

UC Office of the President

UCOP Previously Published Works

Title

Enhancing Print Journal Analysis for Shared Print Collections

Permalink

<https://escholarship.org/uc/item/2b97j1k5>

Author

Jemison, Dana Lucile

Publication Date

2021-06-14

Data Availability

The data associated with this publication are within the manuscript.

Peer reviewed

Enhancing Print Journal Analysis for Shared Print Collections

The Western Regional Storage Trust ([WEST](#)), is a distributed shared print journal repository program serving research libraries, college and university libraries, and library consortia in the Western Region of the United States. WEST solicits serial bibliographic records and related holdings biennially, which are evaluated and identified as candidates for shared print archiving using a complex collection analysis process. California Digital Library's [Discovery & Delivery](#) WEST operations team (WEST-Ops) supports the functionality behind this collection analysis process used by WEST program staff (WEST-Staff) and members.

For WEST, proposals for shared print archiving have been historically predicated on what is known as an [Ulrich's journal family](#), which pulls together related serial titles, for example, succeeding and preceding serial titles, their supplements, and foreign language parallel titles. Ulrich's, while it has been invaluable, proves problematic in several ways, resulting in the approximate omission of half of the journal titles submitted for collection analysis.

Part of WEST's effectiveness in archiving hinges upon its ability to analyze local serials data across its membership as holistically as possible. The process that enables this analysis, and subsequent archiving proposals, is dependent on Ulrich's journal family, for which ISSN has been traditionally used to match and cluster all related titles within a particular family. As such, the process is limited in that many journals have never been assigned ISSNs, especially older publications, or member bibliographic records may lack an ISSN(s), though the ISSN may exist in an OCLC primary record.

Building a mechanism for matching on ISSNs that goes beyond the base set of primary, former, and succeeding titles, expands the number of eligible ISSNs that facilitate Ulrich's journal family matching. Furthermore, when no matches in Ulrich's can be made based on ISSN, other types of control numbers within a bibliographic record may be used to match with records that have been previously matched with an Ulrich's journal family via ISSN, resulting in a significant increase in the number of titles eligible for collection analysis.

This paper will discuss problems in Ulrich's journal family matching, improved functional methodologies developed to address those problems, and potential strategies to improve in serial title clustering in the future.

By Dana Jemison, Lucy Liu, Anna Striker, Alison Wohlers, Jing Jiang, and Judy Dobry

Background

WEST, Western Regional Storage Trust, founded in 2010, is a distributed shared print journal repository program comprised of 68 members, plus five past members who continue to retain journals on behalf of WEST [1]. Member institutions participate in a number of ways depending on local resources and capacity, for example by:

1. Committing to retain titles on behalf of WEST on site (i.e., in campus libraries) or in dedicated storage facilities;
2. Physically validating retained titles for completeness and/or condition;
3. Actively seeking holdings from other members to fill gaps (i.e. missing volumes) in retained titles;
4. Offering volumes to fill gaps in retained titles; and
5. Supporting WEST's objectives by taking part in the program's governance and strategic decision-making.

WEST members, on agreeing to commit to archiving proposals made by the WEST Operations and Collections Council (WEST-OCC), record information about the proposals in their library and consortial catalogs in the form of MARC holdings records following program guidelines. These holdings are submitted to WEST-Ops once a year for inclusion in a local database managed by the California Digital Library (CDL), which supports ongoing collection analysis and local comparison functionality.

WEST archival commitment files in the form of MARC holdings records are supplied by WEST-Ops once a year in the form of MARC holdings records for inclusion in the Print Archives and Preservation Registry ([PAPR](#)) of the Center for Research Libraries ([CRL](#)). The PAPR system supports the archiving and management of serials collections by providing comprehensive information about titles, holdings, and the terms and conditions of archiving for major print archiving and shared print programs. In addition to submitting to PAPR, member institutions also supply these same archival commitments to OCLC as MARC holdings records in order to also surface these holdings in WorldCat.

Shared print programs strive to ensure ongoing, long-term access to the scholarly print record, while also allowing member institutions to optimize campus library space. It is of the utmost importance that the functionality of the process includes as many member held titles for consideration as possible, while still retaining accuracy in the process of pulling together related titles to build journal families.

