

2CUL TSI Batch Processing Working Group Outline

Leads: Mark Wilson (Columbia -- chair), Gary Branch (Cornell)

Other Members: Columbia: Evelyn Ocken, Gary Bertchume

- I. Inventory of batch processing staff and expertise
 - A. Columbia University
 1. Gary Bertchume, Manager, Library Systems, Libraries Information Technology Office (LITO). Digital Programs and Technology Services. 35 yrs. experience.
 - MARCIVE record loads
 - Rapid exports
 - OTF deletes
 - CLIO SOLR ingests
 - Smart barcode projects
 - Global data changes
 - Bibliographic file statistics
 - OCLC batch match
 - Production & Change of Size extracts
 - System deltas to backup/mirror server
 - System data migrations
 2. Evelyn Ocken, Senior Systems Analyst/Programmer, Libraries Information Technology Office (LITO). Digital Programs and Technology Services. 20 yrs. Experience
 - ReCAP accessions
 - OCLC Exports
 - LTI extracts and loads
 - MARC record approval loads (ca. 10 vendors)
 - MARC record firm order loads (ca. 20 vendors)
 - EDI orders and invoices (4 vendors)
 - POOF processing
 - Borrow Direct record deletes
 - Acquisitions reports & EOFY reports
 - Collection commitments and expenditures
 - Audit reports
 - E-book, e-journal, e-resource loads (30+ data sources)
 3. Jennifer Chong, Associate Systems Engineer, Libraries Information Technology Office (LITO). Digital Programs and Technology Services. Transferred from User Services Librarian position as of 7/13.
 - Manually executes many of the procedures and reports scripted by the Batch Programming staff
 - Updates domains in EZ Proxy for e-resources identified as unproxied after batch loading.

- B. Cornell University – Primary staff involved in batch processing.
1. Gary Branch, Library Administrator II, Batch Processing Supervisor, Library Technical Services, 23 years of experience, 5 years supervising Batch Processing.
 - Global maintenance specifications
 - Large scale extracts to third party specifications (HathiTrust, Google)
 - New process and technique development
 - Current and new OCLC projects
 2. Peter Martinez, Technical Services Assistant V, Batch Processing & Metadata Specialist, Library Technical Services, 13 years of experience, 6.5 in technical services, 6.5 in library Rare and Manuscript Archives .
 - MARC data analysis
 - Perl scripting
 - Access Reports
 - Character encoding issues
 3. Natalya Pikulik, Technical Services Assistant IV, Batch Processing & Standing Orders, 7 years in technical services.
 - Load vendor records such as SpringerLink, Books 24x7
 - Batch withdrawals
 - Location flips
 - Vendor record load specifications
 4. Nancy Solla, Technical Services Assistant IV, Batch Processing & Standing Orders, 14 years in technical services, 2 years earlier in Access Services.
 - Load vendor records such as Safari and Knovel
 - Batch withdrawals
 - Location flips
 - Vendor record load specifications
 - Design with emphasis toward usability and needs assessment
 5. Joe McNamara, Technical Services Assistant IV, Batch Processing & Standing Orders, 23 years in technical services.
 - Load vendor records such as Rand and Memso
 - Batch withdrawals
 - Location flips
 - Vendor record load specifications
 - Authority record loads and headings maintenance
 - 490 0 flip to add 830 fields
 6. Pete Hoyt, Programmer/Analyst, Library Systems, 15 years library programming experience
 - Creator and responsible for maintenance of LStools. Used for Voyager harvest, MARC record file modification and bulk importing.
 - Creates Perl scripts for LTS staff MARC record loading via LStools

II. Reporting structure considerations

A. Columbia University

1. Columbia University Libraries and Information Services organizational structure (link to chart) branches into three distinct service groups and one administrative service group headed by Associate University Librarians or Vice Presidents:
 - a. Robert Wolven, Associate University Librarian for Bibliographic Services and Collection Development.
 - b. Rob Cartolano, Associate Vice President for Digital Programs and Technology Services.
 - c. Damon Jaggars, Associate University Librarian for Collections and Services.
 - d. Kris Kavanaugh, Associate Vice President, Finance and Administration.
2. Batch Processing at Columbia is almost exclusively coordinated through the Libraries Information Technology Office. LITO Director Breck Witte reports to the Associate Vice President for Digital Programs and Technology Services.
3. Technical Services Units that rely upon the batch processes and reporting services of LITO report to The Associate University Librarian for Bibliographic Services and Collection Development. These units include:
 - a. ReCAP & Recon Projects – Misha Harnick
 - b. ReCAP coordinator – Zack Lane
 - c. Continuing and Electronic Resources (CERM) – Joyce McDonough, Director ; includes Serials Acquisitions – Alan Schaplowsky, Head
 - d. Original & Special Materials Cataloging – Kate Harcourt, Director
 - e. Monographs Processing Services – (vacant) Mark Wilson, Acting Director ; includes Monographs Acquisitions – Matt Pavlick, Head
 - f. Collection Development – Jeffrey Carroll, Director
4. East Asian Technical Services reports to the C.V. Starr Library Director who in turn reports to the Associate University Librarian for Collections and Services. East Asian Technical Services unit is likewise dependent upon the batch and reporting services of LITO.
5. Barnard College Libraries are likewise dependent upon LITO for record migration and support.
6. Health Sciences Libraries are likewise dependent upon LITO for record migration and support.
7. The Batch Processing staff of LITO is also highly utilized for projects within its own Digital Programs and Technology Services group.
8. The Batch Processing staff of LITO is also highly utilized for projects and support for the Collections and Services group.
9. The Batch Processing staff of LITO is also highly utilized for reports and support for the Finance and Administration group.

