It's Not Just About Weeding: Using Collaborative Collection Analysis to Develop Consortial Collections

Anne Osterman
Virtual Library of Virginia, aelguind@gmu.edu

Genya O'Gara
James Madison University, gogara@gmu.edu

Leslie O'Brien
Virginia Tech, lobrien@vt.edu

Anne Osterman, Genya O'Gara, and Leslie O'Brien, "It's Not Just About Weeding: Using Collaborative Collection Analysis to Develop Consortial Collections" (2014). *Proceedings of the Charleston Library Conference.*
http://dx.doi.org/10.5703/1288284315578

It's Not Just About Weeding:
Using Collaborative Collection Analysis to Develop Consortial Collections

Anne Osterman, Virtual Library of Virginia
Genya O'Gara, James Madison University
Leslie O'Brien, Virginia Tech

Abstract

From fall 2013 to the present, the Virtual Library of Virginia (VIVA) has undertaken a pilot collection analysis project with Sustainable Collection Services (SCS). This pilot has involved analyzing the main stacks holdings of 12 of the VIVA member libraries, a total of almost six million records. As is usual for an SCS analysis, the project involved comparing the pilot libraries' holdings with each other, the consortium as a whole, the state, and the United States, as well as with HathiTrust, the internet archive, and selected peer library groups.

The goals for this project were varied, but unlike most library groups, which have used SCS analysis services to inform collaborative print preservation and deselection projects, a primary interest for VIVA was to use the analysis to inform future collection development. The hope was that learning about titles that had been acquired and used across this representative cross-section of the consortium could be effectively translated into collaboratively acquiring e-books in a more thoughtful, data-driven manner, in addition to other collection development initiatives. This paper presents four different collection development approaches that have been applied to this shared data set.

Introduction

The Virtual Library of Virginia (VIVA) is a consortium of 72 nonprofit, academic libraries in Virginia. It includes both public and private colleges and universities, ranging in size from large doctoral institutions to small, specialized institutions, and the Library of Virginia. Central funding for the consortium is provided by Virginia's General Assembly, but there is also extensive cost-sharing by members to acquire products. The consortium is grounded in the coordinated collection development of online resources and an extensive resource sharing program, and more recently it has undertaken significant analysis of some of the member institutions' physical format materials.

In the fall of 2013, VIVA began a pilot collection analysis project with Sustainable Collection Services (SCS). This pilot involved analyzing the primary circulating holdings of 12 of the VIVA member libraries, a total of almost six million records. The project compared the pilot libraries' holdings with each other, the consortium as a whole, the state, and the United States, as well as with HathiTrust, the internet archive, and selected peer library groups.

The goals for this project were varied. Unlike most library groups that have used SCS analysis services to inform collaborative print preservation and deselection projects, a primary interest for VIVA was to use the analysis to inform future collection development. The hope was that in understanding the makeup of circulating collections, and how they were being used across this representative cross-section of the consortium, the consortium could effectively translate this information into collaboratively acquiring e-books in a thoughtful, data-driven manner, and that this would open up new opportunities for future collaborative collection development.

Background: Data Used

The 5.8 million bibliographic records analyzed in the project included all circulating, English language, LC-classed print monographs in the main stacks of the participating libraries. Not included in this analysis were theses and dissertations, reference materials, government documents, special collections, juvenile literature,

http://dx.doi.org/10.5703/1288284315578

e-books, musical scores, microform, audio-visual materials, serials, withdrawn materials, or those found in specialized libraries on the participating institutions' campuses such as law and medical libraries.

As there is no shared integrated library system (ILS) within VIVA, there were challenges for SCS in working with a diversity of systems (Alma, Voyager, Sierra, Millennium, and Symphony) and of library practices and policies. For example, a Google Book digitization project at the University of Virginia added extra circulation counts that could not be filtered out from regular check outs.

Background: The Collection Development Discussion

As mentioned above, collaborative collection development was of interest in this project from the very beginning. It was represented as one of the five basic project goals:

- Pilot a coordinated, consortial approach to collection assessment.
- Use the data and analysis to inform future, collaborative collection development.
- Identify scarcely held titles in need of protection.
- Begin a discussion about the possibility of reducing unnecessary duplication and saving local space through strategic weeding.
- Provide remediated and enhanced records back to the participating schools.

At first, this collection development goal was explored through conversation at an in-person meeting. Ruth Fischer from SCS was at this meeting, and she helped guide the conversation toward practical possibilities for data analysis. Many areas were determined to merit further analysis after that meeting, including the four that will be discussed in this paper:

- Look for local disciplinary strengths in both uniqueness and general holding levels to inform the possibility of distributed, collaborative acquisitions.
- Examine widely held and highly and recently circulated books to determine shared factors such as key publishers.
- Examine shelf life (how long after purchase/publication date do books in different subject areas tend to be used by patrons) as a means of informing the acquisition model selected for new e-books.
- Focus on the print holdings of publishers recently acquired by the consortium in e-format to see if usage patterns were similar.

Local Disciplinary Strengths

During this analysis, SCS was able to provide VIVA with a comparison of subject collection size by percentage, distributed across the pilot libraries, as well as a comparison of unique-in-Virginia titles held by the pilot libraries that participated in this project. This initial snapshot enabled the task force to ask a few key questions about VIVA's consortial monographic holdings:

- What does the whole collection look like distributed across the pilot libraries?
- What do our uniquely held titles tell us about our collections?

In general, there was a fairly wide (if not even) distribution of LC classes, although there were a few notable exceptions. Figure 1 shows the classes for which the percent distribution of total collections is most equitably shared, and an example of how that SCS presented that data.