Collection Analysis

WEST collection analysis is predicated on titles being organized into clusters, based upon the Ulrich's journal family ID. Ulrich's assigns a unique journal family ID to related titles such as preceding, foreign language parallel, supplemental, and other interconnected titles. Titles that are successfully matched with journal families are candidates for collection analysis at both the family and the individual title level. Titles submitted by member institutions are matched in Ulrich's using ISSN, and assigned a journal family ID. Titles which do not match are considered to be "orphan" records, and are removed from the process stream at that point.

After the matching process, titles having matched with a journal family go through a number of exclusionary filters to identify only those titles which have an optimal value-add to shared print archiving. Records that do not successfully pass through the exclusion filters are dropped from the process stream. Exclusion filters include:

- WEST previously archived titles
- Non-print formats
- Government documents
- LC classification K (Law)
- Title keywords “online” and “monograph”
- Location exclusions as directed by the owning institution. For example, exclude any titles with location “reference desk, current periodicals, rare materials.”

The remaining titles are organized and clustered into their relevant journal families and are output with their summary holdings and other various data points as reports. WEST staff use the reports to further refine for title review and proposal by WEST-OCC to WEST members for shared print archiving.

Internal Processing, Pre-Enhancements

Before internal functionality enhancements were made, titles were matched on limited, selected data fields found in MARC bibliographic serial records furnished by WEST members. WEST-Ops loaded the Bibliographic records and local holdings to a database and the following input fields were isolated as match candidates with Ulrich’s journal family ISSNs:

- Primary ISSN (022\$a)
- Linking ISSN (022\$I)
- Former Title ISSN (780\$x)
- Succeeding Title ISSN (785\$x)

WEST-Ops further enhanced the data by using the [OCLC Search API](#) to harvest 022\$a, 022\$I, 780\$x, 785\$x fields from primary OCLC records (master records) and loading them to a dedicated table (Table 1). This data, in most cases, was a supplement to the supplied data in WEST member records and, in some cases, was a correction to one or more submitted values. It should be noted that data harvested from the OCLC primary record was always assumed to be most current and most correct, and so took precedence over input values in Ulrich’s match search order.

Table 1. OCLC Harvested Supplementary ISSN Values in the ISSN table. ISSN values are harvested using the OCLC Search API. The OCLC_primary number connects data from this table to matching records in the source data table (by way of the OCLC_xref table) which stores full bibliographic details.

OCLC_primary	Primary ISSN	Linking ISSN	Former ISSN	Succeeding ISSN
1102013	1001-3456	1001-3456		3344-9930
9131027	3498-1209	3399-2020	1443-9008	6775-6678

In addition to ISSNs, the OCLC primary number and any cross reference OCLC numbers [2] harvested were stored locally as an OCLC number “set” to prevent the searching of a title more than once (Table 2).

During processing, the primary OCLC number from each contributor record was first searched in a dedicated table containing harvested primary and cross-reference numbers (Table 2.) If a match was found, the OCLC number set would then be associated with that contributor record in the source data table. If no match was found, then a new search would be done using the OCLC API. Any ISSNs and OCLC numbers gleaned from the matching OCLC record would then be harvested to further populate and enhance the ISSN and oclc_xref tables for further searching. This strategy specifically addresses the ubiquitous issue of multiple OCLC numbers (primary and xrefs) associated with the same OCLC record, but which may show up in local catalogs as primary values.

Multiple OCLC numbers associated with a single OCLC record occur when OCLC detects duplicate records in their union catalog. Duplicate records are weighed to determine which is of better quality. Holdings are then moved over to the better record, the “winner”, while the “loser’s” OCLC number is inserted into 019\$a [3] as an OCLC cross-reference number (xref.) Since a bibliographic record in a local catalog is, in essence, a snapshot of the primary record in OCLC, its primary OCLC number may deprecate to a cross-reference in future. It is therefore important to have a complete set of primary + cross-references when searching and matching on OCLC number.

Table 2. OCLC Harvested Supplementary OCLC Values in the oclc_xref table. OCLC number sets are stored so that any unique title isn’t searched more than once by the OCLC Search API. The OCLC primary or x-ref number connects this table’s applicable OCLC number “set” to the source data table which stores full bibliographic details.

