

2016

# Let's Get Technical--A Herculean Task: Cleanup in Preparation for Migrating to a New ILS

Stacey Marien

*American University Library*, [smarien@american.edu](mailto:smarien@american.edu)

Alayne Mundt

*American University Library*, [mundt@american.edu](mailto:mundt@american.edu)

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



Part of the [Library and Information Science Commons](#)

---

### Recommended Citation

Marien, Stacey and Mundt, Alayne (2016) "Let's Get Technical--A Herculean Task: Cleanup in Preparation for Migrating to a New ILS," *Against the Grain*: Vol. 28: Iss. 6, Article 39.

DOI: <https://doi.org/10.7771/2380-176X.7585>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

# Let's Get Technical — A Herculean Task: Cleanup in Preparation for Migrating to a New ILS

Column Editors: **Stacey Marien** (Acquisitions Librarian, American University Library) <smarien@american.edu>

and **Alayne Mundt** (Resource Description Librarian, American University Library) <mundt@american.edu>

## The Situation

In 2014, it was decided that the Washington Research Library Consortium (WRLC), a nine-member consortia of libraries based in the DC area, would be migrating to a new ILS. Even though the timeline had not been officially set and the selection of the new ILS system had not been chosen, the WRLC Metadata Committee (comprised of all of the heads of metadata and cataloging at the WRLC schools) was charged with identifying bibliographic and other data that would need cleanup before migrating. The group also had to identify other areas for potential cleanup, such as patron and checkout information and purchase order data.

Despite not knowing which new ILS system would be selected, the Metadata Committee was able to identify core areas that would need cleanup in order to make a migration go as smoothly as possible. The committee identified four main areas of cleanup work. Those areas were prioritized in terms of importance to their impact on a migration. In areas where the committee did not have the necessary expertise, we identified stakeholders that would need to weigh in separately.

## Areas for Cleanup

The first, most critical area we identified for cleanup was the need for an **OCLC** reclamation. **OCLC** will provide libraries a one-time free reclamation, also known as a data synch, to synchronize a library's holdings against WorldCat. This reclamation will ensure that holdings are up to date and accurate. This can be done again after an initial reclamation, but at a cost. At **American University**, we had performed this reclamation in 2009 as part of an experiment with using WorldCat Local.

The consortia had also recently gone through a project to identify books for retention through **Sustainable Collection Services (SCS)**, which has subsequently been absorbed by **OCLC**. Out of that project, remediation lists of records with potential problems were produced, such as records without **OCLC** numbers and records for which the title and author do not match Worldcat. We identified those records as easy targets for evaluation and cleanup as necessary. **American University (AU)** had several hundreds of these records, which were easy to correct because the problems were straightforward.

Another critical area the committee identified for cleanup was standardized bibliographic data. This area includes standard identifiers and codes, fixed length elements, and other incorrect or outdated information such as:

- OCLC Number (035)
- ISSN (022)
- LCCN (010)
- Leader (LDR) and Control Fields (006, 007, 008)
- Incorrect and/or missing indicators and/or subfields
- Obsolete MARC fields/subfields
- Local notes

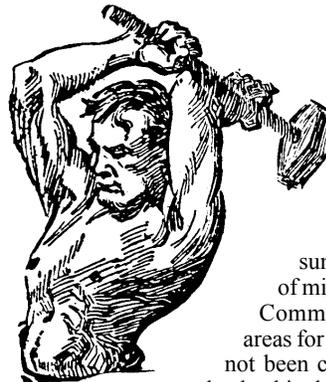
A final area we identified was cleanup or evaluation relating to inventory. This includes:

- Duplicate barcodes
- Items without barcodes
- Item records with empty barcodes or without barcodes
- Empty holdings records
- Bibliographic records without holdings records
- Holdings records without item records
- Print items attached to electronic resource records
- Short records (such as stub records created on the fly)
- Suppressed records

