model; readers may agree or disagree with their choice or placement. The quadrant on the lower left combines existing uses and existing users and is the least promising as a long term strategy, because the existing local catalog's market position has eroded to the point where there is real concern for its ability to weather the competition for information seekers' attention. One can speculate that the quadrant on the lower right, which combines existing users with new uses, will revitalize the catalog to a degree. Strategies like enhancing discovery and delivery for e-resources within the catalog, updating the local catalog's public interface, adding new functionality for browsing, producing new accession lists or subject-specific pathfinders, or adding tools to export citations to bibliographic management software (e.g., RefWorks) belong in the lower right quadrant.

Above the middle horizontal line, the quadrant on the upper left is characterized by more ambitious strategies such as intensive marketing or library instruction campaigns to capture the attention of non-users, especially new freshmen; pushing library data out to university portals and course Web pages; and (to capture new users outside the library community) the regional combination of collections and/or catalogs or the introduction of state-wide or regional unmediated interlibrary lending programs (like Borrow Direct). The quadrant on the upper right is the place where transformative, higher risk, long-term, and typically costly strategies reside. Such strategies might include creating more coherent and comprehensive scholarly information systems, perhaps by discipline; building the necessary infrastructure to permit global discovery and delivery of information among open, loosely-coupled systems (e.g., find it on Google, get it from your library); enabling universal, unmediated access to research library collections; or leveraging catalog data to support mass digitization projects and/or to produce long-tail effects for research library collections. Naturally, blended strategies are also possible.
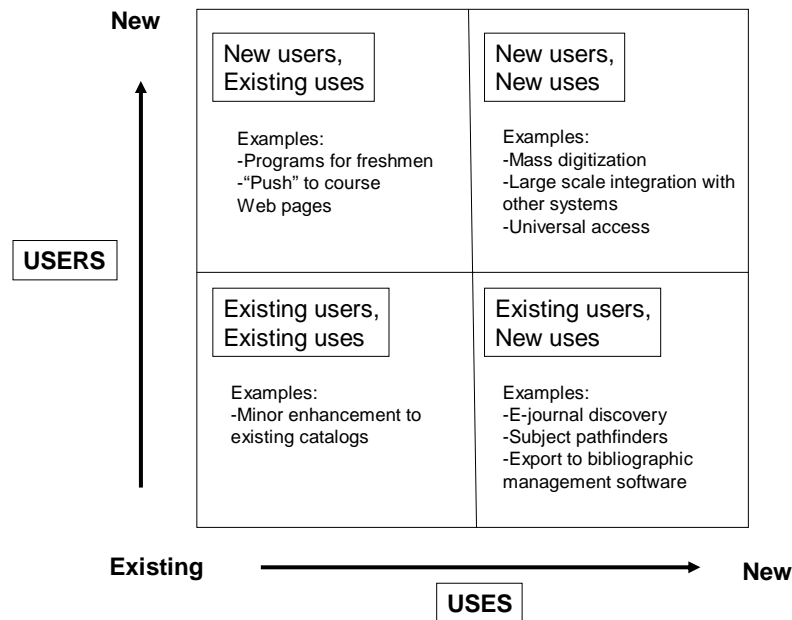
Figure 1. Revitalizing the Research Library Catalog

Michael Porter is a leading authority on competitive strategy. Kathryn Rudie Harrigan writes often on the strategic management of maturing businesses. In their joint article Harrigan and Porter offer options for responding to the problems that declining demand poses [5]. The key problem of declining industries is the effect wrought by falling demand—that is, excess supply and capacity. For example when new technologies began to replace vacuum tubes in televisions, vacuum tube manufacturers had to shut down excess plant capacity as people replaced their old televisions with new ones. In healthy businesses, the capacity to produce a product matches the demand for the product.

Research libraries have invested and continue to invest millions to develop and maintain the capacity to produce local catalogs—in 2004, ARL libraries spent an estimated $239 million on technical services labor alone. As information seekers increasingly turn to search engines, research library leaders need to examine ways to bring the capacity to produce local online catalogs back into line with the demand for them. Carrying out this responsibility is a subtle, rigorous, and complex task, fraught with peril, but as Harrigan and Porter point out, many companies have successfully coped with declining products. As is suggested by the strategies offered by Levitt and laid out in Figure 1, innovations, cost reductions, and changing circumstances can slow or reverse a decline. If however declining demand is pervasive, Harrigan and Porter offer four strategic alternatives: leadership, niche, harvest, or quick divestment.

A library or library service organization pursuing a leadership strategy would seek a prominent position in the market and reduce exit barriers for other organizations—in this

case, one might expect such an organization to offer an alternative that will help others retire their own catalogs. The niche strategy would be characterized by specialization; the research library or service organization would choose one or more segments of users—say, humanists or area studies specialists—in which demand is expected to be reasonably stable and move to serve those segments exclusively. The harvest strategy would involve controlled, gradual disinvestment in existing local catalogs. Harrigan and Porter note the difficulties of the harvest strategy for businesses, due to the risks of maintaining customers' confidence and employees' motivation as investment and service levels are curtailed. In a research library, pursuing the quick divestment strategy would most likely involve getting out of the business of producing a local online public access catalog (OPAC) entirely by outsourcing the work or relying on an alternative service.

Different research libraries and the organizations that serve them will choose different strategies for revitalizing their catalogs. An organization's strategic choice will depend on the organization's position with respect to others who supply or produce catalogs, its financial position, its perception of the likelihood and rate of revitalization or decline of the catalog, the actual strength and nature of remaining demand for the existing catalog, the availability of practical alternatives, and the level of difficulty the organization will have diverting its capacity to new uses. In addition, an individual library or library service organization might blend strategies or pursue different strategies over a period of time. Harrigan and Porter's research suggests that companies that weather declining demand tend to be participants in the substitute industry (in this case, open Web discovery tools).

Thirty-Two Options and Three Strategies

Libraries are unlikely to divest themselves of their catalogs. Most library leaders and scholars would deem such a course of action unthinkable. It is more practical to think about research libraries' divesting themselves of the status quo—that is, the situation in which the research library community would continue offering their existing local online public access catalogs for existing users and uses. In effect, a choice to continue with the status quo is a harvest strategy, complete with its downsides of eroding user satisfaction and deeply frustrated library staff members. At a minimum, research libraries need first to explore extending the life of the catalog through innovation and cost reduction and second, to develop new uses for catalog data for existing catalog users. At the far end of the range of strategic choices open to research libraries, they and the organizations that serve them could develop systems for discovering and delivering library collections and other scholarly information that would advance the progress of knowledge in ways that no one could have imagined a decade ago.

Figure 2 offers thirty-two possible remedies organized by the type of underlying strategy—extend, expand, and lead. The time frame is assumed to be five years. *Extending* the research library catalog involves innovations and cost reductions; these strategies are the building blocks for the next two strategic choices. An example of innovation in the context of the "extend" strategy is the new Endeca-powered catalog at North Carolina State University [6]. *Expanding* involves attracting new users for catalog data and the research library collections they describe. An example of the collaborative
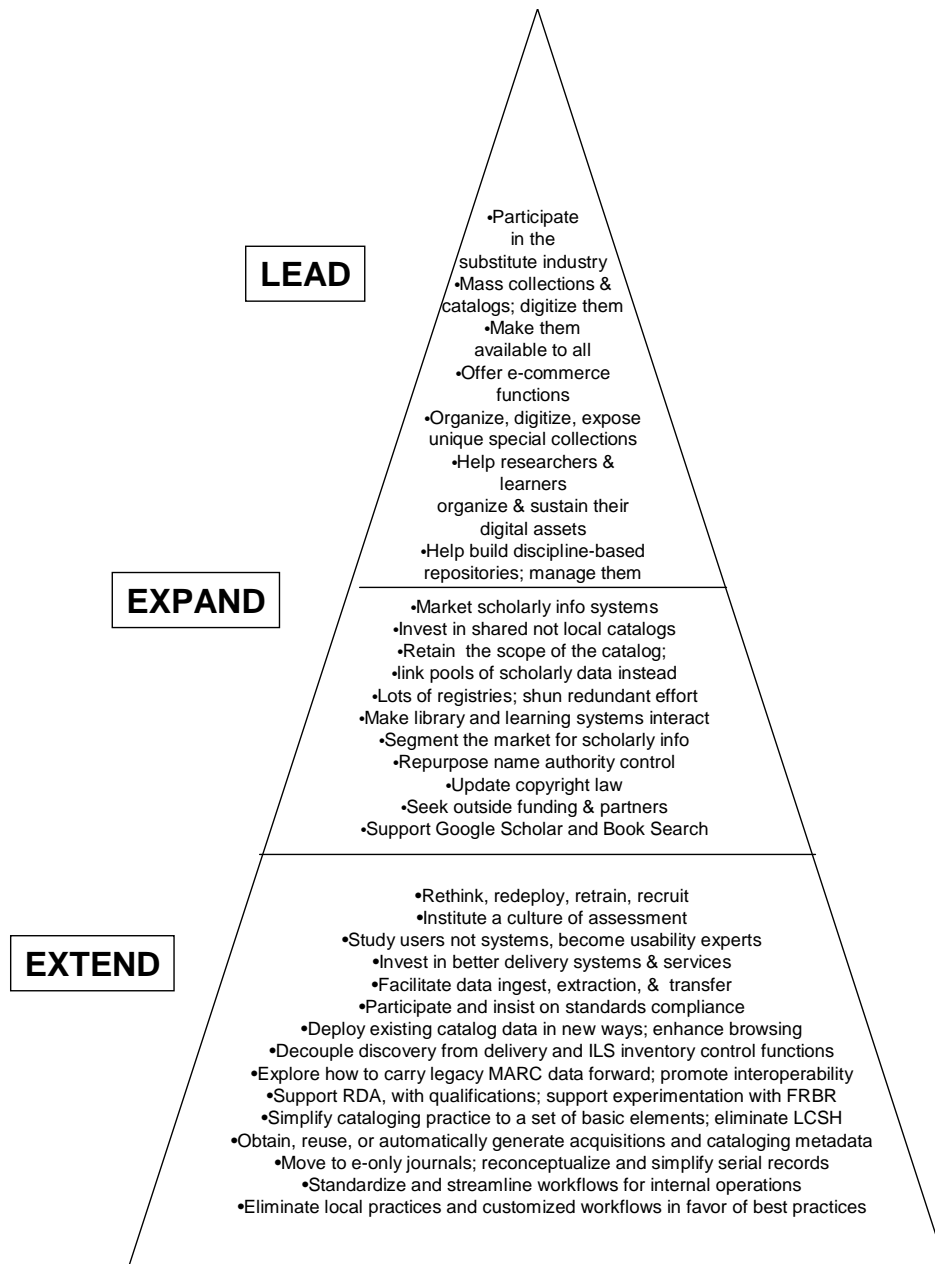
aspect of the "expand" strategy is the new "CalCat," an IMLS-supported initiative to build a "super-sized catalog" for Californians [7]. *Leading* involves significantly expanding the research library's role in developing information systems that support teaching, learning and research on a global scale. To date, there is no fully realized, practical example of the leadership strategy among research libraries, although some library leaders appear to be at the perimeter of this unexplored country. Readers may reach different conclusions about the choice and placement of these possible actions. The figure is offered as a starting point for sparking dialogue, collaboration, and movement toward revitalization of research library catalogs.

