Batch Loading MARC Records for Electronic Resources at the University of Michigan Library
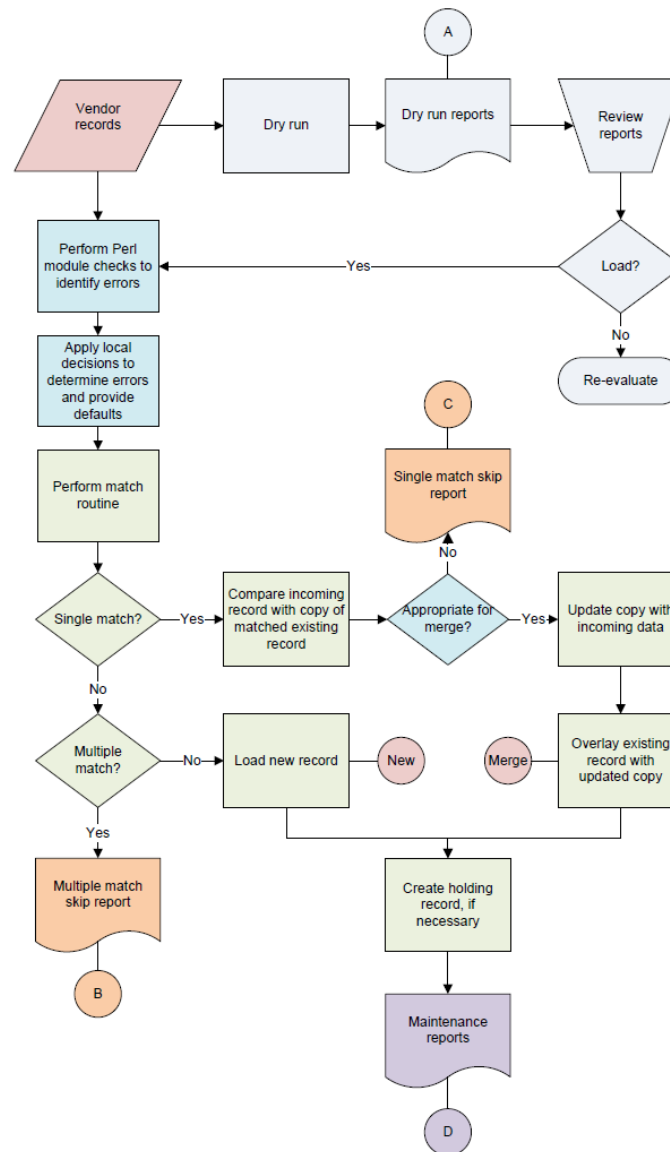
Connie McGuire

Electronic Resources Cataloging

Vicki Dillon

Library Systems

M Library

Batch Loading Procedure Workflow

1/16/10

Goal

Implement a standard validation process for the timely batch loading of MARC records for large sets of electronic resources that we receive from a variety of sources so users are able to retrieve the records when searching the catalog.

M Library

Background Information

The University of Michigan Library:

- Follows the single record approach

- Shares the catalog with three other libraries

- Uses a batch process to load records for large sets of electronic resources

- Has an Ex Libris Aleph catalog

MLibrary

Broad Overview

- Record retrieval

- Initial review

- Validation

- Matching, loading, and merging

- Post-load error resolution

- Dry run

- Vendor profile

- Tracking database

MLibrary

Record Retrieval Options

- FTP from ftp site

- Download from website

- Email attachment

MLibrary

Initial Review - Systems

Uses MarcEdit and/or a Perl program for viewing MARC records to:

- Confirm validity of MARC records
- Identify character sets and issues
- Get record counts

MLibrary

Initial Review - Cataloging

- Unique 035 record identifier
- 710 or 730 vendor added entry
- 856 with a $z note
  - 856 $z Access to the <vendor name> online version is restricted

MLibrary

Local Policy

- Corrections are made to:
  - Access points
  - Data elements that affect record retrieval
- Corrections are not made when the errors do not affect access

M Library

Validation

Perl

- Lint and Errorcheck

- Local program that makes corrections and supplies defaults

M Library

Errors and Actions

Errors are linked to one of the following actions:

- Ignore the error

- Remove the data

- Use a local default

- Load the record and report the error

- Load the record and suppress it from public view

- Terminate processing

1/16/10

969 Samples

- 96999 $$aREPORT preload validation error$$b020: Subfield a has bad checksum, 1212121212(cased)

