

## Creating New Policies and Automating Standardization for Serials Holdings Records

**Jamie Carlstone**

University of Illinois at Urbana-Champaign, Urbana, IL, USA

[jdc6@illinois.edu](mailto:jdc6@illinois.edu)



Copyright © 2019 by Jamie Carlstone. This work is made available under the terms of the Creative Commons Attribution 4.0 International License:

<http://creativecommons.org/licenses/by/4.0>

---

### Abstract:

*The University of Illinois at Urbana-Champaign (UIUC) Library has a large print serials collection consisting of 317,066 bibliographic titles with 474,190 holdings and 3,638,833 items. At UIUC, serials holdings records had two problems potentially causing hindrance to discovery and data interoperability: a lack of consistency in the local notes in the MARC Format for Holdings Data (MFHD) 852 field, and missing standard holdings statements in the MFHD 866 field.*

*Assessment of local notes revealed that there were 58,723 unique notes in the 474,190 holdings records. This was caused by inconsistent policies and incorrect data entry. After categorizing notes into 22 types, the library came up with standardized local notes policies based on three principles: 1) notes do not duplicate information in the bibliographic record, 2) notes do not refer to information in a different holding record, and 3) notes are not created to conform to discovery systems. The policies also reiterated the need for each MFHD to have a summary written to ANSI/NISO Z39.71-2006 (R2011) Holdings Statements standard.*

*The data clean-up work was performed in two ways. First, using Python and OpenRefine the local notes contained in the MFHD 852 field, subfields x and z were batch edited. The updated MFHD 852 fields were batch replaced with notes that conform to the new policies. Second, summary of holdings statements were written to ANSI/NISO standards using Python. The statements were batch added to 105,404 holdings records as the MFHD 866 field. This cut the number of print serials records with missing summary of holdings statements from 277,282 to 171,878.*

*This paper will share how the library created the policies on local serials notes and automated data clean-up work, and will discuss lessons learned from the project.*

**Keywords:** serials, local notes, policies, holdings records, automated data processing

---

## **Introduction**

The University of Illinois at Urbana-Champaign (UIUC) is the state's flagship public university, located in east central Illinois. The university was founded in 1867 and is Illinois' land grant institution. It is the largest university in the state with a total student population of 47,238 (UIUC, 2019) in a community of approximately 207,000 residents.

The UIUC Library is the largest public university academic library in the United States with over 14 million volumes held (UIUC, 2019). The UIUC Library buildings consists of a Main Library and over a dozen departmental libraries and special collections. The largest collections are in the Main Library and the Oak Street Facility, a high density storage facility opened in 2004 with over four million volumes held (UIUC Library, 2019). Over the years, many departmental libraries were closed and merged into new libraries or absorbed into other collections.

The UIUC Library's print serials collection consists of 317,066 bibliographic titles, 474,190 holdings, and 3,638,833 items. The holdings records are in MARC Format for Holdings Data (MFHD). UIUC currently uses the Voyager Integrated Library System with VuFind as a discovery layer. The Library is currently in the process of migrating systems to Alma with PrimoVE as the discovery layer.

At UIUC, assessment of the records revealed two major issues potentially hindering discovery and data interoperability: a lack of consistency in free text local notes and missing summary of holdings statements. This paper will describe work by the UIUC Library standardizing policies and automating clean-up in anticipation of a data migration.

### **Local Notes Challenges**

Free text local notes for all item types are contained in the 852 field of the MFHD in either subfield x: suppressed from public view or subfield z: public note. Reports of the holdings notes revealed that there were 58,723 unique notes in the 474,190 holdings records. Local notes vary widely in scope, including: information about location, related titles, marking patterns, analyzed volumes, bound withs, other copies, or the order status. Furthermore, the documentation for local notes was created in 2007 and was due for an update. The use of local notes varied across the departments of the library, so rules and standards were applied inconsistently.

Analysis of local notes indicated several problems. First, many of the notes were redundant, because the information is also in the bibliographic record (e.g. title ceased) or the purchase order (e.g. cancelled). Although the information could be correct, it was unnecessary to repeat in a note. In the first example (figure 1) the information about the volumes held clearly lists holdings starting in 1979, so the note "Copy 1 has no. 49 (1979) to date" is redundant. The more places that data could be in the catalog, the more opportunity for errors—these notes can start as simple redundancies and become incorrect over time if they are not maintained. Since there was never a workflow to maintain local notes, this was a common occurrence.

