

James Madison University

JMU Scholarly Commons

Libraries

Libraries

2-11-2021

Metadata Analysis for Pre-Migration Cleanup

Rebecca B. French

James Madison University, frenchrb@jmu.edu

Follow this and additional works at: <https://commons.lib.jmu.edu/letfspubs>



Part of the [Cataloging and Metadata Commons](#)

Recommended Citation

French, Rebecca B., "Metadata Analysis for Pre-Migration Cleanup" (2021). *Libraries*. 194.
<https://commons.lib.jmu.edu/letfspubs/194>

This Presentation is brought to you for free and open access by the Libraries at JMU Scholarly Commons. It has been accepted for inclusion in Libraries by an authorized administrator of JMU Scholarly Commons. For more information, please contact dc_admin@jmu.edu.

Metadata Analysis for Pre-Migration Cleanup

Rebecca B. French, James Madison University

Where Are You?

- A. Recently completed a migration
- B. In the middle of a migration
- C. Will be migrating soon
- D. No migration planned





Our Recent Migration

- Migrated from III Sierra to Ex Libris Alma and Primo VE
- January – July 2020
- 11-member Alma Implementation Team
- 1.5+ million bibliographic records
- 1.6+ million item records
- ~5,600 checkin records
- ~980,000 electronic resource bibs

What Is Metadata Analysis?

- Metadata: “a set of data that describes and gives information about other data”
- Analysis: “detailed examination of the elements or structure of something, typically as a basis for discussion or interpretation;” “the process of separating something into its constituent elements”
- Accurately identify problems to inform decision making and action





Overview

- Metadata assessment criteria
- Prioritization
- Metadata analysis techniques
- Cleanup (not covered)

Metadata Assessment Criteria

Criteria for Metadata Quality

- Completeness
- Accuracy
- Accessibility
- Conformance to expectations
- Consistency
- Timeliness
- Provenance



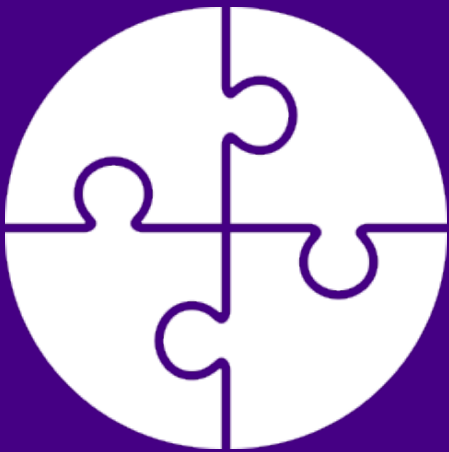
Criteria for Metadata Quality

- **Completeness**
- **Accuracy**
- Accessibility
- **Conformance to expectations**
- **Consistency**
- Timeliness
- Provenance



Completeness

- “The element, property, and/or attribute is present”
- All required elements are present
- Know required fields in the new system





“Target” icon by Libby Ventura
from [The Noun Project](#)

Accuracy

- “Information is correct both semantically and syntactically”
- Metadata means what it is intended to mean
- Metadata has the correct structure

DLF Metadata Assessment Working Group. (2017). *Metadata Assessment Framework and Guidance*.
<http://dlfmetadataassessment.github.io/Framework>

Conformance to Expectations

- “Values adhere to the expectations of your defined user communities (both internal and external)”
- Metadata conforms to standards (e.g., MARC) and local guidelines



Consistency

- “Semantic and structural values and elements are represented in a consistent manner across records. Values are consistent within your domain”
- Values are used consistently across all records