- 96999 $$aSUPPRESS preload validation error$$b245: Field is not repeatable

- 96999 $$aREPORT local routine error$$brecord lacked 007, created default 007

MLibrary

Match Points - Existing

Existing catalog record:

- System control number – 035

- LCCN – 010 $a; 010 $z

- ISBN – 020 $a; 020 $z

M Library

Match Points - Incoming

Incoming vendor record:

- System control number – 035; 776 $w (OCoLC)

- LCCN – 010 $a; 776 $w (DLC)

- ISBN – 020 $a; 020 $z, 776 $z

MLibrary

Match Processing

Aleph matching mechanism

- Match criteria

- Each incoming record is placed in one of three files:

  – Matched more than one existing record

  – Matched only one existing record

  – Matched no existing record

M Library

Multiple Match Processing

The holdings location information associated with all existing matched records is examined to determine if the incoming record can be:

- Treated as a new record

- Treated as a single match

- Rejected and reported

M Library

Multiple Match Skip Report

Flowchart "B"

LDR ^^^^^nam^^2200349^a^4500

006 m^^^^^^^d^^^^^^^^

007 cr^cn|||||||||

008 990514s2000^^^^enkab^^^sb^^^^001^0^eng

010 $$z777888

020 $$a1234567890

035 $$a(EVND)xyz5000328

040 $$aEVND$$cEVND$$dMiU

1001 $$aFine, Matthew.

24513 $$aNation and society$$h[electronic resource]/$$cMatthew Fine.

260 $$aCambridge, UK ;$$aNew York, NY :$$bCambridge University Press,$$c2000.

651 0 $$aRhine River Valley$$xSocial conditions.

7300 $$aElec Vendor Digital Library.

85640 $$zAccess to the Elec Vendor Digital Library online version restricted; authentication may be required:$$uhttp://site.evendor.com/lib/umich/Doc?id=5000328

96999 $$aREPORT preload validation error$$b245: First word, nation, does not appear to be an article, check 2nd indicator (3).

MAT $$a004060332,006060447

M|Library

Single Match Processing

A locally written Perl program:

- Examines and compares a copy of the matched existing record and the incoming record

- Identifies conflicts

- Merges incoming fields into the copy

- Inserts appropriate 969 field(s) into the copy

- Reports and rejects, or prepares the merged record for loading

1/16/10

M Library

Merged Record (without error)

Flowchart "Merge"

LDR ^^^^^nmm^^2200349^a^4500

006 m^^^^^^^^d^^^^^^^^

007 cr||n|||||||||

008 001027s2000^^^^enka^^^^sb^^^^001^0^eng

020 $$a0415236978 (hardcover)

020 $$a9780001234567 (ebook)

035 $$a(EVND)xyz2002502

040 $$aEVND$$cEVND$$dEYM

1001 $$aGannon, Julia.

24510 $$aNetworking children$$h[electronic resource]:$$bcare and family life in later childhood/$$cJulia Gannon.

260 $$aLondon :$$bRoutledgeFalmer,$$c2000.

300 $$avii,232p. :$$bill. ;$$c23 cm.

650 0 $$aChildren$$xFamily relationships.

7300 $$aElec Vendor Digital Library.

85641 $$zAccess to the Elec Vendor Digital Library online version restricted; authentication may be required:$$uhttp://site.evendor.com/lib/umich /Doc?d=2002502

M Library

Single Match Skip

Conditions causing the merge to be skipped:

- Existing record already has URL for vendor

- Existing record is suppressed

- Existing record has no holdings

MLibrary

Single Match Skip Report

Flowchart "C"

LDR ^^^^^nam^^22003254a^4500

006 m^^^^^^^d^^^^^^^

007 cr||n|||||||||

008 000330s2001^^^^enkab^^^sb^^^^001^0^eng

020 $$a1212121212(cased) :$$c45.00

035 $$a(EVND)xyz2002446

040 $$aEVND$$cEVND$$dMiU

1001 $$aWilliams, Martin,$$d1971-

24510 $$aCharting cyberspace$$h[electronic resource]$$cMartin Williams.

260 $$aLondon ;$$aNew York :$$bRoutledge,$$c2001.

650 0 $$aComputers and civilization.

7300 $$aElec Vendor Digital Library.