Figure 1. VuFind Local Holding Display 1

<b>Location:</b>	Oak Street Library [request online]
<b>Call Number:</b>	Q. 480.0705 BU <a href="#">Text me this call number</a>
<b>Copy:</b>	1
<b>Notes:</b>	Copy 1 has no.49 (1979) to date Unbound issues in Classics Library
<b>Library Has (Volumes):</b>	no.112-123 (2000-2003) no.92-111 (Incompl.) (1993-1999) c.1 no.73-91 (1987-1993) c.1 no.67-72 (1985-1986) c.1 no.61-66 and Index (1980-1984) c.1 no.55-60 (1981-1982) c.1 no.49-54 (1979-1980) c.1
<b>Status:</b>	Available

Second, information was simply incorrect. Incorrect local note data was often a result of putting location information in a local note, then moving the item to a different library (most often due to a library merger or closure) without updating the note. The result can direct patrons to the wrong location or even libraries that no longer exist. In figure 1, the location for all holdings is the Oak Street Library, but the note says “Unbound issues in Classics Library.” There are no longer unbound issues since the title died in 2003 and the Classics Library was absorbed into another departmental library in 2016. In this example, the first note is not needed since the volumes are listed and the second note contains outdated information.

Third, notes were also created for specific discovery systems and were no longer useful after migrations. For example (figure 2), the current discovery system, VuFind, cannot perform call number searches. When a serial has an earlier or later title local notes direct patrons to search for the other title by call number, which could only be done in the previous catalog. The information is correct and lets patrons know about other titles, but VuFind does not have the functionality to make it useful. These notes, while technically correct, are not helpful to the patron’s online discovery.

Figure 2. VuFind Local Holdings Display 2

<b>Location:</b>	Main Stacks
<b>Call Number:</b>	Q. 020.5 BUI <a href="#">Text me this call number</a>
<b>Copy:</b>	1
<b>Notes:</b>	For earlier vols. see 020.6ASF For later vols. see 020.5BUI1
<b>Library Has (Summary):</b>	v.2(1949)-v.65(1969)
<b>Library Has (Volumes):</b>	v.58-65 (1968-1969) v.46-57 (1965-1967) v.34-45 (1964-1965) v.22-33 (1957-1960) v.13-21 (1954-1956) v.2-12 (BD W/O NO.3) (1949-1953)
<b>Status:</b>	Available

Fourth, since local notes are free text, they also contained inconsistencies in spelling and format. This looks messy and makes it difficult to run reports or batch clean-up data. There were at least five ways the Social Sciences, Health, and Education Library were formatted in notes:

- Last year in Social Sciences, Health, and Education Library
- Last 3 years in Education & Social Sciences Library
- Last year in Social Sciences, Health & Education Library
- Last 5 years and recent issues in SSHEL Library
- Next yr. in Education Library (Reference)

It was clear from looking at the issues with local notes that changes had to be made to the policies so that new poorly formatted notes would not be added. However, merely updating the policy would not address the legacy data in the catalog. The next goal was fixing existing notes to adhere to new policies through batch processes. The work began with the policy changes, then batch work was done to make existing data compliant with new local standards.

### **Local Notes Step 1: Policy Changes**

The first step in cleaning up the local notes for serials is to create consistent policies. With such a large collection across many departmental libraries, this can be a challenge. A working group made up of public services, cataloging, and acquisitions staff was formed to update the Library's local notes policies, focusing on three principles:

#### 1) Notes do not duplicate information

Assessment of the notes revealed information was often duplicated in the bibliographic record or in purchase orders. This is information like "title ceased" or "cancelled." By the new standards, the bibliographic record or purchase order conveys information about the work and order, while the holdings conveys information about the library's volumes.

However, one exception was made to this principle. The UIUC library is not a participant in the Cooperative Online Serials (CONSER) of the Program for Cooperative Cataloging (PCC), therefore cannot edit PCC records. For records where no more items would be received, regardless of the reason, the note “Library no longer receives new items” is added. This way it is clear the title ceased even if the bibliographic record is not updated.

2) Notes do not refer to information in a different holding record

A common practice in the holdings notes was to refer to another copy’s holdings (e.g. a note in Copy 1 MFHD record that says “Copy 3 has...”). This practice becomes especially problematic when changes are made to the other holding but the note is not updated. At best, the information was redundant and cluttered the display. At worst, the information was wrong; assessment of the notes found that many were referring to copies that were withdrawn or transferred. Figure 3 shows a holding where copy 3-7 were withdrawn, but copies 1 and 2 still reference 3-7 in notes. Each holding, or copy, should describe just that copy. This best exemplifies the Functional Requirements for Bibliographic Records (FRBR) concept of describing the item separately so that relationships between copies can be understood: “Defining item as an entity enables us to separately identify individual copies of a manifestation, and to describe those characteristics that are unique to that particular copy and that pertain to transactions such as circulation, etc. involving that copy. Defining the entity called item also enables us to draw relationships between individual copies of manifestations” (Byrum et al., 2009, pp. 24–25). The bibliographic record allows for the link to be made between copies, so it is not necessary to make this link in the holding note.