Prioritization

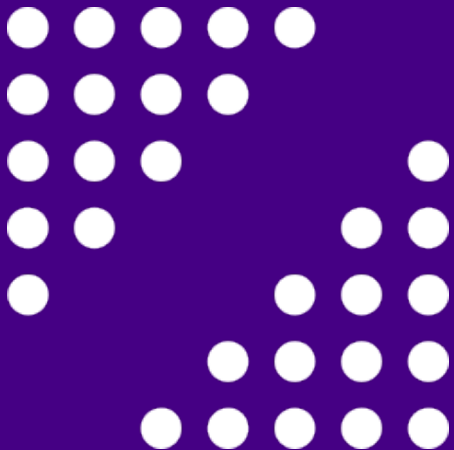


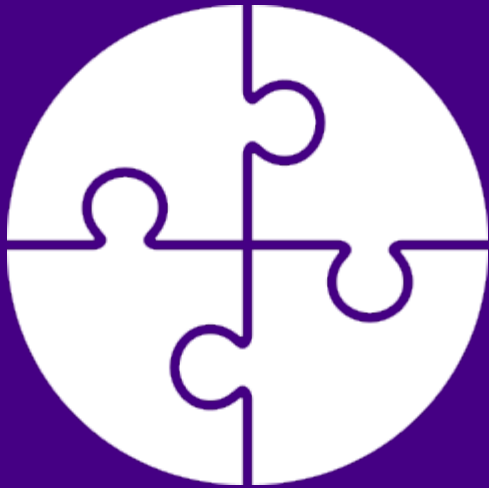
Keep the Goal in Mind

- Data migrates the way you intend
- Meaning of data is not changed
- Minimize data loss

What Records Are Migrating/ Not Migrating?

- What records will you be extracting from the old system? How will they be identified?
- Do not spend time cleaning up records that will not migrate
- Examples
 - Missing, lost, damaged items
 - Withdrawn items
 - Brief records





What Data Is Missing?

- What fields are required in the new system? Are they present in all migrating records?
- Examples
 - 245 field
 - Record identifiers
 - Call numbers
 - Barcodes

What Data Will Be Lost?

- What data cannot migrate in its current format or field?
 - Know your target format/schema
 - Consider locally defined fields and values
- Examples
 - Non-MARC fields
 - Locally defined material type values





Where Will Data's Meaning Be Incorrect or Unclear?

- Data that could be interpreted incorrectly if left in its current format or field
- Field scope creep
- Examples
 - Control number without organizational identifier (001 without 003)
 - Using location codes for things that are not physical locations

What Data Inconsistencies Will Affect Migration?

- Inconsistencies that impact what records migrate or whether they migrate correctly
 - Between parent/child records (e.g., bib/item, bib/checkin)
 - Between sibling records (e.g., item/checkin on the same bib)
- Examples
 - Withdraw codes on bib/item
 - Internet location code and 856
 - Location codes to assign items to the correct holdings record





What Will You Need after Migration?

- Consider what processes will be different in the new system
- Facilitate additional post-migration cleanup
- Examples
 - Match points for loading records
 - Electronic resource collection identifiers

Prioritization

- What records are migrating/not migrating?
- What data is missing? (completeness)
- What data will be lost? (conformance to expectations)
- Where will data's meaning be incorrect or unclear? (accuracy)
- What data inconsistencies will affect migration? (consistency)
- What will you need after migration?





Metadata Analysis Techniques

Search Your System

- Advanced queries in ILS
 - Sierra: Create Lists function
 - Alma: advanced search, indication rules
- Gather a set of records that have a particular problem
- Good for:
 - Finding missing fields
 - Identifying brief records



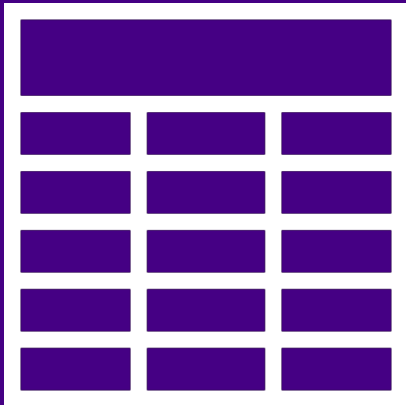
Search Your System – Examples

- Missing fields
 - No 245 field
 - No 001 field
- Brief records
 - No subject headings
 - Receive date but no cataloged date
- Remember to limit the search to only the items that will be migrating