- Discrepancies between holdings and item location codes
- Two items attached to one holdings (which is a Voyager-specific problem given how Voyager displays multiple copies)

The committee also identified other areas that should be evaluated and cleaned up outside of bibliographic data, as well as stakeholders in different modules of the ILS with more expertise in those areas. These included Acquisitions, Circulation, Archives and Special Collections, and Electronic Resources departments, as well as other consortial committees involved with resource sharing and access services. We gave general recommendations for cleanup such as fund codes and old purchase orders to be worked on by colleagues in Acquisitions departments and expired patron data and obsolete statuses to be worked on by colleagues in Access Services. The committee also identified areas that may require collaboration across areas of responsibility, including items without barcodes, duplicate barcodes, and obsolete location codes.

The Resource Description Unit and Circulation Unit at **AU** have a strong working relationship and have collaborated on many projects over past years, including a massive move to storage initiative and various inventory undertakings. As a result of this, the **AU** catalogers have gotten a head start on many of the areas for cleanup that were identified by the Metadata Committee that required cross-unit expertise and responsibility. Although the Metadata Committee identified certain things as priorities to ensure a smooth migration, certain areas already had work in progress. These works in progress include a longstanding project to clean up approximately 7,000 short records created by circulation staff more than a decade ago; cleanup of outdated user/operator id's; and separating out two item records attached to one holdings record. These pre-existing endeavors have been happening on a low-scale basis for some time. This work gave us a bit of a head start on getting our data clean for migration, even though the areas **AU** was working on were already identified as being relatively low-priority.



Subsequently, **AU** has been working to identify and clean up several thousand duplicate **OCLC** Numbers (035s) in the catalog that were incorrectly mapped when we did the **OCLC** reclamation back in 2009. This is going as scheduled and is relatively time-consuming, but is absolutely critical in terms of migrating correctly. When the Metadata Committee was charged with identifying areas for data cleanup, the new ILS system had not been chosen. In August 2016, the WRLC leadership decided that we will be migrating to **Ex Libris's** Alma platform in 2018. The 035 field is the unique identifier used in migrating and mapping records into that platform so it is critical that this field be correct. We are also working to clean up fixed field and fixed length elements, such as the Leader, 006, 007, and 008 fields, which are more complicated to identify errors in due to factors such as whether or not they are indexed and where characters are positioned in fixed length fields. These problems may be complicated but they are far fewer in terms of volume to correct.

Since WRLC has decided to move to Alma, **Ex Libris** has provided us with recommendations for cleanup prior to migrating. **Ex Libris's** recommendations are largely in line with what the Metadata Committee identified previously, though there were some unexpected new recommendations for cleanup. The most notable and complex of these has been separating out records that have been combined from multiple formats