OCLC Primary Number	OCLC X-Refs
1102013	23456, 89076, 10003322
9131027	334467

Following the OCLC API harvesting process, the system then attempted to match records to Ulrich’s journal families, using ISSN values. For each source table title, ISSNs from both the source data table and the ISSN table were searched in Ulrich’s, in the order listed below, until a match was found [4]:

1. OCLC record harvested primary 022\$a
2. OCLC record harvested primary 022\$l
3. OCLC record harvested preceding title 780\$x
4. OCLC record harvested succeeding title 785\$x
5. Source table record primary 022\$a
6. Source table record primary 022\$l
7. Source table record proceeding title 780\$x
8. Source table record succeeding title 785\$x

Following Ulrich’s journal family matching, records went through the title exclusionary filters (as described in the previous section), eliminating government documents, non-print formats, etc.

Slightly less than half of member-contributed records (48%) in the pre-enhancement process matched with an Ulrich's journal family, leaving approximately half (52%) that did not (orphan records). There were several reasons why records did not match:

- No ISSN, primary or otherwise, was assigned to that record. This is especially common in older titles, which unfortunately, are often those most at risk.
- Low-level records [5] sometimes lacked an OCLC number, nullifying the OCLC search API harvesting process, consequently omitting potential, additional ISSN match candidates for Ulrich's.
- Titles were not found in Ulrich's despite OCLC number and ISSNs in source and harvested data. Based on Ulrich's internal criteria, not every title is eligible for inclusion, or it may be that while the title is marked for inclusion, Ulrich's hasn't yet captured that title as part of their latest database update.

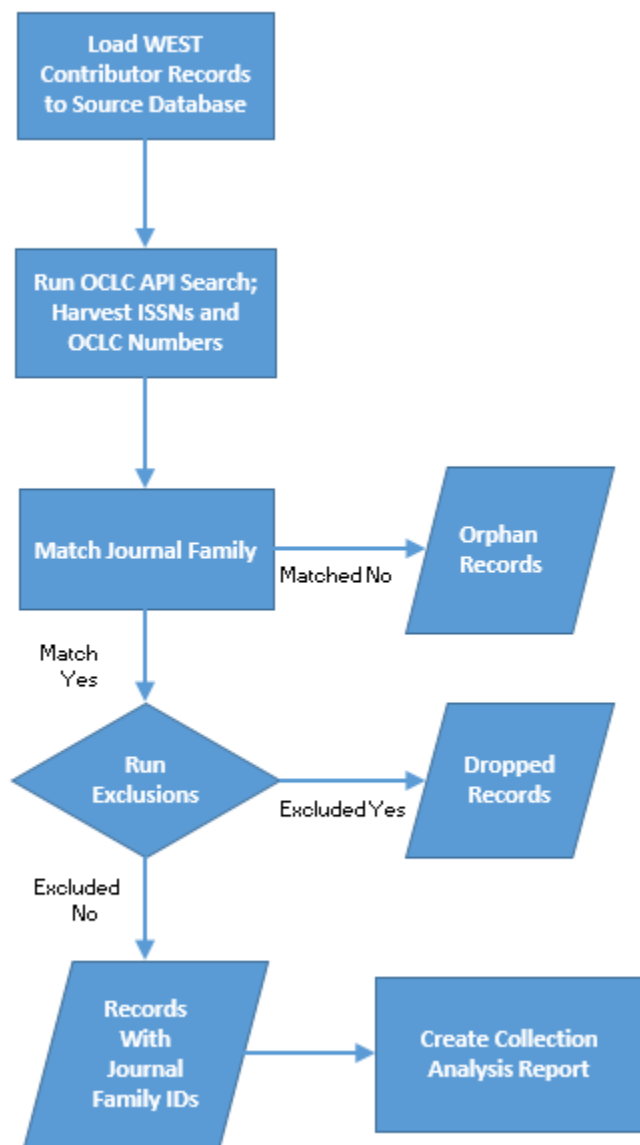


Figure 1. Overview of internal process, pre-enhancements.

Internal Processing Post-Enhancements

To address the approximately 900,000 orphan records comprising a little more than half of member contributed records (52%), the WEST-Ops team devised a more complex match strategy using additional control number values, resulting in a significant improvement in Ulrich's journal family matching. The processing steps are identical, with the addition of several enhancements (figure 2).

