Purdue University

Purdue e-Pubs

Charleston Library Conference

Outsourced and Overwhelmed: Gaining a Grasp on Managing Electronic Resources

Matthew D. Harrington
North Carolina State University at Raleigh, matthew.harrington@duke.edu

Follow this and additional works at: https://docs.lib.purdue.edu/charleston

Part of the Cataloging and Metadata Commons, and the Collection Development and Management Commons

An indexed, print copy of the Proceedings is also available for purchase at:

http://www.thepress.purdue.edu/series/charleston.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences.

Outsourced and Overwhelmed: Gaining a Grasp on Managing Electronic Resources

Matthew D. Harrington, Data Projects and Partnerships, North Carolina State University

Abstract

Outsourcing the management of electronic journals has significantly reduced the autonomy academic libraries have over their collections' metadata, as well as the ways in which that data is collected, organized, and made available to the library. However, the ephemerality of this metadata makes quality control burdensome and costly on the corporate end and necessitates ongoing title tracking and maintenance for the library. As a result, the quality of data in outsourced knowledge bases is often inversely proportional to the library's tolerance of "bad data," as well as its inability to tell the difference. This session demonstrates how an MS Access database was constructed that integrates data from various sources in order to reconcile e-journal and e-book title lists, process yearly subscription changes, and manage the distribution of work to departmental staff. As such, it both serves as a reconciliation tool with administrative functions for linking and displaying summary data about subscribed packages, and it provides a workflow tool with a user interface designed for staff to easily manage ongoing subscription maintenance. Electronic resources are dynamic by nature, and a management system should have the ability to track and respond to these changes. This easily maintained tool offers a model for managing change across the interrelated applications that manage electronic resources.

Introduction

Managing title lists has become an increasingly difficult task in the context of big publisher deals and the various packaging models used to describe these bundled sets of titles. Not only are these title sets too large to manually sift through for inconsistencies, but they change too frequently to justify such labor-intensive title-matching. Contributing to the problem is the outsourcing of data management to commercial products that may or may not communicate with each other. Subscription agents, knowledge bases and publishers may all communicate title lists to a library, but it is often the library's responsibility to ensure those title lists are accurate and uniform.

Therefore, to manage electronic resources, libraries must regularly perform title list reconciliation. What gets paid to the publisher must match what is both discoverable and accessible by the patron. However, electronic resources are not static, and this is where resource management becomes critical. Libraries cannot let go of their management responsibilities simply because commercial management services have been contracted. Though these services can automate many of the daily, monthly, and annual maintenance tasks associated with electronic resources, these are not enough. The library must have a tool that can monitor change and respond to inconsistencies among these systems to ensure title lists reconcile between those services.

Electronic resource management systems, or ERMS, often fail on this level. Though they may maintain records of licenses and perhaps even include some assessment metrics, they rarely flag problems that require staff attention. Therefore, the reports produced by these ERMS rely on the assumption that data are correct but lack the ability look for data discrepancies. IMPART (Integrated Multi-Package Reconciliation Tool), the electronic resource management database described in this paper, attempts to tackle some of the recurring issues that have arisen from the overwhelming amount of electronic resource data that must be managed and the loss of autonomy over that data once the management has been outsourced to commercial services.

Data and Data Integrity

Data are messy, and though everyone involved with data management may want clean and consistent data, there is rarely such thing as a

perfect match or one-size-fits-all approach. Libraries can, however, focus their efforts on the few percentiles that have discrepancies. These data issues range from slight inconsistencies that can be ignored to large-scale issues that lead to misinformation in the reports relied upon to make key collection decisions. Therefore, libraries, and especially technical services departments, must be able to identify which problems can be found, which ones require a solution, and which ones can be solved. These are overlapping but not congruent categories, and the ultimate goal is to continually increase the overlapping area through new tools and new ways of looking at data.