Implementation Issues

This section provides a preliminary assessment of the organizational and technical feasibility of the options for next steps. For the next couple of years, most research libraries will be best positioned to pursue innovation and cost reduction as their principal strategies (i.e., the "extend" strategy). There are already examples of libraries on this path and a number of active writers and consultants providing direction, documentation, and leadership. The challenges to feasibility include:

- Difficulty achieving consensus on standardized, simplified, more automated workflows
- Unwillingness or inability to dispense with highly customized acquisitions and cataloging operations
- Overcoming the "not invented here" mindset preventing ready acceptance of cataloging copy from other libraries or external sources
- Resistance to simplifying cataloging
- Inability to find and successfully collaborate with necessary partners (e.g., ILS vendors)
- Difficulty achieving basic levels of system interoperability
- Slow development and implementation of necessary standards
- Library-centric decision making; inability to base priorities on how users behave and what they want
- Limited availability of data to support management decisions
- Inadequate skill set among library staff; unwillingness or inability to retrain
- Resistance to change from faculty members, deans or administrators

In summary, the implementation issues associated with the innovation and cost reduction strategy include some technical but mostly organizational hurdles. To succeed at this strategy, research libraries will need to master organizational change management and achieve unprecedented levels of collaboration with peers and external partners. The challenges of integrating research library catalogs with other discovery tools will not be solved by individual libraries working alone.

LEAD

•Participate in the substitute industry
•Mass collections & catalogs; digitize them
•Make them available to all
•Offer e-commerce functions
•Organize, digitize, expose unique special collections
•Help researchers & learners organize & sustain their digital assets
•Help build discipline-based repositories; manage them

EXPAND

•Market scholarly info systems
•Invest in shared not local catalogs
•Retain the scope of the catalog;
•link pools of scholarly data instead
•Lots of registries; shun redundant effort
•Make library and learning systems interact
•Segment the market for scholarly info
•Repurpose name authority control
•Update copyright law
•Seek outside funding & partners
•Support Google Scholar and Book Search

EXTEND

•Rethink, redeploy, retrain, recruit
•Institute a culture of assessment
•Study users not systems, become usability experts
•Invest in better delivery systems & services
•Facilitate data ingest, extraction, & transfer
•Participate and insist on standards compliance
•Deploy existing catalog data in new ways; enhance browsing
•Decouple discovery from delivery and ILS inventory control functions
•Explore how to carry legacy MARC data forward; promote interoperability
•Support RDA, with qualifications; support experimentation with FRBR
•Simplify cataloging practice to a set of basic elements; eliminate LCSH
•Obtain, reuse, or automatically generate acquisitions and cataloging metadata
•Move to e-only journals; reconceptualize and simplify serial records
•Standardize and streamline workflows for internal operations
•Eliminate local practices and customized workflows in favor of best practices

Figure 2. Options Organized by Type of Strategy

The "expand" strategy has organizational issues associated with it, but also many more technical, legal and financial challenges. The decision regarding local vs. shared catalogs will be hotly debated. The financial and organizational challenges of sharing catalogs could be complex. The fates of metasearch and Web services as viable technologies for research library information systems will be determined in the context of this strategy. It is difficult to tell if it will be technically possible for library and learning management systems to easily interoperate. There may be legal barriers associated with this strategy; if collections and catalogs begin to be shared, how should a library's user community be defined for the purpose of licensing e-resources and making reproductions of copyrighted works?

For understanding the implementation issues associated with the leadership strategy, it is important to be clear about what is meant by the "excess capacity" of catalogs. Most catalogers would deny there is excess capacity in today's cataloging departments, and they are correct. Library materials continue to flood into acquisitions and cataloging departments and the staff can barely keep up. Yet the key problem of today's online catalog is the effect of declining demand. In healthy businesses, the demand for a product and the capacity to produce it are in balance. Research libraries invest huge sums in the infrastructure that produces their local catalogs, but search engines are students and scholars' favorite place to begin a search. More users bypass catalogs for search engines, but research libraries' investment in catalogs—and in the collections they describe—does not reflect the shift in user demand.

The capacity required to produce a local library catalog today is directly related to the amount spent on printed books and serials—in ARL libraries, the lion's share of $665 million a year. The structure of most ARL budgets privileges the purchase of printed books and serials for locally-housed, locally-circulated collections. Research libraries continue to be ranked primarily by how many things they have in locally-housed, locally-circulated collections [8]. Due to the continuing large annual investment in published materials that need cataloging, the remedies of innovative methods and cataloging cost reduction can only go so far. The declining demand for today's catalogs reflects diminishing interest in already low-use research library collections, at least as they are currently housed, managed, and delivered.

The leadership strategy cannot succeed without investment in new, global information systems that make research library collections more visible and that cover more of the scholarly information universe. Investing in cataloging of unique special collections (which could be quite a costly enterprise) may eventually have equal importance. All this will require large sums of money. A lot of funds are currently locked up in building many parallel, redundant research library collections. Therefore, a first step in implementing the leadership strategy must be helping libraries pool their collection efforts, freeing up materials budget funds for reinvestment [9]. A competing priority—preserving the cultural record for future generations—will make this a difficult, but not impossible task.

Reallocating substantial portions of library funding to building bigger scholarly information environments will probably not be enough; outside funding and many new professional and business partnerships will be required. Pursuing the leadership strategy carries considerable financial, organizational, technical, and legal demands, uncertainties, and risks, but the potential to serve the public good is high. Successful implementation could radically democratize access to scholarly information and boost scholarly productivity to new levels.

A Blueprint for Phased Implementation—Two Year Plan

This concluding section offers a vision for change and a blueprint for revitalizing the research library catalog. The vision looks five years or more out, while the blueprint suggests a set of practical steps to be carried out over the next two years. Libraries can use this blueprint for implementing either individual or shared catalogs and for positioning themselves to pursue any of the three strategic options—extension, expansion, or leadership.

Vision for Change

- Information seekers will have a range of options for discovering, requesting, and obtaining materials from research library collections
- The model for producing and maintaining acquisitions and catalog data will be financially sustainable
- The scope of the catalog will continue to be primarily books, serials, and their electronic counterparts
- Legacy catalog data will be important for supporting mass digitization projects
- The catalog will evolve toward full integration with other discovery tools and with the larger scholarly information universe
- Public and private partners will collaborate with libraries to revitalize the research library catalog
- Support for shared catalogs will grow
- Research libraries and their partners will deploy shared catalogs as a key component of providing affordable global access to larger, richer collections than any single institution could house locally

Blueprint

The author offers the following ten-step planning process with the hope that implementers will further refine each step and develop concrete action plans. While all steps are important, within the context of the catalog's future, the choice of a strategic option (2), preparing for linkages in and out of the catalog (3), innovation/cost reduction (4), improving the user experience (5), managing change (8), and building partnerships (10) may offer the greatest rewards to implement.

1. Define the Community to Be Served
    1.1. Describe precisely the population potentially served by the catalog, including size, geographic location, fields of study, demographics, catalog users/non-users

1.2. Understand the work practices and information seeking behaviors of each segment of this population [10]. (Take advantage of information and statistics already available, undertake new studies, or both)

1.3. Choose a focus: select which segments of the population the catalog will serve

1.4. Choose a strategy: retain existing users, or attract new users while retaining existing ones

1.5. Prepare user requirements

1.6. Evaluate the existing catalog's ability to satisfy the needs of the selected user population

1.7. Describe the desired changes to the catalog [11]

2. Choose a Strategic Option

2.1 Study available products

2.2 Choose an approach: local or shared catalog?

2.3 Choose a strategy: continue with discovery and management in one integrated system or decouple discovery from delivery and inventory control functions? What level of integration into larger scholarly research and learning environment? What level of integration in Web search engines?

3. Prepare for Linkages In and Out of the Catalog and/or ILS

3.1 Define requirements for catalog data ingest, conversion, extraction and transfer

3.2 Identify systems with which catalog data and the ILS must interoperate; define interoperability requirements

3.3 Work with library vendors and other organizations to implement requirements (this task is best done through cooperative programs and standards organizations)

4. Innovate and Reduce Costs

4.1 Obtain Metadata

4.1.1 Simplify catalog records to a set of basic elements to support discovery, browsing, identification, delivery, resource sharing, linking, and inventory control [12]

4.1.2 As much as possible, obtain or reuse data available at the point of selection, or automatically generate this data

4.1.3 Reserve manual data creation for ordering, receiving, claiming, and cataloging for those situations in which it is the only viable approach

4.1.4 Manage acquisitions and catalog data through batch processes; as much as possible, avoid working on one record at a time

4.1.5 Identify local customization (e.g., for call numbers) and record editing practices and eliminate them in favor of accepting as much cataloging copy as possible without review or modification

4.1.6 Monitor RDA, keeping in mind the need for simplicity and the larger role that automated metadata creation, batch processing, linking, and Web services will play

4.2 Support Browsing and Collocation

4.2.1 Use classification data to cluster catalog data for browsing by subject [13]

4.2.2 Explore automatic classification

4.2.3 Abandon the attempt to do comprehensive subject analysis manually with LCSH in favor of subject keywords; urge LC to dismantle LCSH [14]

4.2.4 Encourage research and development in automatic subject analysis, including ways to reuse legacy data containing LCSH headings to support automatic subject analysis

4.2.5 Explore new ways to manage vocabulary for the names of places

4.2.6 Support experimentation with FRBR and urge vendors and library service organizations to implement clustering based on FRBR concepts

4.2.7 Encourage the review of developments in other disciplines on ontologies and taxonomies and their application to library catalogs

4.2.8 Encourage research and development on deploying catalog data in new ways to support discovery and browsing

4.2.9 Continue and expand participation in name authority control cooperative programs

4.2.10 Encourage research and development to align library name authority control methods and data with the requirements of modern machine processing

4.2.11 Encourage a collaborative cost-benefit analysis of series authority control; determine who needs controlled vocabulary for series headings and how/where to provide it at substantially less cost

4.2.12 Encourage research and development for new visualization tools and techniques

4.3 Streamline Workflows

4.3.1   Encourage collaborative development of community-wide best practices or standards for acquisitions and cataloging workflows, then implement them [15]

4.3.2   Ask vendors and library service organizations to support workflow best practices and standards; do not ask them to customize their systems to support non-standard workflows

4.3.3   When designing workflows, consider the entire flow of activity, from selection to access or shelving; streamline the workflows to eliminate unnecessary or redundant effort

4.3.4   Shun workflows that create or perpetuate processing backlogs

4.3.5   Define fast turnaround and delivery of library materials to users as the standard of quality service, not the fullness of catalog data

4.3.6   To the extent possible, support workflows with automated techniques and tools

4.3.7   On campuses with multiple technical processing centers, integrate operations to achieve consistent practices, clear direction, and savings

4.3.8   Mainstream, adequately support, and give high priority to e-resource licensing, discovery, linking, access and management

4.3.9   Support the re-use of catalog data and cooperative development of new workflows and/or data elements to support mass digitization projects

4.4 Adopt Standards and Best Practices
    4.4.1   Participate, and ask vendors and library service organizations to participate, in standards development/compliance in support of linking, metasearch, metadata harvesting, registry development, and Web services
    4.4.2   Comply with these standards in in-house development projects; ensure that every new repository or digital collection that is built can be successfully integrated into the larger scholarly information universe, both locally and globally