1. Additional ISSN values from both the source table data and the OCLC harvested values, 770\$x, 772-777\$x, 786-787\$x, are searched in Ulrich's as part of the match journal family step. This captures additional matches on the following Ulrich's title categories:

- Abridged Edition
 - Abridged Edition of
 - Alternative Frequency Edition
 - Alternative Media Edition
 - Cumulative Edition
 - Cumulative Edition of
 - Issued with Title
 - Parallel Language Edition
 - Partial Translation Edition
 - Partial Translation Edition of
 - Regional Edition
 - Seasonal Edition
 - Seasonal Edition of
 - Special Edition
 - Special Edition of
 - Supplement
 - Supplement to
 - Translation
 - Translation of
2. Additional control numbers (LCCN, CODEN, OCLC) are harvested from both source data and OCLC to form an expanded set of match values. These values are used for a more robust ISSN search in Ulrich's, as well as an additional matching strategy which matches any remaining orphan records with records that have already been assigned a journal family ID; orphan records that match are assigned that same journal family ID.

This expanded match set includes:

- OCLC numbers [6],[7]: 019\$a, 035\$a\$z, 079\$a, 770\$w(OCOLC), 772-777\$w(OCOLC), 780\$w(OCOLC), 785-787\$w(OCOLC)
- LCCN numbers: 010\$a, 770\$w(DLC), 772-777\$w(DLC), 780\$w(DLC), 785-787\$w(DLC)
- ISSN numbers [8]: 022\$a\$l, 770\$x, 772-777\$x, 780\$x, 785-787\$x
- CODEN numbers: 030\$a, 770\$y, 772-777\$y, 780\$y, 785-787\$y

As orphan records are searched and matched, their control numbers are pooled and deduplicated with the target record's control numbers. This provides an expanding pool of control number values from which subsequent orphan records can match. If an orphan record does not match, it is shuttled to the "end of the line," in hopes that as additional orphans match and expand the pool of match targets, these remaining orphans will eventually find matches with target records. When all orphan records in line have failed to match with a target record pool, this match processing step ends. The unmatched orphan "line" may cycle through numerous times until matches are no longer found.

Added control number values, both primary (019\$a, 035\$a\$z, 079\$a, 010\$a, 022\$a\$l, 030\$a), and from relevant linking data fields (770\$wxy, 772-777\$wxy, 780\$wxy, 785-787\$wxy), allow for multiple, highly reliable matching options, barring member

submitted record content errors [9]. Matching on records that previously successfully matched with Ulrich's provides a back-door method to garner new matches for records which otherwise would have been dropped from collection analysis.

3. The OCLC Search API is employed to make a final pass over orphan records to harvest any additional OCLC numbers and ISSNs that may have otherwise been missed. A last attempt is then made to match any of these remaining orphan records with records previously matched in Ulrich's. Orphan records that match with a record with a journal family ID are assigned that same journal family ID.

In this final step, OCLC-harvested OCLC numbers and ISSNs from remaining orphans are searched in the following order:

- OCLC numbers in the following order: primary, 019\$a.
- If no match, then ISSN numbers in the following order: 022\$a, 022\$l, 780\$x, 785\$x.

It should be noted here that the OCLC Search API was set up in previous development cycles to harvest a limited set of additional, or corrected, ISSN values 022\$a\$l, 780\$x, 785\$x. Ideally, this would have been refactored to include the full, expanded control number set of 010\$a, 019\$a, 022\$a\$l, 030\$a, 035\$a(OCOLC)\$z(OCOLC), 079\$a, 77[02-7]\$w(DLC)\$w(OCOLC)\$x\$y, 78[05-7]\$w(DLC)\$w(OCOLC)\$x\$y. However, because of the significant development time this expansion would require, we were unable to fold this particular enhancement into this project. We hope to add this expanded set into post-enhancement step 4 (see following section) during a future development cycle, anticipating that it will further improve our match rates.