To manage so much data, libraries often rely on commercial services to maintain that data and help organize it, and while these can ease the burden for certain tasks, they can also lead to data integrity problems if there is no communication of data among these different services. Electronic resources pose a particularly messy problem because the library has the least amount of autonomy over both the content and the metadata. There may be multiple systems managing details related to an item, none of which communicate directly with each other. The result is redundant data stored in varied systems, each managed by a different organization or service, and when a change occurs, such as a title transfer or title change, the data must be updated in every system separately, which never happens simultaneously. Furthermore, the ILS often depends upon data integrity, so this places the library in a middle-man position in which it must receive information from one source and communicate that information to another source in order to maintain the integrity of the data that
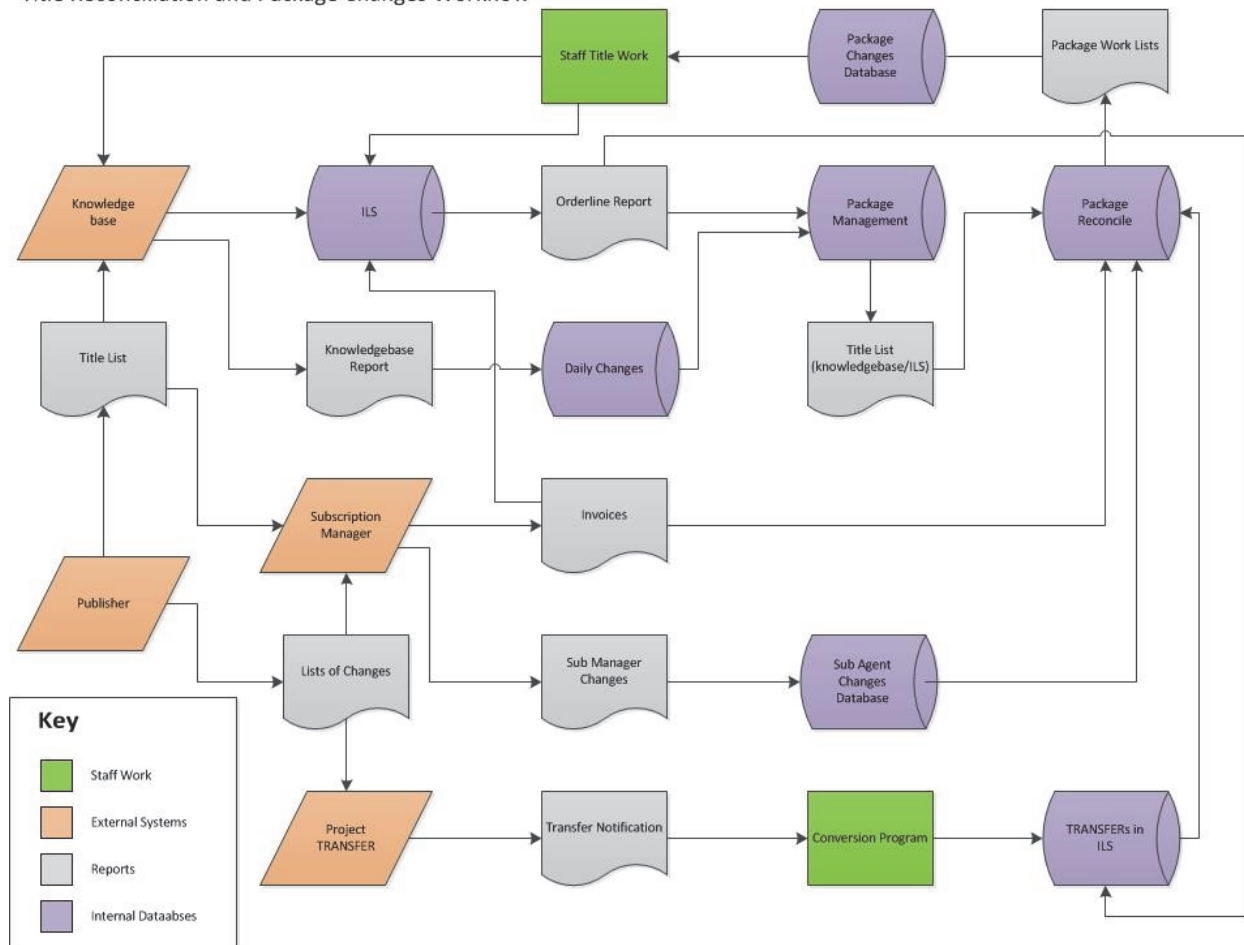

Figure 1. Data flow for title list management.

is stored by each. As an example of this, figure 1 illustrates the flow of data from different commercial systems, through reconciliation databases, to staff for processing, and finally back to the commercial knowledge base.

One of the problems encountered in this middle-man position is linking data together. Outsourcing the management of electronic resources to multiple services leads to multiple identifiers, which may present problems when forming links between that data. In addition to the identifier used to locate materials on shelves, there are international identifiers for books and serials, OCLC numbers, subscription identifiers, knowledge base title IDs, proprietary identifiers supplied by publishers, and even locally created identifiers in the ILS. These may not all refer to the same type of object, but they all relate to each other, and creating links between them allows the library to manage resources more efficiently.

Some of these links may already be in place to automate matching invoices to orders or linking data in an ERM. However, unless all services are supplied through the same company, it may be more difficult to see a link that connects the title ID assigned to your order by a subscription agent to the database code your knowledge base may have assigned to denote the package containing that title. Finding links between these sets of data often falls upon the library. Therefore, understanding the nature and context of these identifiers is key to managing the links between data sources, and consequently, the integrity of the library's data.

