

Reconciling E-Book Packages at NCSU Libraries

Christee Pascale
North Carolina State University Libraries, cpascal@ncsu.edu

Xiaoyan Song
North Carolina State University Libraries, xsong9@ncsu.edu

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



Part of 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>.

Christee Pascale and Xiaoyan Song, "Reconciling E-Book Packages at NCSU Libraries" (2015).
Proceedings of the Charleston Library Conference.
<http://dx.doi.org/10.5703/1288284316267>

Reconciling E-Book Packages at NCSU Libraries

Christee Pascale, Associate Head, Acquisition and Discovery Department, NCSU Libraries

Xiaoyan Song, E-Resource Librarian, Acquisition and Discovery Department, NCSU Libraries

Abstract

As e-books become an increasingly large part of our collection, the NCSU Libraries acquisition and discovery department created an e-book reconciliation database to ensure that all of our purchased e-book package content is available in the ILS and throughout the Libraries discovery layers and to create definitive title lists that associate and articulate e-book titles with package purchases. This tool compares vendor title lists against ILS metadata in order to identify missing titles and generate reports. The paper will discuss what prompted the development of the database; present the e-book data flow in NCSU Libraries and e-book reconciliation workflows designed based on the data flow; report on our approach on how to collect e-book title lists, normalize metadata and identify matching point for the reconciliation; present our findings by analyzing the data; and discuss the most common e-books issues found in the reconciliation process, the causes and solutions for these issues.

Introduction

The NCSU Libraries acquisitions and discovery (A&D) department is currently staffed by twenty-four full-time employees divided into three units. The monographs and serials units each have nine staff members, and our data projects and partnerships unit (DPP) contains five staff members. We are a highly centralized, merged acquisitions and cataloging department. All unit staff perform acquisitions and cataloging tasks related to their functional areas. For example, staff in the monographs unit order, purchase, catalog, and troubleshoot non-continuing resources in all formats. Acquisitions and discovery handles purchasing, licensing, and access across all NCSU libraries and branches.

In 2014–2015, the Libraries acquired and managed approximately 34,000 combined monographic and serial e-resources. From an e-book perspective, this number includes e-book packages, firm e-book orders, e-book DDA/PDA titles, and e-book approval materials. We are a Serials Solutions customer and we typically manage a combined 555,000 unique serial and monographic titles in our Serials Solutions knowledge base.

NCSU Libraries purchased seventy-seven e-book packages in 2014–2015. This statistic represents

the total number of packages acquired, as opposed to the total number of providers, as we purchased multiple packages from some vendors. From those seventy-seven e-book packages, we managed approximately 16,000 e-book titles in this fiscal year.

Last year the monographs unit team tackled a project to create an e-resource hub for our e-book packages. The e-resource hub is a Wiki-based tool that captures administrative metadata about our e-book packages. This includes data like: licensing, acquisitions information, access, and cataloging information. The e-resource hub was our first step toward capturing all known information about all our e-book packages in one space and helped inform a needs assessment for a possible e-book package electronic resource management (ERM) solution. Since an e-book ERM solution is still somewhere in our future, we realized we needed a tool to help us address more immediate concerns. We created the eBook Reconciliation Database, a Microsoft Access database, as this interim solution.

Currently our eBook Reconciliation Database includes title list data sourced from vendors and MARC record data about our e-book packages sourced from our ILS. As part of the design, we intentionally focused on e-book packages and left firm, DDA/PDA, and approval e-book orders out of

the equation. We have a solid workflow for these e-book resources and, as a result, feel more confident in our management of these materials. The scope and volume of our e-book packages, however, demanded that we consider a tool that would help automate some identified problem areas.

Project Overview

We designed the eBook Reconciliation Database to initially address three specific concerns: access verification, title list certification, and a package/provider lookup tool. We needed a tool that would help us definitively describe what e-books we owned, their vendor and package association, and allow us to ensure we were providing access to those titles throughout all our discovery layers.

In our existing landscape, we had three elements:

1. Publisher title lists that represented purchased e-book content
2. Corresponding e-book packages activated in the NCSU Libraries Serials Solutions knowledge base
3. MARC records for each individual e-book title through Serials Solutions 360 MARC Record Service

Due to the sheer volume and scale of e-book titles purchased, we found we were unable to perform basic quality control for our e-book packages. We were unable to manually check that every e-book title in a publisher title list was activated in Serials Solutions. Instead we spot-checked titles when activating an e-book package in the Serials Solutions knowledge base. We were unable to follow up e-book front file content the year following to codify a final title list and ensure each title was properly activated in Serials Solutions. Finally, we were not able to check that Serials Solutions provided a MARC record for each title activated in the Libraries knowledge base.