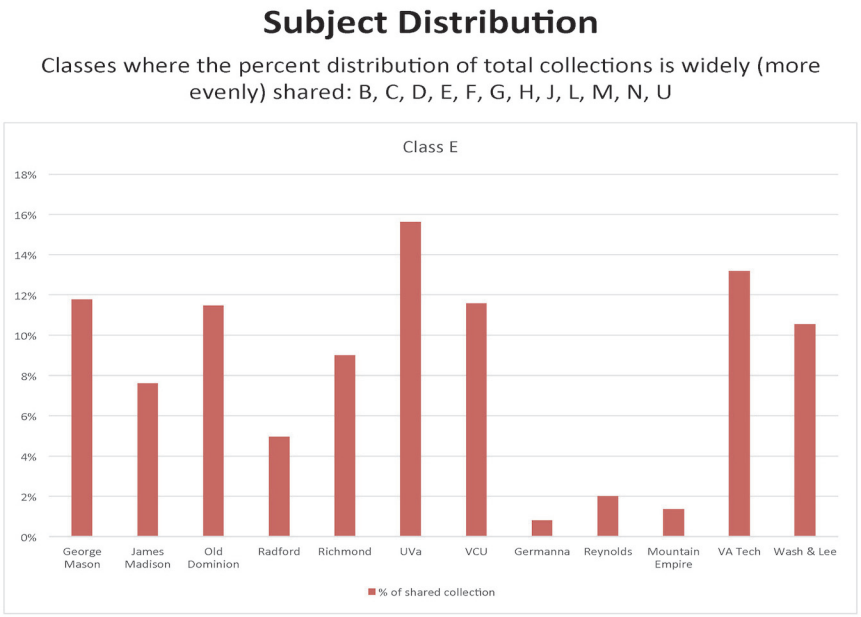

Figure 1.

For the classes where the total distribution of collections was not as widely shared, as detailed in Figure 2, there was often an explanation. For example, although some unlikely schools showed great relative strength in R (Medicine), the medical libraries at a number of participating schools had not been included in the analysis. Similarly, since reference collections were not included in the analysis, the A (General Works) category could be distorted by institutions that had moved more of their traditional reference materials into their main stacks.

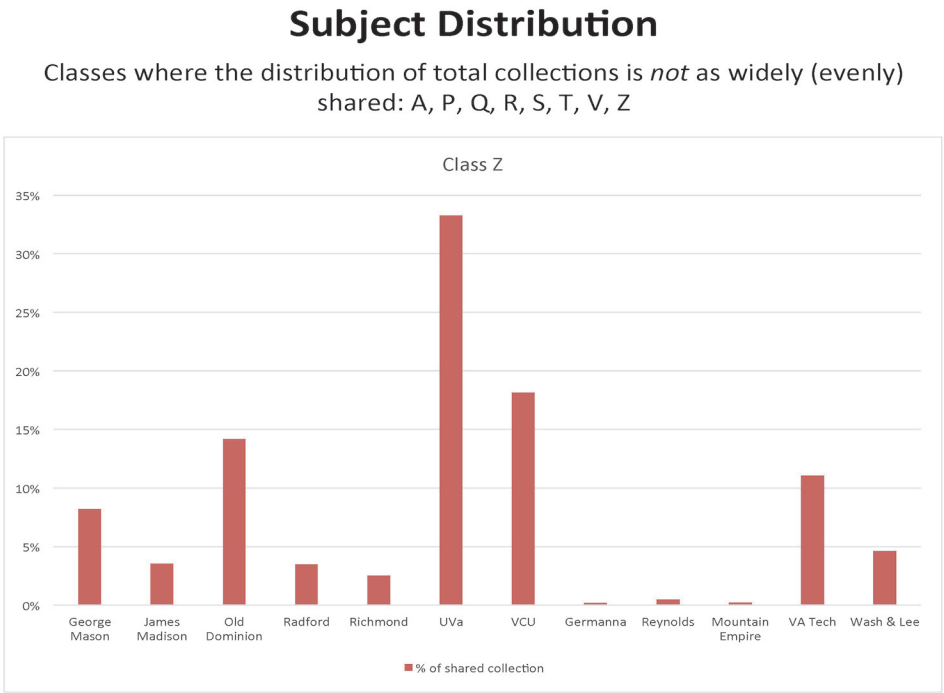

Figure 2.

Finding that the pilot libraries had a wide distribution of general subject strengths was critical to understanding VIVA's capacity to embark on future collaborative collection projects. The second piece of the puzzle was to examine how the depth of collection strength was distributed across the state. To look at this, the task force analyzed the percentage of unique titles at the pilot institutions and compared this to the percentage of total holdings in a given subject among the libraries. This tested the assumption that the collections were deepest where they were expected to be, e.g., where libraries had historic or current institutional disciplinary strengths.

Some things were quickly obvious. For example, in looking at all the data it was clear that the University of Virginia (UVA), one of the oldest and largest public research institutions in Virginia, held the majority of the unique titles across all of the collections analyzed even though the distribution of collections was more evenly distributed across the institutions. This can be seen in Figures 3 and 4, which illustrate the collection distribution by institution across the B (Philosophy, Psychology, Religion) and C (Auxiliary Sciences of History) classifications compared to the percentage of unique titles held by individual pilot libraries. In class B, for example, although there is a wide distribution of general holdings across institutions, UVA has 70% of the unique holdings.