5. Improve the User Experience

5.1 Enrich the catalog with services (e.g., "more like this," "get it" options, new book lists, etc.), and data (cover art, reviews, TOCs)
5.2 Enable much better browsing and organization of large retrieval sets
5.3 Enable best-match retrieval (no search dead ends)
5.4 Provide relevancy ranking of search results
5.5 Continue working with available technologies (but look for better ones) to federate discovery and delivery of books, journals, and journal articles
5.6 Link the user to full text whenever possible
5.7 For items that cannot be delivered instantly, offer a range of unmediated, quick delivery options
5.8 Provide and market extended-hour hotlines, rush, and troubleshooting services to help users quickly solve delivery or connection problems
5.9 Push library metadata and links out to course Web pages and portals
5.10 Take advantage of e-commerce functions to serve non-members of the library community

6. Make Good Decisions

6.1 Institute a "culture of assessment" [16]
6.2 Train library staff to conduct and apply work practice and usability studies to service development or enhancement
6.3 Capture and maintain usage data; support queries and report generation
6.4 Track and maintain other data to support evidence-based decisions about user services and internal operations

7. Market the Library

7.1 Within the context of the library's entire marketing plan, institute annual library publicity campaigns that promote the catalog and collections, both in and outside the library building, and on Web pages that students and faculty view frequently
7.2 Promote the catalog and collections differently to different segments of users
7.3 Set and track annual objectives to increase or maintain awareness of the catalog and collections
7.4 Emphasize the unique benefits of the catalog and collections and the services built on top of them

8. Manage Change

8.1 Complete stakeholder analyses; prepare and carry out two-way communication plans customized for each stakeholder group

8.2 Train managers and staff to understand and cope with the dynamics of personal and organizational transition [17]

8.3 Recruit and train change agents

8.4 Reward flexibility and openness

8.5 Create transitional roles

8.6 Involve stakeholders, managers and staff in transition planning and execution

9. Develop, Retrain, and Recruit

9.1 Offer staff development services

9.2 Identify future skill requirements

9.3 Complete a skills gap analysis

9.4 Prepare training programs and necessary documentation

9.5 Retrain or recruit to address skill gaps

9.6 Modify job descriptions and expectations as needed

9.7 Manage performance

10. Find Funding and Partners

The financial and organizational challenges of revitalizing and integrating research library catalogs with other discovery tools will not be solved by individual libraries working alone. A new era of unprecedented levels of collaboration and partnerships is about to begin. Librarians will need a great deal more business acumen than in the past. Some specific steps include:

10.1 Expand the number of staff members who can write effective grant proposals, including preparing realistic cost estimates and budgets

10.2 Expand the number of staff members who are familiar with funding agencies and foundations

10.3 Teach managers how to evaluate market opportunities, make capital budgeting decisions, and prepare business plans and related documents

10.4 Teach managers how to seek out, establish, and manage partnerships and joint ventures

10.5 Enhance managers' negotiation and conflict management skills

10.6 Introduce a new product/service innovation program and process

10.7 Encourage and reward an entrepreneurial spirit

Acknowledgements

Appendix A: Current Situation

Market Analysis

The number of Association of Research Libraries members (123) may be taken as an approximation of the number of online research library catalogs in use today in North America. Of the 123 ARL members, 113 are university libraries and 10 are public, governmental, and nonprofit research libraries [18]. In 2004 total expenditures of these libraries came to more than $3.4 billion. All use the shared cataloging systems of OCLC and/or the Research Libraries' Group to obtain cataloging copy. LC plays a central role as the most important supplier of catalog data, setting the bar for cataloging standards and contributing over six million records to the OCLC database over the life of the shared cataloging service [19].

The prevailing model for catalog creation and maintenance is for libraries to contribute original cataloging records to the shared systems (RLG's or OCLC's) and to download copies of records for all their holdings. Each library maintains its own local catalog. Thus data that is stored centrally is replicated, with certain extensions, locally. This model for cataloging began in 1971, when the first shared cataloging program using a central computer system came into being at OCLC (then the Ohio College Library Center). Since the founding of OCLC (and before that), individual libraries have been concerned about loss of control and autonomy in their cataloging practices, and while the benefit of standardized practices is widely recognized, research libraries continue to maintain local practices.

Online catalogs serve a segment of the global online information industry. For most of the last thirty-five years, university students, faculty, researchers and staff have been well served by research library catalogs and their complements, abstracting and indexing services. Until perhaps the mid-1980s, the catalogs and information services provided by academic research libraries and A&I firms had few substitutes. At that time, the emergence of a variety of online information services began to change the landscape for scholarly information discovery and retrieval. Since then, dramatic impacts on the demand, supply, and exchange of online information have been wrought by the emergence of the Internet and World Wide Web, widespread ownership of personal computers, and increasing computer literacy. These developments have lowered the opportunity costs of obtaining information, increased self-sufficiency, and kicked off an accelerating decline in information seeking methods based on printed sources. During this process, which is in keeping with the operation of a free market, numerous organizations have been shaken out of the online information industry, primarily due to intense competition or failure to keep pace with innovations.

Viewed from a microeconomic perspective, the raw material of research library online catalogs is scholarly output and other publications deemed to have value for teaching, study, or research. Scholars, authors, governments, and the wide array of individuals and organizations that participate in the creation and dissemination of

knowledge serve as manufacturers, and until the last fifteen years or so, libraries, archives, scholarly publishers and societies, other content providers, and A&I services provided the primary distribution services. New entrants—both suppliers and distributors—now crowd this space.

The Scholarly Information Universe

Today's research library catalogs—even those that include records for thousands of scholarly e-journals and databases—reflect only a small portion of the expanding universe of scholarly information. Library catalogs manage description and access for mostly published resources—tangible materials such as books, serials, and audiovisual media, plus licensed materials such as abstracting and indexing services, full text databases, and electronic journals and books [20]. The abstracting and indexing community has long covered the domain of description and access for individual journal articles [21].

In contrast, the stuff of cultural heritage collections, digital assets, pre-print services and the open Web, research labs, and learning management systems remain for the most part outside the scope of the catalog. Scholarly information objects now include digitized rare and historical materials, textual primary source materials, graphical images, materials described in institutional and disciplinary repositories, conference Web sites, scholarly Web sites such as those identified by services such as INFOMINE [22], data sets, software, simulations, a rising array of multimedia resources, learning objects and courses—the list goes on.

In his article on the multi-type, multi-format information objects that make up the academic internet, Norbert Lossau promotes "a concerted initiative of the library community to pick up state-of-the-art search technology and build reliable, high quality search services for the research and teaching community" [23]. Herbert Van de Sompel provides another perspective, promoting a highly collaborative, data-intensive and networked scholarly communication system that supports many types of "units of scholarly communication" including not only documents but also data sets, simulations, software and multi-type or compound information objects [24]. In a joint white paper created on behalf of the IMS Global Learning Consortium and the Coalition for Networked Information, Neil McLean and Clifford Lynch discuss the challenges of greater interaction between library information services and instructional management system (e-learning) environments [25]. Lorcan Dempsey and his colleagues offer a schematic for thinking about the wide-ranging categories of scholarly information objects. They argue that libraries will become more engaged with research and learning materials produced by faculty and note that such materials involve higher levels of digital content management expertise. Metadata management and knowledge organization approaches will need to mature in order to handle multiple metadata and repository environments [26].

Information Seeking Behavior

One frequently hears librarians claim there is a dearth of understanding of library users and their behavior. On the contrary, a great deal (perhaps too much!) is known

about information seeking behavior from the fields of communications, learning theory, sociology, psychology, consumer research, human-computer interaction, and elsewhere. As Donald Case points out in his book on information seeking, "there is no shortage of theory from various disciplines that might be applied to the search for, and use of, information" [27].

Social science researchers have employed many paradigms in information seeking research [28]. The Principle of Least Effort, attributed to philologist George Zipf, is probably the best known in libraries. Marcia Bates' report to LC on action item 2.3 (improving user access to catalogs and portals) contains a helpful review of information seeking literature. Speaking about the Principle of Least Effort, Bates notes "people do not just use information that is easy to find; they even use information that they know to be of poor quality and less reliable—so long as it requires little effort to find—rather than using information they know to be of high quality and reliable, though harder to find." She concludes "despite heroic efforts on the part of librarians, students seldom have sufficiently sustained exposure to and practice with library skills to reach the point where they feel real ease with and mastery of library information systems" [29].

Case points out that much information seeking research focuses on information sources (e.g., books or newspapers) and systems (e.g., catalogs) rather than on the needs, motivations, and behavior of information users [30]. In other words, much research has emphasized information systems over people. In contrast, usability experts have recognized the importance of designing systems contextually—that is, conducting "work practice" studies and using that information to drive information system design [31].

The Catalog and Information Seekers

The recent library literature contains numerous articles on the need for change in online catalogs to better satisfy the expectations of information seekers who are accustomed to easy-to-use Web search engines, online bookstores, and seamless linking to full text. In his August 2005 paper for the International Federation of Library Associations (IFLA), LC's John Byrum wrote of the need for library catalogs to provide access to more content and to offer significantly enhanced functionality based on the features of popular search engines [32]. Speaking of the limited scope of the catalog and its emphasis on print, Norm Medeiros wrote "more and more, users want, expect, and pursue full text. In increasing numbers they look past the catalog when searching for e-journals, databases and Web sites" [33]. In a January 2005 presentation at ALA, Dale Flecker asked the audience "How often have you heard 'Why can't I find journal articles along with books in the catalog?'" [34].

Although it is eleven years old now, Charles Hildreth's 1995 report to the then Council on Library Resources [35] continues to provide an authoritative analysis of the shortcomings of Boolean online catalogs and the problems users have with them:

- Many failed searches
- Frustrating navigation
- Unfamiliar subject indexing policy and vocabulary

- Confusing search and retrieval methods (e.g., pre-coordinate phrase searching, post-coordinate keyword/Boolean searching)
- Poorly organized search result sets (e.g., little or no relevance ranking)

Hildreth concluded his report with an outline for a new kind of library retrieval system based on actual information seeking behaviors. Major functional improvements he proposed include natural language searching, automatic term conversion and matching aids (like "did you mean?"), best-match retrieval (no search dead ends), ranked output of search results, relevance feedback (like "more like this"), linking to related information, clustering tools, and expanded coverage and scope.

Unfortunately, the next generation online catalogs that Hildreth envisioned have been built, but not by libraries. In 2004, Holly Yu and Margo Young, in their introduction to a new online catalog transaction log analysis, noted "in spite of many studies and articles … over the last twenty-five years, many of the original ideas about improving user success in searching library catalogs have yet to be implemented. Ironically, many of these techniques are now found in Web search engines" [36]. Yu and Young conclude their paper with recommendations for future OPACs including federated access to the OPAC and scholarly databases, more effective interfaces, enhanced browsing and display of search results, relevance ranking, and tools like "did you mean?" [37] Their list agrees in principle with Hildreth's older one, even the recommendation for federated access to the catalog and databases.

