Notes on Operations
Shared Resources, Shared Records

Letting Go of Local Metadata Hosting within a Consortium Environment

Charles Pennell, Natalie Sommerville, and Derek A. Rodriguez

Charles Pennell (cpennell@ncsu.edu) is Principal Cataloger, North Carolina State University; **Natalie Sommerville** (natalie.sommerville@duke.edu) is Head, Monographic Cataloging Section, Duke University Libraries; and **Derek A. Rodriguez** (drodriguez@ searchtechnologies.com) is Technical Consultant, Search Technologies, Herndon, Virginia.

Many libraries share regional, statewide, or even national union or consortium catalogs to enable consolidated search and display of participant holdings. These catalogs typically duplicate search capabilities provided by individual libraries' local catalogs. Search TRLN is a discovery layer built to support both group and individual library catalog interfaces for the four member institutions of the Triangle Research Libraries Network. In 2010, the Shared Records Pilot Task Group extended this shared catalog concept to the individual bibliographic record level. In this model, individual member libraries assume responsibility for building and maintaining record sets for commonly held electronic collections on behalf of the consortium. Today the program includes more than 220,000 shared records representing 765,000 individual library holdings. This has resulted in considerable savings in staff costs, processing costs, and metadata storage and suggests an evolving model for catalogers as managers, rather than as creators and curators, of metadata. This article discusses the evolution of this project, the development of staff trust necessary to let go of proprietary metadata, and the systems logic needed for implementation. The article closes with criteria for assessing the success of the program, including improvements in catalog display, throughput and timeliness, time savings, and elimination of duplicated maintenance activities.

In 2007 the four member institutions of the Triangle Research Libraries Network (TRLN) implemented a shared platform for discovery called Search TRLN (http://search.trln.org). Based on groundbreaking work performed by the North Carolina State University Libraries, TRLN used Oracle's Endeca Guided Search enterprise search application to support discovery and delivery services across the consortium's collections of more than 16 million volumes. Library patrons immediately took advantage of Search TRLN, and resource sharing within the consortium increased 70 percent in the first year after implementation. While member libraries immediately recognized the public service advantages of the new shared search platform, it took somewhat longer to recognize, and indeed to accept, the technical services advantages that might be gained through shared effort. The Search TRLN project exposed many cataloging processes, practices, and expenditures that were duplicated two, three, and even four times across the consortium's campuses and integrated library systems (ILSs). This article describes TRLN's Shared Records Program

(www.trln.org/endeca/shared-records .html) which leverages the Search TRLN system to share cataloging expertise and reduce duplicate cataloging activities within the consortium.

The Search TRLN Project

TRLN is a consortium of academic libraries in North Carolina. TRLN's history can be traced to the 1930s when the libraries at Duke University (Duke) and the University of North Carolina at Chapel Hill (UNC-Chapel Hill) began cooperative collection development activities and shared use of library collections. North Carolina State University (NCSU) entered into these cooperative agreements in 1955. In June 1980, the consortium formally adopted Triangle Research Libraries Network as its name. North Carolina Central University (NCCU) joined the consortium in 1995. The TRLN libraries currently collaborate in the areas of collection development, joint licensing of electronic resources, reciprocal borrowing and document delivery services, library automation, digital preservation, collaborative print retention, and various human resources initiatives.

Cooperative approaches to library automation have deep roots at TRLN. Beginning in 1976, the consortium provided early leadership in the development of shared online systems for maintaining bibliographic records and holdings for library collections. When the system known as the Bibliographic Information System (BIS) came online in 1985/1986, it was the earliest example of an online library catalog providing federated search across multiple databases of library holdings.[1] In 1993, the TRLN libraries ceased development of BIS and adopted vendor-provided ILSs and online catalogs.

Local innovation in this area resumed in 2006, when the NCSU Libraries implemented the first faceted library catalog, based on a commercial search engine provided by Endeca.[2] NCSU's next-generation catalog harvested MARC and item records from its local SirsiDynix ILS for indexing. No longer tethered to the data structures and indexes within the static framework of the ILS, the NCSU Libraries' Endeca catalog provided patrons with a much richer discovery experience than traditional library vendor-provided catalogs. NCSU's Endeca catalog inspired rapid development of "next generation catalogs" throughout the industry, forever changing patrons' expectations of library search.

The idea of federated searching across the holdings of all four TRLN collections was revisited in the Search TRLN project initiated in 2007. Led by a steering committee and several task groups, the project's goals were to provide Endeca-driven search capabilities across all of the consortium's holdings to facilitate discovery and delivery of library materials.[3] In this implementation, the Search TRLN system harvests MARC and item records from the ILSs of all four institutions and generates a single shared index. By March 2008, library patrons were searching the holdings of the entire consortium from a single web interface called Search TRLN. By 2009, TRLN's Endeca implementation supported locally scoped Endeca-based catalogs for all four institutions and indexed metadata in a wide variety of formats and schemas, including MARC, MARC/XML, Encoded Archival Description (EAD), Data Documentation Initiative (DDI/XML), and Dublin Core (DC) (see figure 1).

Union catalogs are typically created by merging bibliographic records for identical titles during metadata preparation and ingest. For instance, Coyle reported on the methods used to merge records during the implementation of the University of California Melvyl system; more recently, this practice has been used by the HathiTrust Digital Library.[4] The Search TRLN project takes a different approach. Instead of merging records during ingest, the Search TRLN system harvests and indexes the entire bibliographic database for each TRLN institution. Campus-specific Endeca catalogs can then scope searches to portions of the index corresponding to each institution's bibliographic database. Records that share common numeric identifiers such as OCLC numbers or Serials Solutions control numbers are merged "on-the-fly" in the consortium catalog, Search TRLN (see figure 2).