Scrutinize Spreadsheets

- Export metadata from ILS and review in table form
- Good for finding:
 - Inconsistent values
 - Multiple occurrences of non-repeatable fields
 - Coding mismatches within a record
 - Coding mismatches between parent/child records



Scrutinize Spreadsheets – Example 1

Reviewing OCLC numbers

A	B
BIB #	OCLC # (001)
b13269173	1057098555
b11914634	1086845312
b13567822	1126787498
b19481743	2005051284
b11604943	
b19400238	374276c
b1941481x	4499803c
b19418292	5014459c
b19672135	5396559c
b19783139	5750156c
b1307779x	8720569;"1519526"
b10460512	
b10585527	



Scrutinize Spreadsheets – Example 1

Reviewing OCLC numbers

A	B
BIB #	OCLC # (001)
b13269173	1057098555
b11914634	1086845312
b13567822	1126787498
b19481743	2005051284
b11604943	
b19400238	374276c
b1941481x	4499803c
b19418292	5014459c
b19672135	5396559c
b19783139	5750156c
b1307779x	8720569;"1519526"
b10460512	
b10585527	

↓ A
Z

← Not an OCLC number

← 001 – field contains spaces

Not OCLC numbers

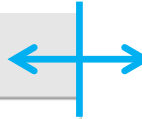
← Duplicate 001 field

No 001 field

Scrutinize Spreadsheets – Example 2

Finding multiple barcode fields

	A	B
1	Item #	Barcode
2	i15017965	1000813995
3	i15038993	1000814261
4	i15047970	1000814574
5	i15050622	1000814506
6	i15051717	1000827712
7	i15058487	1000474591
8	i1506332x	1000814202
9	i13719282	1000648991
10	i13719865	1000814275
11	i13721227	1000814947
12	i13726158	1000815221
13	i13727023	1000814377
14	i13727035	1000814386

 Resize column to fit data



Extra space shows some values are longer than standard barcode

Scrutinize Spreadsheets – Example 2

Finding multiple barcode fields

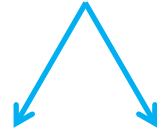
	A	B
1	Item #	Barcode
8323	i10009486	1000656427;1000151386
10589	i10071866	1000663582;1000031207
16726	i10145564	1000815527;1000115351
31677	i10265806	1000726357;1000068190
46625	i11887928	1000726379;1000191121
49097	i1137035x	1000726752;1000338743
68019	i10508995	1000664013;1000182962
74793	i10562047	1000663979;1000035251
83069	i12287854	1000663572;1000198988
90018	i10604613	1000660414;1000023656
118476	i11527274	1000688902;1000075026
121603	i11436943	1000174237;1000451760
133746	i17191403	1000737908;1000555578

← Filter for “;”
to show
repeated fields

Scrutinize Spreadsheets – Example 3

Checking withdraw code
consistency

Values should be the
same in both columns



	A	B	C	D	E
1	BIB #	BCODE3	ICODE2	049 n	NOTE(ITEM)
2	b18368578	y	n	WITHDRAWN 5/29/2019 PER LIAISON REQUEST	Wiley Online Library
3	b10037305	z	-	WITHDRAWN AND REPLACED 7/23/2019	
4	b27133461	z	-		Emerald Business
5	b32689263	z	-		Brill E-books
6	b33745031	z	-		Gale Virtual Reference Library
7	b33745043	z	-		Gale Virtual Reference Library
8	b27902857	z	-		
9	b30962729	z	-	WITHDRAWN, STREAMING RIGHTS EXPIRED 3/18/2020	
10	b12980316	p	n	WITHDRAWN PER FY20 CDC BOUND JOURNAL REVIEW P	
11	b12980316	p	n	WITHDRAWN PER FY20 CDC BOUND JOURNAL REVIEW P	



“WITHDRAWN” note should
be present in both columns

Scrutinize Spreadsheets – Example 3

Checking withdraw code
consistency

=EXACT(B2,C2)