Prospects for the Research Library Catalog

In summary, research library online catalogs reflect a small portion of the universe of scholarly information. Information seeking studies in libraries have tended to focus on information sources and systems rather than the people who use them. Useful research has been done to identify catalog improvements, but they have not been implemented. Catalogs are hard to use and their interfaces seem increasingly out of date.

Stated in business terms, the library catalog can be said to be in a declining stage of the product life cycle [38]. Newer and more appealing products (like Google or Amazon) have entered the information market and while loyal library catalog users remain so, other users have begun to shift partly or completely to the products they find more appealing. For example, the ARL's 2004 LibQual+™ results suggest that more than two-thirds of undergraduates, graduate students, and faculty get information daily from popular search engines, compared to less than two-fifths of any group (and for undergraduates, slightly over a tenth) using library Web pages daily [39]. Library online catalog use would account for only a subset of the use of library Web pages. Similarly, OCLC research on perceptions of libraries and information resources confirmed "the library is not the first or only stop for many information seekers. Search engines are the favorite place to begin a search …" [40]

The Catalog and Research Library Collections

It is easy to understand why some dispute the claim that library catalogs (or rather, the records they contain) will remain important for supporting discovery and

delivery of the world's library collections. Indeed, as early as 2003, OCLC forecasters predicted widespread expansion of digitization projects—commercial, national, state and local [41]. Since that time, not only have large-scale projects like Google Book Search, the Million Book Project, the Open Content Alliance, and a project funded by the European Commission entered the field or continued to ramp up, but also e-book publishing continues to evolve [42, 43, 44, 45]. E-journals, e-books and the full-text and image files produced through digitization projects give users new options besides catalog data for discovery and retrieval. Yet barring a technological miracle (and certainly they have occurred in the past twenty years), it will take some time for the world's library collections to be fully converted to digital forms.

The following scenario contains many flaws and unexamined issues but it can serve as a crude guide to what would need to be accomplished. Considering only books, OCLC researchers estimated the January 2005 size of the "system-wide library print book collection" at 32 million books [46]. Assuming an average of 300 pages per book, and that digitization could occur at the speed of the fastest robotic scanners available today (about 1,000 pages per hour), and given 10 of these scanners operating 24 hours a day for 365 days a year, it would take 110 years to digitize 32 million books [47]. Digitization projects of course do not achieve speeds of one thousand pages per hour per scanner (scanning is only one of the activities that must take place). While the scale of current mass digitization projects is impressive, even if digitization occurs at many times the speed calculated here, it may be safe to say that catalog records will have a role to play in discovery and retrieval of the world's library collections for at least a couple of decades and probably longer.

In addition, research libraries continue to invest heavily in printed library materials. In 2004, the ARL libraries spent more than one billion dollars on library materials [48]. Subtracting the average 32% the ARLs expended on e-resources, ARL libraries expended an estimated $665 million on traditional materials in 2004, mainly printed books and serials. While A&I and full text services provide access at the article level to many serial titles, and experimentation with ONIX or publisher-supplied data shows promise for providing for discovery for some of the books, the most reliable method for providing for discovery for printed materials continues to be to catalog them.

One may predict, then, that the legacy of the world's library collections is for the time being tied to the future of catalogs. The declining market position of the research library catalog puts research libraries on the horns of a dilemma. The problem of the catalog is acute and cannot be ignored. The ARL libraries have invested and continue to invest huge amounts in their catalogs' creation and maintenance—in 2004, an estimated $239 million on technical services labor alone [49]. On the one hand, this investment seems justified; catalogs enhance the value of the world's library collections, and it can be predicted that catalog data will fill this role for years to come. On the other hand, a large and growing proportion of students and scholars are bypassing the catalog in favor of other information sources, and the catalog represents a shrinking percentage of the scholarly information universe.

Appendix B: Interview List and Questions

Name	Title	Affiliation
Bill Arms	Professor, Computer Science	Cornell University
Christine Borgman	Professor, Information Studies	UCLA
Peter Brantley Laine Farley Daniel Greenstein	Director, Digital Library Technologies; Director, Digital Library Services; University Librarian and Executive Director	California Digital Library
Beth Picknally Camden	Director, Goldstein Information Processing Center	Van Pelt-Dietrich Library Center, University of Pennsylvania
Lorcan Dempsey	Vice President and Chief Strategist	OCLC Office of Research
Carol Diedrichs	Dean of Libraries	University of Kentucky
Dale Flecker	Associate Director for Planning and Systems	Harvard University Library
Carl Grant	President and COO	VTLS Inc
Sandy Hurd	Director of Strategic Markets	Innovative Interfaces, Inc.
Bruce Kingma	Professor and Associate Dean	School of Information Studies, Syracuse University
Amos Lakos	Librarian	Rosenfeld Management Library, UCLA
David Lindahl	Director, Digital Library Initiatives	University of Rochester Libraries
Judy Luther	President	Informed Strategies, Inc.
Clifford Lynch	Executive Director	Coalition for Networked Information
Norm Medeiros	Coordinator for Bibliographic and Digital Services	Haverford College Libraries
John Miller	Senior Special Projects Librarian	University of Kansas Enterprise Academic Systems Information Services
John Price Wilkin	Associate University Librarian for Library	University of Michigan

Name	Title	Affiliation
	Information Technology and for Technical and Access Services	
Roy Tennant	User Services Architect	California Digital Library
Sherry Vellucci	Assistant Professor	SCILS, Rutgers University
Ann Wolpert	Director	MIT Libraries
Bob Wolven	Director of Library Systems and Bibliographic Control	Columbia University Libraries

1. How should the library online catalog change over the next five years to maximize its utility for the *communities* served by major research libraries? How should the library online catalog change over the next five years to maximize its utility for managing *internal* library operations?

2. How should library management system (a.k.a. ILS) vendors be positioning themselves and their products for the future? In the future, what will be the role of the library catalog database in a library management system?

3. By tradition, the library catalog has focused on one particular level of description— for books, the edition; and for serials, the title. In the future, what should be the scope of the library online catalog—that is, what kinds of information objects should it contain, and why?

4. What are the future roles of MARC, MARC cataloging data, and cataloging content rules for description and access, including subject access?

5. What if any are the library online catalog's unique benefits to information seekers, compared to other types of discovery systems?

6. What are the one or two most important challenges to the successful integration of the library online catalog with other discovery tools available to information seekers?

Appendix C: Key Findings from the Interviews and Literature Review

        The author and Martin Kurth conducted the interviews between October and December 2005. Twenty-three individuals responded and were interviewed. Nearly all interviews took place in hour-long phone conversations; three were completed via e-mail. The interviewers transcribed the notes from each phone interview and e-mailed the transcript to the interviewee for approval or changes. Once complete, the transcripts became the basis for the analysis that follows. The appendix lists the interview questions and the names of those interviewed. The following sections offer highlights of the findings and also discuss various issues gleaned from the literature review.

The Catalog's Unique Advantages

        Discovery and management in one package. When asked to identify the catalog's unique benefits to information seekers, many interviewees pointed to the classical functions of bibliographic control—both its retrieval functions (to enable a person to find, identify, and select an item of interest, then use the data to obtain the item) and its management functions (recording identifying information for each item, inventory control). The notion of bibliographic control is centered on surrogates—that is, catalogs work best as they were designed, with online metadata pointing to offline materials. Since bibliographic data is purposely chosen for collocation (by author, title, subject, etc.), retrieval and evaluation, the catalog is consistent and precise; it offers ways to search that are impractical in other settings. As one interviewee put it, "A user who knows how to search the catalog gets excellent results."

        Support for browsing; **consistency and collocation.** The catalog's support for browsing was another oft-mentioned advantage. The predictable and consistent structure of catalog records can facilitate serendipitous discovery and offer a wide variety of ways to browse. The catalog contains good metadata, in the sense that it uses authority control, classification, and content standards to describe and collocate related materials. The catalog can play a role in structuring the research library communities' deepest collections (e.g., five thousand versions of *Hamlet*). One interviewee noted that AACR/MARC catalog records "are a unique benefit of the catalog; they are a huge asset that libraries are not taking advantage of." Along these lines, it is worth mentioning the unveiling in early January 2006 of the new Endeca-powered online catalog at North Carolina State University. The press release notes the new catalog "provides the speed and flexibility of popular online search engines while capitalizing on existing catalog records. As a result, students, faculty, and researchers can now search and browse the NCSU Libraries' collection as quickly and easily as searching and browsing the Web, while taking advantage of rich content and cutting-edge capabilities that no Web search engine can match" [50].

        Identification and inventory control. This was another characteristic response to the question about the catalog's advantages for information seekers. The catalog gives

detail about what and where items are, their status (on shelf, checked out, etc.) and how to get them. The catalog is uniquely useful for managing large collections (an internal function) and for supporting delivery of those collections to users for free (or close to it). Finally, for some users, it is important to be able to identify a manifestation (for example, a particular edition of *Hamlet*); the catalog has a unique advantage in this respect.

Delivery of local holdings. One interviewee captured this notion by saying "the catalog has a unique benefit when it provides access to information not available via search engines, that is, information available only inside libraries." The catalog provides the link to materials owned locally; in this sense the catalog may be viewed as "last-mile technology," carrying signals from the broader network along the last mile to a home or office, and providing infrastructure at the "neighborhood" level to complete the discovery to delivery value chain.

Cataloging Tradition and Catalog Data

MARC. Interviewees found little to be satisfied with but the consensus was that MARC is not going anywhere. Many recognize that the MARC communications format created much that was unique but is now out of date, and that the future lies in the convergence of MARC data within the global information network. Only two interviewees fell into what might be deemed a "MARC must die" camp. Other interviewees were more sanguine about MARC encoded data, noting that data transformations between structures are routine matters. It may be necessary to change how MARC data is packaged (e.g., in XML) to more easily pass data around with the rest of the information world. It will be important to consider how MARC data can be reused. Since libraries have such a large base of MARC data, library system migrations will remain a big deal and libraries will need to figure out how to take their stores of MARC data forward.

Two additional insights about MARC encoded data are worth reporting here: one interviewee argued that "libraries should be using their cataloging data more aggressively than in the past, processing it more, passing it around more." Another mused "it is difficult to imagine the costs of converting millions and millions of MARC records in thousands of databases around the world to new metadata structures." Along these lines, Roy Tennant's article on a bibliographic metadata structure for the 21st century describes a collaborative approach that will enable libraries to absorb their legacy data while also supporting interoperability. He writes "what must die is not MARC and AACR2 specifically … but our exclusive reliance on those components as the only requirements for library metadata." Speaking of system migration, Tennant notes "the changes proposed here must clearly be fostered by cooperation at a national, and perhaps international level and carefully staged" [51].

Alternatives to cataloging records. Interviewees would generally agree that there must be good accepted ways to rely on or reuse metadata from outside the library and having this metadata co-exist with manually created records. Some interviewees favored starting over with a core package of metadata elements.

Cataloging practice. There was some consensus around the position that cataloging needs to be simpler, faster, and less expensive. There is "too much hand wringing and worrying about each record: this is extreme and wasteful." There was sharp criticism of local variations in cataloging records "with little or no added benefit … Such 'fruitless variation' now looks a bit precious. In the 80s and 90s such variation seemed defensible, but no longer."