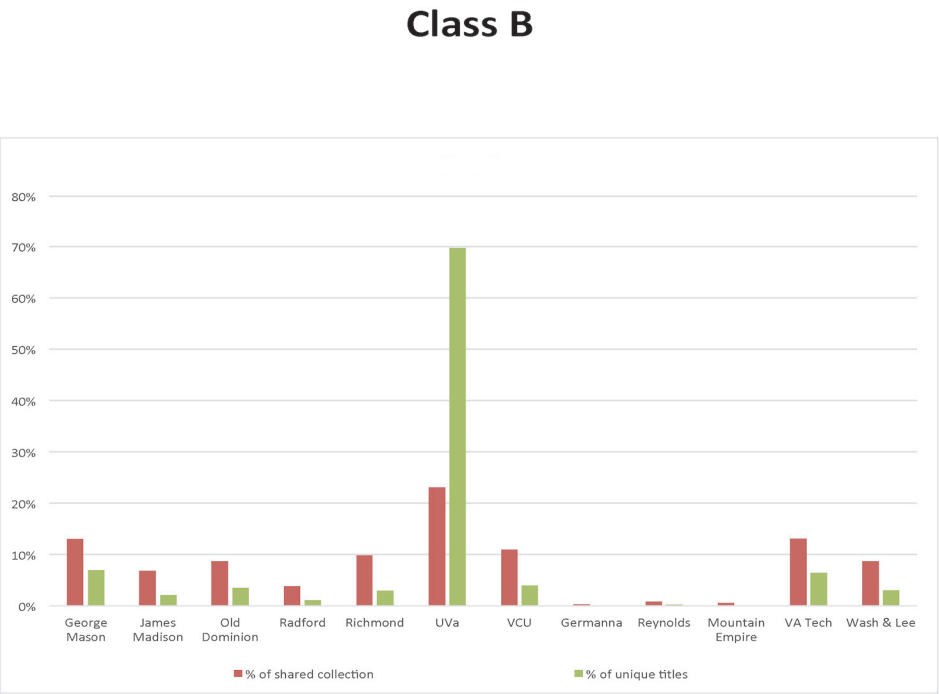

Figure 3.

Even though UVA holds the majority of the unique titles across all institutions, there were many examples of other institutions with high percentages of unique titles by LC class. For example, as seen in Figure 5, Virginia Commonwealth University (VCU) and UVA together hold around 60% of the unique art titles in the state, a total of over 20,000 titles. This was not surprising considering VCU's disciplinary and historical institutional strengths in art.

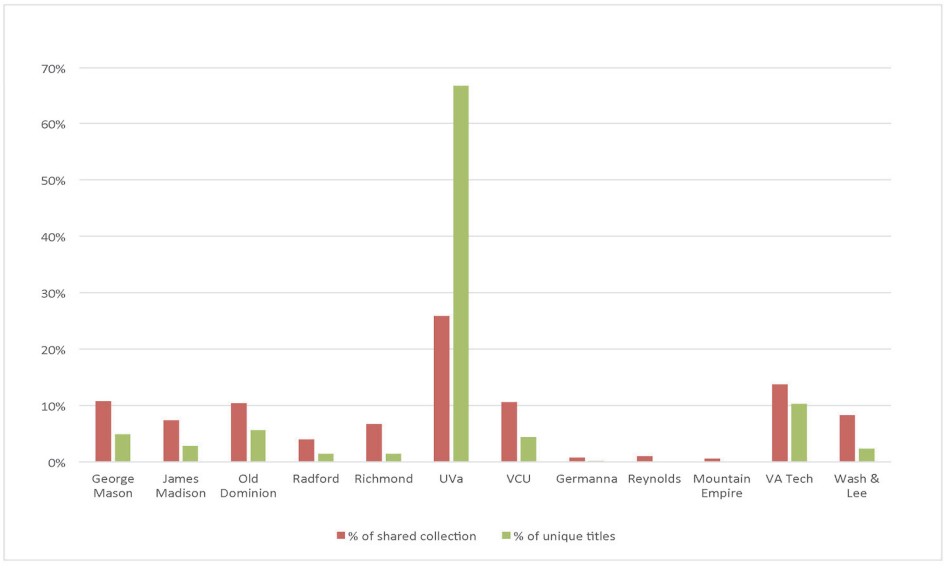

Figure 4.

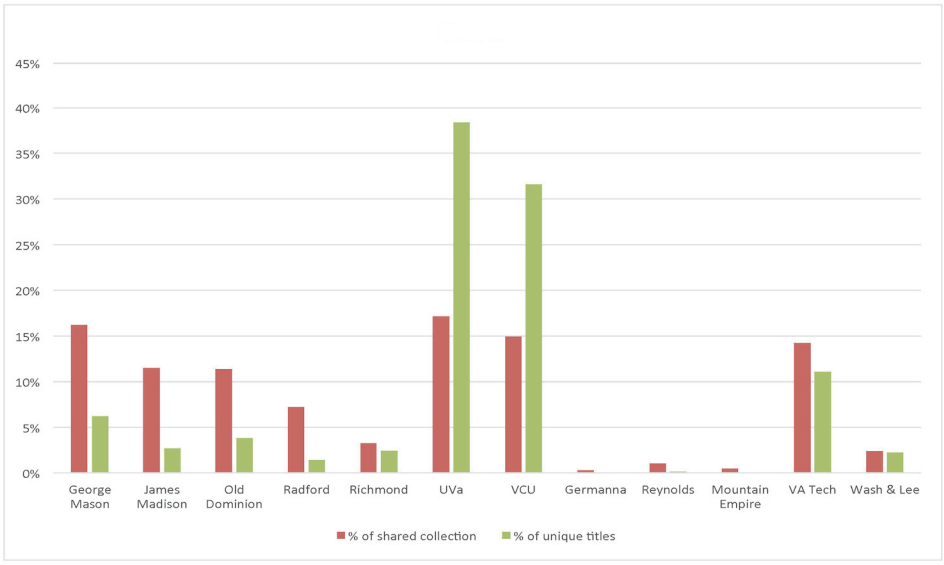

Figure 5.

Similarly, Virginia Tech (VT), as seen in Figure 6, shows depth in agriculture by holding 11,000, or over 75%, of the unique agriculture titles in circulating collections within the state. Again, this was in line with historical and current institutional strengths, but important to be able to visualize.

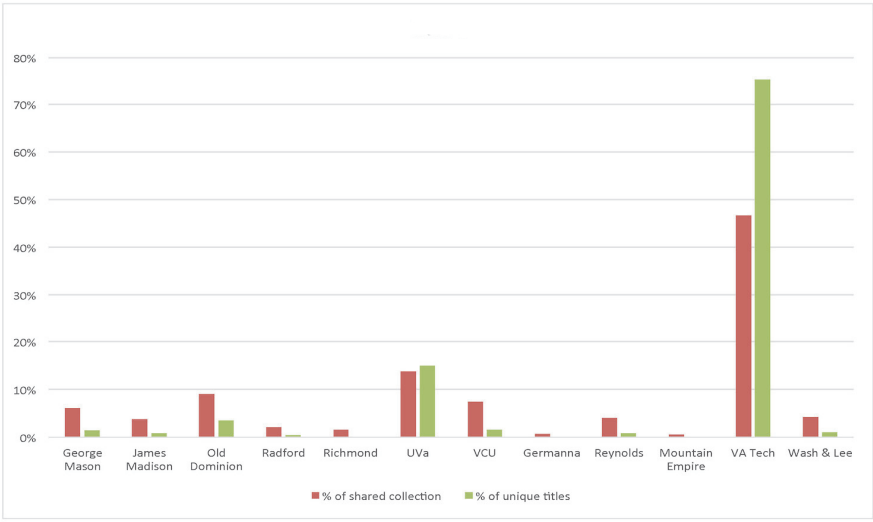

Figure 6.