	A	B	C	D	E	F
1	BIB #	BCODE3	ICODE2	B&C MATCH?	049 n	E HAS WITHDRAWN?
2	b18368578	y	n	FALSE	WITHDRAWN 5/29/2019 PER LIAISON REQUEST	TRUE
3	b10037305	z	-	FALSE	WITHDRAWN AND REPLACED 7/23/2019	TRUE
4	b27133461	z	z	TRUE		FALSE
5	b32689263	z	-	FALSE		FALSE
6	b33745031	z	-	FALSE		FALSE
7	b33745043	z	-	FALSE		FALSE
8	b27902857	z	-	FALSE		FALSE
9	b30962729	z	-	FALSE	WITHDRAWN, STREAMING RIGHTS EXPIRED 3/18/2020	TRUE
10	b12980316	p	n	FALSE	WITHDRAWN PER FY20 CDC BOUND JOURNAL REVIEW P	TRUE
11	b12980316	p	n	FALSE	WITHDRAWN PER FY20 CDC BOUND JOURNAL REVIEW P	TRUE



=ISNUMBER(SEARCH("WITHDRAWN",E2))



Delve into Databases

- Use available capabilities for direct querying of the database (e.g., SierraDNA)
- Can also export data and create your own database
- Good for:
 - Matching up data that is not connected through built-in system searching capabilities
 - Checking code consistency between sibling records (item/checkin)

Delve into Databases – Example

Checking location code consistency
between item and checkin records
on the same bib

	A	B	C
1	BIB #	CHECKIN_LOCATIONS	ITEM_LOCATIONS
2	b10599307	cl =	cl
3	b10608412	cl,clrf =	cl,clrf
4	b10614631	cl,clrf,ecrf =	cl,clrf,ecrf
5	b22246381	clp	clpe
6	b12978279	clp	clpeb
7	b12985144	clp	clpeb
8	b13112958	clp	clpeb
9	b21565120	clp	ecp
10	b10579564	cl	clpeb,ecl
11	b11828213	cl,clsp,ecrf,scrf	cl2a,clsp,ecrf
12	b13433635	clci,clpe,ecp	oss
13	b13651882	clci,clsp	clsp,oss
14	b13047231	clofc	clpeb

≠

Delve into Databases – Example

Checking location code consistency
between item and checkin records
on the same bib

BIB #	CHECKIN_LOCATIONS	ITEM_LOCATIONS
b22246381	clp	clpe

← Mismatched checkin
and item locations

Delve into Databases – Example

Checking location code consistency
between item and checkin records
on the same bib

BIB #	CHECKIN_LOCATIONS	ITEM_LOCATIONS
b22246381	clp	clpe

← Mismatched checkin
and item locations

Created a second holdings record with no items

	<input type="checkbox"/>	ID	▲ Library	↕ Location	↕ Call Number	No. of Items	Available	
1	<input type="checkbox"/>	22144904130006271	Carrier	Carrier-Browsing Periodicals	PS501 .C63	1	1	...
2	<input type="checkbox"/>	22144904140006271	Carrier	Carrier-Periodicals	-	0	0	...

Delve into Databases – Example

Checking location code consistency between item and checkin records on the same bib


BIB #	CHECKIN_LOCATIONS	ITEM_LOCATIONS
b22246381	clp	clpe

← Mismatched checkin and item locations

Created a second holdings record with no items

	ID	Library	Location	Call Number	No. of Items	Available		
1	<input type="checkbox"/>	22144904130006271	Carrier	Carrier-Browsing Periodicals	PS501 .C63	1	1	...
2	<input type="checkbox"/>	22144904140006271	Carrier	Carrier-Periodicals	-	0	0	...

Resulting in a misleading public display →

LOCATIONS 

- Carrier
Available , Browsing Periodicals ; PS501 .C63 >
- Carrier
May be available , Periodicals >

Metadata Analysis for Pre-Migration Cleanup

Metadata assessment criteria – Prioritization – Metadata analysis techniques

Rebecca B. French
frenchrb@jmu.edu