Library of Congress Subject Headings (LCSH) and traditional subject access. Interviewees had a lot to say about LCSH and library tradition for providing subject access. Opinions ranged from the strongly critical to an attitude akin to quiet resignation. There were no strong endorsements for LCSH. The table below offers a summary of interviewees' comments.

Less critical	**More critical**
There is need for subject cataloging in the context of clustering related content. LCSH is not ideal but it offers a readily available means of labeling clusters.	If you put the money you're spending on LCSH in automatic classification, you might get something more competitive in the Google world and get better subject access too.
For subject access, is the technology good enough so we can move away from manually assigned terms from controlled vocabularies?	Dedication to LCSH is ridiculous. We must deal with this.
I'm not ready to give up on LCSH. Subject analysis during cataloging is important; it produces a call number and controlled terms. It is hard to say how well LCSH serves as a source of controlled terms, but it is better than relying on keyword alone.	How much subject searching is actually done? Now with the ability to search full text or even TOCs, do we need subject analysis for textual materials?
LCSH headings will be around because of the economic value of legacy data and because users will continue to need access to library materials in this way.	LCSH requires too much behind-the-scenes understanding to be useful.
Subject access will be and has already been marginalized by keyword searching; this could change if clustering finds favor with ILS vendors and customers.	Traditional subject access prospectively will become less and less heavily used. Providing traditional subject access is a big part of the investment in cataloging and the economics will not be attractive to continue it.
	There is a real question whether LCSH is cost effective.

Authority control. In general there was more support for authority control of names and places. Several noted that the Web community's adoption of ontologies, taxonomies and folksonomies could spur new interest and vigor in these set of library practices. One interviewee noted that libraries are particularly skilled at organizing deep collections and cautioned "we should not walk away from library history too quickly." Another predicted "there will be great value in (computationally) recognizing people's names and place names in full text and normalizing them; gazetteers and name authority files will take on more prominence."

Cataloging content rules, RDA and FRBR. Again, interviewees had a lot to say on these topics. The following chart summarizes the comments.

Content rules*	RDA**	FRBR***
The intellectual thinking behind AACR and related content standards should not be lost; however many the details are not well suited for computer processing.	The 3-fold initiatives—RDA, FRBR and MARC21—are now moving along in parallel, but they should be more closely and deliberately attuned and aligned. All three initiatives are dominated by the worldview of a single monolithic local catalog— very likely a false assumption.	What needs to be part of catalog data to support FRBR catalogs, new kinds of delivery services, data mining, mass digitization, offsite storage, etc.? Libraries should be using their catalog data more aggressively, processing it more, and passing it around more.
Generally speaking there is value in tracking information at the edition level; and for digital objects, we need version control.	Those responsible for updating the cataloging rules must keep in mind the larger role that automated metadata creation will play.	We need FRBR to rethink where we are; I am not sure FRBR is the right solution but it is an attempt to address the right problem.
There will be less reliance on metadata with the availability of full text, but this has limits.	Those who shape RDA need to examine how metadata standards work and how metasearch standards (NISO) fit in.	We should FRBRize our data.
There is enormous value in a structured approach to discovery using metadata, particularly for unique special collections.	The more important issues related to the cataloging rules; how do the content rules change in a world where more content is available in computable form?	FRBR has a lot of potential; its structure is more like users think.
The content rules are more important than the tagging system; AACR and the coming RDA set the groundwork for sharing data.	RDA will be a big step forward. If there's any hope of other communities using it, I can't guess.	The catalog is being rethought in light of FRBR—people want to see clusters rather than long lists.
The content rules and other data constraints have benefited the Google Book Search project.	It's nice to think that we can expand the content rules and subject taxonomies beyond libraries, but I'm not sure it is likely to happen.	A FRBR view of collections by work is good; I'm a little nervous because it pushes catalogers into more content analysis.

Content rules*	RDA**	FRBR***
I see an ongoing role for content rules for a subset of library materials.	RDA is visionary; it could have a global impact on how communities code their metadata and on cross-community information retrieval.	A FRBRized display would allow us to more easily expand to add new types of resources. For example, flat MARC has never worked well for AV.
The content rules should evolve but they will handle traditional cataloging objects, not specialized materials.	It's important that RDA is integrating the FRBR structure. The ability to link records hierarchically is important.	We can't implement FRBR-like enhancements in the current metasearch environment.

*Anglo American Cataloging Rules
**Resource Description and Access
 ***Functional Requirements for Bibliographic Records

The Scope of the Research Library Catalog

No consensus. The interviews did not yield a clear preference for the proper scope of the research library catalog. One set of responses clustered around the sentiment that, to be useful, the catalog must hold all the information that the university community wants to search on. One interviewee sat on the fence, remarking "I see virtue in both integration and segregation; there are not a lot of convincing models for approaching this question." Further along the continuum of responses, another set clustered around statements like "I have never been a fan of smashing everything into the catalog" and "the catalog may be less likely to contain everything we want people to know about." At the end of the continuum of opinion, one interviewee noted "within five years, we'll be past the notion that the online catalog is the way you find things in libraries."

User expectations. There was more consensus around what users expect the scope of a research library discovery system to be. Many interviewees commented that users don't understand the difference in scope between the catalog and A&I services (or the catalog, databases, digitized collections, and free scholarly content). Interviewees tended to agree that a more seamless approach to the serial literature is needed, allowing users to look in one place for books, serials and articles. A couple interviewees emphasized that users absolutely want individual serial articles discoverable from the catalog. One of these individuals continued "For serials, librarians and faculty may be the only users who think in terms of the container [the serial title]." An interviewee who is involved in the Google Library Project remarked on the importance of including digital texts in the catalog. Another summed things up with "users don't get the idea of the catalog; they just want results."

Linked pools of data. The question about the scope of the catalog was a complex one that prompted many qualified answers. One respondent encouraged the separation of function from architecture when thinking about the future catalog's scope. Others envisioned many pools of data linked together for searching, rather than one data store for every kind of information object. Many suggested emphasizing interoperability and using federated access methods. Others noted the rising need for automatic indexing and

metadata creation. Speaking of a student's expectation to discover a faculty member's course Web pages, DSpace entries, and published works in one place, one interviewee remarked "the catalog doesn't necessarily need to hold everything, but users want a search engine or portal to bring them all together."

The trouble with metasearch. There was some hope and many fears about metasearch as a technology, but no consensus. Comments ranged from "metasearch is a fatally flawed technology" to "metasearch may not be the right solution but it is addressing the right problem" to "metasearch has enough promise that we should go forward with it." Among the many interviewees who talked about metasearch, there was agreement that the NISO MetaSearch Initiative is critically important to the future of this technology [52]. Quoting from the Initiative's Web site: "The absence of widely supported standards, best practices, and tools makes the metasearch environment less efficient for the system provider, the content providers, and ultimately the end user." The problems with metasearch are well documented in the literature [53, 54, 55]. Some writers, like Marshall Breeding, are beginning to point to Google Scholar as an example of a better approach (i.e., searching based on a centralized index). Besides the absence of shared standards, which was interviewees' most frequent complaint about metasearch, they cited problems with the time commitment required for local and vendor work with metasearch engines and to keep connectors working, the absence of needed relevance ranking in search results, and the nascent state of metasearch technology.

Integration with Other Discovery Tools

The issue of surrogates and full text. Several interviewees noted that the catalog is based on the assumption of surrogates. Over time, catalog construction has been constrained by the high cost of cataloging, the small size of a catalog card, and the scarcity of full text. As mass digitization projects advance, more full text searching of books will be possible. One interviewee noted "the catalog needs to function in the context of full text searching"; another said "cataloging practices going forward need to think about operating in a world where full content is available; it is foolish to replicate things that can be done computationally; there are only a few cases in which you'd want to build and index and assign subject headings [manually]." At the same time, this individual recognized "there remains a big inventory control problem to deal with; there will continue to be a sizable amount that is only represented by surrogates."

Google and Google Book Search. During November 2005, Google searches accounted for 40% of all search engine traffic on the Internet in the U.S., further increasing Google's lead over Yahoo. The search engine market can be characterized as an oligopoly, with just a few firms controlling nearly all usage. Between them Google and Yahoo accounted for 69.3% of the total 5.15 billion searches that U.S. residents conducted last November, followed by MSN (14.2%), AOL (8.7%), and Ask Jeeves (6.5%) [56].

Google's popularity is immense and growing. Google Book Search and its Library Project continue to make progress [57]. The most recent OCLC report presents compelling evidence that college students begin looking for information on search

engines—89% of this group said they begin searches with a search engine vs. 2% who start their searches on library Web pages [58]. The Open WorldCat program exposes library-owned materials to popular search engine users and provides links that lead to local library collections [59]. From Google Book Search, when users click on "Find in a Library," Open WorldCat data enables the identification of appropriate local libraries. While it is not so easy for now to find library versions of books on Google, there is an expectation that library materials will gradually become more visible as the Google Library Project progresses [60].

It is not surprising, then, that for a number of interviewees, the question of the catalog's integration with other discovery tools orbited around getting a Google user from Google to library collections. Several noted the importance of the interface between the library and Google. One remarked, "In the best of all possible worlds, people could search Google and library resources together [on Google]." Another noted "data about a library's collection needs to be on Google and other popular sites as well as the library interface." One interviewee, however, was cautious about such an approach because of the extent to which catalogs contain surrogates pointing to physical locations. This interviewee said that indexing library catalogs for Google searching "would be antisocial, because it would introduce millions of records of noise into Google. OCLC and others have experimented with exposing union catalogs on the Web … we're still very early in learning how to do this effectively. Google can deliver instant gratification. Libraries don't typically do that, especially with their physical holdings."

Notwithstanding the varying perspectives on Google, the suggestion that currently, "integration is outward rather than inward, with libraries seeking to use their components in new ways" is highly characteristic of the interview data.

Local vs. shared catalogs. Interviewees agreed that today's research library catalogs are not the right finding tools for users. Catalogs are one tool among many in a user's information universe. The local catalog provides insufficient coverage of the scholarly information universe. An information seeker's first exploratory point is highly likely to be outside the catalog. Today, the research library's task is to switch users in their communities from where they find things to library collections. Perhaps one interviewee asked the key question: "how does the catalog become manifest in the user's environment?"

Along these lines, the interview data suggest that librarians are beginning to question the prevailing model of catalog creation and maintenance, that is, replication of data in shared cataloging systems and in thousands of local catalogs. The approach of aggregating catalog data regionally or nationally is increasingly attractive to some. "Should there be one catalog or many?" is an emerging question. Two interviewees suggested exposing WorldCat (presumably, with scoping) to local library users instead of the local catalog. Another noted "the idea that every library has to have its own catalog is problematic. Libraries are starting to collaborate on collection development; why continue to have single library catalogs?" Speaking of the growing importance of interlibrary lending, another wondered "will ILL among research libraries be structured

around local catalogs or around union catalogs like WorldCat?" Two interviewees noted the advantage of emerging "on-demand" services like salesforce.com [61], which is a hosted service that simplifies operations and lowers costs for firms that use it.