B. Cornell University

1. The library is divided into six organization units under the University Librarian.
 - a. Dean Kraft, Associate University Librarian for Information Technology

- b. (currently vacant), Associate University Librarian for Teaching, Research, Outreach, and Learning Services
 - c. Xin Li, Associate University Librarian for Central Library Operations
 - d. Oya Reger, Associate University Librarian for Digital Scholarship and Preservation Services
 - e. John Saylor, Associate University Librarian for Scholarly Resources and Special Collection
 - f. Lee Cartmill, Associate University Librarian for Administrative Services
2. Batch Processing is dependent on library systems programming for all scripted loads. However we are able to do our own MARC file manipulation and loading through the Library Systems and Discovery Services, under Chris Manly's supervision, using LStools as well as other tools such as MARCedit and Strawn tools such as Location Changer and Record Reloader through Voyager's batchcat.dll utility. We also have a staff member who does some in house Perl scripting utilizing batchcat.dll for Voyager loading and manipulation of bib, mfhd, and item records. Many of the LTS reports via Microsoft Access against the Voyager database have been developed and generated by Lydia Pettis of Library Systems. She also led classes where LTS and Batch Processing staff in particular generate their own queries and reports for batch processing projects.
- III. Inventory of Policies and Workflows
- A. Columbia University
1. MARC record loads for print materials
- a. Vetting records – Vendor supplied records for both print and electronic resources are vetted by technical services units prior to implementation. OSMC for e-journal and e-resource records and MPS for print records.
 - b. EOD – Embedded Order Data in vendor supplied records must match certain local criteria (e.g. decimal point must be present in the price field)
 - c. Voyager Operator ID's – Voyager Operator ID's are created and used to reflect the batch record source in the record history. This is an invaluable forensic tool.
 - d. MARC Load Reports
 - i. Duplicate resolution – MARC record load reports indicate any duplication by ISBN. These are combined to a single bib and line items and duplicate bibs are deleted and then relinked to the new bib by an MPS supervisor.
 - ii. The purchase order number appearing in the load reports are noted to then write onto the original invoice.
 - iii. Shelf ready materials load with item records and immediately appear as "Not Checked Out" after loading. A separate report of Item barcodes is provided by LITO in addition to the standard load report. The barcodes are scanned to a status patron in order to change the OPAC display to "In Transit to Library".
 - iv. Gobi Orders – YBP/LINDS firm order confirmation records placed via GOBI will add holdings to an existing record based on match criteria. If the

matched record is an ebook, the record does not load but kicks out for manual inputting handled by the Order Unit in MAS.

- v. Casalini orders – Catalog level MARC records replace/merge with the record used for the order placement. If the Publication status of the MARC record is “m” for multiple the incoming record does not merge/replace but kicks out for manual review or input by cataloging staff in MPS.
 - e. EOFY exceptions – some batch routines must either be suspended or amended during the Fiscal Year close period. Formerly it was a practice that after a fixed date, line items for print approval loads would set all line item prices to zero in order to stop encumbering funds. After an internal university audit we were told we could not roll over zero'd out purchase orders with multiple (per title) line items, the loading of purchase orders and line items is therefore suspended until the ledgers are open in the next FY.
 - f. Cron job record loads – YBP/LINDS , Harrassowitz, Aux Amateurs , Casalini record files are retrieved and loaded by a cron script, therefore records are in the catalog before the books are received in the department – sometimes (rarely) vendors can mistakenly repost a file that then gets automatically reloaded. Procedure: <https://wiki.cul.columbia.edu/display/cliogroup/ApprovalRecordLoads>
 - g. Manual loads for East Asian – multiple records representing firm order for East Asian are loaded manually with PO's and Line Items; these records are received as email attachments rather than retrieved by FTP cron jobs.
2. E resources
- a. 965 – Bibliographic field marker 965 is placed into many e-resource MARC records in order to identify record sets for batch selection and to enforce policies in place for excluding record sets from uploads to OCLC.
 - b. Resolving URL's – E-resource (and some print) records are provided with resolved links built on the Libraries' base URL and the Voyager bibliographic record key. These are formed with a batch editor once bibliographic keys exist for the newly loaded records.
 - c. Serials Solutions vs 3rd party record providers – For efficiency and centrality of e-resource management, Serials Solutions records are preferred. Upon occasion, superior record sets are available from the content vendor. These are vetted by OSMC staff. Promotion of this exception is not encouraged.
3. Monthly extracts for statistical gathering
- a. Production statistics – a cron routine selects 948 field data with date markers from the preceding month. This data is imported into an Access database by an MPS supervisor who corrects erroneous entries and compiles statistics for MPS, OSMC and the Starr East Asian libraries.
 - b. Change in size of collection – a cron routine extracts data for all items created from the preceding month. This data is imported into an Access database by an MPS supervisor who then extracts only the desired items to compile ARL change in collection size contributions for the MPS and OSMC production units.