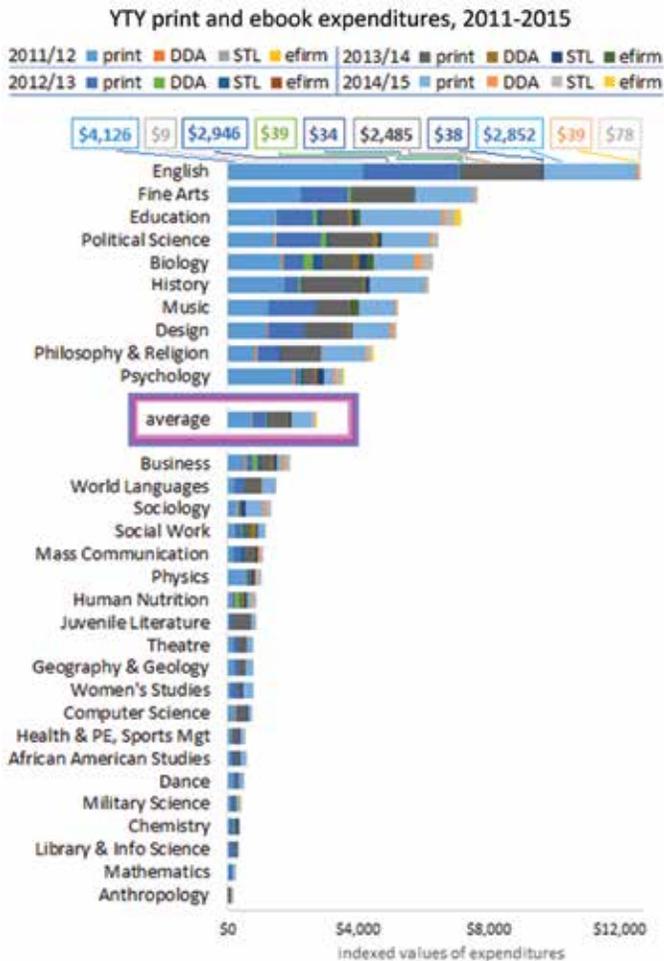
*continued on page 69*

**Print and eBook Expenditures by Discipline, 2011-2015**

Similarly to the last issue's article on broad eBook and print use and expenditure trends, the findings showcased here present eBook and expenditure trends by discipline. This section presents cumulative data for years 2011/12 through 2014/15 and preliminary comparison data from the new ILS for the 2015/16 operating year through February 18, 2016.

**Subject-specific expenditures for eBooks compared to hardcopy:**

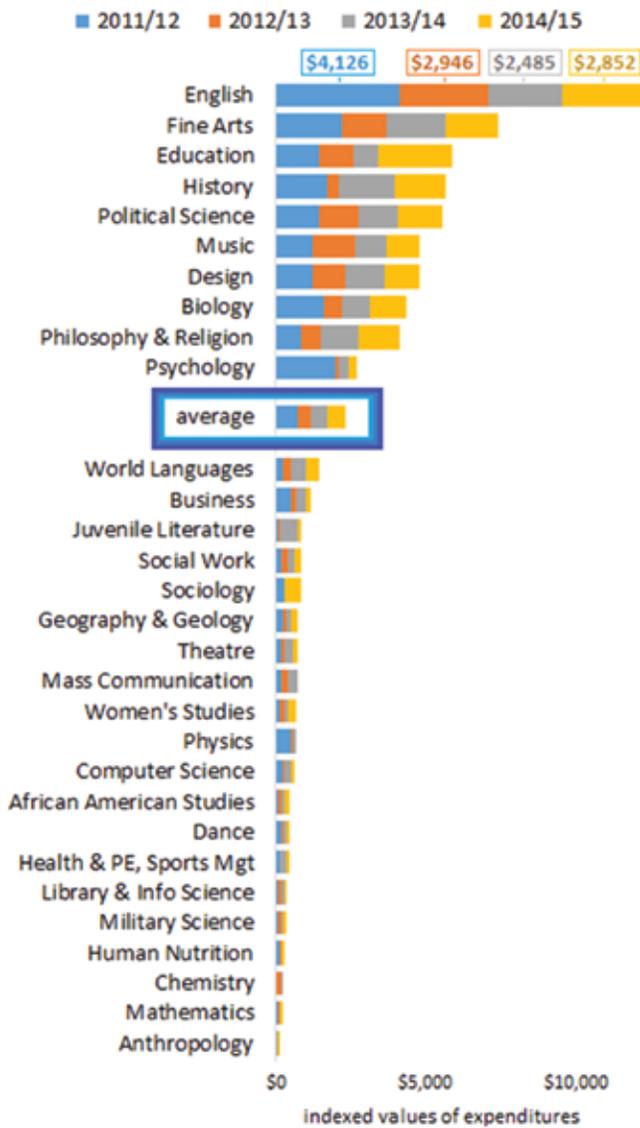
The chart below shows the 30 examined disciplines by each format's expenditures for each of the four years from 2011/12 through 2014/15 (ending June 30). The top five purchasers over the four-year period for print and eBooks combined are (1) English, (2) Fine Arts, (3) Education, (4) Political Science, and (5) Biology.



