

Data, Data Everywhere: Migration and System Population Practices

<http://www.niso.org/news/events/2009/datasystems09/>



• Introduction

– **Todd Carpenter**, Managing Director, NISO

• Data quality, policy, and large-scale data flows

How do regional consortia establish and implement policies to allow them to cope with increasing amounts of data in a widening variety of formats?

– **Hilary Newman**, Vice President of Implementation Services, Innovative Interfaces, Inc.

• Academic library perspective

What are the successes, stress points, and failures from the perspective of the academic library?

– **Maribeth Manoff**, Coordinator for Networked Service Integration, University of Tennessee, Knoxville

– **Robert McDonald**, Associate Dean for Library Technologies, Indiana University

NISO Webinar • November 11, 2009

Wednesday, November 11, 2009



Upcoming NISO Events

<http://www.niso.org/news/events>

2009

- December 9 (NISO Webinar): **ONIX for Publication Licenses: Adding Structure to Legalese**
- December 14 (NISO Open Call): **ISO TC46, Subcommittee 11 (Archives/Records Management) Update - Free!**

2010

- January 13 (NISO Webinar): **From ILS to Repository and Back: Data Interoperability**
- February 10 (NISO Webinar): **What It Takes To Make It Last: E-Resources Preservation**
- **March Two-Part Webinar: Identifiers: New Problems, New Solutions**
 - March 10: What's in a Name? Latest Developments in Identifiers
 - March 17: Content Identification: What's New with ISO Identifiers, ISBN and ISTC
- March 23 (Atlanta, GA): **Discovery to Delivery: Creating a First-Class User Experience**

- Buy 4 webinars, get 2 free
- Buy 6 webinars, get 7 free

NISO Webinar • November 11, 2009

Wednesday, November 11, 2009

Data quality, policy, and large-scale data flows

NISO Webinar
November 11, 2009

Hilary Newman

Vice President, Implementation Services
Innovative Interfaces, Inc.

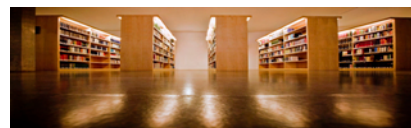
Contact information: hilary@iii.com; 510-655-6200

Innovative
interfaces

Wednesday, November 11, 2009

My Background

- Work with libraries to merge and populate bibliographic and other system data
- 15 years at Innovative
- MLIS University of North Texas



Copyright Innovative Interfaces. Not to be duplicated or distributed without permission of Innovative Interfaces.

Innovative

Wednesday, November 11, 2009

Agenda

- To Merge or Not To Merge?
- Challenges
- Quality Implementation Practices
- Strategic Policy Decisions
- Data Quality and Standards
- Project Management

Copyright Innovative Interfaces. Not to be duplicated or distributed without permission of Innovative Interfaces.

Innovative

Wednesday, November 11, 2009

To Merge or Not To Merge: Merge!

- Initial Data Population Planning
- Ongoing Maintenance and Growth
 - Authority Control
 - Outsourcing
 - Institutionalizing Practices
- Duplicate Records, Who Wins?



Copyright Innovative Interfaces. Not to be duplicated or distributed without permission of Innovative Interfaces.

Innovative

Wednesday, November 11, 2009

Challenges

- Quality of Data or Cataloging Practices
- Maintaining Encoding Levels?
- Record Source Quality
- Local Practices and Needs
 - Item identification, barcodes, call numbers
 - Licensing Information
 - ISSN/ISBN Cataloging Practices
 - 856 and other URL Link Uses
 - Etc....

Copyright Innovative Interfaces. Not to be duplicated or distributed without permission of Innovative Interfaces.

Innovative

Wednesday, November 11, 2009

Quality Implementation Practices

- Following Standards
- Set Local Control Policies
- Match Points and Control Numbers
- Consider End User, Is More Data Better?
- User Interface Considerations

Copyright Innovative Interfaces. Not to be duplicated or distributed without permission of Innovative Interfaces.

Innovative

Wednesday, November 11, 2009

Strategic Policy Decisions

- Consider Data Alive and Evolving
- Consistency and Simplicity Win
- Don't Make Policies Based on Today's Needs
 - User Interface Changes
 - Future Data Mergers
 - Possible Data Splits?
 - Staff Costs to Maintain Customized Practices



Copyright Innovative Interfaces. Not to be duplicated or distributed without permission of Innovative Interfaces.