4. Other Sample Batch Procedures

a. LTI Authorities:

<https://wiki.cul.columbia.edu/display/cliogroup/LtiQuarterlyProcedures>

Workflow includes that a message goes out to Tech Services not to edit records from the extract period during the several hours the records are out for editing as edits will not be saved when the records are replaced. A follow up message is sent announcing editing may resume. Post processing reports of unpaired 880 fields are provided to managers in MPS and Starr East Asian tech services for review and handling.

b. Hathi Trust Harvesting:

<https://wiki.cul.columbia.edu/display/cliogroup/HATHITRUST>

Suspended pending investigation of OCLC's WorldShare Metadata Collection Manager to identify target records.

c. POOF loads: <https://wiki.cul.columbia.edu/display/cliogroup/POOF+-+2CUL>

d. For a full list of detailed batch procedures see:

<https://wiki.cul.columbia.edu/display/cliogroup/Ev+Procedures>

B. Cornell University

1. MARC record loads for print

- a. Vendor supplied records -- are vetted by technical services CMS Original Cataloging Unit for correctness and completeness of MARC data.
- b. EOD-- Embedded Order Data in vendor supplied records must match certain local criteria (e.g. decimal point must be present in the price field)
- c. Size of shipment -- Vendor whose shipments do not consist of a minimum size are not automated, but are manually processed.
- d. Voyager Operator IDs -- For those loads using Bulk Import only batch as an operator id is present.
- e. MARC load reports.
 - i. No duplicate resolution at time of load. A daily ISBN report is generated for resolution the following day.
 - ii. Purchase order number is on the load report which is written on the invoice to aid processing.
 - iii. Shelf ready (Chinese approval) does nothing to item statuses. We only create the holdings and item records on reloading of the bibs.
- f. EOFY exceptions -- we only stop loading approval orders for a few days prior and restart as soon as new fiscal year starts, usually within 48 hours of previous years closing.
- g. Currently no cron approval record loads. (In planning stages.) Bulk Import scripted jobs are queued and the import must then be enacted by staff as a separate step.

2. Electronic records

- a. 899 fields are used to identify vendor packages. (also used to identify certain print groups)
- b. Resolving URLs. For most loads a resolver string is prepended to the 856 link string during the batch processing prior to loading.
- c. Serials Solutions vs. other record providers. For many packages we utilize Serial Solutions provided records, however because of poor metadata quality for some packages we will utilize vendor provided or OCLC records. While we do not discourage using vendor records, we prefer Serial Solutions when the metadata is of good quality.

IV. Dependencies and Limitations

A. Columbia University

1. Monographs Processing&Acquisitions and CERM are dependent upon LITO for the processing of MARC record loads.
2. Selectors are dependent on LITO and technical services to implement MARC records when desired from a specific vendor
3. Monographs Recon Projects, the ReCAP Coordinator and Collection Development are dependent on LITO both for batch processes and reports.
4. The implementation of Technical Services batch processes that are coordinated through LITO are in competition with other priorities.
5. LITO depends on CERM to select the correct records from the ERM KB to load.
6. LITO and MP&A depend on vendors to supply MARC records at the time of their arrival and to avoid error files, duplicate postings or omissions.
7. LITO and MP&A depend on Cornell developers to make required adjustments/improvements to POOF functionality and output.
8. LITO depends on MP&A and CERM to identify errors in MARC loads that are not detected by load routines.
9. HSL depends on Morningside operations for the timely loading of newly acquired e-resources in order to avoid ordering duplicate materials – often weighing options between committing limited funds on duplicate material vs providing access for high demand content.