The decision to avoid merging records before indexing provided a straightforward method for individual member libraries to implement locally scoped catalogs and likely decreased implementation time for the entire project. This decision, however, laid bare the immense scale of the duplication of catalog records in the TRLN bibliographic databases. As an example, all four institutions independently maintained MARC records for US federal documents in electronic format. Essentially identical records were loaded into four separate ILS databases, sent out for authority processing by each library, then replicated in the Endeca indexes four times. This resulted in redundant and unnecessary staff effort, authority processing expenses, record storage costs, and processing costs. As the TRLN libraries implemented their Endeca-driven catalogs, it became clear that the Search TRLN platform provided an opportunity to share metadata and distribute the costs for its maintenance among the TRLN institutions.

Literature Review

Sharing bibliographic records has been a common interest among libraries for many years, beginning with the distribution of Library of Congress
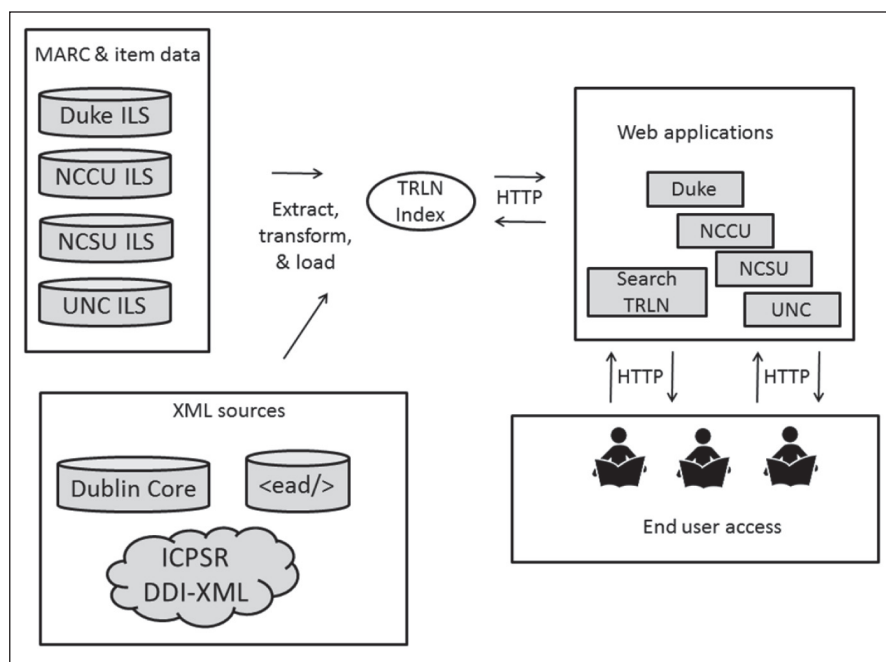
Figure 1. Search TRLN System Architecture

Figure 2. A Merged Print Record Display Rendered from Four Bibliographic Records in Search TRLN

(LC) catalog cards, but it was certainly accelerated by automation efforts that started with the introduction of the MARC record in 1968. OCLC, the Program for Cooperative Cataloging (PCC), its cooperative serials component (CONSER), and its Name and Subject Cooperative Programs (NACO and SACO) can all be viewed as institutionalized means for sharing the effort and cost of building our current bibliographic infrastructure. However, these still hew to the

model of the central metadata store that libraries draw on to populate physically separate catalogs at the local library site. There is very little literature on libraries sharing metadata at the record level.

Shared/Cooperative Cataloging

Shared records projects take many forms throughout the library world. In the United States, consortium members most commonly determine

model cataloging standards and practices (AACR2/RDA, MARC21, LCSH or MeSH subjects, successive or latest entry serials cataloging, separate or merged records describing multiple versions), but they continue to add and maintain institutional records within their own ILS. MARC records are harvested from each consortium member's ILS and merged in a separate centralized database, often maintained by the same ILS vendor individual consortium libraries use. This is the model described by Moeller at Prospector, the Colorado Unified Catalog, and is the model used by several other state consortia, most notably the University of California's California Digital Library (CDL), the Illinois Library Computer Systems Organization (ILCSO), and OhioLink.[5]

Consortia have other record-sharing models available, like TRLN, whose members do not share a common integrated library system. The Virtual Library of Virginia (VIVA) Project allowed individual libraries with disparate ILSs to voluntarily for host, maintain, and distribute particular collections of set records to consortium participants.[6] Unlike the Colorado, California, and Ohio model, which initially encompassed all formats, VIVA focused on electronic resources, which presumably required little local editing other than perhaps customizing the MARC 856 field to provide information for proxy server and display text.

The CDL also created a separate Shared Catalog Program (SCP) for managing e-resource metadata. In the SCP's centralized model, University of California-San Diego library creates and maintains metadata and contributes these records to Melvyl, which then further distributes those records to nine campuses. In a 2002 article, French, Culbertson, and Hsiung delineated several factors that lead to success in shared cataloging projects, including common descriptive

standards, high-quality metadata, timeliness, acceptance of records without local modification, use of holdings records for localized metadata, and good communication.[7]

Batch Loading Records and E-Resources Issues

Batch loading records for large collection sets is a strategy that libraries use to provide access to titles that are beyond the scale of current staffing levels. By relying on vendor metadata, libraries are able to make content discoverable much quicker than would be possible through manual copy cataloging procedures. Timeliness has its tradeoffs, however, including the potential for poor source metadata and the logistics for keeping current with vendor releases of record sets. Martin and Mundle discuss efforts to enhance the quality of vendor-supplied MARC metadata.[8] They ask readers to consider the staff cost of these efforts and urge libraries and consortia to work with vendors up front to enhance the initial quality of metadata. In his 2009 article on batch loading records for open-access books, Beall also discussed the poor quality of third-party supplied records but described how timeliness of access to 100,000 titles can outweigh poor metadata quality.[9]

Most recently, Mugridge and Edmunds surveyed large academic libraries to assess current practices and issues associated with batch loading MARC records.[10] Timeliness of record loads was an issue identified by a majority of the responding libraries. The authors also found that the eighteen libraries surveyed anticipated an increase in the importance of batch loading in the next five years, as long as the ILS continues to be identified as the database of record for a library's holdings. Further, they identified the use of discovery-layer software as a factor that may affect batch loading workflows.