Revitalizing Catalogs for the Communities They Serve

Interviewees seem to agree that however it is done, catalogs must blend into the user's environment and engage users more. The user's experience needs to be more seamless, with easy movement between the services that occupy students and scholars—course pages, commercial databases of e-content, repositories, search engines, and so on. Although they may have expressed it differently, interviewees find linking in and out of the catalog a crucial component of what is needed to move ahead. The catalog interface (whether a local or shared catalog) should look and work like Google's. Searchers expect instant gratification and positive feedback from the systems they use. On behalf of research communities, libraries should be contributing to bigger scholarly environments through data exchange, collaboration and partnerships to aggregate more of the expanding universe of scholarly content.

Discovering, requesting, and getting. As discussed earlier in this appendix, today's catalogs offer discovery and management in one package. A unique benefit of catalogs is their provision of bibliographic control—both its retrieval functions (to enable a person to find, identify, and select an item of interest, then use the data to obtain the item) and its management functions (recording identifying information for each item, inventory control). Some interviewees shared the insight that the Internet has caused these functions to diverge—that is, finding can happen in one system, identifying and selecting in another, and getting (that is, delivery of the item) in still another. As for bibliographic control's support for management functions, these functions can and increasingly do happen in separate but linked systems.

Juha Hakala, in his 2004 article on the emerging heterogeneous system environment, wrote "in one vision of the future, even the monolithic ILS will be split into smaller pieces … integrated systems will be replaced by modular ones, which by definition are a perfect fit for networked or consortial environments" [62]. Recently Roland Dietz and Carl Grant, presidents of Endeavor and VTLS respectively, described the "dis-integrating world of library automation" and called for library systems to be more compatible with other systems [63].

The future will require the kind of catalog that is one link in a chain of services enabling users to find, select, and obtain the information objects they want. One requirement of this future catalog is thus to ingest and disperse data from and to many systems inside and outside the library. It would be helpful to reconsider what needs to be part of catalog data—and where catalog data needs to be present—to facilitate the user's process of discovering, requesting, and getting the information they need.

The long tail. Chris Anderson, editor in chief of *Wired*, wrote in 2004 about the Internet's impact on the economics of the entertainment industry [64]. Freed by online distribution methods from the need to sell only to local audiences in the theaters of cities

and towns, firms like Netflix can profitably offer a huge selection of films—both mainstream hits and special interest titles. Having so many titles to choose from has spurred the rise of numerous niche markets and eroded the economic impact of the 80-20 rule (only twenty percent of releases account for eighty percent of sales). The eighty percent of films that appeal to a smaller (rather than mass) audience represent the "long tail." Because Internet distribution of content reduces costs and greatly enhances the visibility of both "hits" and "long tail" titles, the entire demand curve shifts outward, spurring interest in many more titles than would otherwise have been possible. Anderson calls this "the power of the long tail" and he has applied its lessons to libraries [65]. Because libraries are increasingly connected—and could become more connected through the Internet—there is an opportunity for connecting individual library collections into "a vast supercollection that can go further down the Tail than any single institution could afford" [66]. If research libraries could find a way to make their collections more visible on the Web to the global scholarly audience, such an achievement has the potential to aggregate now dispersed audiences and launch a new era of scholarly research and learning. A huge question is, can research libraries and their systems become sufficiently connected on the network to tap into the benefits of the long tail?

Web services and interoperability. Several respondents talked at length about Web services— technologies allowing applications to communicate across platforms and programming languages using standard protocols based on XML—to connect catalogs and other library resources to search engines, e-learning systems, portals, Amazon, etc. with the goal of providing a more seamless and satisfying experience for information seekers in research institutions. NISO's VIEWS (Vendor Initiative for Enabling Web Services) Initiative has been working since 2004 on the issues of enabling web services between the disparate applications used in libraries [67]. Most recently, VIEWS has evolved into the NISO Web Services and Practices Working Group, which began work in late fall 2005 to produce and maintain best practices and interoperability documents [68]. The Working Group is defining use cases for library web services including discovery, locating, requesting, delivery and administrative services. Interoperability, the core of web services, is a crucial factor for designing the kinds of information systems that research libraries need to operate in the larger scholarly information universe.

Usability. There is a lot of evidence that users are quite frustrated with research library catalogs. In his oft-mentioned diatribe "Burn the Catalog," Swarthmore faculty member Tim Burke concludes "I'm to the point where I think we'd be better off to just utterly erase our existing academic catalogs …" [69] Besides reiterating the improvements Hildreth proposed (now typically found in Web search engines), interviewees also suggested enrichment of the catalog with title page or jacket images, reviews, tables of contents and such—services that appear to be gaining some ground [70]. Respondents also discussed newer ideas for improving catalog usability through FRBR concepts, visualization techniques, and interactive features.

As noted earlier, applying FRBR concepts to improve the user's experience with catalogs was often mentioned by interviewees. Much is appearing in the library literature about deploying FRBR concepts [71, 72, 73, 74, 75]. There is excitement around the

Research Library Group's RedLightGreen [76] and OCLC's work-based catalog investigations such as Curioser [77].

A couple of interviewees urged enhancing catalogs by making them more visual and browsable, for example with concept maps, either word-oriented or visual. Along these lines, Stanford has experimented with Grokker [78], which groups search results from a variety of sources (including the Stanford catalog) and presents them in an interactive visual map. Naomi Dushay, working on behalf of the National Science Digital Library project, reports a wealth of ideas from her research to apply technology to "provide a user experience analogous to walking among well-organized, well-stocked bookshelves" [79].

Interviewees also suggested more interactive catalogs—letting users give feedback (such as reviews), giving users more power to control transactions (such as interlibrary loan or payments), offering RSS feeds or canned queries (such as for new books), permitting social bookmarking, and providing new output options. OCLC is one organization that is exploring the feasibility of user-contributed data through its WikiD research project [80].

Communities of users. Interviewees recognized that for catalog functionality, one size does not fit all. The catalog serves different communities with different needs. Differences exist by type of user (student, scholar, librarian), by discipline (humanist vs. scientist) and by generation. One interviewee remarked "students and junior faculty have different expectations than our older generation of users who are deeply embedded in library traditions and trained in traditional bibliographic tools."

Revitalizing the Catalog for Internal Operations

Inventory control. Interviewees agreed the local catalog would remain important for acquisitions and for supporting storage, interlibrary services, delivery, digitization, preservation, and circulation of the library's physical assets. There was a sense that the catalog will remain core for behind-the-scenes work and that "we are not at a point where we can walk away from ILS support for internal operations." Interviewees had a variety of suggestions for enhancing behind-the-scenes functionality: better management of versions of things (formats and editions); better support for item-level management of digitized resources (esp. in connection with mass digitization projects); support for "clumped" or multi-library collections. One interviewee speculated that "the catalog's support for internal operations will take on more weight … We need to connect descriptions to financial transactions, to patron use information, rights transactions, asset management systems, and storage systems. Libraries will head toward not bundling everything into one set of programs but toward assembling interacting services, toward multiple formats and multiple databases to handle multiple operations."

Automated metadata production. Some interviewees noted that today's catalogs are put together mainly by humans and that this approach doesn't scale. Several urged building or expanding the scope of catalogs by using automated methods.

External sources of metadata; data ingest and extraction. New and emerging library workflows will require easier ways to gather, load, extract, manage and maintain catalog data from a variety of sources.

Collection level records. Two individuals discussed the need to be more flexible about defining collections or clusters of information objects at various levels due to the problems of scale and economics. One suggestion was "for collections of images with limited metadata, think about ways to describe them in clusters of some kind." Another was "offer a hierarchical approach to works—link from collection level records to surrogates/objects in other databases."

Reporting functions and management data. Several respondents suggested that library decision-makers need more and better statistics and management data for making decisions about internal operations.

We're (not) special. Libraries currently have many custom processes and workflows for accomplishing the same internal operations. This diversity makes it extremely costly for ILS vendors to support libraries, because they have to design so much flexibility into their systems. Interviewees urged standardization of library workflows, cutting out redundant work, and generally streamlining library operations.

Integrated Library System (ILS) Vendors

The ILS market. One interviewee's comment "the [ILS] market is more broken than ever before" captures the general sentiment. Respondents agree that ILS systems are out of whack with customer expectations and pointed to the difficulty ILS vendors have serving the traditional needs of libraries at the same time they are attempting to evolve into information services firms. There are few vendors, poorly capitalized, and libraries are a small and demanding market with, relatively speaking, little to invest in new ventures. Speaking of ILS vendors' and libraries' potential for supporting scholarly portal development, one interviewee noted "there is not enough money to make it happen." Speaking of how difficult the library market is to serve, another remarked "libraries want a ton of customization; this is ridiculous and must stop."

The characteristic tension between competition and collaboration is particularly strong in the ILS market at this time. At the "competition" end of the spectrum, some interviewees noted that vendors are interested in "stickiness"—that is, the integrated system is monolithic, and the costs of switching to another provider are so high that libraries are reluctant to undertake migrations or change the way they do things. At the "collaboration" end of the spectrum, some interviewees urged ILS vendors to be "deeply interoperable," to participate in the standards development and implementation process, and to position themselves through partnerships with each other and with complementary firms (for example publishers and aggregators, campus learning systems, search engines, etc.). Interviewees want vendors to sell components and to position themselves so that libraries can offer alternatives to the local catalog for discovery—for example WorldCat or Grokker or Endeca.

Decoupling discovery and the ILS. The recent achievement of North Carolina State University, with its Endeca-powered online catalog decoupled from the ILS, has created a good deal of excitement. NCSU's Web catalog represents a significant step toward what the interview data suggests that libraries want, not only in terms of a Google-like interface and better support for browsing, but also in terms of system modularity. As Lorcan Dempsey put it recently in his blog, "we are seeing a growing need to be able to interact with an ILS in various ways, so that some functionality can be placed in another interface—to show status of an item, to place a hold, to do lots of things"[81]. Dempsey suggests that rather than extending the variably implemented protocol-based approaches and local fixes that are emerging in the market today, library service providers define a new service layer to the ILS "that would allow some of these things to be done in more routine ways" to increase the portability of current solutions.

At the same time, the new NCSU catalog is limited in scope to NCSU's library collections; it has not diversified its functions to cover more of the scholarly information universe. It does not merge the ILS finding function and metasearch, nor does it support a variety of metadata types. It does not interoperate with the campus learning management system or enable users to search library data directly from external search engines or portals. This is not to be critical of NCSU's highly praiseworthy achievement, but to suggest the scope of the problems that remain to be solved.

Modularity, the ERMS and the ILS. There was a good deal of consensus around the notion that the catalog database and the ILS would continue to be needed as inventory tools and that library systems need to become more modular and more open to outside data and functionality. As one respondent put it "ILSes should think in terms of linking rather than building." A new kind of management system—the ERMS (e-resource management system)—is perhaps the beginning of the trend toward modularity. ERMSes are taking hold in libraries to support their growing e-resource collections. One interviewee offered the insight that given the direction that the future catalog may take as an inventory management system, together with the growing importance of licensed e-resource management, "it is possible that the ERMS and the ILS will collapse into a single system within five years."

Challenges and Obstacles

The fault is not in our stars but in ourselves. The final interview question dealt with the most important challenges to successfully integrating the online catalog with other discovery tools. Many considered the greatest challenge librarians' own narrow views and lack of vision. The following chart offers a small sample of interviewees' more provocative comments. Respondents urged librarians to take a fresh approach, articulate and market the value of libraries to their communities, build new professional skills, listen better, and give users what they want. A couple of interviewees advised ILS vendors not to ask librarians what systems should do, but to find out what libraries need to do for their users (and forget the long enhancement lists from librarians).