Data Reconciliation

Reconciliation is all about maintaining data integrity. Whether payments are to be matched to invoices or orders to holdings, reconciliation involves ensuring two sets of data link correctly and match as expected. Problems occur when a portion of one dataset does not link to a portion of the other dataset or vice versa, and these often stem from changes in the data that affect the relationships between them. Therefore, reconciliation is also a continual process of relationship maintenance, and those relationships are often created from the data supplied by the

various vendors managing one element of the maintenance process. The key to adapting a model to a specific group of vendors, commercial products, and ILS reports is to analyze the common elements that link together the varied sets of data from these systems. Sometimes this requires changing the way data is maintained in the catalog, and sometimes new procedures are needed to periodically collect data.

The database at NC State University was constructed incrementally as needs developed. Its original purpose was to reconcile serial title lists from large electronic journal packages, which ultimately meant ensuring the publisher lists matched orders in the catalog. This naturally grew to encompass title lists from our knowledge base and subscription agents, as well. With four possible lists to compare, it was evident that an important piece of the process would be to continually maintain matching title lists between the catalog and knowledge base while reconciling those as a whole against title lists received annually from various publishers and the subscription agent managing those subscriptions. Therefore, the first priority was to automate the reconciliation between the orders in the ILS, which meant relying on daily order reports provided through the ILS's reporting module, and the daily holdings report from our knowledge base, which was provided for download each day.

Although the ISSN was an early candidate for linking these datasets, using the proprietary identifier from the knowledge base proved most useful because the majority of the title records in the catalog for large e-journal packages were delivered through the MARC record service provided by the same company, and that meant nearly all of the titles in need of reconciliation contained the knowledge base identifier in their 001 field. This identifier was also available in the daily report of all serials holdings. After quickly cleaning up the 001 fields, which resulted in a shared identifier between titles in the ILS and titles in the knowledge base, all that was left was matching specific subscriptions to their appropriate holdings. Database names from the knowledge base had already been added to a specific field of the order record, so logically this

field was normalized and linked to the holdings in the knowledge base through their database details report, which included a database code. This created a shared composite key identifier for both the knowledge base holdings and ILS orders.

The initial database ran simple queries that compared the two lists and combined attributes from each source of data. Not only could a single line of data include costs from the ILS and coverage dates from the knowledge base, but it was possible to find all records that appeared in one list but not the other, which is ultimately the premise for reconciliation. And because the process of linking records from each source had been automated, updated reconciliation lists could be run as quickly as the source reports were made available by the ILS and knowledge base. This allowed regular reconciliation reports to be generated and distributed to staff, who could then review the unmatched titles to determine the cause.

Matching to publisher lists meant using a different key, though. The only identifier provided by most e-journal publishers is the ISSN, and because there is so much inconsistent use between the print ISSN and electronic ISSN, conversion to an ISSN-L almost always yields fewer broken links. By adding the ISSN-L to the existing reconciled list of titles from the catalog and knowledge base, it becomes possible to link these to publisher-supplied title lists. All inconsistencies can then be sent to staff for review. For regular reconciliation reports, this worked very well. However, the management cycle for electronic resources involves more than quick, reactive reconciliation.

Managing electronic serials means processing annual changes, and this entails not only processing those changes, but also discovering those changes and distributing them to staff. This had previously been managed through spreadsheets and Google Docs, but for the 2015 subscription year, a new workflow tool was designed to manage this process. Title transfers and subscription changes were collected from sources such as Project TRANSFER and publisher sites, and these were then compiled into a list of changes, which could be formatted into a database table capable of tracking the renewal

process. One column in the table held a staff person's initials so that tasks could be distributed. Each staff member had a queue, which was simply a filtered version of the table by the initials column, and throughout the process of making modifications to orders and knowledge base holdings, they checked off boxes in the database to keep a record of what had been done and what tasks remained. Staff had previously used databases for regular title maintenance and cancellation projects, so they were already accustomed to working in a database environment. The success of this database prompted a proposal to consider the possibility of integrating it and the reconciliation tool into a single database capable of collecting and reconciling title lists while also providing a dashboard for staff to manage the individual title work assigned to them within the database.

Integrating Databases and IMPART

By the beginning of 2015, there were four databases involved in the serials reconciliation process at NC State University. The two primary tools were the reconciliation database and title work database already described. In addition to these, a separate database was created that simply looked for daily changes occurring in the library's knowledge base holdings, and that involved comparing holdings lists from different days. Updated title lists from publishers appear in the knowledge base without warning, and with no means for proactively searching for these updated holdings, those changes may not be known to the library until a problem is discovered by a patron. Even large-scale changes may be missed if no process for finding these updates is in place. Maintaining a database that compared each day's holdings avoided this issue. When updates were made to the knowledge base, this database highlighted them so that corresponding updates could be made to the holdings displayed in other systems. Mistakes also happen, and when updated title lists contain bad data, those errors were easily identified using this database, and immediate corrections could be made.

