Next Generation Library Catalogs: Local developments and research opportunities

Derek Rodriguez, TRLN
September 26, 2008

Center for Research and Development of Digital Libraries
UNC Chapel Hill, School of Information and Library Science, September 26, 2008
Overview

- Introduction to TRLN
- Scope and goals of the TRLN Endeca Project
- Project challenges and system architecture
- Discussion of opportunities for SILS / TRLN collaboration
An Academic Research Library Consortium

- Duke University Libraries
- NC Central Libraries
- NC State Libraries
- UNC Chapel Hill Libraries

Founded in 1977, cooperation dates to 1930s
TRLN: Oversight and Governance

- Governing Board
- Executive Committee
  - Council of Directors
    - TRLN Councils
      - Collection Development
      - Human Resources
      - Services
      - Technology
    - Committees and Task groups as needed

- TRLN Staff (Director, two Program Officers, and an Administrative Assistant)
Program focus

- Collaborative collection development
  - Print and electronic
- Services to extend access to materials to affiliated patrons
  - Document Delivery and Reciprocal Borrowing
- Digital TRLN
  - Endeca project
  - Single Copy Archive
  - TRLN Management Academy
Next Gen Catalogs

Generations of OPACs

- Online ‘Card’ Catalog
- + exact-match Boolean keyword searching
- + Web enabled catalogs offered search OR browsing of pre-coordinated headings

Next Generation Library Catalogs

- Not bound to legacy ILS
- Take advantage of latest web technologies
- Should be ‘metadata agnostic’
- Exhibit an open system architecture

Center for Research and Development of Digital Libraries
UNC Chapel Hill, School of Information and Library Science, September 26, 2008
TRLN Endeca Project: Phase 1

Why Endeca?

- Support for integrated search and browse
- ‘Forgiving’ features (spell correction, term suggestion, etc.)
- Can ingest ‘metadata of all types’
- NCSU’s previous experience with Endeca

Phase 1 Goals

- Search and request from one interface
- Support scoped interfaces at the campuses

Timeline

- June 2007 – August 2008
**TRLN Endeca Project: Phase 1**

- **Steering Committee and four task groups**
  - User Interface
  - Data and Indexing
  - Document Delivery
  - Usability

- **Distributed Implementation Team**
  - Programmers, designers, data specialists

- **Commercial partners**
  - Endeca professional services
  - MCNC
  - Syndetic Solutions

---

Center for Research and Development of Digital Libraries
UNC Chapel Hill, School of Information and Library Science, September 26, 2008
Phase 1 Timeline

- March 2007 – Contract with Endeca is signed
- April 2007 – Task Groups are formed
- May 2007 – UITG and DITG begin work
- June 2007 – Staff attend Endeca training
- July 2007 – Requirements Definition Meeting
- August 2007 – Order servers
- September 2007 – Complete wireframes
- October 2007 – Complete data and indexing recommendations, install servers, conduct focus group interviews
Phase 1 Timeline cont’d

- November 2007 – Milestone: TRLN collections are indexed in Endeca (~10 million records)
- December 2007 – Begin testing Requests feature
- January 2008 – Build Search TRLN User Interface
- February – Conduct usability testing, refine data extracts, complete Requests, refine user interface, begin contracts with Syndetic Solutions
- March – Final testing, phase 1 launch of Search TRLN
- May – Refine relevance ranking / performance tuning
- June – August – Duke, UNC, and NCSU interfaces launch
- July 2008 – New oversight group convenes
Project Challenges

- Coordinating the efforts of four task groups, an implementation team, and relationships with three vendors
- Gather user perspective to support design
- Create a data model to fit application logic
- Create extract and data harvesting routines
- Define indexes and relevance ranking
- Build 4 user interfaces
- Integrate with 9 local document delivery offices
- Performance tuning
- Ongoing governance and evaluation
Create a data model

Data model should …

- Support needed functionality
- Accommodate but not reproduce MARC
- Be flexible and accepting of multiple metadata types
- Respect local needs where possible

Stakeholders represent 4 institutions, 10 libraries, and numerous cataloging units

Task: Define fields / facets, MARC to Endeca mappings, extract rules, and indexing rules
Endeca Properties (like database fields)

- Can be indexed, displayed, or used for sorting
- Can serve as unique ids or keys for joins
- Can support linking via hyperlinks or calls to external services

Examples:
- Main Title: Trout Magic
- Location: Davis Library
- Call Number: SH687 .T73 1989
- Uniqueld: UNCb2260544
Endeca Dimensions (Facets)

- Values can be pre-set or data-driven
- Hierarchies and ranges are possible

Example:

Format

Book
  Record id 1
  Record id 2

Journal
  Record id 3
## Endeca Record: Dimensions

<table>
<thead>
<tr>
<th>Facet</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>6xx $a or $x</td>
</tr>
<tr>
<td>Genre</td>
<td>6xx $v / 008</td>
</tr>
<tr>
<td>Time Period</td>
<td>6xx $y</td>
</tr>
<tr>
<td>Region</td>
<td>6xx $z</td>
</tr>
<tr>
<td>Medical Subject: Topic</td>
<td>MESH</td>
</tr>
<tr>
<td>Author</td>
<td>Author fields</td>
</tr>
<tr>
<td>Language</td>
<td>MARC 008</td>
</tr>
<tr>
<td>Format</td>
<td>Item record / MARC 008</td>
</tr>
<tr>
<td>Publication Year</td>
<td>MARC 008</td>
</tr>
<tr>
<td>Call Number Range</td>
<td>LC and NLM Call Numbers</td>
</tr>
<tr>
<td>Location (Library)</td>
<td>Item Record</td>
</tr>
<tr>
<td>Availability</td>
<td>Item status</td>
</tr>
<tr>
<td>New Titles</td>
<td>Date cataloged</td>
</tr>
</tbody>
</table>

Center for Research and Development of Digital Libraries  
UNC Chapel Hill, School of Information and Library Science, September 26, 2008
Mapping and extract policies

MARC
LEADER 00000nam 2200337Ia 4500
001 187110658
003 OCoLC
005 20080312084320.0
008 080103s2007 mumuc r s001 0deng d
035 (OCoLC)187110658
040 NOC|cNOCC|dNOC
043 n-us-nc
049 NODK
090 Z669.U537|bU55 2007
099 9 CC378|aUZL
110 2 University of North Carolina at Chapel Hill. |bSchool of Information and Library Science.
245 10 Illuminating the past : |ba history of the first 75 years of the University of North Carolina's School of Information and Library Science.
300 119 p. : |bill. (some col.), ports. : |c24 cm.
500 Includes index.
505 0 The library school in Chapel Hill : an essay / David Carr
-- Imagining the future : profiles of the 12 deans who have led the School of Information and Library Science --
Illuminating the past : photos and events from the first 75 years of the School of Information and Library Science
-- Alumni : seventy-five years of graduates from the School of Information and Library Science.
610 20 University of North Carolina at Chapel Hill. |bSchool of Information and Library Science|xHistory.
610 20 University of North Carolina at Chapel Hill. |bSchool of Information and Library Science|xHistory|vChronology.
610 20 University of North Carolina at Chapel Hill. |bSchool of Information and Library Science|xAlumni and alumnae

Center for Research and Development of Digital Libraries
UNC Chapel Hill, School of Information and Library Science, September 26, 2008
Mapping and extract policies

Record UNCb5319029

PROPERTIES:

Call Number: Z669.U537 U55 2007|CC378 UZL|
Item Due Date: ||
Item Notes: ||
Item Types: Book|Book|
Libraries: |||wb|
LocalId: b5319029
Location property: ||da|wbnal
Main Author: University of North Carolina at Chapel Hill. School of Information and Library Science.
Main Title: Illuminating the past : a history of the first 75 years of the University of North Carolina's School
OCLCNumber: 187110658
Published: c2007.
Rollup: 187110658
Serial Holdings Summary: wb|wbnal|CC378 UZL||||
Statuses: Available|Available (library use only)|
Title Sort: Illuminating the past : a history of the first 75 years of the University of North Carolina's School

Center for Research and Development of Digital Libraries
UNC Chapel Hill, School of Information and Library Science, September 26, 2008
### DIMENSION VALUES:

<table>
<thead>
<tr>
<th>Subject</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alumni and alumnae</td>
<td></td>
</tr>
<tr>
<td>Deans (Education)</td>
<td></td>
</tr>
<tr>
<td>History</td>
<td></td>
</tr>
<tr>
<td>Information science</td>
<td></td>
</tr>
<tr>
<td>Library schools</td>
<td></td>
</tr>
<tr>
<td>University of North Carolina at Chapel Hill</td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>Book &gt; Book</td>
</tr>
<tr>
<td>Location</td>
<td>UNC Chapel Hill &gt; Information &amp; Library Science Library</td>
</tr>
<tr>
<td></td>
<td>UNC Chapel Hill &gt; North Carolina Collection</td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
</tr>
<tr>
<td>Publication Year</td>
<td>2000 to present</td>
</tr>
<tr>
<td>Genre</td>
<td>Biography</td>
</tr>
<tr>
<td>Genre</td>
<td>Chronology</td>
</tr>
<tr>
<td>Genre</td>
<td>Registers</td>
</tr>
<tr>
<td>Region</td>
<td>Chapel Hill</td>
</tr>
<tr>
<td>Region</td>
<td>North Carolina</td>
</tr>
<tr>
<td>Availability</td>
<td>Available</td>
</tr>
<tr>
<td>Source</td>
<td>Catalog</td>
</tr>
</tbody>
</table>
Mapping and extract policies
(examples)

MARC 245 -> Endeca Main Title
MARC 100, 110, 111 -> Endeca Main Author
MARC 100, 110, 111, 700, 800 -> Author Facet
MARC 6xx -> Pre-coordinated Subjects

eg. Library schools – North Carolina – Chapel Hill --History

MARC 6xx -> Subject Facets

Subject (Topic): Library schools
Subject (Topic): History
Region: North Carolina
Region: Chapel Hill

Center for Research and Development of Digital Libraries
UNC Chapel Hill, School of Information and Library Science, September 26, 2008
Mapping and extract policies (examples)