**Expenditures by Discipline: Print, 2011-2015**

The top five print book purchasers are (1) English, (2) Fine Arts, (3) Education, (4) History, and (5) Political Science. Of these, two areas are also among the top 5 eBook purchasers: Education (2nd place) and Political Science (3rd).

**YTY print expenditures, 2011-2015**



*continued on page 70*

**Let's Get Technical**  
from page 66

onto one record. Merging multiple formats of continuing resources onto one bibliographic record had been a longstanding local policy in the consortia, and will require collaboration among cataloging, electronic resources, and acquisitions units to resolve, since it also involves moving purchase order data.

**Recommendations**

With planning for the data cleanup taking place over the past year and the work already in progress, I would recommend the following general principles for an ILS migration:

1) Plan early. Even if your library hasn't identified which platform they are moving to or when, certain areas are all but universal to clean up when migrating, such as 035s, fixed-length fields, and unique identifiers. If nothing else, it will help improve your bibliographic data and improve discovery for users, will make the migration process go more smoothly, and will ease the burden of additional work by spreading it out over time. We began actively discussing cleanup in 2015, and are planning to migrate in 2018.

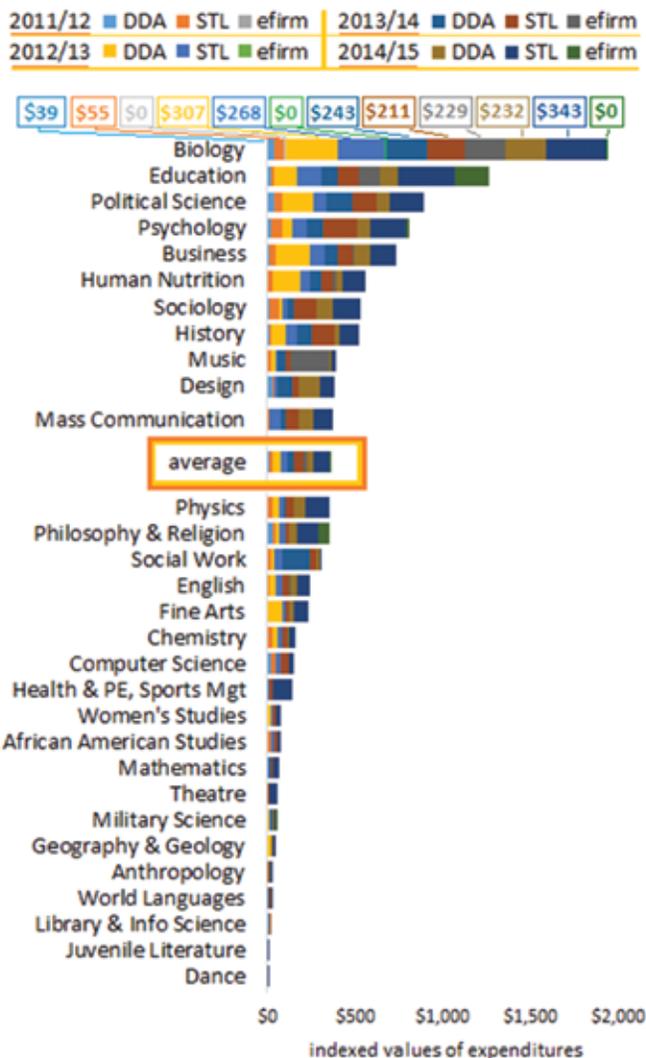
2) Communicate with colleagues. As one member of a nine-member library consortia with a shared catalog, it has been imperative that the heads of cataloging stay in communication with one another concerning data cleanup and sharing expertise. Additionally, within one library, there are so many cross-functional workflows and practic-