Figure 3. VuFind Local Holdings Display 3

<b>Location:</b>	Main Stacks
<b>Call Number:</b>	510 M433M <a href="#">Text me this call number</a>
<b>Copy:</b>	2
<b>Notes:</b>	COPY 1 HAS 1 1950/1960. CEASED COPY 2 HAS 1 1950/1960. CEASED. COPY 3 HAS 1 1950/1960. CEASED. COPY 4 HAS 1 1950/1960. CEASED. COPY 5 HAS 1 1950/1960. CEASED. COPY 6 HAS 1 1950/1960. CEASED. COPY 7 HAS 1 1950/1960. CEASED. FOR LATER VOLS ENTER DSC/510M433M1
<b>Library Has (Summary):</b>	v.1(1950-1960)
<b>Library Has (Volumes):</b>	v.1 (1950-1960) c.2
<b>Status:</b>	Available

3) Notes are not created to conform to discovery systems

Local notes were written over the years to conform to the functionality of the current discovery system. If a discovery system did not display information in a functional way, a local note would be created as a workaround. Issues arose after migrating to new discovery systems and the note was no longer useful or necessary. Often, policies also carried over into new systems despite the irrelevance of the note. Now, the library strives to have discovery systems comply to the metadata, not the other way around.

In the original documentation for serials notes from 2007, there were 22 types of notes. After the policy review the number of note types went down to eight:

1. Library no longer receives new items
2. Gift or exchange cancelled on records without a purchase order
3. Back issues not retained
4. Bound with
5. Spine marking
6. Preservation action
7. Analyzed
8. Bound with other titles (where no relationship exists between titles besides the binding)

The notes were limited to information about the copy that is not included in the bibliographic record, purchase order, or elsewhere in the MHFD record. The reduction in the number of local notes cleans up the display of holdings information, adheres better to cataloging concepts such as FRBR, and streamlines serials workflows for staff.

### **Local Notes Step 2: Batch Changes**

In addition to creating new policies, automated work was done to get the existing records cleaned up and better compliant with new local note policies.

Through committee discussions, it was decided that there were three note types to be deleted:

1) “Recent issues...” notes

These refer to new issues of serials on other holdings records. The syntax and wording of this note varies, but generally it points to another MFHD’s unbound issues.

2) “For later volumes see UIUC online collection”

This was added to print serials holdings records after a cancellation project in order to direct patrons to online access. The note is not maintained well and the UIUC library uses the separate record approach for cataloging online serials.

3) “Indexed by...”

For a time, this note indicated whether the serial was indexed in Wilson or other publishers. The note is no longer relevant, so it was decided it should be removed along with other local notes.

Two note types were suppressed instead of deleted because they contained information that could be useful to staff but did not need to be in the public displays:

1) “Copy [#] has...”

Copy notes refer to another holding and was often out of date after deduplication projects. However, copy note information could also contain information about withdrawn issues, which is data the library wants to keep. This was suppressed so the library has a record of all copies once held.

2) “For earlier/later issues”

This note is used for titles changes. It is information in the bibliographic record and was not needed in the holding. However, the information was typically correct and could let staff know about title changes, so it was kept in the catalog but suppressed.

### Automating the Note Clean-up

The UIUC Library Voyager system is managed by the Consortium of Academic and Research Libraries in Illinois (CARLI). UIUC works with CARLI in order to make batch changes to the catalog. CARLI was able to send a file of all the serials MFHD records in the catalog. The MARCedit program allows for pulling data out of MARC files and into Comma Separated Values (CSVs). Using MARCedit, a spreadsheet was created with the MFHD identifier and the 852 field.

The process for changing the notes had several steps. First, Python was used to help build regular expression patterns. Second, the regular expression patterns were applied to the MFHD 852 fields in OpenRefine. Finally, the file of new 852 fields was returned to CARLI and replaced the old 852s.

“Recent issues” is the most challenging of the notes because the format and wording had so many variations. It would often also contain the name of the library where recent issues were received (e.g. Recent issues in Communications Library).

After looking at the notes, it was clear there were hundreds, or even thousands, of ways the note was worded in the catalog. The note typically began with some combination of the words in the three columns in table 1.