Project TRANSFER e-mails also alerted the department to publisher changes, and these e-mails were collected, formatted into a

spreadsheet, and periodically matched against orders in the catalog using a fourth database. The resulting list of transfers was then directly added to the package changes database for staff to review during the renewal process. Eventually this process was simplified by downloading data directly from the Project TRANSFER ETAS system.

These four databases provided the foundation for IMPART's design. The original premise was to use a single database tool to perform all of the reconciliation functions in the department. It needed the ability to access updated data reports to remain current, it needed to compile archival title lists for reconciled title collections, and it needed to route all reconciliation problems to the staff member assigned to manage that collection of titles. To ensure IMPART met the first of these requirements, its source data was collected into a single space and linked to the database through the use of linked tables. Queries within the database could then format the data directly from the source file, update any necessary identifiers, and create formatted, local tables from which to run additional reconciliation queries. These formatting queries can be captured using macros so that updates can be made easily using a button from a database form. Once the data is updated, simple queries that search for matches based on key identifiers were run to produce lists of all matched records, as well as lists of all unmatched records. The former were exported and archived once the problem titles from the latter query were resolved.

This model is not limited to e-journals, though. In addition to the reconciliation functions for online serials, an additional set of queries was designed to reconcile e-book packages. These operate from source files exported from the ILS, the commercial knowledge base, and the publisher, and like all other source data for the database, the files can be overwritten with updated title lists in order to rerun update queries within IMPART. Therefore, producing a new reconciliation list is almost fully automated.
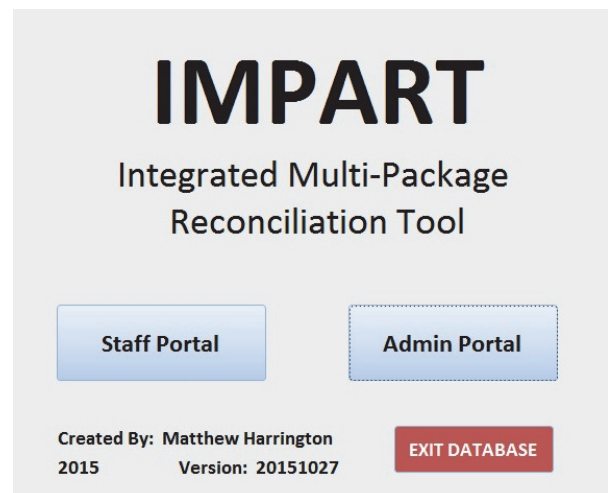

Figure 2. IMPART main menu.

As figure 2 illustrates, IMPART is divided into two portals, one for staff to access title work that has been routed to their queues and one for the administrator to perform all of the necessary reconciliation functions that discover that title work. Therefore, once reconciliation queries have been run, additional functions allow these to be assigned to staff according to the name of the package by linking to a table containing all of the staff managing packages. Assigning title work creates a record in a separate table, which staff access when they open their assigned queue, and these are then filtered by staff initials. Figure 3 shows the staff view within those queues. Work tasks have three statuses; they can be marked as in process, completed, or if newly assigned, neither. Instructions for specific workflows can be automatically added based on the type of task, and linked documentation explains how to process each update. These queues can also be filtered by staff in different ways, such as limiting the display to tasks marked "in process."

Beyond IMPART

Once IMPART had the ability to fulfill each of its original three goals, additional functions were added. Using files exported from Project TRANSFER's ETAS database, functions that match these against order records in the ILS to create a list of all publisher transfers matching a

Figure 3. IMPART staff queue.

subscription were added to IMPART, and it could then assign the transfers to staff for processing. This simplified and automated a potentially time-consuming process. Invoices from the subscription agent could also be matched against orders in the ILS to find items that do not match correctly or to compare encumbered money in the ILS orders to the invoice amount provided by the subscription agent. The daily knowledge base changes previously monitored were incorporated through an additional function, and like other reconciliation processes, problems can be routed to the appropriate person. Additionally, if links were made to COUNTER usage reports, usage data could be imported and linked to the costs included in the ILS orders to produce cost per use reports. With so much data to organize and interpret, especially in technical services

departments, there are many potential uses for tools such as IMPART. The key is finding where data intersects and how those intersections can be exploited.

Managing electronic resources is becoming more complicated as the multitude of those resources increases and their stability decreases. Management tools such as IMPART are now necessary to keep track of what has been purchased and what should be accessible. However, it is just one of many local databases created to address maintenance issues at NC State University. There is no one-and-done solution; this maintenance is a perpetual process, and libraries will have to continue to discover how electronic resources can be managed more effectively in this environment.