Occasionally there were surprising results. For example, as seen in Figure 7, although James Madison University (JMU) is far from being the largest holder of education titles by LC class among the pilot libraries, it holds the largest number of unique titles in L—almost 30%. Not a surprise from a disciplinary point of view, in that education is historically a flagship program of that university, but surprising in number of unique titles and the resulting implied depth.

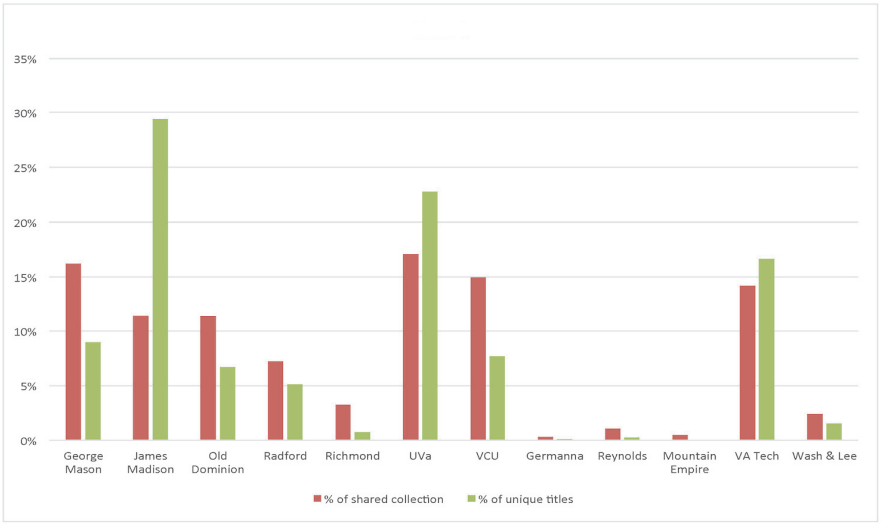

Figure 7.

These initial results gave the task force confidence that a potential future project of building on existing subject strengths within the consortium, where particular institutions could more formally become collectors for specific subjects on behalf of the other institutions, would have merit.

Widely Held and Highly and Recently Circulated Books

As there is no shared ILS or discovery system within VIVA, this collection analysis presented the special opportunity to look at a representative sample of the consortium and see which books were both widely held and highly used. E-books are still a relatively new acquisition approach for VIVA, and the hope was that this analysis might show patterns that would inform what kind of e-books would be useful for VIVA in general.

In order to do this, the task force set some benchmarks for these criteria and asked SCS to generate a list of ISBNs (and other data, such as titles, publishers, and publication years) of books that were owned by 10 or more VIVA libraries (in any edition), had more than 10 recorded uses, and had a last charge date after 2007. This list of just over 175,000 "widely held and highly and recently used" titles was then used in a variety of ways including:

- The ISBNs were sent to ProQuest's Title Matching Fast service to see which products matched up as good fits for VIVA.
- The ISBNs were matched to a standardized list of publishers using an in-house approach that had been used for similar studies in the past.

This second approach was useful in seeing patterns of publishers that might be appropriate for broad acquisition within VIVA. Over 3,200 publishers were matched, but, as can be seen in Figure 8, only around 150 had more than 200 titles in the list, fewer than 40 had over 1,000 titles, and only seven had more than 3,000 titles.

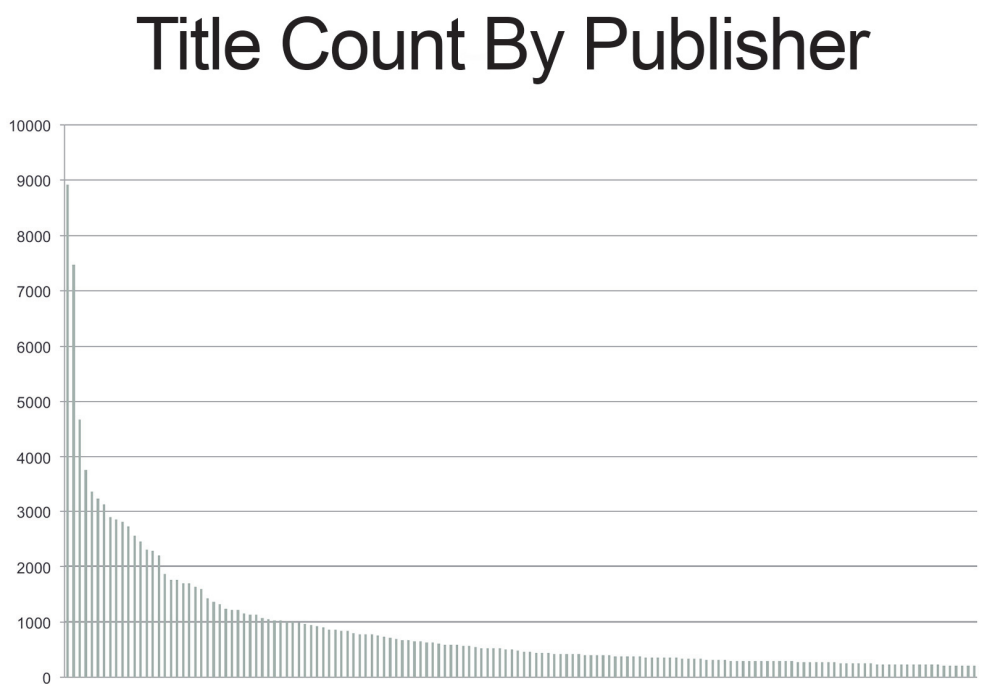

Figure 8.