Starting Over	Vision and Value	Skills and Resources	Listening and Acting
The biggest challenge is history. There is great investment in skilled people, standards, catalog records, and technology. Yet in many cases they represent the past, not the future.	Find people who can articulate a vision for what the library is trying to achieve, to envision how users will interact with a given service, to answer questions like 'What should the front page of the library site include?'	The lack of skill sets and financial resources is a challenge. Libraries can't build the systems they need, which means they will be built externally.	Librarians can no longer remain wrapped up in a self-defined profession, supported by the few professors who want to keep working with bibliographic tools.

The catalog and its boundaries. As noted in an earlier section, no consensus emerged about the proper scope of the catalog. The interview data suggest that defining the boundaries of the library's coverage of the scholarly information universe is a critical challenge, on at least two levels. The first level has to do with the comprehensiveness and coherence of the collections for which a research library claims or accepts stewardship. Today's catalog covers an important core collection (mostly books and journals, electronic and print), but a shrinking proportion of what students and scholars want to find and use.

The second level has to do with how and where scholarly information seekers find things on the Web and then get hold of them—these are questions of integration and aggregation of data. Already, in the case of Web-based citation indexes, users can start on a citation, pass through a library's information systems, then move on somewhere else (to the full text of the article, one hopes). Could this model be expanded to include many starting points on the Web and multiple types of scholarly information objects? As one interviewee put it "there needs to be a middle layer that gets users from description to access. The OPAC can play a role in that middle layer." Whether library catalog data is aggregated for discovery in popular search engines, in WorldCat or RedLightGreen, or in group or single library catalogs, the catalog must be open to various agents to retrieve and display the appropriate metadata within a variety of interfaces.

Standards development and implementation. Implementing Web services requires standards definition and compliance. Metasearch will not endure as a technology without broad-based adherence to the standards that are emerging. The proliferation of metadata types has not yet reached its zenith. Most current library systems are fairly closed, when they need to be open. The pace of standards development and implementation is painfully slow, because the process requires competitors to collaborate. For some firms, the benefits to their own firms of complying with standards are not clear. When standards are approved, compliance is voluntary, so the community ends up with

many variants of the same standard. Some worry that interoperability, for now, is more a myth than a reality.

Copyright. The U.S. copyright law contains certain exceptions and limitations that apply to libraries and archives. In general libraries have certain permissions related to making reproductions of copyrighted works, but there is a lack of clarity around digital reproductions, as opposed to photocopies. Various mass digitization projects are impeded to one degree or another by this lack of clarity. Google has been sued over its Google Library Project [82]. A Section 108 Study Group was convened by the Library of Congress in April 2005 and "charged with updating for the digital world the Copyright Act balance between the rights of creators and copyright owners and the needs of libraries and archives" [83]. Depending on the eventual interpretation of what constitutes fair use of copyrighted materials for educational purposes, the climate may range from favorable to hostile for the development of the kinds of open Web access and aggregated scholarly collections suggested by those interviewed for this report.

Who pays? How to build it? Interview responses demonstrate that the challenges of integrating research library catalogs with other discovery tools will not be solved by individual libraries working alone. Instead, research libraries need to commit to concerted action on an unprecedented level, and major players need to play the role of integrators. Research libraries have a spotty record developing and sustaining the kind of large scale partnerships that are required. With respect to building cooperative cataloging and resource sharing systems with a national or international reach they have been reasonably successful; with respect to collaborative collection development or other kinds of shared services, less so. An encouraging development in January 2006 was the release of the University of California Libraries Bibliographic Services Task Force's final report [84]. One recommendation for rearchitecting the OPAC was to create a single catalog interface for the whole UC system. Other regional or state-wide shared catalogs (e.g., OhioLink) have enjoyed success.

Endnotes

[1] For a summary of the conference see Calhoun, Karen, Carol Hixson and Sherry Kelley. 2001. "Highlights of the Library of Congress bicentennial conference on bibliographic control for the new millennium." *Library Collections, Acquisitions, and Technical Services* 25(2): 223-233.   Full conference proceedings are available in *Proceedings of the Bicentennial Conference on Bibliographic Control for the New Millennium* (Washington, DC: Library of Congress, Cataloging Distribution Service, 2001).

[2] Marcum, Deanna B. 2005. "The future of cataloging." Presented at the EBSCO Leadership Seminar, January 16, 2005, Boston MA. Available: http://www.loc.gov/library/reports/CatalogingSpeech.pdf

[3] Library of Congress. 2001. "Bibliographic Control of Web Resources: A Library of Congress Action Plan." Available: http://www.loc.gov/catdir/bibcontrol/actionplan.html

[4] Levitt, Theodore. 1965. "EXPLOIT the product life cycle." *Harvard Business Review* 43(6): 81-94.

[5] Harrigan, Kathryn Rudie, and Michael E. Porter. 1996. "End-game strategies for declining industries." In Porter, Michael E. *On Competition* (Boston MA: Harvard Business Review), p. 99-116.

[6] North Carolina State University. "NCSU libraries unveils revolutionary, Endeca-powered online catalog" [press release] January 12, 2006. Available: http://www.ncsu.edu/news/press_releases/06_01/007.htm. The catalog is located at http://www.lib.ncsu.edu/catalog/. Information about Endeca ProFind™ is available at http://endeca.com/solutions/enterprise_search.html.

[7] California Library Association. "Search California: discover the world" [press release]. California State Library, January 24, 2006. Available: http://www.cla-net.org/weblog/mt/archives/cat_californialibrarynews.php. See also the catalog itself: http://www.calcat.org/

[8] The Association of Research Libraries "New Measures Initiatives" respond to the rising need for libraries to shift to outcome-based measures, rather than mere size of collections or number of dollars spent, as determinants of excellence. A list of projects may be found at http://www.arl.org/stats/newmeas/index.html

[9] Atkinson, Richard C. "A new world of scholarly communication." *Chronicle of Higher Education* November 7, 2003. Available: http://chronicle.com/free/v50/i11/11b01601.htm

[10] The University of Rochester library has conducted notable work-practice studies to understand what users want and need. See for example Foster, Nancy Fried and Susan Gibbons. 2005. "Understanding faculty to improve content recruitment for institutional repositories." *D-Lib Magazine* 11 (1). Available: http://www.dlib.org/dlib/january05/foster/01foster.html

[11] University of California Libraries. Bibliographic Services Task Force. 2005. "Rethinking how we provide bibliographic services for the University of California." Available: http://libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf The UC report is an excellent example of analyzing and defining needed changes to the catalog. The UC team was doing their research and writing up their results at the same time as the author of this report. The UC report became available to the author in January 2006 just as she was completing her first draft.

[12] LC's recent work to define "access level" records for certain categories of materials may serve as a starting point. For example, see the brief report on an access level record for serials at http://www.loc.gov/acq/conser/Access-level-chargelc-pccaug17.pdf

[13] Chandler and LeBlanc describe an interesting experiment using classification to create subject browse categories. See Chandler, Adam and Jim LeBlanc. 2005. "Exploring the potential of a virtual undergraduate library collection based on the Hierarchical Interface to LC Classification (HILCC)." [preprint] Available: http://dspace.library.cornell.edu/handle/1813/2223

[14] Gross and Taylor made the case in their 2005 article that more than a third of records retrieved by keyword searches would be lost if subject headings were not present. See Gross, Tina and Arlene G. Taylor. 2005. "What have we got to lose? The effect of controlled vocabulary on keyword searching results." *College & Research Libraries* 66 (3): 212-30. As the UC report points out, automated enriched metadata such as TOCs can supply additional keywords for searching. For further research on the value of summary and contents data see Drabenstott, Karen Markey, and Karen Calhoun. 1987. "Unique words contributed by MARC records with summary and/or contents notes." In *ASIS '87* (Learned Information), p. 153-62.

[15] An excellent place to begin such an investigation is Farrell, Katharine Treptow and Marc Truitt. 2003. "The case for acquisitions standards in the integrated library system." *Library Collections, Acquisitions & Technical Services* 27: 483-92.

[16] For an introduction to the topic see Lakos, Amos, and Shelley Phipps. 2004. "Creating a culture of assessment: a catalyst for organizational change." *Portal: Libraries and the Academy* 4 (3): 345-61. For an overview of data-driven management in libraries see Nicholson, Scott. 2006. "Proof in the pattern." *Library Journal* (January 15). Available: http://www.libraryjournal.com/article/CA6298562.html

[17] Bridges, William. 2003. *Managing Transitions: Making the Most of Change* (Cambridge MA: Perseus).

[18] Association of Research Libraries. 2005. "ARL Statistics, 2003-04." (Washington DC: ARL), p. 5. Available: http://www.arl.org/stats/pubpdf/arlstat04.pdf

[19] OCLC. 2006. "Bibliographic records by source of cataloging." Available: http://www.oclc.org/worldcat/statistics/source.htm

[20] OCLC. "Bibliographic records by source …" Records for books and serials make up nearly 89% of the close to sixty million record WorldCat database. As for records for digital materials, Lynn Connaway, Brian Lavoie and Ed O'Neill estimated that digital resources made up 1.4% of the WorldCat database in July 2004. These digital resources were chiefly the kinds of resources that libraries have generally collected--books, government documents, serials, and computer files. See "Mining for digital resources: identifying and characterizing digital materials in WorldCat." In *Currents and convergence: navigating the rivers of change: proceedings of the ACRL 12th National Conference, April 7-10, 2005, Minneapolis MN.*

[21] See for example the discussion of catalogs and bibliographies in Calhoun, Karen. 2000. "Redesign of library workflows: experimental models for electronic resources description." Prepared for the LC Bicentennial Conference on Bibliographic Control for the New Millennium, Nov. 15-17, 2000. Available: http://www.loc.gov/catdir/bibcontrol/calhoun_paper.html

[22] INFOMINE : Scholarly Internet Resources Collections. Available: http://infomine.ucr.edu/

[23] Lossau, Norbert. 2004. "Search engine technology and digital libraries: libraries need to discover the academic internet." *D-Lib Magazine* 10 (6). Available: http://www.dlib.org/dlib/june04/lossau/06lossau.html

[24] Van de Sompel, Herbert, et al. 2004. "Rethinking scholarly communication: building the system that scholars deserve." *D-Lib Magazine* 10 (9). Available: http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html

[25] McLean, Neil and Clifford Lynch. 2004. "Interoperability between library information services and learning environments: bridging the gaps." Joint white paper on behalf of the IMS Global Learning Consortium and the Coalition for Networked Information. Available: http://www.imsglobal.org/digitalrepositories/CNIandIMS_2004.pdf

[26] Dempsey, Lorcan, Eric Childress, Carol Jean Godby, et al. 2005. "Metadata switch: thinking about some metadata management and knowledge organization issues in the changing research and learning landscape." In *E-Scholarship: A LITA Guide*, ed. Debra Shapiro. (Chicago IL: ALA)

[27] Case, Donald O. 2002. *Looking for information: a survey of research on information seeking, needs, and behavior* (San Diego CA: Academic Press), p. 138.

[28] Case, p. 140.