We were relying on publishers to provide Serials Solutions with the correct information for each e-book package and for Serials Solutions to manage

our e-book collections for us. We knew, however, that just because we did not have the staffing to manage the volume of e-books purchased did not mean we could rely on publishers and Serials Solutions to do this work for us.

For these reasons, we created a tool that would help automate e-book package processes, allowing staff to focus on problems by weeding out the titles that were correct. The volume of e-book packages purchased and titles included in each individual e-book packages is simply too great for us to manually ensure:

1. Every title included in an e-book title list is correctly activated in the Serials Solutions knowledge base
2. A MARC record is present in our library catalog for each title
3. We have access to every title on the title list at the publisher's site

The access verification functions of the eBook Reconciliation Database are designed to address these concerns. We load publisher-provided title list data and we load data from our ILS into the eBook Reconciliation Database; then write queries to compare the two sets of data. Through this reconciliation process, we ensure that every title included in an e-book package publisher title list is correctly activated in Serials Solutions and that we have a MARC record in our library catalog. The scale of e-book titles means we are not able to check access at the publisher's site for every single e-book title in every package, but staff do check access for problem titles and we have discovered and corrected patterns of access issues by this method.

We have learned that perpetual access as a licensing concept and perpetual access as an acquisitions concept do not necessarily agree. As companies are bought and sold, content changes platforms, or publisher's perform backend metadata cleanup, access to e-book titles we own "in perpetuity" is sometimes lost. Discovering titles that fall through the cracks and reestablishing access to those owned titles requires the Libraries be able to:

1. Demonstrate we own the title or titles in question through a license, a title list, or some other means
2. Provide proof of purchase

The eBook Reconciliation Database allows us to certify a title list, stamping it with a date and time giving us a title list of record. This title list shows we owned a title on a specific date and negotiate restored access. Acquisitions and discovery staff troubleshoot e-book access issues reported by public services and the first questions we ask are: do we own this title and should we have access to this e-book title?

Performing access verification, title reconciliation, and creating title lists of record are critical to answering this question quickly and efficiently. We are halfway to our answer when we can say with certainty that, yes,

1. The title is included on a title list of record
2. We had access to it on a known date and time

Knowing this information allows us to use our e-resource hub to track down the last pieces of information we need to contact a vendor.

Workflows and Tools

We started the e-book reconciliation project with the pilot phase in December 2014. Through the pilot phase we built up the database with some sample data, identified different roles and tasks for staff with various levels of knowledge and skills, and established a set of workflow and documentation.

In May 2015, we rolled out Phase 1. During Phase 1, we applied the one-way reconciliation strategy, which identified titles shown in the publisher list but not in our ILS. Staff were trained with adequate e-book knowledge and skills to complete the assigned tasks. We completed 21 e-book provider/packages reconciliation.

Phase 1 was an incredibly valuable learning experience for us. It allowed us to ensure titles in the publisher provided list were activated correctly in our system and to identify common

errors existed in our e-book data and to find solutions to correct those errors. We also discovered there was a significant delay for new e-book titles to appear in our knowledge base and realized it would be beneficial to create a reconciliation calendar for timely and efficient reconciliation. Moreover, to capture titles showing in our ILS but not in the publisher provided list, a two-way reconciliation strategy needs to be in place.

With the lessons learned and findings discovered during phase 1 we were very well prepared to move into Phase 2 in October, 2015. We modified the database to perform a two-way reconciliation, created the reconciliation schedule, streamlined the workflows, and provided more training to staff with new knowledge and skills. For the time being, we are working on reconciling 2014 e-book packages. And the report function to produce a certified title list is being implemented in the database. We will continue develop the lookup function in the database as our next phase.