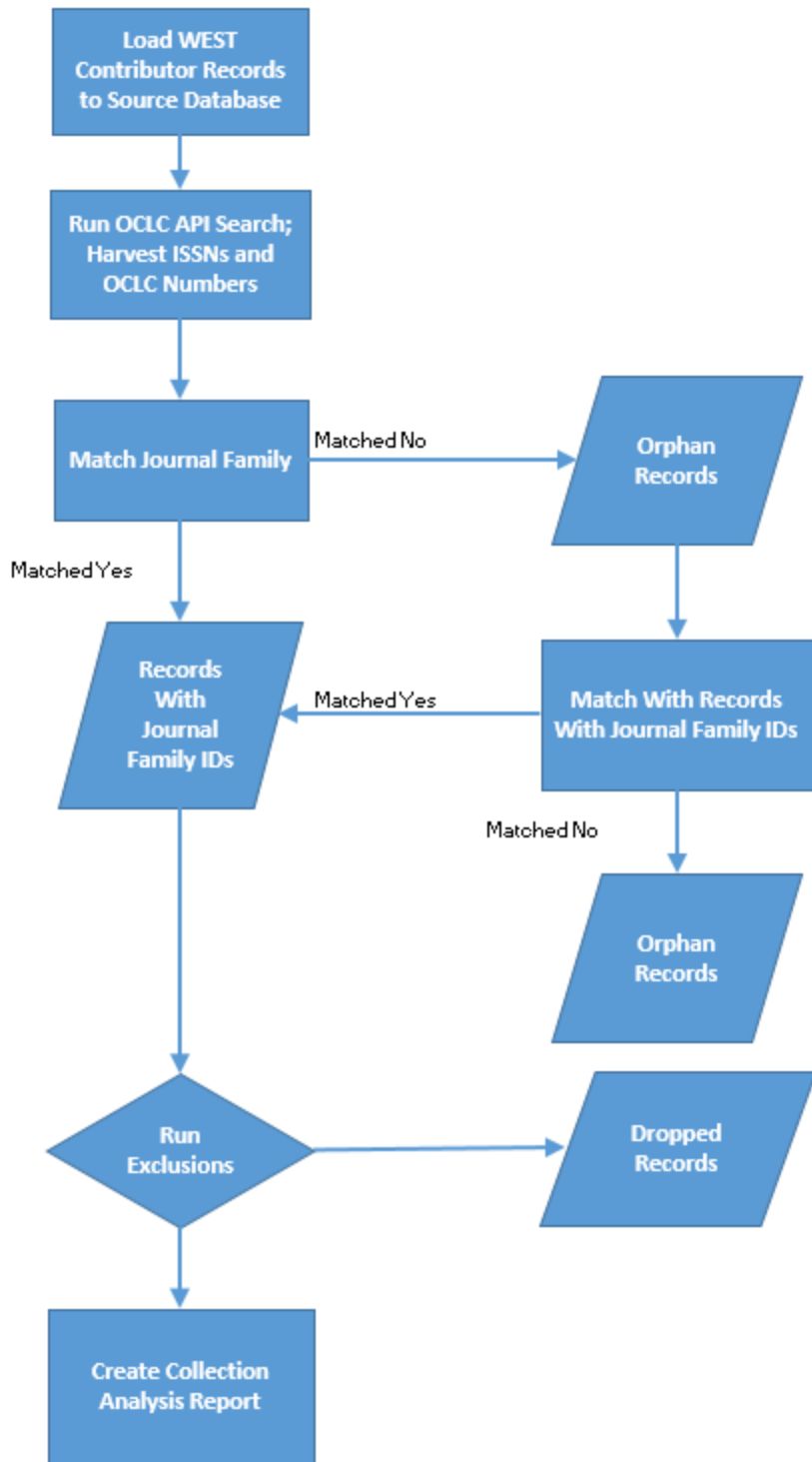


Figure 2. Overview of internal process, with enhancements..

Results Post-Enhancements

Table 3. Processing Post-Enhancements Results. Steps 3 and 4 were done as part of the processing enhancement project only, illustrating the improvement in journal family matching using the expanded set of control numbers for Ulrich’s matching, and matching orphans to records with journal families as an added strategy using an expanded control number set.

	Baseline(a)	Step 1(b)	Step 2(c)	Step 3(d)	Step 4(e)
Total Records	1,721,416	1,721,416	1,721,416	1,721,416	1,721,416
Matched Journal Family	0	783,983	825,419	920,619	921,127
Did Not Match Journal Family	1,721,416	937,433	895,997	800,797	800,289
Matched Journal Family					
... And ISSN(s) in Source Data	0	783,983	789,214	815,273	815,409
... And No ISSN(s) in Source Data	0	0	36,205	105,346	105,718
Matched Journal Family					
... And Source Data LCCN(s) <i>Only</i>	0	0	0	681	683
... And Source Data CODEN(s) <i>Only</i>	0	0	0	3	3
... And Source Data OCLC#(s) <i>Only</i>	0	0	0	39,599	39,833