*continued on page 70*

**Expenditures by Discipline: eBooks, 2011-2015**

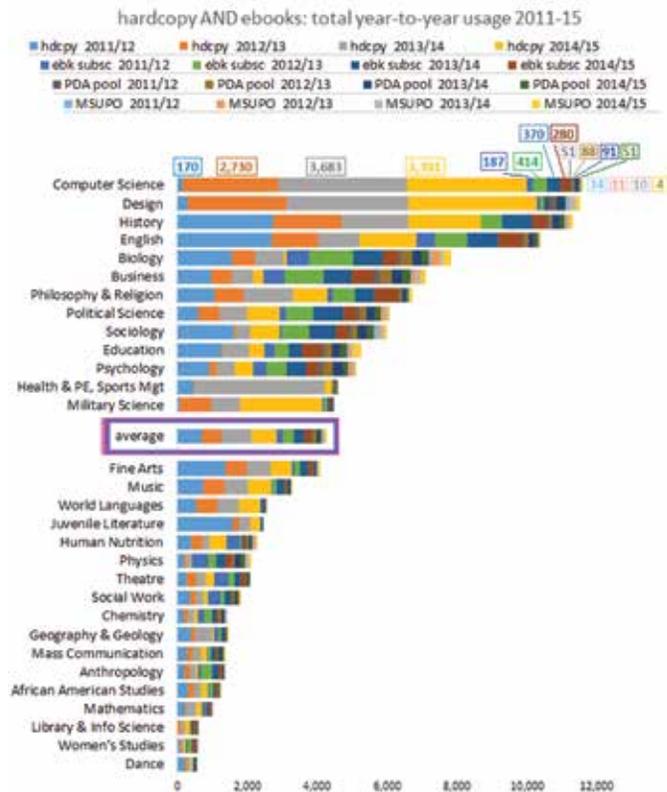
The following chart depicts **eBook expenditures** over the four-year period examined in this study. It shows each area's percentage of total eBook expenditure year-to-year. The **top five eBook purchasers** are (1) Biology, (2) Education, (3) Political Science, (4) Psychology, and (5) Business. Of these disciplines, not all are the largest print purchasers: Education is in 3rd place, Political Science 4th, followed by Biology (8th place), Psychology (10th), and Business in 12th place.

**YTY ebook expenditures, 2011-2015**



**Print and eBook Usage by Discipline, 2011-2015**

**Discipline-specific usage for eBooks compared to hardcopy:** The chart below shows year-to-year hardcopy and eBook usage and depicts the proportions of use between these two formats. While a few areas' eBook usage exceeds their hardcopy circulation (notably Physics, Social Work), eBook usage is not on a consistently upward trend among all disciplines.



**Usage by Discipline: Print, 2011-2015**

Of the top five **print** users, most are also among the higher-ranked purchasers as measured in expenditures, with some notable surprises: The top print users are (1) Design (7th among print purchasers), (2) Computer Science (in 21st place among print buyers), (3) History (in 5th place as print buyer), (4) English (1st-ranked print buyer), and (5) Health & PE (in 24th place among print purchasers). Two top five print purchasers were not among top five users: Fine Arts, the 2nd-ranked purchaser, placed 8th among users; while 4th-placed buyer Political Science placed 10th among users. The most dramatic usage increases occurred in Computer Science and Design, likely owing to the growth in course-taking and research projects in web and software design, interior design, and illustration.

*continued on page 71*

es, colleagues across different silos can provide insight and valuable perspectives on why certain things are the way they are and how this should inform different parts of the ILS cleanup and migration. Internally, AU has just started to have scheduled cross department meetings to start discussions, gather questions, and provide updates about the migration process.

3) Become an expert in running reports to pull data out of your ILS. For those of us on Voyager, we use Access to query data across

the different Voyager modules. Subscribe to relevant listservs to see what others are doing, and start getting familiar with SQL, Regular Expressions, or other relevant tools to help you pull the data you want from your ILS to work on cleanup. Much of what you need will already be out there in listservs, blogs, and articles, so you don't need to become an expert in these tools, but it helps to be able to identify what you need to know. Befriend your systems librarian and ask for help from the community. 🌸