Measuring the Effect of Record Sharing

Stalberg and Cronin reported on the 2009 work of the "Task Force on Cost/Value Assessment of Bibliographic Control."[11] The task group was charged by the Association for Library Collections and Technical Services (ALCTS) to "develop and articulate metrics for evaluating the cost and value of cataloging activities."[12] Following an extensive review of the cost/benefit literature of cataloging, they identified seven operational definitions of value including discovery success, use of collections, improvements in display, interoperability of library bibliographic metadata, support for Functional Requirements for Bibliographic Records (FRBR) user tasks, throughput and timeliness, and support for library administrative goals.[13]

Stalberg and Cronin proposed that "the extent to which data-creation processes facilitate timeliness in resource availability" can be used as a measure of value.[14] Furthermore, they argue that resources that are uncataloged are undiscoverable, and library patrons cannot use undiscoverable resources. These observations are consistent with studies indicating a negative correlation between cataloging backlogs and circulation of print materials.[15] Howarth, Moore, and Sze identified a major cause for cataloging backlogs to be a mismatch in the quantity of cataloging work to be done and the capacity to complete it.[16] They provided several suggestions to reduce backlogs including optimizing workflows, reallocating staff, using automated processes, and outsourcing cataloging tasks.

Writing in 2004, Fischer, Lugg, and Boese provided a ten-point checklist of business practices for reducing backlogs of print materials to release staff time for describing electronic resources.[17] Though Fischer and her colleagues focused on cataloging backlogs of print materials, their

recommendations to "control the Expert Mentality," "automate and outsource where possible," and "trust vendor-provided metadata" remain relevant in the context of cataloging electronic resources. Fischer, Lugg, and Boese noted that the expert mentality results in overly complex and often-manual procedures to solve problems in cataloging.[18] Stalberg and Cronin echoed this sentiment when stating that "time spent on low-value activities (no matter which operational definition is used for 'value') is time not spent on high-value activities."[19]

Stalberg and Cronin isolated several costs associated with managing bibliographic metadata including staff salaries, tools and systems, and database maintenance, which are inherent to any evaluation of library work processes.[20] Efforts to increase cataloging efficiency and timeliness should therefore be judged by their effect on controlling if not reducing the costs of cataloging and releasing expert staff to work on more complex issues or deferred processing projects.

Making the Case for Record Sharing

Challenges with providing access to electronic resources provided the impetus to pursue this record sharing initiative at TRLN. In particular, the practice of displaying institution-specific information accompanying URLs for open access titles in Search TRLN was confusing and misleading to patrons. As these titles were freely available to all users, it struck TRLN's Technology Council as counterproductive to have these appear in the catalog with institutional identifying information, discouraging any but patrons of the loading agency from using the metadata.[21] This information included proxy-server URLs, restrictive notes (e.g., "Available to NCSU users only"), and inconsistent URLs. The Technology Council charged the Electronic

Resources Access Restrictions Task Group in 2009 to examine the display of e-resource links in Search TRLN.[22] While this group did make recommendations for clearer link displays, the Technology Council was particularly engaged with the Task Group's final recommendation for TRLN member libraries to consider sharing records, especially for large, commonly held collection set titles. With that in mind, the Shared Records Pilot Task Group was charged in 2010.[23]

The Shared Records Pilot Task Group began meeting just as TRLN was considering the purchase of additional storage space to accommodate growth in the number of records contained in the consortium catalog. The task group conducted an inventory of electronic resources collections held by institutions within the consortium and found two or more institutions subscribed to several large collections such as the Early American Imprints (Evans), Early English Books Online (EEBO), and Eighteenth Century Collections Online (ECCO). These records, not to mention the possibility of loading records from the Open Content Alliance (OCA), Google Books, and HathiTrust, when multiplied by just two or three institutional subscribers, could place a significant burden on shared storage space. Further, these shared sets represented unnecessarily duplicated staff time and expenses for purchase, record loads, and maintenance. The Search TRLN platform provided an opportunity to reduce this duplication and distribute metadata maintenance costs across the consortium.

The Shared Records Task Group developed a model, described below, to enable record sharing and recommended conducting a pilot project using record sets for three collections of electronic resources. Those collections were the NC LIVE Video collection of streaming videos from PBS, the Marcive "Documents Without Shelves" collection of online US federal documents, and the Inter-university Consortium for Political and Social Research (ICPSR) dataset metadata. The Task Group's recommendations were accepted in 2010, and work soon began to make them a reality.

The Shared Records Model and Workflow

The Shared Records Pilot Task Group defined a Shared Records model including a mutually acceptable set of rules and expectations to guide the program. This was easily achieved within the task group, which was made up of technical service representatives from each campus along with a TRLN representative. It was a bit more difficult to achieve buy-in at the campus level, at least initially, for reasons that are discussed in detail below.

In the TRLN Shared Records Program, a single institution volunteers to assume responsibility for maintaining metadata for a given record set in a local ILS or other metadata repository. Those metadata records are harvested for indexing in the Search TRLN system and shared virtually with partner libraries through that system. As of November 2012, five record sets for electronic resources have been added to the Shared Records program: The NC LIVE Video collection, Marcive's Documents Without Shelves (DWS), ICSPR Codebooks (ICPSR), EEBO, and records for e-books from the Oxford University Press Scholarship Online (UPSO) platform.