The top publishers in this list were then included in a survey to collection development contacts that asked which e-book publishers they would like to have VIVA negotiate with for packages. This combination of evidence through analysis and institutional opinion gave clear direction toward a few publishers that merited further exploration, and some interesting discoveries arose from this analysis. For example, when the top ten publishers were reviewed more closely regarding holdings and usage (Figure 9), the data showed that although average holdings were higher for university press (UP) publishers, average recorded uses were higher for the commercial (Comm) publishers, at least relative to their holding levels.

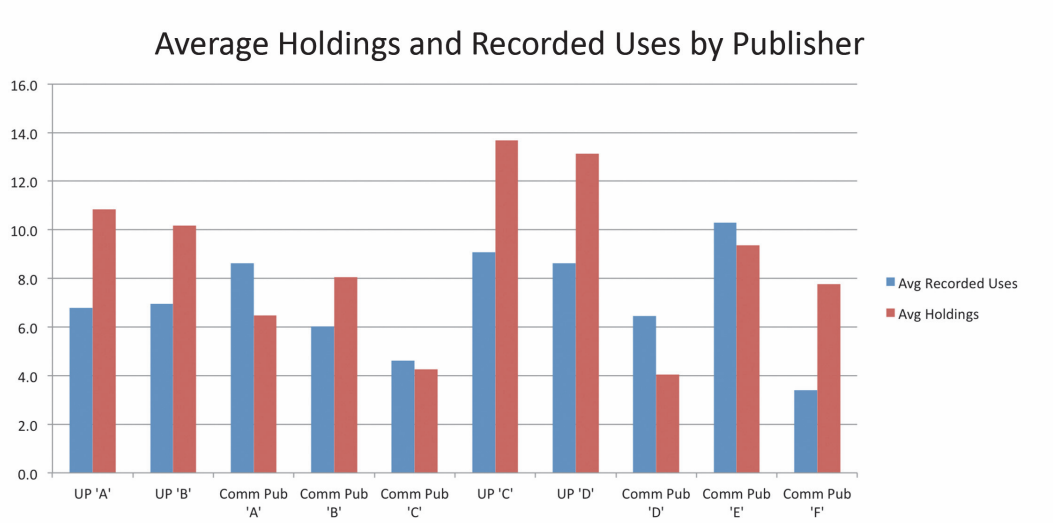

Figure 9.

Because the data included VIVA-wide holdings, not just the pilot library holdings, it was also demonstrated that this data could foster a discussion about how many copies have historically been held by VIVA in print, which could be used in pricing negotiations. A specific publisher is shown as an example of this in Figure 10. This publisher shows an overall average of 7.5 holdings in VIVA that becomes, with a general decline in print holdings over time, an average of only 6 holdings between 2008 and 2012.

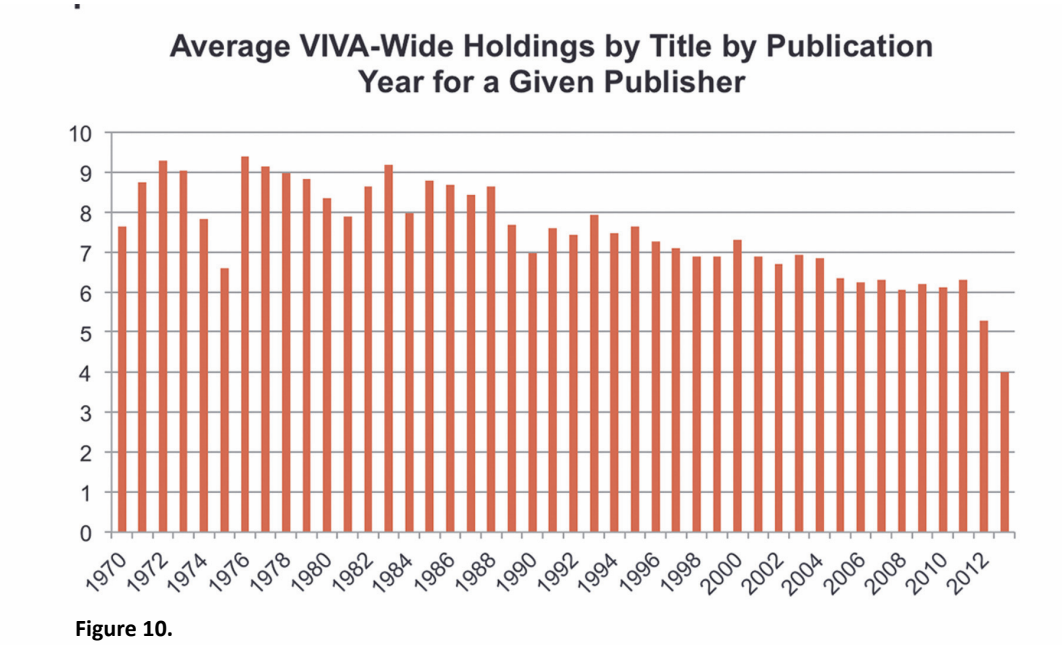

Figure 10.

Visualizing this level of duplication also led to deeper discussions about consortium-wide print holdings. It has been shown by other studies that academic libraries that rely on approval plan purchasing typically buy many of the same titles, and this did seem to be the case for VIVA. Data about duplication was used to guide recommendations about a distributed print repository or archive and establishing a voluntary threshold for new print copies, discussed further at the end of this paper.

Shelf Life

As e-books are still relatively new to the consortium as a shared purchase, also new has been the decision of what kind of acquisition model to use—perpetual access purchase, subscription, or demand driven. Although many factors play into this kind of decision, most notably pricing, the task force wanted to see if this analysis could inform a consortial preference of acquisition model, particularly for different subject areas of e-books. In order to do this, the task force focused on shelf life—or how long books are considered to be useful by patrons.