1. Baseline preprocessing source data table captures counts of primary control number values 022\$a\$1, 780\$x, 785\$x in input MARC record.
2. Match Ulrich’s journal family using source data ISSNs. MARC fields 022\$a\$1, 780\$x, 785\$x.
3. Match Ulrich’s journal family using additional OCLC API harvested ISSNs. MARC fields 022\$a\$1, 780\$x, 785\$x.
4. Match Ulrich’s journal family using expanded control numbers from source data, then match remaining orphan records to records with journal families using expanded control numbers. MARC fields 010\$a, 019\$a, 022\$a\$1, 030\$a, 035\$a(OCOLC)\$z(OCOLC), 079\$a, 77[02-7]\$w(DLC)\$w(OCOLC)\$x\$y, 78[05-7]\$w(DLC)\$w(OCOLC)\$x\$y.
5. Harvest any remaining control numbers from OCLC to match remaining orphans to records with a journal family. MARC fields: 019\$a, 022\$a\$1, 035\$a(OCOLC), 780\$x, 785\$x.

A set of 1,721,416 contributed bibliographic serial records was used as a baseline (Table 3) Pre-enhancement processing steps 1 and 2 were run, resulting in an Ulrich’s match rate of 48% (825,419). 895,997 records (52%) failed to match with Ulrich’s data. Pre-enhancement processing step 1 used source data ISSNs from 022\$a\$1, 780\$x, 785\$x, while step 2 used additional, or corrected, ISSNs where available, from OCLC (022\$a\$1, 780\$x, 785\$x) were used to match with Ulrich’s ISSNs.

Following the enhancement of the process, and using the same baseline set of bibliographic source data, steps 3 and 4 were run following steps 1 and 2. Fifty-four percent (54%) of records

(921,127) matched with a journal family, a positive increase of +6%, leaving 46% of records (800,289) still orphaned. If we review improvements based on a percentage of increased records ((enhancement total – pre-enhancement total)/pre-enhancement total), we gained +11.6 percent (95,708 records.)

Step 3, matching records to Ulrich’s journal family using an expanded set of control numbers from the source data and, following this, attempting to match any remaining orphan records using an expanded control number set to records with an Ulrich’s journal family, garnered the most significant gain in journal families, +95,200 (~10%.)

Step 4, a final search of OCLC for a limited set of ISSNs and OCLC numbers (019\$a, 022\$a\$, 035\$a(OCOLC), 780\$x, 785\$x), gained very few additional titles, +508 (~.06%), but in light of the total number of journal titles proposed, (Table 4), this may have had some significance.[\[10\]](#)

In examining matching and prevalence of controls numbers at each step in the updated processing routines, we found that expanding our control number match set from the source data had a significant effect where the records had no, or limited, ISSN values in the source data:

- 105,718 records lacked base set ISSNs (022\$a\$, 780\$x, 785\$x) in the source data. 36,205 of these found matches following the first OCLC API harvest of base set ISSNs. An additional 69,513 records were added when their additional linking data field ISSNs (77[02-7]\$x, 78[67]\$x) were used to match with Ulrich’s journal families or match with those records previously assigned a journal family ID..
- 40,519 titles were gained where there was no available ISSN for matching in Ulrich’s. These records had one or more LCCN or, one or more CODENs or, one or more OCLC numbers only. These control numbers were used to match with a record having already been assigned a journal family ID.

Conclusion

Following enhancements, a total of 921,127 records (54%) were assigned a journal family ID. This was an improvement of 6% over the previous collection analysis cycle, pre-enhancement (Table 4)

We theorize that this gain is primarily comprised of older titles lacking ISSNs and/or more obscure titles which have a tendency to be represented by brief cataloging records in local catalogs,[\[11\]](#) titles that are, furthermore, more likely at risk due to age and rarity.

MARC records for older and more obscure titles can be lacking the expansive metadata of more recently published titles, as they may lack access points that were not implemented and used until decades following the development of MARC standards; older title metadata, for example, may be limited to data transcribed from catalog cards to MARC when a library first moved to an automated library catalog. Where a retrospective catalog conversion [\[12\]](#) would otherwise upgrade these types of records, they still lurk in many library catalogs.

Performing OCLC API searches on these records can be thought of as an abbreviated retrospective conversion. This search fleshes out control numbers from primary and linking data fields garnered from the OCLC primary record, the “record of record”, which generally speaking, is the fullest and most accurate version. These additional control numbers then allow for enhanced matching capabilities with Ulrich’s, and on records having already matched with Ulrich’s, using the full array of, and most correct values known to an individual title as expressed in the OCLC “record of record.”