The task group defined characteristics to determine whether a record set was eligible for the program. A worthy candidate for the Shared Records program would be a collection held by two or more member libraries for which the consortium has access to acceptable bibliographic records or updates available in appropriate metadata formats such as MARC, XML, .txt, or a fielded database or spreadsheet.

First, record license rights for consortium use of metadata must be secured from the publisher or metadata provider load. The TRLN libraries gained access to the NC LIVE Video collection metadata through their membership in NC LIVE, North Carolina's statewide online library. Duke, NCSU, and UNC each licensed the EEBO metadata from Chadwyck-Healey independently. TRLN secured a consortium license to share the DWS records and the UPSO records were provided as part of TRLN subscription to the UPSO e-book collection. Second, descriptive cataloging standards for any set should be agreeable to all sharing institutions. These standards may be less than full AACR2 or RDA if agreed on by all participants. Finally, URLs in the metadata should be easily made institution-neutral through minor editing such as removing local proxy server prefixes and local use notes.

The size of a record set, possible savings in processing costs, and opportunities to standardize procedures and workflows were also criteria for determining eligibility. For instance, the EEBO record set was considered a viable candidate because of the large number of records involved (123,521) and the high expense of performing authority control at three institutions. Further, the use of non-unique control numbers in the EEBO source records had already generated numerous duplicate records in local catalogs, which were then carried over to Search TRLN.

The task group defined a set of responsibilities for institutions serving as record hosts. The host institution for a given record set is expected to take responsibility for maintenance of record sets in its ILS. In some cases, stewardship responsibilities have grown out of existing commitments. For instance, UNC-Chapel Hill, as the Regional Depository Library for North Carolina, was the logical candidate to maintain the consortium's DWS

Table 1. Shared Record Set Attributes

	NC LIVE Videos	MARCIVE's Documents Without Shelves	ICPSR	EEBO	UPSO E-Books
Format	MARC	MARC	DDI/XML	MARC	MARC
Shared by	4 institutions	4 institutions	3 institutions	3 institutions	4 institutions
Host institution	NCSU	UNC-Chapel Hill	ingest into Endeca directly from source	NCSU	Duke
Host institution provides authority control?	yes	yes	no	yes	yes
Set holdings at OCLC?	no	yes	no	no	no
E-resource access restrictions	IP-restricted	Open Access	Mixed open access and restricted content	IP-restricted	IP-restricted

subscription. In this case, TRLN served as purchasing agent and invoiced member libraries for their share of the costs. NCSU was an obvious choice for the NC LIVE and EEBO record sets since NCSU was already creating metadata for streaming videos on behalf of NC LIVE and had been maintaining the EEBO record set for many years. Duke recently took responsibility for maintaining the UPSO e-book record set. As described in more detail below, the ICPSR are not maintained locally at all; instead they are harvested directly from the ICPSR server and ingested into Endeca directly.

Host institutions also have responsibilities to maintain URLs for electronic resources and provide authority control of name and subject access points through local or vendor processes. Where appropriate the host institution should also set holdings at OCLC on behalf of the consortium. For instance, UNC-Chapel Hill uses an OCLC group profile and associated batch update services to set holdings at OCLC for the DWS records on behalf of all four institutions.

The task group also planned for a future possibility when the Shared Records Program may come to a close or when an individual TRLN institution might choose to withdraw and migrate to a new discovery platform.

In these cases, host institutions are expected to be able to supply a current version of set records in an appropriate communication format (e.g., MARC) if needed by another member library for migration to a new discovery system.

Preparing Shared Records for Search TRLN

The TRLN Shared Records Task Group devised two models for managing shared metadata: a hosted model with a TRLN member library hosting records in their local ILS and a direct load model where metadata are harvested directly from an external source.

Hosted Model

In the hosted model, one TRLN institution assumes responsibility for hosting and maintaining metadata for a commonly held collection in their local ILS, and the host library shares the metadata with partner institutions through the Search TRLN indexes. Non-host institutions may also choose to maintain records for the set locally for reasons internal to that institution, but their records will be prevented from loading into Search TRLN through ingest filters in the metadata pipeline. In most cases, these record

sets were removed from the local ILS of all but the responsible host library, although some federal document e-journals are still managed in serials knowledge bases.

Managing metadata in the hosted model includes three activities: local metadata preparation, indexing, and supporting display for library patrons. Some of these tasks are conducted by the host institution; others are managed by consortium staff.

Local Metadata Preparation

The first step involves loading records into the host library's ILS using typical batch loading processes. A good example is the Marcive Documents Without Shelves (DWS) record set acquired through a consortium license in 2011. UNC-Chapel Hill serves as the host institution for this collection. Each month, a UNC-Chapel Hill staff member downloads the monthly notification service file from the vendor and loads the records into their Innovative Millennium ILS. UNC-Chapel Hill takes responsibility for three maintenance tasks that were formerly handled by staff at each TRLN campus, including authority processing with an external vendor, setting holdings at OCLC for all four institutions using an OCLC group profile, and
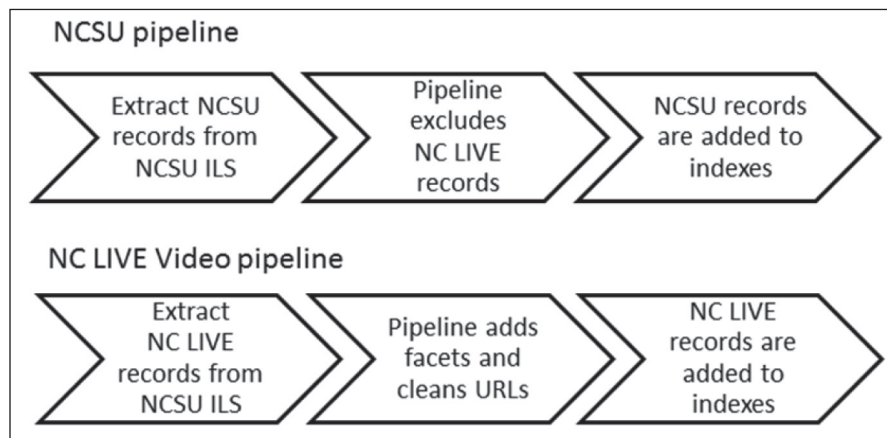
Figure 3. Pipelines for Preparing Shared Records for the Endeca Indexes

correcting URLs as needed.