[29] Bates, Marcia J. 2003. "Improving user access to library catalog and portal information: final report." Prepared for the Library of Congress. For information on the Principle of Least Effort, see section 2A, "General information seeking behavior." Available: http://www.loc.gov/catdir/bibcontrol/2.3BatesReport6-03.doc.pdf

[30] Case, p. 6.

[31] See for example Holtzblatt, Karen, and Hugh Beyer. 1996. "Contextual design: using customer work models to drive systems design." In *Conference companion on human factors in computing systems* (New York NY: ACM Press), p. 373-374.

[32] Byrum, John D. 2005. "Recommendations for urgently needed improvement of OPAC and the role of the National Bibliographic Agency in achieving it." Paper prepared for the 71st IFLA General Conference and Council, Oslo, Norway, August 14-18. Available: http://www.ifla.org/IV/ifla71/papers/124e-Byrum.pdf

[33] Medeiros, Norm. 1999. "Driving with eyes closed: the perils of traditional catalogs and cataloging in the Internet age." *Library Computing* 18 (4): 300-305.

[34] Flecker, Dale. 2005. "OPACs and our changing environment: observations, hopes, and fears." [PowerPoint file] Presented at the PCC Participants Meeting, Boston, MA, January 16. Available from: http://www.loc.gov/catdir/pcc/archive/pccpart05m.html

[35] Hildreth, Charles R. "Online catalog design models: are we moving in the right direction?" 1995. Report prepared for the Council on Library Resources. The summary of user problems is in section 3.4, and the functional improvements are in section 6. Available: http://www.ou.edu/faculty/H/Charles.R.Hildreth/clr-opac.html

[36] Yu, Holly and Margo Young. 2004. "The impact of Web search engines on subject searching in OPAC." *Information Technology and Libraries* 23 (4), p. 168.

[37] Yu and Young, p. 176-178.

[38] Levitt, Theodore. "EXPLOIT the product life cycle."

[39] Lippincott, Sarah. 2004. "How ARL university communities access information: highlights from LibQual+™." *ARL Bimonthly Report* 236. Available: http://www.arl.org/newsltr/236/lqaccess.html

[40] De Rosa, Cathy, et al. 2005. *Perceptions of libraries and information resources* (Dublin OH: OCLC Online Computer Library Center), p. 6-5. Available: http://www.oclc.org/reports/2005perceptions.htm

[41] OCLC Library and Information Center. 2003. "Five year information format trends." Available: http://www5.oclc.org/downloads/community/informationtrends.pdf

[42] "About Google Book Search." 2006. Available: http://books.google.com/intl/en/googlebooks/about.html

[43] Carnegie Mellon University Libraries. 2006. "Frequently asked questions about the Million Book Project." Available: http://www.library.cmu.edu/Libraries/MBP_FAQ.html

[44] Open Content Alliance. 2006. "FAQ." Available: http://www.opencontentalliance.org/faq.html

[45] European Commission. 2005. "2010: digital libraries: communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions." Available: http://europa.eu.int/information_society/activities/digital_libraries/doc/communication/en_comm_digital_libraries.pdf

[46] Lavoie, Brian, Lynn S. Connaway, and Lorcan Dempsey. 2005. "Anatomy of aggregate collections: the example of Google Print for Libraries." *D-Lib Magazine* 11 (9). See the section entitled "The system-wide print book collection." Available: http://www.dlib.org/dlib/september05/lavoie/09lavoie.html

[47] Kirtas Technologies, Inc. claims a capture rate of 2400 pages an hour for its APT 2400 robotic scanner; see http://www.kirtas-tech.com/apt_2400.asp. In 2004, Stanford University Libraries' reported the robotic scanner they used for mass digitization of bound print materials could produce high quality digital images at throughput rates as high as 1160 pages an hour. See http://www-sul.stanford.edu/depts/dlp/bookscanning/

[48] ARL. 2005. "Yearly Increases in Average: Electronic Resources vs. Total Materials Expenditures, 1993-2004." Available: http://www.arl.org/stats/arlstat/graphs/2004/ematbar.pdf. In 2004 ARL libraries spent an average of 32% of their materials budgets on e-resources. A query of the 2004 "ARL Statistics, Interactive Edition" (http://fisher.lib.virginia.edu/arl/index.html) indicates ARLs spent a little over one billion dollars on library materials in 2004. Subtracting $335 million (32%) of this for e-resources leaves a crude estimate of $665 million for everything else.

[49] "ARL Statistics, Interactive Edition." This is a crude estimate. In 2004 ARL libraries spent $1.2 billion on salaries and wages. A recent informal survey conducted by the author of the members of the ALCTS Technical Services Directors of Large Research Libraries Discussion Group suggests that about 20% of library staff work in technical services. Assuming that the percentage of technical services staff can be taken as a proxy for labor expenditures, then an estimated $239 million (20% of the total) was expended on technical services in ARL libraries in 2004.

[50] "NCSU libraries unveils revolutionary, Endeca-powered online catalog" [press release]

[51] Tennant, Roy. 2004. "A bibliographic metadata infrastructure for the 21st century." *Library Hi Tech* 22 (2): 175-81.

[52] National Information Standards Organization. "NISO MetaSearch Initiative." Available: http://www.niso.org/committees/MS_initiative.html

[53] Breeding, Marshall. 2005. "Plotting a new course of metasearch." *Computers in Libraries* 25 (2): 27-9.

[54] Luther, Judy. 2003. "Trumping Google? Metasearching's promise. *Library Journal* (October 1). Available: http://www.libraryjournal.com/index.asp?layout=article&articleid=CA322627

[55] Research Libraries Group. 2006. "Metasearch survey report." Available: http://www.rlg.org/en/page.php?Page_ID=20750

[56] "Google increases lead over Yahoo in search." January 9, 2006. *ComputerWorld.* Available: http://www.computerworld.com/developmenttopics/websitemgmt/story/0,10801,107591,00.html The statistics reported in the news story are from comScore Media Metrix Search Engine Ratings, available from http://searchenginewatch.com/reports/

[57] At the time of this writing, information about Google Book Search was available at http://books.google.com/intl/en/googlebooks/about.html. Information about the Library project could be found at http://books.google.com/googlebooks/library.html

[58] De Rosa/OCLC report *Perceptions of libraries* … p. 1-17, A-14.

[59] "Open WorldCat." 2005. Available: http://www.oclc.org/worldcat/open/default.htm

[60] "Google Book Search: not so easy to find the library link." 2005. *Library Journal* (December 12). Available: http://www.libraryjournal.com/article/CA6290431.html

[61] Salesforce.com is associated with "on demand" computing and Web services (http://www.salesforce.com/). On-demand computing refers to innovative outsourcing services that allow businesses to save the costs of owning and maintaining their own IT infrastructures. Firms pay only for what they use, not a fixed fee. Salesforce.com takes the trend further by offering Web services rather than providing software that firms install on their own computers. The *Wall Street Journal* ran stories about Salesforce.com and on-demand computing on January 17, 2006 and March 22, 2004.

[62] Hakala, Juha. 2004. "Resource description in a heterogeneous system environment." *Library Collections, Acquisitions, and Technical Services* 28 (3): 275-82.

[63] Dietz, Roland, and Carl Grant. 2005. "The dis-integrating world of library automation." *Library Journal* (June 15). Available: http://libraryjournal.com/article/CA606392.html

[64] Anderson, Chris. 2004. "The long tail." *Wired Magazine* 12 (10). Available: http://www.wired.com/wired/archive/12.10/tail.html

[65] "Q & A with Chris Anderson." 2005. *OCLC Newsletter* 268. Available: http://www.oclc.org/news/publications/newsletters/oclc/2005/268/interview.htm

[66] "Mining the Long Tail: Libraries, Amazoogle and Infinite Availability. " OCLC symposium at the 2005 ALA annual conference. Streaming video is available by clicking on the link to the 2005 symposium at http://www.oclc.org/education/conferences/presentations/default.htm#launch

[67] National Information Standards Organization (NISO). 2005. "VIEWS Initiative." Available: http://www.niso.org/committees/VIEWS/VIEWS-info.html

[68] NISO Web Services Working Group. 2006. Wiki: Draft Document Outline. Available: http://research.talis.com/2005/wswg/wiki/Main/DraftDocumentOutline

[69] Burke, Timothy. "Burn the Catalog" [Weblog entry]. *Easily Distracted*. Swarthmore College. January 20, 2004. Available: http://www.swarthmore.edu/SocSci/tburke1/perma12004.html

[70] Tennant, Roy. 2004. "Enriching the catalog." *Library Journal* (June 15). Available: http://www.libraryjournal.com/article/CA423795.html

[71] Bennett, Rick, Brian Lavoie, and Edward T. O'Neill. 2003. "The concept of a work in WorldCat: an application of FRBR." *Library Collections, Acquisitions, and Technical Services* 27 (1): 45-59.

[72] Madison, Olivia M.A. 2005. "The Functional Requirements for Bibliographic Records: a tool for defining library services." Presented at the EBSCO Leadership Seminar, January 16, 2005, Boston MA. Available by request from http://www.ebsco.com/home/printsubs/vantagepoint.asp

[73] Bowen, Jennifer. 2005. "FRBR: coming soon to your library?" *Library Resources and Technical Services* 49 (3): 175-188.

[74] Tillett, Barbara B. 2005. "FRBR and cataloging for the future." *Cataloging and Classification Quarterly* 39 (3/4): 197-205.

[75] Jones, Ed. 2005. "The FRBR model as applied to continuing resources." *Library Resources and Technical Services* 49 (4): 227-242.

[76] RedLightGreen is available at http://www.redlightgreen.com. A description and additional information is at http://www.rlg.org/en/page.php?Page_ID=433. RLG's achievement is considered by many an important breakthrough in the usability of catalogs and the first practical and large scale, production application of FRBR.

[77] Curioser is a project of the OCLC Office of Research; information is available at http://www.oclc.org/research/projects/curiouser/default.htm. The project deploys techniques for displaying records in a FRBR context and for the purpose of enhancing the utility of Open WorldCat.

[78] "For your information: Stanford Grokker." 2005. Available: http://www-sul.stanford.edu/about_sulair/special_projects/stanford_grokker_faqs.html

[79] Dushay, Naomi. 2004. "Visualizing bibliographic metadata: a virtual (book) spine viewer." D-*Lib Magazine* 10 (10). Available: http://www.dlib.org/dlib/october04/dushay/10dushay.html

[80] WikiD is a project of the OCLC Office of Research; information is available at http://www.oclc.org/research/projects/wikid/default.htm. See also "The Wiki world of WorldCat." 2005. *OCLC Newsletter* 268. Available: http://www.oclc.org/news/publications/newsletters/oclc/2005/268/labs.htm

[81] Dempsey, Lorcan. "A palindromic ILS service layer" [Weblog entry]. *Lorcan Dempsey's Weblog*. OCLC Office of Research. January 20, 2006. Available: http://orweblog.oclc.org/archives/000927.html

[82] Band, Jonathan. 2005. "The Google Print Library Project: a copyright analysis." *ARL Bimonthly Report* 242. Available: http://www.arl.org/newsltr/242/google.html

[83] Library of Congress. 2006. "Section 108 Study Group." Available: http://www.loc.gov/section108

[84] University of California Libraries. Bibliographic Services Task Force. 2005. "Rethinking …"