Table 4. Record Total Overview Comparison, Pre- and Post-Enhancement, Journal Families.

	Pre-Enhancement Number of Records	Percent of Total	Post-Enhancement Number of Records	Percent of Total
Total Contributor Records Loaded	1,922,108	100%	1,721,416	100%
Records Loaded with ISSN	1,068,059	56%	977,937	57%
Records Matching Ulrich’s Journal Family	917,032	48%	921,127	54%

The collection analysis report, seeded by journal titles which successfully matched with an Ulrich’s journal family, and subsequently passed exclusion criteria, was likely improved. After correcting for duplicate titles shared across contributors, 191,750 unique titles were included in the 2020 collection analysis report. Of these, 8,051 titles were proposed for archiving in 2020 (Table 5)

Table 5. Record Overview, Post-Enhancement, Collection Analysis Report and Titles Proposed.

	Pre-Enhancement Number of Records, 2018	% of Total	Post-Enhancement Number of Records, 2020	%of Total
Total Contributor Records Loaded	1,922,108	100%	1,721,416	100%
Number of Titles in Collection Analysis Report [13]	197,956	10%	191,750	11%
Number of Individual Titles Proposed	8,074	.42%	8,051	.47%

Of significance is the relatively small number of records which are ultimately included in the collection analysis report, and of these, the even smaller number of titles that are proposed for archiving. This would indicate that while overall improvement was only a positive gain of 11.6%, these 95,708 records comprise a significant pool from which titles may be gleaned for a relatively smaller set of archiving proposals.

The number of individual titles included in the collection analysis report, and proposed after enhancement, improved proportionally between 2018 and 2020 by a small margin (table 5.) However, due to changes in exclusion criteria and other considerations between collection analysis cycles, it's impossible to draw any sort of concrete conclusions at such a granular level when comparing across cycles. Unfortunately, due to the complexity of programming efforts it would take to track individual field and process step matching, we were unable to determine how many additional titles were captured and added to the final set of archive proposals. So, while the overall number of records with matches to journal families may have increased with enhanced matching routines by 11.6%, we do not know how many of these specific additions may have dropped out during the exclusion process, or, having passed, subsequently failed to be proposed for archiving. Ideally, we'd be able to identify which titles were added post-enhancement, if any, to determine how successful this project was.

To improve tracking and matching, further enhancements might include:

- Expanding on control numbers harvested from OCLC in step 4 for inter-record matching, including LCCN and CODEN numbers, both base and linking data values.
- Using tightly controlled and normalized author/title string matching between records using primary values from 1XX/2XX fields with appropriate linking data field contents 77X-78X.
- Adding tracking fields to database records to trace MARC field and subfield match values in steps 3 and 4 for further analysis as to which fields produce the highest match rate.
- Adding tracking fields to database records see which records having gone through step 3 and step 4 matching processes were included in the collection analysis report, and were subsequently nominated for archiving.
- Move from the Ulrich's model to an entirely in-house model, where control number and author/title string matching are used outside of Ulrich's to cluster titles together for "WEST journal families."

Bibliography

California Digital Library. 2019. Discovery & Delivery. California Digital Library, services and projects. [Online] May 2019. <https://cdlib.org/services/d2d/>

California Digital Library. 2020. WEST membership. [Online] June 2020. <https://cdlib.org/west/about-west/west-membership/>

California Digital Library. 2020. WEST: Western Regional Storage Trust. [Online] May 2020. <https://cdlib.org/west/>

Center for Research Libraries. 2020. Center for Research Libraries. Global Resources Network. Center for Research Libraries. Global Resources Network. [Online] 2020. <https://crl.edu>

Center for Research Libraries. 2020. PAPR – Print Archive and Preservation Registry. [Online] 2020. <http://papr.crl.edu/>

Library of Congress. Network Development and MARC Standards Office. 2020. MARC code list for organizations. MARC standards. [Online] May 2020. <https://www.loc.gov/marc/organizations/>

Library of Congress. Network Development and MARC Standards Office. 2020. MARC standards. [Online] March 2020. <https://www.loc.gov/marc/>

Library of Congress. Network Development and MARC Standards Office. 2003. The LCCN namespace. MARC standards. [Online] November 2003. <https://www.loc.gov/marc/lccn-namespace.html>

OCLC. 2007. Technical bulletin 254, OCLC-MARC format update 2007 and institution records to accommodate the RLG Union Catalog. s.l. : OCLC, 2007. 1097-9654.