Table 1. Recent Issues Note

First Word	Second Word	Third Word (Optional)
Bound	ed	at
Current	ed.	held
Earlier	edition	in
Earliest	editions	located in
Early	eds	on
Last	eds.	[blank]
Last	issue	
Last [#] years and recent issues	issues	
Later	no	
Latest	no.	
Most recent	nos	
New	nos.	
Newer	vol	
Next	vol.	
Next to last	vols	
Previous	vols.	
Previous	volume	
Previous [#] years and recent issues	volumes	
Recent	year	
Recent	years	

Recent and unbound	yr	
Unbound	yr.	
Unbound and recent	yrs	
	yrs.	

Rather than attempting to find each form of the note, regular expression (regex) patterns were used. Regex allows for scripts to capture multiple possible patterns in text and can be applied in different programs to do data clean-up. For example, if you want to capture both “cat” and “cot,” a regex pattern can be used that captures both: c[ao]t.

In the case of serials notes, it is far more complex since finding every possible combination of words is a challenge in itself. In order to do this, Python was used to write regex. A Python script was written that printed each combination of the three words in table 1 in that order, separated by a pipe (e.g. Later nos.|Later ed.|Later yr., etc.). The pipe functions as an “or,” so the regex captures either wording of the note e.g. Later nos OR Later ed. OR Later yr.

The regex patterns for the other notes were not as complicated, since the wording was far more consistent. For the two notes getting suppressed, it was a simple replacement of the subfield z with a subfield x, which makes the note staff only.

Applying the regex patterns to the file of 852 fields was the next step. It is important to be specific in the formatting of the regex because the notes are only part of the 852. A bad regex pattern can edit or delete good data if it captures too many or too few characters in the note. The regex patterns were designed to only edit information that fits the pattern in subfield x or z (notes subfields). This is the regex pattern for finding all records with “UIUC Online Collection” in the notes subfields:

```
(.+) (\\$[xz]) (.+) (UIUC [Oo]nline [Cc]ollection)
```

The software OpenRefine allows for regex to batch edit spreadsheets using GREL (General Refine Expression Language). For example, this is the regex expression in a GREL statement that removes “\$zUIUC Online Collection”:

```
grel:value.replace(/(.+) (\\$[xz]) (.+) (UIUC [Cc]ollection)/, "\\$1\\") [Oo]nline
```

There were multiple regex patterns needed for this project. JSON (JavaScript Object Notation) allows for multiple GREL statements to be applied to a spreadsheet. This is the JSON that was applied to the OpenRefine spreadsheet of 852s fields for removing the “UIUC Online Collection” note:

```
{
  "op": "core/text-transform",
  "description": "Text transform on cells in column 852 using expression
grel:value.replace(/(.+) (\\$[xz]) (.+) (UIUC [Cc]ollection)/, "\\$1\\")",
  "engineConfig": {
    "mode": "row-based",
    "facets": [
      {
        "mode": "text",
        "caseSensitive": false,
        "query": "UIUC Online Collection",
```

```

        "name": "852 Unedited",
        "type": "text",
        "columnName": "852 Unedited"
    }
  ]
},
"columnName": "852",
"expression": "grel:value.replace(/(.+) (\\$[xz]) (.+) (UIUC [Oo]nline
[Cc]ollection)/,\\\"$1\\\")",
"onError": "keep-original",
"repeat": false,
"repeatCount": 10
},

```

Overall there were 18 JSON scripts designed to clean-up the five notes getting deleted or suppressed. After the JSON scripts were applied to the over 400,000 serials MFHD records, 145,158 fields were changed. The newly edited 852 fields were sent back to CARLI in a CSV file and replaced the data in the existing records, matching on the MFHD ID.

Due to human error, there are still notes in the catalog that no longer follow our policy. Many have typographical errors that make them difficult to capture using regular expression patterns. However, the automation process provided a mechanism to remove bad data and unclutter displays in the catalog.

### **Batch Creating Summary Statements**

It is the UIUC Library's official policy for serial records to always contain a summary of holdings statement in the 866 field, but many are missing this data. Holdings summaries may at first appear redundant, since the volumes held also appears in item records and are listed in the MFHD 863 fields. Historically, the summaries were used in union lists of serials that were printed, and therefore it was necessary to condense holdings to save space. Although generally people no longer use print union lists, the summaries are still useful in our current practices. They are valuable in shared collection building because it makes it easier to compare holdings between libraries. In union catalogs like WorldCat.org the summary allows for the user to see if an institution holds a volume—this is helpful for interlibrary loan. Furthermore, not every institution or discovery system uses or displays 863 fields, so the summary statements ensure holdings data is still displayed. The summary statements should be written to ANSI/NISO Z39.71-2006 (R2011) (NISO) standard, so summaries have the benefit of consistency between systems and institutions.