Host institutions are responsible for adding identifying fields to each record in a given set so they can be isolated in the local ILS for global operations or extract. For instance, UNC-Chapel Hill adds a MARC 919 field including the text string "dwsgpo" to each of the GPO DWS records loaded into its ILS. UNC-Chapel Hill technical services staff can use this field to isolate these records for extract, for batch editing, and for archiving purposes. NCSU uses Sirsi-Dynix Symphony's "Item Cat2" to identify EEBO and shared open-access records. Similar procedures are used to identify the other hosted record sets managed in the Shared Records program.

As noted above, each institution uses automated processes to send metadata extracts of their MARC, EAD, and DC records to the Search TRLN system for indexing. Similar processes are used to provide sets of shared records with Search TRLN for indexing. Again drawing on the DWS record set as an example, UNC-Chapel Hill provides regular maintenance of this dataset including URL corrections, monthly DWS bibliographic updates, and authority processing. UNC-Chapel Hill then extracts all of the DWS MARC records and sends them to the Search TRLN servers for indexing on a weekly basis.

Indexing

At this point, processing is handed off to the Endeca applications, called pipelines, which prepare metadata for the indexes. The pipelines make several changes to each Shared Record set to support expected functionality in the user interfaces.

First, e-resource URLs must be made institution-neutral in the indexes. This typically involves removing a proxy server string from the URL as found in the extracted records. For instance, the record set for the NC LIVE Videos is hosted and maintained by NCSU and the record 856 field (Electronic Location and Access) for these records contains NCSU's proxy server string, http://proxying.lib.ncsu.edu. The NC LIVE Video pipeline removes the NCSU proxy prefix from each MARC 856 field and stores an institution-neutral URL for each record in the index. Proxy information is restored later for display, if it is appropriate for the set.

Second, several new facet values are added to the records to support needed functionality in the user interfaces. A Shared Records flag is set to "true" for these records so the user interfaces can detect Shared Records and render them properly for end users. Additional facet values are added for each sharing institution. For example, facet values for institution

(Duke, NCCU, NCSU, and UNC-Chapel Hill) and format (Internet resource and streaming video) are added to each of the records in the NC LIVE Video record set. An Access facet is also used to indicate whether or not the user interfaces should render the records as open access or IP-restricted resources. Additional logic removes these records from the host institution's main pipeline, eliminating the possibility of creating duplicate records in the indexes.

Supporting Display for Library Patrons

The user interface code that drives Search TRLN and the local catalogs of each consortium member library needed minor modifications to display shared records. The most significant change was to restore the institution-specific URLs to support off-campus authentication for IP-restricted resources. UNC-Chapel Hill, Duke, and NCSU all use EZproxy (www.oclc.org/ezproxy). to provide remote access to IP-restricted resources. So in these cases, institution-specific proxy URLs are prepended to the institution-neutral URLs on-the-fly in the user interfaces. NCCU provides off-campus access to electronic resources through VPN access, so no additional processing is needed to render these URLs properly. The link text for IP-restricted (prepended proxy) and open access resources (no prepended proxy) is adjusted appropriately when records are rendered for patron use as shown in figures 4 and 5.

Direct Load Model

In the direct load model, metadata records are harvested from a vendor or third-party source and loaded directly into the Search TRLN indexes. This model includes three processes: harvesting metadata, indexing, and supporting display for library patrons. Metadata in DDI/XML format from

The Berlin airlift (electronic resource)			
Format:	Internet resource; Online Video		
Published:	(Boston): WGBH Educational Foundation; (Alexandria, VA) : Distributed by PBS Home Video, c2007.		
Location:	Duke	Available	Online (Duke only)
	NC Central	Available	Online (NCCU only)
	NC State	Available	Online (NCSU only)
	UNC Chapel Hill	Available	Online (UNC only)

Figure 4. Shared NC LIVE Record Treated as an IP-Restricted Resource in Search TRLN

100 GPO years, 1861-1961 (electronic resource) : a history of United States public printing	
Author:	United States, Government Printing Office
Format:	Internet resource
Published:	Washington, D.C. : U.S. Govt. Printing Office : For sale by the Supt. of Docs., U.S. G.P.O., 2010
Online Access:	Open Access resource

Figure 5. Shared MARCIVE DWS Record Treated as an Open Access Resource in Search TRLN

ICPSR fall into this category and provide examples for discussion.

Harvesting Metadata

ICPSR generates metadata about its datasets using the Data Documentation Initiative (DDI) metadata specification (www.ddialliance.org) and makes it available on the ICPSR website in XML format. ICPSR currently uses the DDI Codebook 2.1 schema and Document Type Definition (DTD) to structure these documents. Once each week the entirety of the ICPSR XML corpus is downloaded to a TRLN server and prepared for indexing.

Indexing

As in the hosted-record model, a specific Endeca pipeline prepares the DDI/XML for indexing. The first step transforms the codebooks into indexable documents. TRLN adapted an ICPSR-provided Extensible Stylesheet Language (XSL) stylesheet to transform each DDI/XML codebook into a format that could be indexed by Endeca.[24] The remainder of the pipeline adds facets appropriate for these records including the Duke, UNC-Chapel Hill, and NCSU institutional facets and the Statistical Dataset and Internet Resource format facets.

Supporting Display for Library Patrons

As with the hosted-record model, records loaded in the direct-load model also need institution-specific proxy URLs. This was not necessary for the ICPSR metadata because access control for these metadata is managed at the ICPSR website through individual logins associated with licensing institutions.