B. Cornell University

1. Highly dependent on library systems for scripting and support when tools do not operate as expected.
2. Dependent on other units to identify problems with the potential for batch correction.
3. Dependent on cataloging for quality control of MARC data.
3. Units dependent on Batch Processing include units within technical services and other departments.
 - a. Ordering Unit for ITSO/POOF firm order processing.
 - b. Monograph Receiving and Documents for loading of approval MARC records.
 - c. Database Quality for batch maintenance, reports, batch withdrawals.
 - d. Cataloging and Metadata Services for reports.
 - e. Annex Library for record cleanup of accessioned volumes.
 - f. Cornell Weill for prepared MARC records for shared Springer packages.

g. Law Library's Technical Services for batch record load specification for library systems programmer.

V. Baseline Statistics (Data reflects FY 12/13 activity)

A. Columbia University:

1. Print Approval MARC Loads:

Approval Vendor	Line Items	Purchase Orders	Invoices w/o PO's FY close	No line items FY close	Invoice total	Item totals
AUXA/APP	2794	151	10	304	161	3098
HRR/APP	4298	266	27	529	293	4827
CLL+ART/APP	4340	345	48	666	393	5006
YBP07	76	45	3	8	48	84
YBP11	1272	79	7	117	86	1389
YBP14	246	50	5	36	55	282
YBP17	715	61	4	49	65	764
YBP21	4578	223	12	298	235	4876
YBP22	1656	110	6	123	116	1779
LINDS63	4	4	1	1	5	5
LINDS66	1216	111	7	84	118	1300
LINDS69	16	13	1	2	14	18
LINDS72	131	46	2	6	48	137
SRBICA	416	21	0	0	21	416
SOFIABOOKS	218	3	1	86	4	304
EASTVIEW	435	55	8	58	63	493
SULMAN/APP	365	5	1	159	6	524
Totals	22,776	1,588	143	2,526	1,731	25,302

Print Approval Summary:

- 22,776 records loaded with linked line items
- 1,588 Invoices loaded with linked purchase orders
- 143 Approval invoices loaded without purchase orders owing to FY close policies
- 2,526 approval titles loaded without line items owing to FY close policies
- 1,731 approval invoices loaded with MARC records
- 25,302 approval titles loaded with MARC records

2. Print Firm Order Loads

Print Order Vendor	Order Line Items	Rec'd Line Items
CLL	N/A	1178

CLLART	N/A	180
YBP05	63	53
YBP06	3	2
YBP09	4768	4720
YBP10	1993	1857
YBP12	261	222
YBP13	38	30
YBP15	755	701
YBP16	242	219
LINDS64	158	171
LINDS65	207	187
LINDS67	3	1
LINDS68	1	1
LINDS70	49	50
LINDS71	23	29
Total	8,564	9,601

Print Firm Order summary

- 8,564 Order confirmation line items and bibliographic records loaded
 - 9,601 Cataloging level MARC records were loaded for received firm orders
3. Ebook Firm Order Loads

Ebook Firm Orders	Gobi Orders & Confirmation Records
YBP50 (Ebrary)	415
YBP55 (Ebscohost – formerly Netlibrary)	54
YBP57 (Wiley)	6
Total	475

E-book Firm Order summary :

- 475 Order confirmation line items and suppressed bibliographic records loaded.

4. Electronic Resource Loads:

E-Journal Records	E-Book Records	E-Resource misc	Record deletes
80,502	444,880	4,131	146,156

- 5. 49 CU uploads to OCLC numbering 170,919 records.
- 6. 49 EA uploads to OCLC numbering 19,051 records.
- 7. 16,873 records sent to OCLC for Precat/Offprecat batch job, 9,916 records replaced.

B. Cornell University

1. Print approval loads – Except for Harrassowitz and Casalini all approval loads are without purchase orders being created.

TOTAL, YEAR-TO-DATE		
2010/11	2011/12	2012/2013
27304	27637	29743

2. Print firm orders – Vast majority created by WorldCat Selection and POOF MARC record loads.

TOTAL, YEAR-TO-DATE		
2010/11	2011/12	2012/2013
27657	24735	23518

3. Ebook firm orders are all manually created, cataloged, and visible in Voyager. No batch processing work at all at this time. (in development)

4. Electronic Resource loads

Ebooks (includes other monographic formats)	270,785 records
Ejournals (primarily from Serial Solutions)	141,804 records
Electronic record deletions	70,957
OCLC daily cataloging export	252,254
OCLC manuscript and archives export	7,930
OCLC institutional export	222,032
OCLC Local Holdings export	1,751,930
OCLC Batchmatch	40,856

5. Various maintenance jobs

OCLC master number mapping into Voyager	24,335
Initial article and title tag correction	474
Misformatted and duplicate OCLC 035 fields	22,896
490 0 flip to 830 field	4454
Authority record loads	928,600
948 Cleanup	185,023
9xx removal	220,228
Batch single monograph withdrawals (approx.)	19,350
Location/item flips (For closures/ Annex)	171,230