The original National Information Standards Organization (NISO) standard for serials was published in 1989, so the catalog contains many pre-1989 statements (NISO, 2011). The most recent update to the holdings standards was published in 2011. Although it is the policy to always have summaries, 277,282 print serials holdings records were missing the 866 field. Typically, holdings statements can only be done by human processing because they require the volumes listed in the 863 fields to be condensed. However, for some holdings, the condensing is not necessary, or very simple to do, because there is only one or two lines. This created an opportunity to automate a workflow previously done one record at a time.

The data needed for the summaries already existed in the record between the 853 field: captions and patterns and the 863 field: volumes. Using Python scripts, the data from the 853 field and 863 fields was combined and reformatted to build an 866 field that complies to NISO. Once

again, the MARCedit program allows for the easy extraction of MARC data converted to CSV to make this process possible.

For example, in this record a summary of holdings can easily be created by pulling in the caption (v.) from the 853 subfield a and the single-line holding (1) from the 863 subfield a to make a summary in the 866 subfield a:

```
853 23 $8 1 $a v.  
863 40 $8 1.5000 $a 1  
866 40 $8 1 $a v.1
```

Automated summaries were also made for 2 line holdings:

```
853 23 $8 1 $a v.  
863 40 $8 1.4998 $a 2  
863 40 $8 1.5000 $a 1  
866 40 $8 1 $a v.1-2
```

If there is a gap in the holdings, the summary could still be created with a comma between non-consecutive volumes:

```
853 23 $8 1 $a v.  
863 40 $8 1.4998 $a 3  
863 40 $8 1.5000 $a 1  
866 40 $8 1 $a v.1,v.3
```

Volumes marked including dates were also created with Python scripts:

```
853 23 $8 1 $a v. $i (year)  
863 40 $8 1.4998 $a 3 $i (1985)  
863 40 $8 1.5000 $a 1 $i (1983)  
866 40 $8 1 $a v.1(1983),v.3(1985)
```

Serials can also be enumerated by year only. Automated summaries were made for holdings of any length where there were no gaps:

```
853 23 $8 1 $a (year)  
863 40 $8 1.4996 $a 1985  
863 40 $8 1.4998 $a 1984  
863 40 $8 1.5000 $a 1983  
866 40 $8 1 $a 1983-1985
```

With the high volume of records missing summary statements, creating 866s automatically cuts down on the manual work needed to be done by staff. Staff can focus on creating summary statements that are more complex. Overall, 28 Python scripts were written to build holdings summaries, resulting in the addition of 105,404 summary statements to the catalog. This reduced the number of missing summaries from 277,282 to 171,878. Since the records never had 866 fields, it was easy for CARLI to add them to existing MFHDs.

## Conclusion

Clean and consistent serials local notes ensures quality user services. To improve user experience for browsing and acquiring serials volumes, the UIUC Library updated the local policy on serials notes and cleaned up legacy notes by exploiting existing tools and software. From this work, more than quarter million serials holdings records were updated with standardized local notes and summary notes. By creating local standards, the library hopes that users can find, identify, and acquire accurate volumes of serial titles and staff can focus on creating quality bibliographic records.

## References

Byrum, J., Jouguelet, S., McGarry, D., Williamson, N., Witt, M., Delsey, T., ... Tillett, B. (2009). Functional Requirements for Bibliographic Records. Retrieved from <http://www.ifla.org/VII/s13/frbr/>

Carlstone, Jamie. (2019). Serials Notes JSON Scripts. Retrieved from <https://github.com/jdcar/serialsnotesjson>

National Information Standards Organization. (2011). ANSI/NISO Z39.71-2006 (R2011). Holdings Statements for Bibliographic Items. Retrieved from [https://groups.niso.org/apps/group\\_public/download.php/18121/Z39.71-2006\\_\(R201%E2%80%A6\)](https://groups.niso.org/apps/group_public/download.php/18121/Z39.71-2006_(R201%E2%80%A6))

University of Illinois at Urbana-Champaign. (2019, January 29). UIUC Student Enrollment by Curriculum and Student Level. Retrieved May 20, 2019, from <http://www.dmi.illinois.edu/stuenr/class/enrsp19.htm>

University of Illinois at Urbana-Champaign Library. (2019a). General Information. Retrieved May 20, 2019, from <https://www.library.illinois.edu/geninfo/>

University of Illinois at Urbana-Champaign Library. (2019b). Oak Street High Density Storage. Retrieved May 20, 2019, from <https://www.library.illinois.edu/oak/high-density-storage/>