Benefits to the Direct Load Model

Before implementing the Shared Records Program, Duke, UNC-Chapel Hill, and NCSU independently loaded and maintained MARC records prepared annually by ICPSR. Once per year catalogers at each institution updated the ICPSR MARC records in each respective ILS. The ICPSR portion of the Shared Records Project generated several benefits. The first benefit is timeliness now that the ICPSR metadata are updated in the Search TRLN indexes on a weekly basis. Second, TRLN was able to access the entire codebook for each dataset. This allowed TRLN to index a greater proportion of the metadata about each dataset than could be done when relying on the MARC records. Third, all of the processing for the direct load model is automated and monitored by consortium staff. This allowed Duke, NCSU, and UNC-Chapel Hill to eliminate all ICPSR dataset records from their ILSs and eliminate three formerly duplicated workflows, allowing cataloging staff to address other projects.

Obstacles Overcome

The technical and workflow issues

Ithaka S+R Faculty Survey 2009: Key Strategic Insights for Libraries, Publishers, and Societies			
Format:	Internet resource; Statistical Dataset		
Published:	Ann Arbor, MI: Inter-university Consortium for Political and Social Research 2011		
Location:	Duke	Available	Online (Duke only)
	NC State	Available	Online (NCSU only)
	UNC Chapel Hill	Available	Online (UNC only)

Figure 6. A Record Derived from an ICPSR DDI/XML Codebook as Rendered in Search TRLN

behind the TRLN Shared Records model necessitated careful planning and management across the consortium. At the Duke University Libraries, a shift in staff perspectives was necessary for the initiative to succeed. Cataloging staff had to move from crafting metadata to managing it, and to achieve this transition had to trust external sources of metadata. Paradoxically, this meant giving up local control as the need to expose metadata to users on a large scale increased. New methods for managing e-resource holdings and the adoption of the Endeca-based catalog helped change perspectives and facilitated a wholesale adoption of the shared record model.

New Methods for Managing E-Resources

Before 2010, the Duke University Libraries (DUL) managed e-resource holdings in two different systems, the ILS and a vendor-provided knowledge base. In late 2010, e-resource management functions were largely consolidated into a single knowledge base provided by a new vendor. This brought about a deepening comfort with managing resources and their associated bibliographic and administrative metadata outside of the ILS. The scale of electronic resource holdings made management of these resources through the ILS impossible. At the point of migration to the current knowledge base, the total number of unique electronic resources managed

was 252,000. By the end of the fiscal year the following summer, 544,800 unique titles were being tracked in the knowledge base.[25] By November 2012, 1,059,795 unique titles were being tracked in the knowledge base. In addition to tracking resources, the knowledge base serves as a repository for details about the terms of access and workflow. Even discovery of these resources became mediated through the knowledge base, with the knowledge base provider supplying MARC records for tracked titles. Timeliness of access was also a feature that gave the knowledge base an advantage over the ILS. On the back-end, as soon as the platform and its associated titles could be tracked, metadata about those resources became available for reporting and documenting workflow decisions. More importantly, the gap between library access to a title and its discovery by the public shrank to twenty-four hours for the A–Z list compared to up to two weeks for the catalog. Thus the ILS ceased to be the database of record for electronic resource holdings, and the stage was set for managing discovery of resources via other means.

Declining Importance of Vendor-Provided Online Catalog

At the same time that a proliferation of often-transient electronic resources changed the staff's perspective toward the back-end ILS as a collection management tool, advances in discovery interfaces led to abandoning of the

traditional ILS-based OPAC as Duke's primary discovery tool. By 2010, Endeca stood firmly as the library's catalog. Two years later, DUL implemented a web-scale discovery tool. This meant that metadata sources were no longer confined to MARC records in the ILS, and that catalogers had to develop new, large-scale understandings of metadata and how it fuels discovery.

Perceiving Benefits

Changing perspectives and reconceptualization had created a willingness to test the Shared Record model with the Marcive DWS portion of the project. Actually seeing the model work in practice, while addressing known metadata and workflow needs, was critical for wholesale adoption. Over the course of various staffing and organizational changes at DUL, the workflow for the loading of MARC records for electronic government documents no longer followed Duke's standards and practices for other electronic resources. An ILS migration in 2004 further added to the issues with this workflow, and staff had a metadata cleanup project to implement along with workflow changes. The workflow changes were easier to implement, though they still required staff time and maintenance. Untangling the metadata issues and fitting cleanup among other priorities was more complex. The proposal to share the DWS records between all TRLN libraries came at exactly the right time—just as staff members were working to shift

Table 2. Shared Records Efficiencies

Collection	Titles	Held by	Local Cataloging Workflows Eliminated or Avoided	Record Savings
MARCIVE's Documents Without Shelves	87,143	4 institutions	3	261,429
NC LIVE Videos	428	4 institutions	3	1,284
ICPSR	8,471	3 institutions	3*	25,413**
EEBO	123,521	3 institutions	2	247,042
OUP E-books	1,568	4 institutions	3	4,704
Total	221,131	--	14	539,872

* Moving to the direct load model for ICPSR allowed Duke, NCSU, and UNC-Chapel Hill to eliminate their local cataloging workflows for this collection.

** Authority processing is not conducted for the ICPSR records.

priorities and address the outstanding metadata issues. To confirm the number of records that needed to be excluded through the ingest filters in the metadata pipeline, staff were able to prioritize this metadata cleanup project, gain a deeper understanding of the issues involved, and plan for making the metadata uniform with other metadata for electronic resources. The time saved from ongoing maintenance of DWS record loads allows staff to focus on metadata cleanup and refining the workflow for managing discovery of all US documents. The way in which the shared DWS records addressed such an immediate need hastened the transformation from a willingness to try the shared record model to adoption of it and taking advantage of all its benefits.