The task force examined the average number of years between publication year and last charge year for titles with the following criteria:

- Added to catalogs during or after 1990.
- Published in 1980 or later.
- Having a charge date.

Three LC-based subjects were considered: H (Social Sciences), N (Fine Arts), and Q (Science). These were chosen as they could potentially provide basic guidelines for the acquisition of e-books as either leases or perpetual access purchases in the three major divisions of Social Science, Humanities, and Science.

As shown in Figure 11, the global results of this were in line with general industry expectations—N had the longest shelf life, followed by H, and then Q. It is often the assumption made that Science titles “expire” in their usefulness sooner.

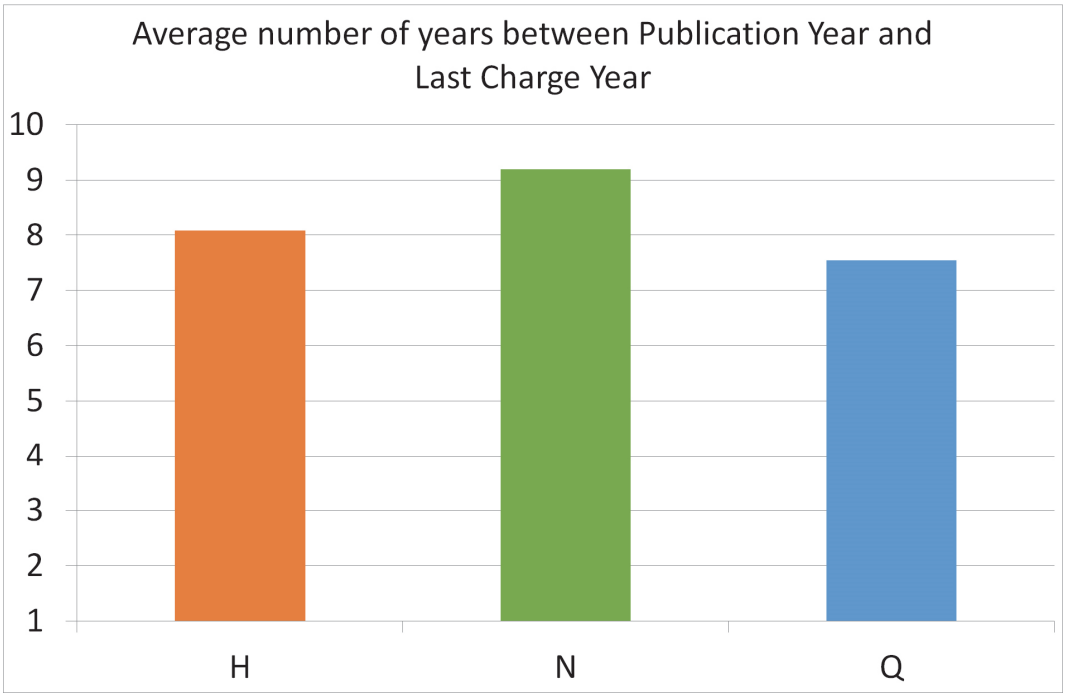

Figure 11.

When looked at in more granular detail (Figure 12), a distinctive higher pattern of titles with extensive years of usefulness in the N class can be seen, while a much greater higher percentage of titles in the Q class had only been used in their first year on the shelves.

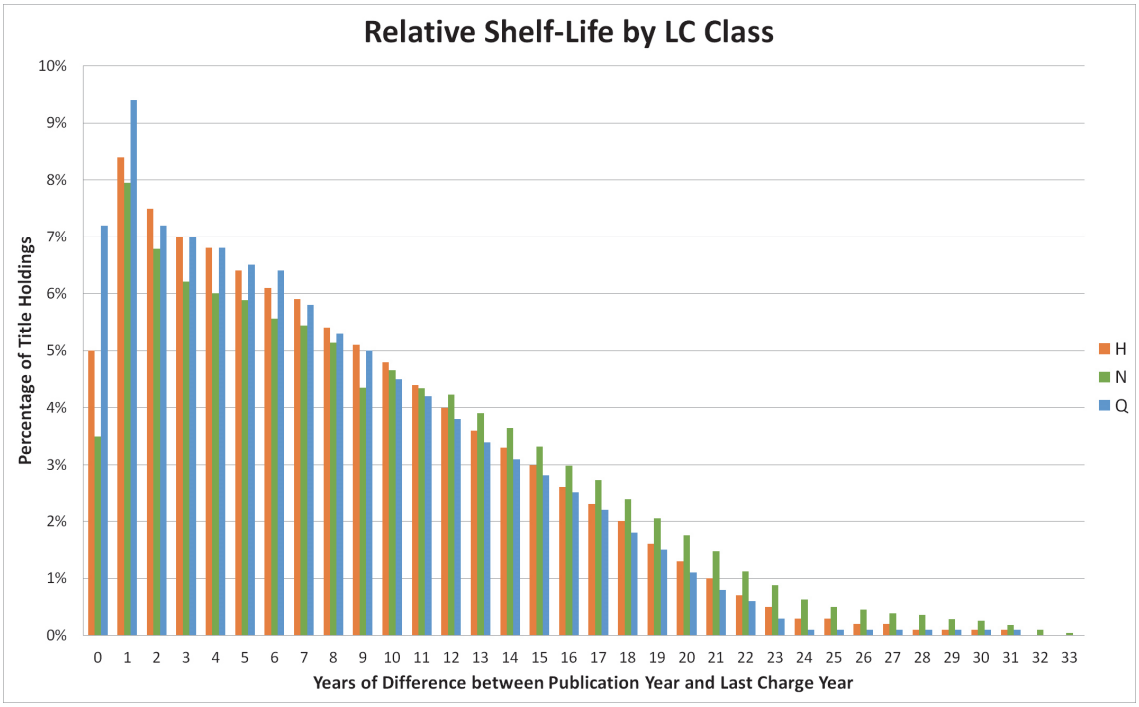

Figure 12.

When viewed at the subclass level, however, as in Figure 13, it can be seen that there is great variety within these broader classes. Some of the subclasses within H such as H, HA, HQ, HS, and HX had long shelf lives. One of them, HS (Societies), was above any other, including the N subclasses.