Innovative

Wednesday, November 11, 2009

Data Quality and Standards

- Cataloging Quality
- Bibliographic Records
- Electronic Resources
- Metadata
- Beware the Cleanup Project
- Leverage Computing Power



Copyright Innovative Interfaces. Not to be duplicated or distributed without permission of Innovative Interfaces.

Innovative

Wednesday, November 11, 2009

Project Management

- Who Decides and Who Does the Work?
- Reconsider Local Workflows
- How Do Others Do It Successfully?
- Plan Ahead
- Don't Plan for Perfection



Copyright Innovative Interfaces. Not to be duplicated or distributed without permission of Innovative Interfaces.

Innovative

Wednesday, November 11, 2009

Conclusions

- Plan for the Life of the Data
- Standards!
- Decision Making and Teamwork
- Make Computing Power Work For You
- Be Realistic



Copyright Innovative Interfaces. Not to be duplicated or distributed without permission of Innovative Interfaces.

Innovative

Wednesday, November 11, 2009



Thank You!

Copyright Innovative Interfaces. Not to be duplicated or distributed without permission of Innovative Interfaces.

Innovative

Wednesday, November 11, 2009

DATA, DATA EVERYWHERE AND CONSTANTLY MOVING: THE CHALLENGE OF ONGOING MIGRATION

MARIBETH MANOFF
UNIVERSITY OF TENNESSEE (UT) LIBRARIES
NISO WEBINAR – NOVEMBER 11, 2009

Wednesday, November 11, 2009

DATA MIGRATION AND SYSTEM POPULATION AT THE UT LIBRARIES

- Large, “one-time” system migration
- Implementation of a new Integrated Library System (ILS)
- Ongoing migration / constant system population
 - The link resolver “knowledge base”
 - A “next generation” discovery and delivery system

Wednesday, November 11, 2009

LARGE, “ONE-TIME” SYSTEM MIGRATION

- Migration to a new ILS (2003)
- Moved to Aleph from 3 separate systems:
 - 1) Cataloging and Circulation - Horizon
 - 2) Acquisitions and Serials - GEAC
 - 3) Web OPAC – OCLC SiteSearch (Z39.50)
- Data migrated was in known, or at least familiar, formats

Wednesday, November 11, 2009

MIGRATING DATA IN KNOWN FORMATS

1. Bibliographic data in MARC format
 - 1.6 million bib records “straight” MARC to MARC
 - Brief order bib records from acquisitions system
 - Some matching and mapping required for orders to be attached to the correct bibliographic record
 - Course reserves records
 - Also in MARC format; stored in a different database table in the new system
2. Item / Holdings data
 - Translation algorithm for item records considering location, collection, and call number (from either bib or item) to holdings records in MARC format
 - Import data stream of item records from Horizon to create holdings and item records in Aleph

Wednesday, November 11, 2009

KNOWN FORMATS (CONT.)

- For Non-MARC data – formats are not standardized, but data elements within the tables are familiar
- Circulation data from Horizon
 - Patrons (4 tables)
 - Loans
 - Hold requests
- Acquisitions and Serials data from GEAC
 - Orders
 - Subscriptions
 - Budgets
 - Vendors, Vendor addresses

Wednesday, November 11, 2009

PROCEDURE FOR ILS POPULATION – NON-MARC DATA

1. Export data from Horizon or GEAC into tab-delimited file
2. Set up local MySQL database with fields equivalent to database tables in Aleph
3. Import tab-delimited files into database, moving data elements into correct fields for Aleph
4. Import from local database into Aleph database tables

Wednesday, November 11, 2009

ILS MIGRATION - THE UT EXPERIENCE

- Labor intensive and time consuming; required extreme and exclusive focus
- Export -> import -> export -> import process used with non-MARC data took a number of iterations to get right
- Familiar formats made it relatively easy to see when the pieces fallen into the right places, a satisfying experience
- Internal workflows and some patron empowerment features (e.g. online renewal) were improved with one integrated system

Wednesday, November 11, 2009

ILS MIGRATION - THE UT EXPERIENCE (CONT.)