A later project to share EEBO records further underscored how beneficial the Shared Records model is to timeliness of access and discovery. The proposal to share EEBO records came at a time when past record loading workflows were transitioning across departments, and the budget for automated authority control was being examined and restructured. Once again, Duke received the benefits of a shared records project supporting timely discovery that also allowed for a restructuring of workflows and budgets. By the time the EEBO project

was complete, shared records became an ingrained part of workflow planning for facilitating timely discovery of electronic resources. The most recent shared records project, which facilitates discovery of consortium-held Oxford University Press/University Press Scholarship Online eBooks, is managed at DUL and provides a concrete example in which changes in perspective transformed local perceptions of the ILS and management of access and discovery.

Effect

As of September 2012, over 220,000 titles in six collections were managed in the TRLN Shared Records program. The effect of the program can be measured in terms of throughput and timeliness (making metadata discoverable faster), saving time and reallocating effort by eliminating duplicate technical services workflows, and financial benefits through reduced licensing costs and the sharing of authority control costs.

Throughput and Timeliness

As previously noted, Stalberg and Cronin identify "the extent to which data-creation processes facilitate timeliness in resource availability" as a measure

of metadata value.[26] Since the inception of the Shared Records project, TRLN has observed improvements in metadata timeliness for several collections. The ICPSR and OUP E-book projects provide good examples. The ICPSR metadata are currently updated in our indexes weekly. Before implementation, these records were updated once per year. The schedule for metadata processing of OUP e-books at DUL is driven by objectives of the broader e-book pilot, decreasing the time between e-resource availability and discoverability. The automation of all metadata harvesting and indexing processes also improves timeliness of metadata availability.

The TRLN Shared Records project has delivered efficiencies in throughput as well. That a single institution can take responsibility for managing metadata on behalf of two or three partner libraries delivers efficiencies immediately through the elimination of workflows. As seen in table 2, fourteen discrete local cataloging workflows have been eliminated or avoided in the TRLN libraries.

Saving Time and Reallocating Effort

The elimination of duplicate workflows has created time and energy at each institution for other projects and new initiatives. For instance, before

the implementation of the Shared Records program, cataloging staff at NCSU, Duke, and NCCU spent time maintaining URLs in bibliographic records for government documents. UNC-Chapel Hill, as regional depository and host institution for the TRLN DWS collection, has taken responsibility for these activities, freeing up staff at the other institutions for different activities. Duke, for instance, reallocated technical services staff to work on deferred metadata management activities related to government documents.

Sharing Costs

The TRLN Shared Records program allows the TRLN libraries to share the costs of licensing records and authority processing. For instance, a consortium license for DWS reduced NCSU's annual Marcive record subscription costs by $650. Centralized processing yields other savings as well. Before implementing the Shared Records Program, the consortium's libraries paid Marcive to "set holdings" at OCLC for DWS titles. This process has been centralized and UNC-Chapel Hill staff use OCLC batch processes and a group profile to set holdings, virtually eliminating this expense. As of September 2012, 221,131 records were in the Shared Records program, which removed 514,459 records from authority control at the four institutions eliminating associated processing costs.

Conclusion

The era of the library catalog as a motley collection of discrete and static records reflecting decisions made at particular points in time and under differing sets of rules and local practices is rapidly drawing to a close. The technological hurdles began falling when LC automated the production of its card sets in the 1960s and continues to this day with the advent of

simple but powerful personal metadata manipulation tools like MarcEdit and MARC Report, commercial MARC record notification services, serial and e-resource knowledge bases, and the promise of linked metadata.

In the last ten years, the pace of this change has finally reached the social, political, and personal spheres, as cataloger retirements and the economics of technical services operations have run head-on into other, even more powerful, movements affecting the bibliographic universe. Principal among these has been the merging of traditional reference discovery tools (indexes, bibliographies, citation analysis) with full-text databases to create a compelling competitor to the more mundane library catalog. Electronic resource management, with its huge package deals, complicated license agreements, knowledge bases, and link monitoring has necessitated a similar deflection of library attention away from the catalog and local ILS toward outsourced bibliographic record creation and maintenance. Finally, the renaissance of special collections has led to additional competition for metadata expertise and discovery layer development. It is not too much of an exaggeration to say that experiments to bring together the new reference tools, electronic and digital resources, and archival finding aids with the catalog have consumed much of the library world's energy over the last decade. This has forced libraries to invest less toward institutionally specific catalog records and more toward customized information discovery tools.

These broad movements in library stewardship have also changed the expectations of library management toward technical services, and particularly cataloging personnel. The ideal cataloger is no longer the person with the deepest knowledge of AACR2, RDA, or LCSH, but rather the person who is most adept at batch metadata manipulation, vendor contract management, liaison with library or

campus IT, and training and motivation of support staff. She is also expected to be on the lookout for synergies with interested third parties and anything else that could potentially reduce institutional processing costs and efforts.

The Shared Records Program has allowed the TRLN libraries to build experience and expertise to address these concerns. TRLN libraries are now putting their trust in metadata that is both vendor-generated and maintained by a single partner library. In the process, the consortium has developed additional expertise in batch metadata manipulation. Removing manual tasks as much as possible from Shared Records processing procedures has decreased the time to disseminate metadata to discovery applications. The elimination of duplicative metadata workflows at member libraries has released cataloging staff to work on more pressing metadata maintenance efforts and has reduced authority processing costs.