There are a wealth of library systems including knowledge base, ILS, and discovery systems for libraries to choose from. These systems form a unique e-book data chain for different libraries. To design the reconciliation project, it's highly recommended to have a good understand of the data chain in your library system. At NC State Libraries, e-book titles are activated in the desired databases in Serials Solutions; MARC record service is activated in Serials Solutions and then loaded into our ILS, Sirsi. Finally the e-book data are pulled into our discovery system, Endeca, for patrons to discover. The data chain suggests that errors can be introduced while data are traveling from one platform/system to another one. The e-book reconciliation database has been designed to source data from the systems and capture errors presented in the chain.

The project is a collaborated effort between two units, Monograph unit and DPP unit, in our A&D department. E-resource librarians (ERLs) and staff are assigned with different roles and assignments. To manage and streamline the process, a seamless workflow with role assignments is indispensable. It starts with monograph ERL

identifying the e-book packages that need to be reconciled. Staff then collect title lists from the publishers and Sirsi. The two title lists for a specific publisher are then compared and the mismatches are identified. Staff then identify and mark the issue with each title and take corresponding actions to resolve the issue. The reconciled final lists are then generated and sent to publishers for confirmation. This completes the whole reconciliation process.

An intuitive project management (PM) tool has proved to be critical to the success of the project. We've been using Trello as our PM tool. Trello is a free cloud-based PM tool. It contains three essential elements: board, list, and card. We built an e-book reconciliation board, and it included several lists as tasks/steps, and created a card for each provider package to be reconciled. The package cards were moved around between the lists as team members moved through the tasks. With Trello, we have kept all the communication in one place, eliminating the use of multiple excel files and e-mails, and the process has been transparent to all team members.

Staff were assigned to identify and mark status in the database. This can be challenging for inexperienced staff due to the complex nature of e-books issues with various scenarios involved. The decision tree workflow maps shown in Figure 1 proved to be helpful to staff. A decision tree is a decision support tool, which uses a tree-like graph to display decision points and chances. It turns complicated if-then scenarios into a straightforward tree-like graph. Following the decision tree workflow map, staff were instructed to mark the status for each title and to take corresponding actions.

The Database and the Reconciliation Approach

The database is built with Microsoft Access. There are two reasons why we chose Microsoft Access. First of all, as part of the Microsoft Office suite, we have access to the Access in our library. Secondly, our DPP unit has an ERL and staff who have experience and skills of developing databases in Access. They have built and maintained several databases for the department.

The database has gone through evolution as we move along with the project.

The database has a dashboard shown in Figure 2, which includes a list of publishers that have e-book packages to be reconciled. Each staff member is assigned to one publisher. Clicking on each publisher brings up a table with a list of mismatched titles. The table includes several important elements including Type, ISBN, URL, Package, Status, and Publication Year. Type indicates the mismatch categories: No Sirsi Match or No Vendor Match. No Sirsi Match means the titles not showing in Sirsi, and No Vendor Match means they are not showing in the publisher lists. The Completed column is checked when all work is completed for the title. Package specifies the

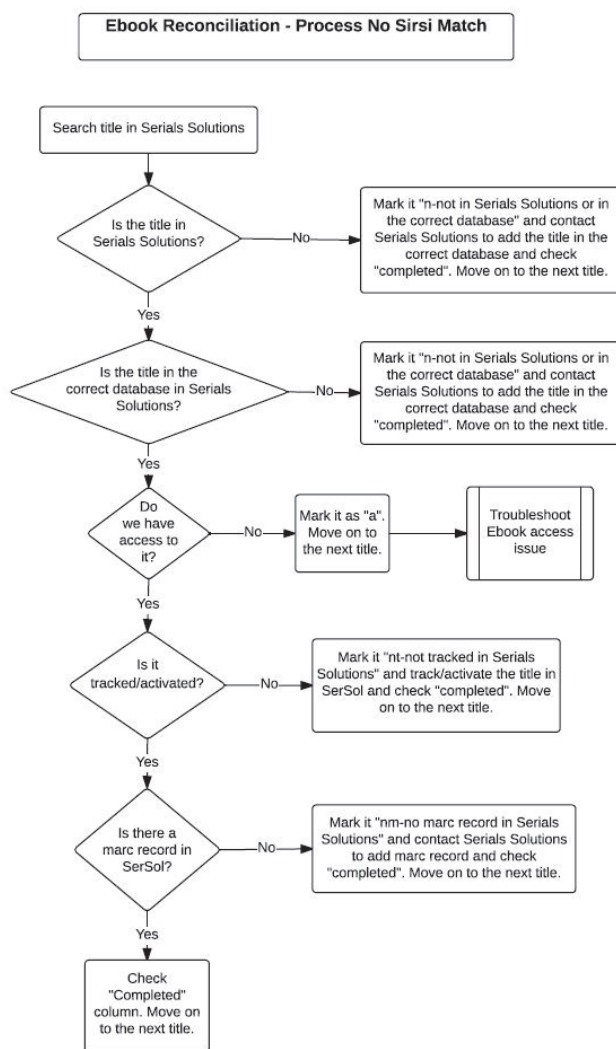


Figure 1. Decision tree map.