85640 $$zAccess to the Elec Vendor Digital Library online version restricted; authentication may be required:$$uhttp://site.evendor.com/lib/umich/Doc?id=2002446

96999 $$aREPORT preload validation error$$b020: Subfield a has bad checksum, 1212121212(cased)

96999 $$a**SKIP** MERGE $$bmatched sys rec 003724124 already has an 856 for this vendor, skipped

M Library

New Record (with errors)

Flowchart "New"

LDR ^^^^^nmm^^2200349^a^4500

001 xyz2002502

003 EVND

006 m^^^^^^^d^^^^^^^

007 cr||n|||||||||

008 001027s2000^^^^enka^^^^sb^^^^001^0^eng

010 $$z987654

020 $$a9780415236977 (hardcover)

035 $$a(EVND)xyz2002502

040 $$aEVND$$cEVND$$dMiU

1001 $$aGannon, Julia.

24510 $$aNetworking children /$$h[electronic resource]$$cJulia Gannon.

24510 $$aNetworking children$$h[electronic resource]:$$bcare and family life in later childhood/$$cJulia Gannon.

260 $$aLondon :$$bRoutledgeFalmer,$$c2000.

650 0 $$aChildren$$xFamily relationships.

7300 $$aElec Vendor Digital Library.

85640 $$zAccess to the Elec Vendor Digital Library online version restricted; authentication may be required: $$uhttp: //site.evendor.com/lib/umich/Doc?id=2002502

96999 $$aSUPPRESS preload validation error$$b245:Field is not repeatable.

96999 $$aSUPPRESS local routine error$$bLDR/06 is m

96999 $$aREPORT local routine error$$brecord lacked 007, created default 007

1/16/10

Load Reports

Flowchart "D"

Suppressed Record Report

004870199 24510 $$aNetworking children /$$h[electronic resource]$$cJulia Gannon.

004870199 24510 $$aNetworking children$$h[electronic resource] :$$bcare and family life in later childhood/$$cJulia Gannon.

004870199 96999 $$aSUPPRESS preload validation error$$b245: Field is not repeatable.

004870199 96999 $$aSUPPRESS local routine error$$bLDR/06 is m

004870199 96999 $$aREPORT local routine error$$brecord lacked 007, created default 007

Non-Suppressed Record Report

006060450 24510 $$aAnimals in space$$h[electronic resource]$$cMary Jones.

006060450 96999 $$aREPORT preload validation error$$b020: Subfield a has bad checksum, 1765748590(cased)

M Library

Load Report Review

- Catalogers decide:
  - Load as new
  - Merge with existing Aleph record
  - Do not load
  - Unsuppress the loaded record
- Catalogers communicate back to Systems
- Systems processes previously unloaded records through a second pass

1/16/10

M Library

Vendor Profile

- Record containing vendor specific parameters
- Resides on server
- Accessed by all load steps
- Sample profile entries:

  – vprof_7xx   710 or 730, which to add including indicators

  – vprof_856_subz   value for 856$z, e.g. Access to the [vendor] online version restricted; authentication may be required:

  – vprof_match_routine   match routine from tab_match for p_manage_36, e.g. MULT1

M Library

Dry Run

- Identifies and addresses problems before the actual load
- Reports
  - Flowchart item A, Error Summary Report (with counts)
    - 17 96999 $$aSUPPRESS local routine error$$bLDR/06 is m
    - 2 96999 $$aSUPPRESS preload validation error$$b245: Field is not repeatable.
    - 1 96999 $$aSKIP MERGE $$bmatched sys rec 003724124 already has an 856 for this vendor
  - Sample Records Reports
    - Some to be loaded as new
    - Some to be merged with existing
- Review
  - Reports delivered to Cataloging staff
  - Catalogers communicate findings to Systems

M Library

Tracking Database

- Each load has a record in the database
- The database tracks information about the:
  - Dry run
  - Loading of records into production
  - Number of records loaded as new and number of records that were merged

MLibrary

Discussion Topics

- What are the strengths and limitations of the tools that are available for retrieving, loading and reviewing large record sets?

- Given that some record sets require a significant amount of effort to load, how "free" are the records that vendors include in license agreements?

- What level of staff expertise is required to manage and monitor batch record loads?

- How can we monitor the quality of the records in large sets?

- How can librarians collaborate with vendors to improve the quality of their record sets?

1/16/10

MLibrary

THANK YOU

Connie McGuire

cmcguire@umich.edu

Vicki Dillon

vjdillon@umich.edu

M Library