- One-time nature of the migration and familiarity with the data may have led to less innovation
 - ✦ Indexing and display of MARC fields was mostly taken from previous system
 - ✦ “We got it right” feeling persisted
 - ✦ Didn’t have enough time to spend on the user experience side (the OPAC)
 - ✦ Following completion of migration, changes that required reindexing were often postponed

Wednesday, November 11, 2009

TOWARD ONGOING MIGRATION AND NEW DATA FORMATS

- UT has developed a number of ongoing processes and services moving data in and out of the Aleph system
 - Load patron information with exports from campus LDAP system
 - Export of invoices for ingest and processing in SAP
 - SQL queries for custom reporting
- Newer, more modular systems require ongoing migration of data in less familiar formats

Wednesday, November 11, 2009

THE LINK RESOLVER “KNOWLEDGE BASE”

- Emphasis of OpenURL link resolver software (SFX from Ex Libris at UT) is on the user experience - simplifying access
- Ongoing data updates are essential for linking accuracy
- Software providers collect and distribute large amounts of data from electronic resource vendors – an invaluable service
- Processes for updating the KnowledgeBase to reflect current subscription information are equally important

Wednesday, November 11, 2009

LINK RESOLVER KNOWLEDGEBASE (CONT.)

- Data elements in the KnowledgeBase are keyed to access
- Constant vigilance in keeping the KnowledgeBase updated, and different data formats, helps us think in new ways, e.g.:
 - * What are the most important data fields to enable efficient and effective searching for journals?
 - * What is the best way to express relationships between serial titles?

Wednesday, November 11, 2009

THE “NEXT GENERATION” DISCOVERY AND DELIVERY SYSTEM

- Another module with a focus on the user experience (Primo from Ex Libris at UT)
- Catalog data is “transformed”
 - MARC Export from ILS
 - Import / transformation into an XML format
- Other data sources, e.g., records from an OAI repository, can be added

Wednesday, November 11, 2009

“NEXT GENERATION” DISCOVERY AND DELIVERY (CONT.)

- XML data elements are keyed to discovery and delivery, e.g.,
 - Facet fields for navigation of search results created from author, subject heading, classification, availability information, ...
 - Dedup and FRBR fields to facilitate streamlined display of duplicate or “like” records
- “Rules” for data transformation are understandable and configurable
- Reindexing does not require system downtime

Wednesday, November 11, 2009

“NEXT GENERATION” DISCOVERY AND DELIVERY (CONT.)

- Ongoing data migration, into new and configurable data format, encourages innovation, e.g.:
 - * How much does “format” matter, and how does it play into faceted navigation?
 - * New ways to bring authority data into the mix without having to browse an index

Wednesday, November 11, 2009

THE CHALLENGE OF ONGOING DATA MIGRATION

- A challenge in the best sense of the word, fostering new thinking and innovation
- Requires procedures that are both rigorous and flexible
- More to look forward to on this front, as other, modular, systems come online at the UT Libraries (e.g., ERMS)

Wednesday, November 11, 2009

THANK YOU!

Wednesday, November 11, 2009

Libraries and Data an IU Perspective

Robert H. McDonald

Associate Dean for Library Technologies
Associate Director Data to Insight Center

NISO Webinar 11.11.09



Wednesday, November 11, 2009

Hello – Greetings - Welcome – Bonjour - Tag

- Tagging
 - #niso
 - #data
 - @mcdonald
- GoogleWave
 - with:public niso



Robert H. McDonald

OVERVIEW

- Institutional Background
- Data at IU
 - Discovery
 - Curation
 - Storage
 - Scale
- Data Futures
- Above Campus

Indiana University Libraries

- Indiana University
- Statewide Location, Indiana
- Large, Research Oriented Public University
 - 101,000+ students
- SIRSIDynix Symphony 3.2.1 Library Management System
 - 6.5 million items
- 9 Regional Campuses
- ILS Managed by Libraries/Enterprise Applications

DATA @IU

- Data Discovery
 - Bibliographic Data
- Data Curation
 - Local
 - Consortial
- Data Storage
 - Local
 - Consortial

DISCOVERY DATA

- SIRSIDynix Symphony Web OPAC
- WorldCat Local Discovery Interface
- Testing Ebsco Discovery Interface
- Mobile Discovery
 - Boopsie Mobile Search
 - MIT Mobile Framework (campus)
- Multiple Instances
 - Dspace – IU ScholarWorks
 - Open Journal System – IU ScholarWorks
 - Fedora (modern DLP infrastructure)
 - Legacy Digital Library Systems

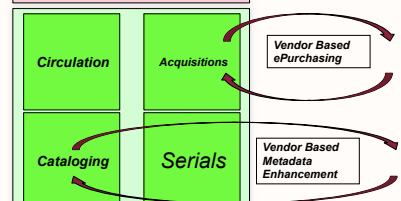
LEGACY ILS

Public User Interface:



Legacy ILS Functionality

Functional modules:



Data Stores:



Staff Interfaces:

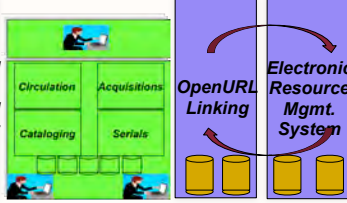
Legacy ILS+Current Management Tools



Learning Mgmt./Campus Portal Feeds

Proxy/VPN for eContent Access

Federated Search/
Advanced
Discovery

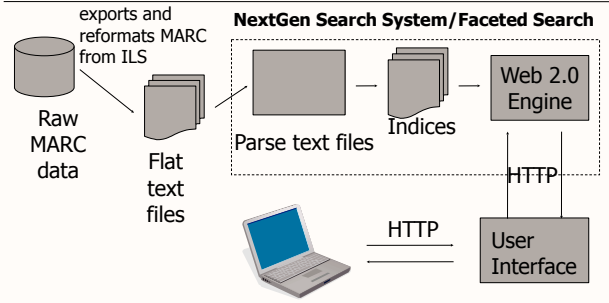


Legacy ILS
Functionality
+
eContent
Module
+
Advanced
Discovery
Interface

Unified Workflow Interface for Staff

Wednesday, November 11, 2009

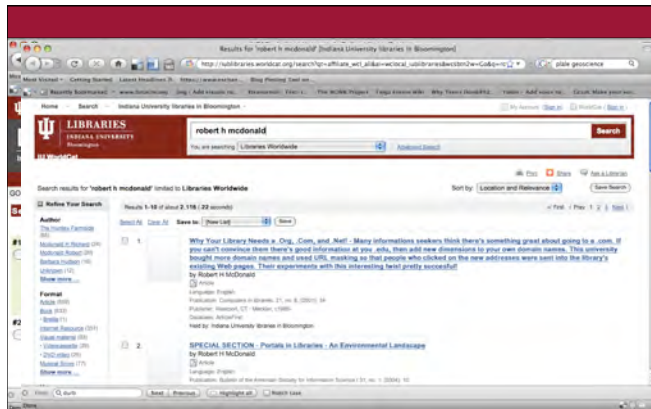
Good Search – Decouples Discovery from ILS



DIGITAL LIBRARY PROGRAM
INDIANA UNIVERSITY
A Maxwell Forestry Extension & Paducah/Paducah

pervasivetechinstitute
AT INDIANA UNIVERSITY

Wednesday, November 11, 2009



DIGITAL LIBRARY PROGRAM
INDIANA UNIVERSITY
A Maxwell Forestry Extension & Paducah/Paducah

pervasivetechinstitute
AT INDIANA UNIVERSITY

Wednesday, November 11, 2009

IU DISCOVERY MOBILE

- MIT MOBILE FRAMEWORK
- DATA APIS
- SAKAI
- IU KNOWLEDGBASE
- IUCAT SEARCH



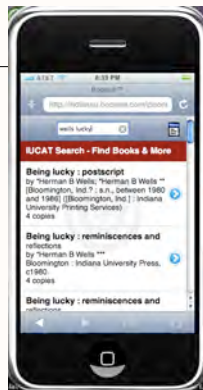
DIGITAL LIBRARY PROGRAM
INDIANA UNIVERSITY
A Maxwell Forestry Extension & Paducah/Paducah

pervasivetechinstitute
AT INDIANA UNIVERSITY

Wednesday, November 11, 2009

IUCAT DISCOVERY MOBILE

- BOOPSIE
- MARC RECORD EXTRACT
- AUTOMATED UPDATES
- IU INTELLIGENT INFRASTRUCTURE



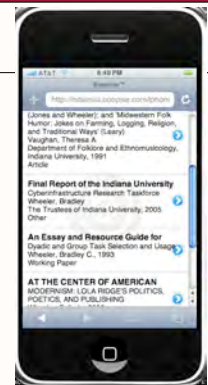
DIGITAL LIBRARY PROGRAM
INDIANA UNIVERSITY
A Maxwell Forestry Extension & Paducah/Paducah

pervasivetechinstitute
AT INDIANA UNIVERSITY

Wednesday, November 11, 2009

SCHOLARWORKS DISCOVERY MOBILE

- BOOPSIE
- DSPACE
- OAI HARVESTING
- AUTOMATED UPDATES
- IU INTELLIGENT INFRASTRUCTURE



DIGITAL LIBRARY PROGRAM
INDIANA UNIVERSITY
A Maxwell Forestry Extension & Paducah/Paducah