Figure 2. E-Book reconciliation dashboard.

Serials Solutions database that the title is or should be in, and it's manually ingested into the publisher title list by staff before the title list is pulled into the database. Staff rely on the package data to determine in which database the title should be activated. Status is used to categorize discrepancies between the two title lists. Several statuses have been identified during the reconciliation process and will be elaborated in the "Findings" section in this paper.

The reconciliation approach entails comparing two title lists for each publisher and identifying mismatches between the two lists.

Publishers were e-mailed to request the title lists for desired packages. Several metadata elements were required including: title, URL, ISBN, publication year, and package/collection. A staff member was designated to prepare the title list ensuring these essential data elements were included in the publisher list. The column names were standardized so that queries could be reused for all publishers/packages.

A custom bibliographic report was built in Sirsi to extract records from specific providers. The report was based on the provider name contained in the 856 field subfield z, which we requested as part of our MARC record service. See example below:

/zSpringer-Verlag SpringerLink ebooks— Behavioral Science (2009)

The URL or a portion of the URL was used as the identifier for matching titles on both lists. These are two examples of the URLs from the publisher list and the ILS list with the matching points underlined:

Publisher:
<http://dx.doi.org/10.1061/9780784478851>

ILS:
<http://ascelibrary.org/doi/book/10.1061/9780784413654>

This matching approach has been successful for many reconciled title lists. However, there are some publisher provided lists containing URLs with redirected DOIs. For example:

Publisher: <http://dx.doi.org/10.1117/3.547461>

ILS:
<http://ebooks.spiedigitallibrary.org/book.aspx?bookid=149>

OpenRefine Jython script was used to find the ending location of redirecting URLs. ISBNs did not prove to be useful because it's complicated to extract from Sirsi, and multiple URLs are often stored in MARC records and publisher lists.

Findings

Several most commonly seen issues have been identified and solutions have been provided to

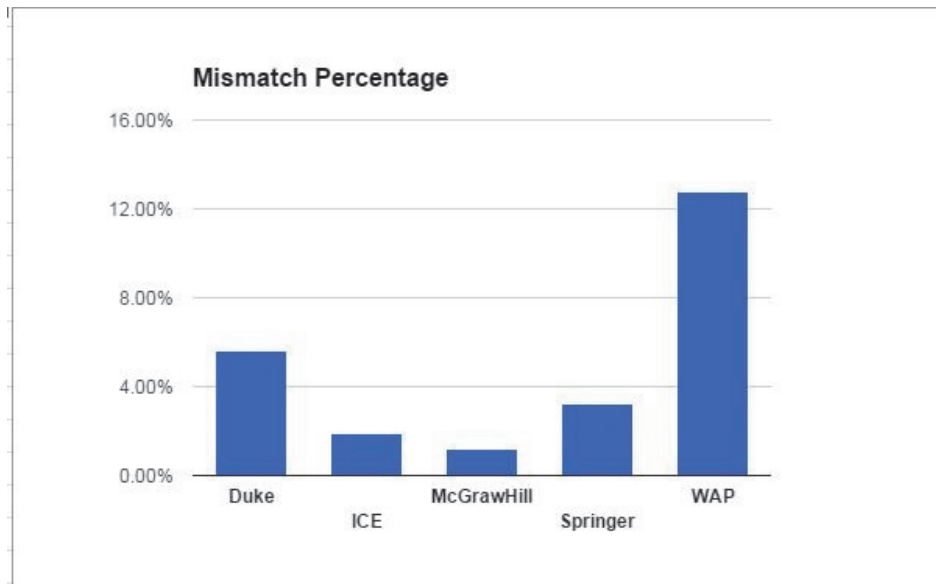


Figure 3. Mismatch percentage.