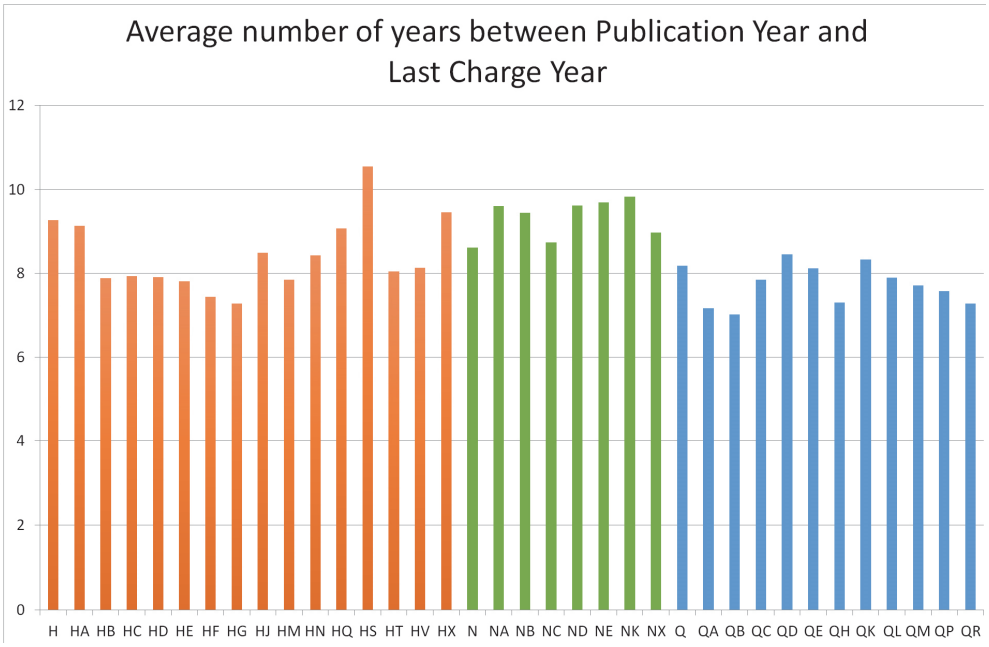

Figure 13.

It was generally recognized that this shelf life approach could be useful in informing future acquisition model decisions. For example, although the consortium's demand driven acquisition e-book pilot had been discontinued after a year due to a lack of state funding, if it were to begin again, different trigger-to-purchase levels could be set for different subjects. Similarly, as the publisher-based discussions progress, the subjects areas that a publisher is strongest in could inform a lease versus purchase decision.

E and Print Usage Comparison

STEM-H e-books have been of key interest in VIVA for the past few years, largely because the consortium received new General Assembly funding for STEM-H e-books in the FY13-FY14 biennium. This print analysis seemed like an excellent opportunity to look more closely at how the books from these publishers had been used in print and how that compared to their usage in the electronic format.

In order to enable this analysis, SCS provided individual records based on keyword searches in the publisher field. (It was fortunate that the three publishers of interest had names conducive to this process.) The task force then matched these holdings up to the shared electronic holdings using the ISBN and sometimes an intermediary match of print to electronic ISBNs. In order to maintain a fair playing field, only the 2013 resources were used in both formats.

As Figure 14 shows, the electronic format had higher levels of the proportion of available titles used. Likely due to the wider availability (across the consortium compared to a presence in only a few libraries), even in a small window of time the electronic format's impact was larger.

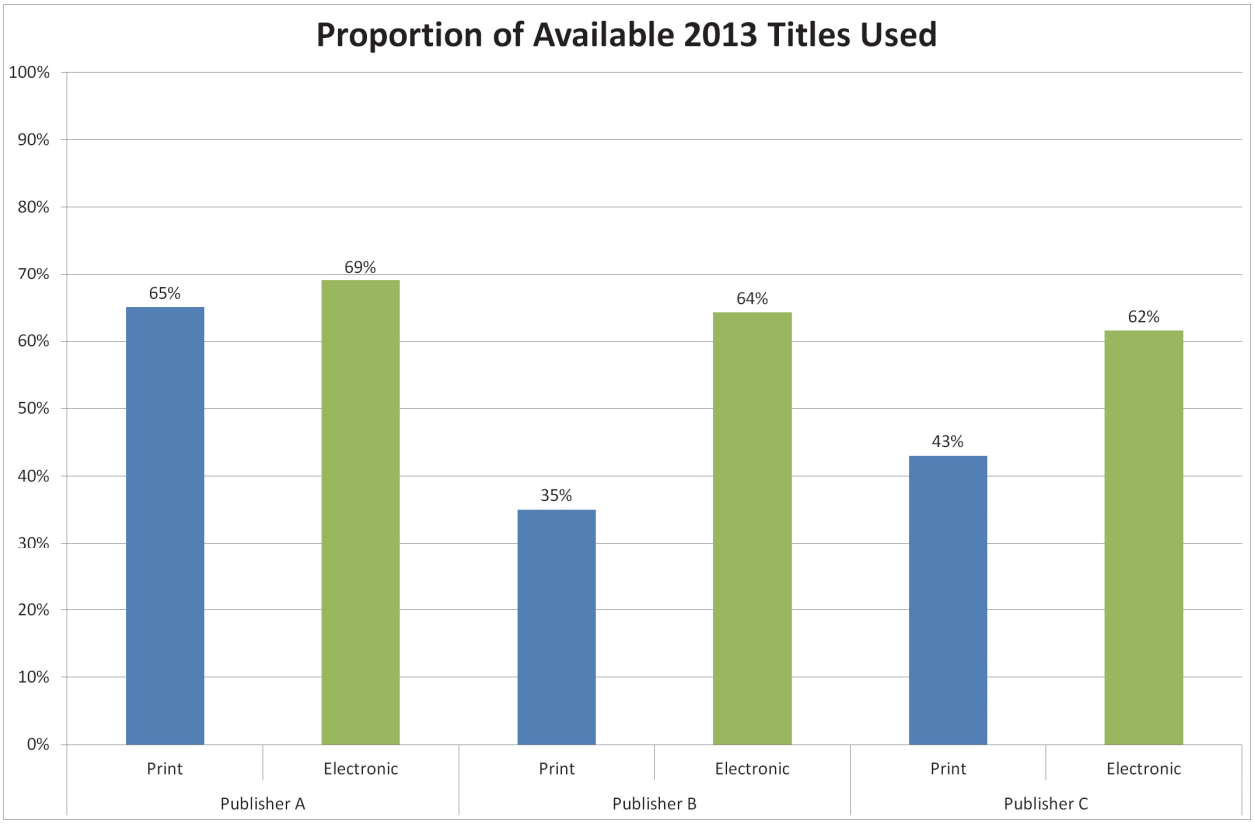

Figure 14.