pervasivetechinstitute
AT INDIANA UNIVERSITY

Wednesday, November 11, 2009

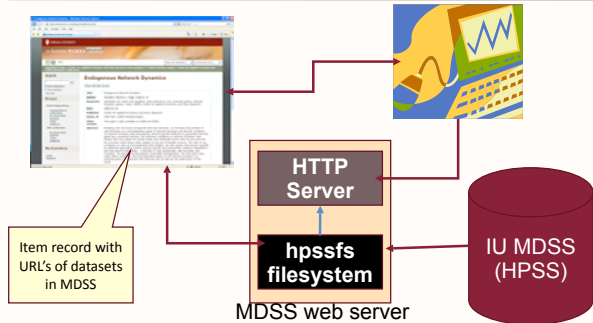
DATA DELUGE

- As volume of digital scientific data increases so increases opportunities for data-driven science
- Funding agencies beginning to mandate that research data products be made publically available
 - “The National Science Foundation is committed to the principle that the various forms of data collected with public funds belong in the public domain.”
 - <http://www.nsf.gov/sbe/ses/common/archive.jsp>

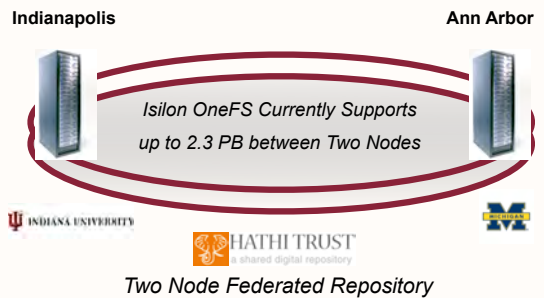
DATA CURATION

- IU Scholarly Data Services
 - IUScholarWorks Data - Publishing Services
 - Consortial Data Curation/Publishing (HathiTrust)
 - Consortial Data Curation (CIC Research File System)

IU SCHOLARWORKS DATA



CONSORTIAL DATA - HATHITRUST



CONSORTIAL DATA – CIC RESEARCH FILESYSTEM

- 7 Node Wide Area File System
 - Indiana University
 - Michigan State University
 - Ohio State University
 - Penn State University
 - University of Illinois
 - University of Iowa
 - University of Minnesota
- Fedora Curation Services
- GPFS-WAN – Global Parallel File System Wide Area Network
- SoFS – Scale Out File System

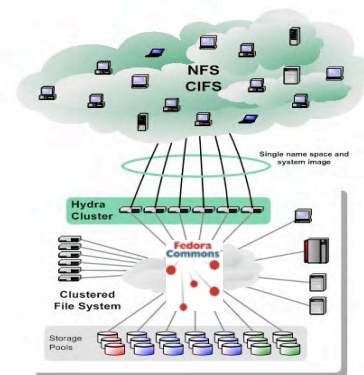


Diagram from IBM



ABOVE CAMPUS

Above-Campus Services: Shaping the Promise of Cloud Computing for Higher Education
 Brad Wheeler and Shelton Waggener – Educause Review 44(6)

Illustration by Randy Lyhus ©2009
 EDUCAUSE Review, Nov/Dec 2009

Ψ DIGITAL LIBRARY PROGRAM
INDIANA UNIVERSITY
A Maxwell Foundation Institution of Pervasives@IndianaU

pervasivetechinstitute
AT INDIANA UNIVERSITY

Wednesday, November 11, 2009

IS YOUR UNIVERSITY OR LIBRARY INVESTIGATING CLOUD SERVICES FOR DATA?

To vote: Text a **CODE** to **99503** or go to <http://poll4.com>
or tweet @poll and a CODE

YES **110352**

NO **110353**

NOT SURE **110354**

http://www.polleverywhere.com/multiple_choice_polls/MTk1NzE4NTU1Ng

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

Ψ DIGITAL LIBRARY PROGRAM
INDIANA UNIVERSITY
A Maxwell Foundation Institution of Pervasives@IndianaU

pervasivetechinstitute
AT INDIANA UNIVERSITY

Wednesday, November 11, 2009

DISCUSSION



Ψ DIGITAL LIBRARY PROGRAM
INDIANA UNIVERSITY
A Maxwell Foundation Institution of Pervasives@IndianaU

pervasivetechinstitute
AT INDIANA UNIVERSITY

Wednesday, November 11, 2009

**Data, Data Everywhere:
 Migration and System Population Practices**

Questions?

All questions will be posted with presenter answers on the NISO website following the webinar:

<http://www.niso.org/news/events/2009/datasystems09>

NISO
How the information world CONNECTS

Wednesday, November 11, 2009