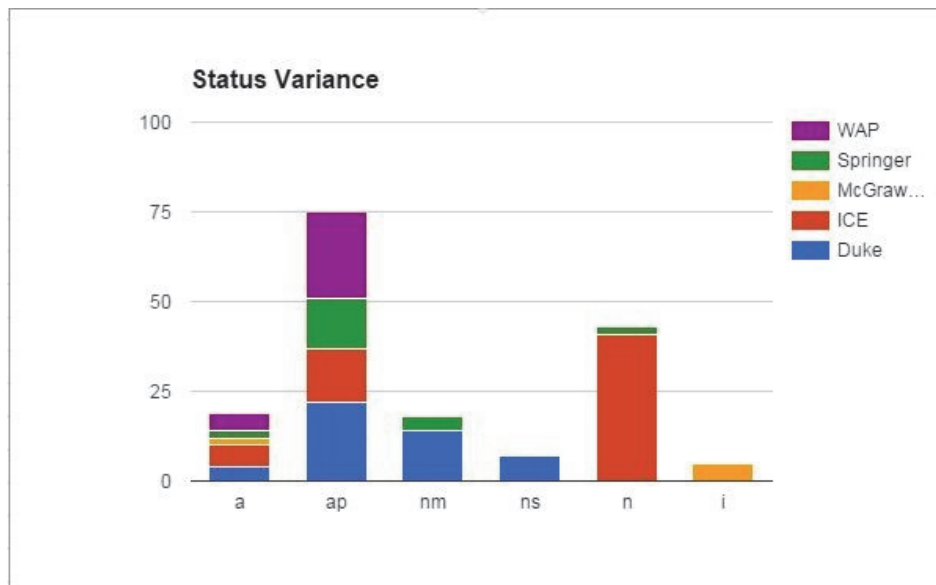


Figure 4. Issues/status variance.

resolve these issues. There were titles found not tracked in the correct database in Serials Solutions or not shown in Serials Solutions. Staff would activate these titles in the correct database in Serials Solutions or contact Serials Solutions to add them to the desired database. For titles without a MARC record in Serials Solutions, resulting in no Sirsi match, staff would request Serials Solutions to add a MARC record to the title. For no vendor match titles with access issues, we

would delete the record in Sirsi. For other no vendor match titles with proper access, publishers would be contacted to confirm our rights, and if rights confirmed, these titles would be added to the publisher list. The reconciliation process allowed us to identify titles with access issues. The solutions to these titles were variant depending on the causes. For issues caused by incorrect URL, the URLs were corrected either in Serials Solutions, Sirsi, or in the publisher list. For vendor platform issues, publishers were contacted to resolve them.

To justify the value of the project, let's take a look at the percentage of mismatched titles for some publishers, shown in Figure 3. It's interesting to see that the number varies among the publishers with ranges from 1% to 13% for these publishers.

Furthermore, we are very much intrigued by the difference among all the found issues. The chart below brings our attention to the discrepancies for the mismatches. It shows the number of titles for each of the issues for these five publishers. The top one is "ap," which means add to the publisher list. These are the titles not included in the publisher list, but we have rights and access to them. This is caused by the fact that publishers keep adding titles to the packages after we collect

titles from them. We feel fortunate to capture these constant changes, and meanwhile, it suggests we should perform the reconciliation soon after we collect the title list and the reconciliation process should be regularly conducted. The second large number of issues are titles either not in Serials Solutions or not in the correct database in Serials Solutions. This demonstrates that publisher and Serials Solutions are not consistent regarding what is included in a collection.

These discrepancies/issues reveal errors, legitimate or not. It also suggests the inconsistency among different segments in the e-book data supply chain. The reconciliation process allows us to correct these errors and maintain data integrity in the data supply chain.

Conclusion

NCSU Libraries eBook Reconciliation Database is an example of a homegrown tool that successfully facilitates management of e-book packages. The scale and volume of e-book package purchases necessitated that we create a tool to help address specific quality control concerns: access verification to ensure title-level representation throughout the Libraries's discovery layers; and, title list certification to definitively describe e-book package entitlements. By sourcing and comparing publisher title list data and ILS e-book MARC record data, we are able to take management of our e-book packages into our own hands for the first time. The local efforts we put forward may initiate a larger conversation with stakeholders in the e-book data supply chain on how to maintain data integrity in the system.