An examination of the titles held in both print and electronic format by discipline (Figure 15) showed the overlap of what kind of usage was present by format. One of the most interesting results was an especially strong preference for the electronic format within R (Medicine). Only a few shared titles in this discipline had usage only in print.

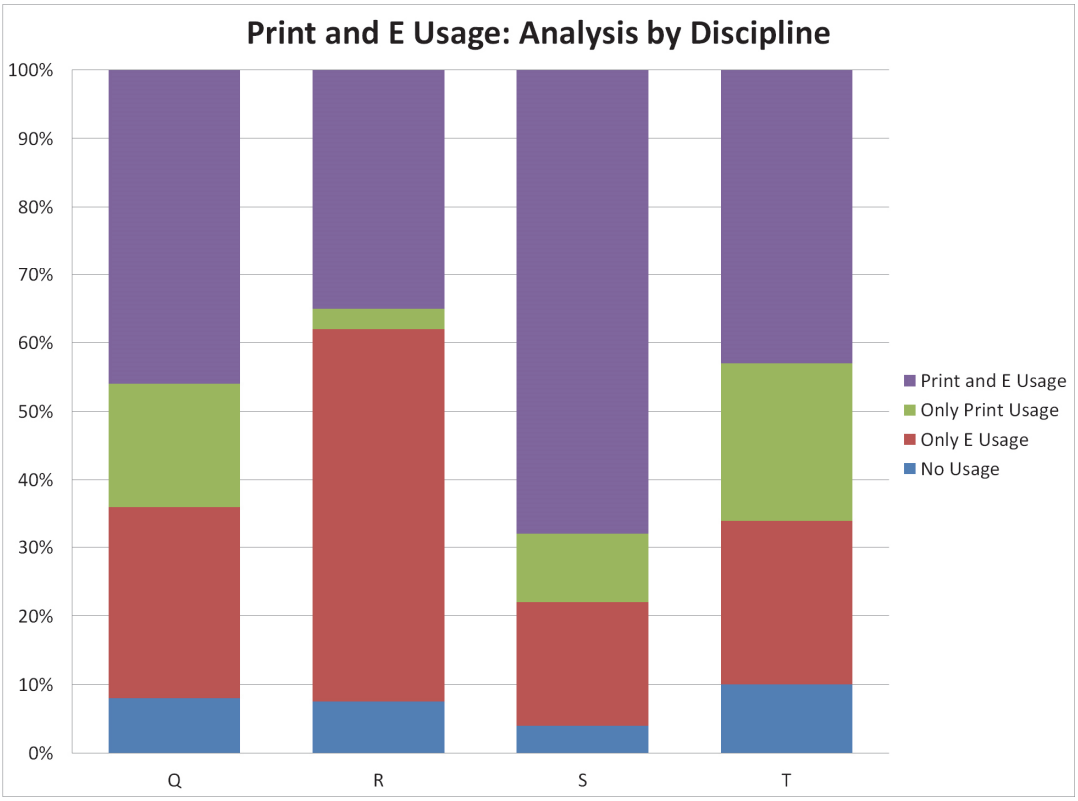

Figure 15.

Recommendations

Based on these collection development and analysis discussions, three major recommendations have gone forward from the task force and have been approved by the VIVA Collections and Steering Committees for further study:

Recommendation 1: Collaborative Retention of Widely Held Monographs

One of the early outcomes of this pilot had been the formation of a Memorandum of Understanding (MOU) for the Cooperative Retention of Rare and Unique Monographs. During the project, it was determined that there were over 72,000 main stacks books (in any edition) held by the 12 pilot libraries that were unique in Virginia and held by fewer than 10 libraries in the United States. Based on the MOU, each library will review the list of their titles and identify those worthy of retention, then add a note to the catalog for these titles to ensure their continued protection.

The discussion about collection strengths and duplication of titles across the pilot groups led to conversations about extending the retention project to widely held monographs to allow for safe deduplication within the consortium. If specific copies were set aside from individual institutions and protected from weeding, it would allow other institutions to more safely weed and reduce concerns about getting rid of the last copy or copies available in the consortium of particular titles. Such a project could incorporate the subject strengths seen in the collections by taking these strengths into account when allocating retention copies.

Recommendation 2: Establish a Recommended Threshold for VIVA Holdings as New Purchases

The duplication of title holdings across the consortium also led to discussions about how to prevent this level of duplication in the future. In line with similar projects done at the Orbis Cascade Alliance and OhioLINK, it has been recommended that VIVA member libraries buy print monographs on an individual basis, but in consultation with each other, to cut down on the number of holdings per title across the participating institutions. A common acquisition system, such as YBP's Gobi or Coutts' OASIS, could make this cross-consortium view possible, and at this time both systems are being explored.

Recommendation 3: Collaborative Publisher-Based E-Book Acquisition

The widely and highly and recently used title analysis showed strong patterns of key publishers that would likely be relevant across VIVA. This led to a recommendation that VIVA focus its collaborative e-book acquisitions on content from particular publishers as identified by the collection analysis as well as the survey of the VIVA collections contacts.

Conclusion

Overall, the pilot has provided VIVA with a wealth of data to mine. For a consortium without a shared library system, this project has enabled a view into print monographic data that was simply not possible before.