Proquest. 2020. [ulrichsweb.com\(TM\)](http://ulrichsweb.com(TM)) — Frequently asked questions. UlrichsWeb global serials directory. [Online] 2020. <https://www.ulrichsweb.com/ulrichsweb/faqs.asp>

Notes

[1] As of 08/2020. <https://cdlib.org/west/>.

[2] Found in 019\$a and 035\$z.

[3] 019\$a contents in OCLC are duplicated in 035\$z in exported records. Therefore, these are found in either, or both 019\$a and 035\$z in records evaluated during these processes.

[4] OCLC primary record ISSNs were searched first, as these were deemed to be most current and therefore most correct.

[5] A low-level record is considered to be one that is brief in nature, often input “on-the-fly” in the local catalog, and will often lack an OCLC number and other metadata descriptive elements.

[6] Note that for matching purposes, the organizational code as supplied in parentheses, is used in tandem with the control number, to identify control number type. <https://www.loc.gov/marc/>

[7] 079\$a is the location of the OCLC primary number in OCLC institution records. Institution records (IRs) were created to accommodate RLG clustering practices when RLG merged with OCLC in 2006. IRs are no longer in use, but 079\$a’s containing an OCLC master number may still exist in legacy records residing in local catalogs. (OCLC, 2007)

[8] LCCNs were normalized for optimal matching using the LCCN Namespace documentation. <https://www.loc.gov/marc/lccn-namespace.html>

[9] Garbage in/garbage out rule.

[10] Unfortunately, due to the complexity and programming efforts it would take to track field and process step matching, we are unable to determine how many additional titles captured with updated processing were added to the final set of archive proposals.

[11] Because, at this time, we are unable to track titles matching under the new processing routines, we have no way of knowing the composition of the gained set of title matches.

[12] A process used to upgrade a library's catalog of MARC records, generally done with OCLC by matching and supplying up-to-date cataloging records from OCLC to replace low-level records in the local catalog.

[13] Note that the number of titles in the collection analysis report will include duplicate entries across owning institutions as these are submitted as individual MARC records. In other words, if institution A and institution B both own that title, that title will show up twice in the report.

About the Authors

Dana Jemison (dana.jemison@ucop.edu) is the Principal Metadata Analyst for the Discovery and Delivery Group at California Digital Library. She works primarily with the Western Regional Storage Trust (WEST), the CDL, CRL, and the HathiTrust Shared Print Collaboration (CCH Collaboration.) She holds her M.L.I.S. from University of Texas, Austin and a B.S. in Botany from California State University, Long Beach.

Lucy Liu (lucy.liu@ucop.edu) is the Application Programmer for the Discovery and Delivery Group at California Digital Library. She works primarily on the AGUA and PAPR project to support the collection analysis for Western Regional Storage Trust (WEST), the comparison tool for CDL, CRL, the HathiTrust Shared Print Collaboration (CCH Collaboration).

Jing Jiang (jing.jiang@ucop.edu) is a Programmer Analyst for the Discovery and Delivery Group at California Digital Library. She works primarily on library metadata management projects including the HathiTrust Metadata Management System (Zephir) and the Western Regional Storage Trust (WEST) Shared Print programs.

Anna Striker (anna.striker@ucop.edu) is the Shared Print Operations and Collections Analyst at the California Digital Library. She works primarily with the Western Regional Storage Trust (WEST) and UCL Shared Print programs. She also works with the Rosemont Shared Print Alliance and contributes to the CDL, CRL, and HathiTrust Shared Print Collaboration (CCH Collaboration). She holds her M.L.I.S. from San José State University and a B.A. in English from the University of Virginia.

Judy Dobry (Judy.Dobry@ucop.edu) is Technical Team Manager for the Discovery and Delivery Group at California Digital Library. She works on a variety of shared print projects including the Western Regional Storage Trust (WEST), the CDL, CRL, and HathiTrust Shared Print

Collaboration (CCH Collaboration), and the Partnership for Shared Book Collections. She holds her B.A. from University of North Carolina at Chapel Hill.

Alison Wohlers (Alison.Wohlers@ucop.edu) is the Shared Print Program Manager at California Digital Library. As such, she supports and coordinates collaborative print management efforts in the UC Libraries system, the Western Regional Storage Trust, and connects that work to national and North American collaborations.