Several factors were essential to the program's success. The shared technical infrastructure provided by the Search TRLN system delivered a platform that enabled record sharing in this way. Deep expertise at member libraries in the areas of batch processing of metadata was critical to the project's success. TRLN's formal council and committee structures provided a vehicle for gaining support for a pilot project and eventually a framework for implementing the program. The Electronic Resources Access and Shared Records Pilot task groups received clear charges with well-defined goals, objectives, and timelines for completion ensuring that the project would stay on track and in scope. Broad representation on these task groups also ensured that appropriate input from throughout the consortium would be gathered and that the program would have wide support upon implementation. The most important factor, however, is the presence of a deep

trust between the libraries of TRLN built through decades of collaboration. Similar conditions undoubtedly exist in other small library consortia to replicate this model, a relatively simple extension of the resource-sharing model that has guided library technical services for decades.

References

1. Will Owen, "The Triangle Research Libraries Network: A History and Philosophy," *North Carolina Libraries*, 47, no. 1 (1989): 43–51.

2. Kristin Antelman, Emily Lynema, and Andrew Pace, "Toward a 21st Century Library Catalog," *Information Technology & Libraries* 25, no. 3 (2006), accessed December 21, 2012, http://eprints.rclis.org/handle/10760/8177.

3. Kristin Antelman and Mona Couts, "Embracing Ambiguity . . . Or Not: What the Triangle Research Libraries Network Learned about Collaboration," *College & Research Libraries News* 70, no. 4 (2009): 230–33, accessed December 21, 2012, http://crln.acrl.org/content/70/4/230.full.pdf.

4. Karen Coyle, *Rules for Merging MELVYL Records* (Technical Report No. 6), revised (Berkeley, CA: Division of Library Automation, University of California, 1992); HathiTrust, Getting Content into HathiTrust, 2012, accessed December 21, 2012, www.hathitrust.org/ingest.

5. Paul Moeller, Wendy Baia and Jennifer O'Connell, "Cataloging for Consortium Catalogs," *Serials Librarian* 44, no. 3/4 (2003): 229–35; Chew Chiat Naun and Susan M. Braxton, "Developing Recommendations for Consortial Cataloging of Electronic Resources: Lessons Learned," *Library Collections, Acquisitions, and Technical Services* 29, no. 3 (2005): 307–25, doi:10.1016/j.lcats.2005.08.005.

6. Karen Cary and Joyce L. Ogburn, "Developing a Consortial Approach to Cataloging and Intellectual Access," *Library Collections, Acquisitions, and Technical Services* 24, no. 1 (2000): 45–51, doi:10.1016/S1464-9055(99)00095-0.

7. Patricia Sheldahl French, Rebecca Culbertson, and Lai-Ying Hsiung, "One for Nine: The Shared Cataloging Program of the California Digital Library," *Serials Review* 28, no. 1 (2002): 4–12, doi:10.1016/S0098-7913(01)00169-1.

8. Kristin E. Martin and Kavita Mundle, "Cataloging E-Books and Vendor Records: A Case Study at the University of Illinois at Chicago," *Library Resources & Technical Services* 54, no. 4 (2010): 227–37, accessed December 21, 2012, http://alcts.metapress.com/content/h1455767637633x8fulltext.

9. Jeffrey Beall, "Free Books: Loading Brief MARC Records for Open-Access Books in an Academic Library Catalog," *Cataloging & Classification Quarterly* 47, 5 (2009): 452–63.

10. Rebecca L. Mugridge & Jeff Edmunds, "Batchloading MARC Bibliographic Records." *Library Resources & Technical Services* 56, no. 3 (2012): 155–70, accessed December 21, 2012, http://alcts.metapress.com/content/m835665768854833fulltext.

11. Erin Stalberg and Christopher Cronin, "Assessing the Cost and Value of Bibliographic Control," *Library Resources & Technical Services* 55, no. 3 (2011): 124–37, accessed December 21, 2012, http://alcts.metapress.com/content/mn57629h584r87l1fulltext.

12. Ibid., 124.

13. Ibid., 130–34

14. Ibid., 132

15. David E. Gleim, "The relationship between Cataloging Delay and the Circulation of Books at a Large Research Library" (unpublished dissertation, University of North Carolina at Chapel Hill, 1992).

16. Lynne Howarth, Leslie Moore, Elisa Sze, "Mountain to Molehills: The Past, Present, and Future of Cataloging Backlogs," *Cataloging & Classification Quarterly* 48, no. 5 (2010): 423–44, doi:10.1080/01639371003767227.

17. Ruth Fischer, Rick Lugg, and Kent C. Boese, "Cataloging: How to Take a Business Approach," *The Bottom Line: Managing Library Finances* 17, no. 2 (2004): 50–54, doi:10.1108/08880450410536062.

18. Ibid, 51.

19. Stalberg and Cronin, "Assessing the Cost and Value of Bibliographic Control," 133.

20. Ibid.

21. "TRLN Technology Council," Triangle Research Libraries Network, accessed December 21, 2012, www.trln.org/committee/TechnologyCouncil/index.htm.

22. "TRLN Electronic Resource Access Restrictions Task Group (2009)," Triangle Research Libraries Network, accessed December 21, 2012, www.trln.org/endeca/task-groups/restrictions/index.htm.

23. "TRLN Shared Records Pilot Task Group (2010)," Triangle Research Libraries Network, accessed December 21, 2012, www.trln.org/committee/TechnologyCouncil/TaskGroups/SRTG.htm.

24. The TRLN DDI to Endeca crosswalk was based on the DDI to Dublin Core crosswalk. "Mapping to Dublin Core (DDI Version 2)," Data Document Initiative, accessed December 21, 2012, www.ddialliance.org/resources/tools/dc.

25. Beverly Dowdy and Rosalyn Raeford, "Electronic Resources Workflow Analysis and Process Improvement" (presented at annual Electronic Resources & Libraries conference, Austin, Texas, April 2–4, 2012), accessed December 21, 2012, http://prezi.com/fkabhs5_quql/electronic-resources-workflow-analysis-process-improvement.

26. Stalberg and Cronin, "Assessing the Cost and Value of Bibliographic Control," 132.