MARC 008 (bytes 35-37) -> Language facet
MARC 008 (bytes 33-34) -> Genre facet

Item Data

Library -> Location property (for display)
Library -> Location facet
Status -> Status property (for display)
Status -> Availability facet
Call Number -> Call number and call number sort properties
Call Number -> Classification facet
<table>
<thead>
<tr>
<th>MARC Field</th>
<th>Extract rule</th>
<th>Endeca Label</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>245 abfghknps</td>
<td>Strip trailing /</td>
<td>Property: Main Title</td>
<td>Index as title and keyword, display on brief and full record screens</td>
</tr>
<tr>
<td>245 abfghknps</td>
<td>Strip non-filing characters</td>
<td>Property: Title Sort</td>
<td>Do not index or display, enable sort</td>
</tr>
<tr>
<td>6xx subfield a or x</td>
<td>Where first indicator = 0,1,2,3,4, or 7). Trim punctuation.</td>
<td>Facet: Subject</td>
<td>Facet</td>
</tr>
<tr>
<td>Item status</td>
<td>Create pipe delimited field of all item statuses attached to title</td>
<td>Property: Statuses</td>
<td>Used for display</td>
</tr>
</tbody>
</table>

Center for Research and Development of Digital Libraries
UNC Chapel Hill, School of Information and Library Science, September 26, 2008
MARC data from ILSs, 4 institutions, 10 libraries, 3 vendors, 11 million titles

-ICE= ToCs

Non-MARC data

- EAD ?
- Dublin Core
- Other ?

• Features
  • Full ILS extracts occur as needed
  • Cataloging changes are updated daily
  • Circulation status is updated every 30 minutes
  • ICE Table of Contents data is updated weekly
  • Index is rebuilt every night

Center for Research and Development of Digital Libraries
UNC Chapel Hill, School of Information and Library Science, September 26, 2008
One index, many interfaces

Center for Research and Development of Digital Libraries
UNC Chapel Hill, School of Information and Library Science, September 26, 2008
Some usage statistics

Search TRLN only, Sept 1 – 21, 2008

7437 sessions
- 7146 (97%) performed text search
- 1461 (19%) performed facet navigation
- 3702 (50%) began as local catalog searches

Searches by index
- Keyword 48.40%
- Title 33.99%
- Author 8.90%
- Subject 3.88%
- ISBN 3.21%
- Multi-index queries 1.62%
Some usage statistics

Other interesting data points

- 18,833 record views
  - 3801 sessions (51%) included at least one record view
- 1956 item requests
  - 1124 sessions (15%) included an item request

Future Analysis

- Most popular facets
  - Subject, Location, and Format seem most popular so far
- Session Level analysis
- Establish benchmarks for considering and evaluating enhancements

Center for Research and Development of Digital Libraries
UNC Chapel Hill, School of Information and Library Science, September 26, 2008
TRLN – SILS Collaboration

Cory Lown, SILS MSLS ’08, Brad Hemminger
  - Transaction Log Analysis with NCSU Endeca,
    http://hdl.handle.net/1901/488

Dre Orphanides, SILS MSLS ‘08
  - Field Experience, Spring ’08

Tessa Sullivan, SILS PhD Student
  - Transaction Log Analysis, Fall ’08

Brad Hemminger, Associate Professor SILS and Sara Ramdeen, SILS PhD Student
  - End-user study, Fall ‘08

Kathy Wisser, Tessa Sullivan, Derek Rodriguez (SILS PhD Students)
  - Metadata Quality Project

Center for Research and Development of Digital Libraries
UNC Chapel Hill, School of Information and Library Science, September 26, 2008
Projects under consideration (next 6 – 12 months)

- Indexing Non-MARC data (EAD, DC, etc.) in Endeca
  - Data modeling
  - Data harvesting
  - User interface issues

- Creating a user interface for NC Central
  - User needs assessment
  - User interface design and implementation
  - Evaluation

- Evaluation (ongoing)
  - Log analysis / Usability Testing
Research Opportunities

Name Authorities
- Can we integrate LC Name authorities into our Endeca interfaces?

Faceted Access to Subject Terminology (FAST)
- Simplified schema based on LCSH to better support post-coordinated retrieval and simplify heading assignment
- See http://www.oclc.org/research/projects/fast/

FRBR / Record Rollup
Research Opportunities

Find similar / recommender services

- How do we leverage metadata about a record to help the user find similar items? Can facets drive ‘find similar’ functionality? Can we use circulation / usage data to recommend other titles to users?

Personalization

- Do users want a personalized experience with a system like this? What about tagging?

Support for small devices like the iPhone, etc.
Possible forms for collaboration

- **Field experiences for Master’s students**
  - **Examples**
    - Shadowing a project
    - System building

- **Independent studies**
  - **Examples**
    - Writing a literature review / background paper
    - Transaction log analysis, user studies

- **Joint research and development projects**
Derek Rodriguez, TRLN
919-962-8022
derek@trln.org
http://www.trln.org

Useful links:
Search TRLN: http://search.trln.org
Endeca: http://www.endeca.com