Representation of Local Collections in Web-scale Discovery Tools -Do Discovery Tools Promote Local Holdings?

Sharon Q. Yang
Moore Library, Rider University, 2083 Lawrenceville Rd, Lawrenceville NJ, 08648 USA.
E-mail address: yangs@rider.edu

Patricia H. Dawson
Moore Library, Rider University, 2083 Lawrenceville Rd, Lawrenceville NJ, 08648 USA.
E-mail address: pdawson@rider.edu

Abstract:

Web-scale discovery tools are the top-notch search engines that provide a single search across a library's catalogue, databases, and digital collections. Most academic libraries have deployed one or two web-scale discovery layers by now. Many libraries acquired discovery tools in the hope of boosting dwindling local borrowings. However, anecdotal evidence seems to indicate a rise in the usage of journal articles and ebooks and a decrease in the circulation of local print collections after the implementation of a discovery service. Quite often we heard complaints that discovery tools promote journal articles and ebooks more than print collections. In spite of the vendors' assurance that their products promote library collections, we know very little about how library materials are represented in the central index and relevancy algorism. This paper will discuss the findings from a research project by two academic librarians who attempted to exam local collections in web-scale discovery services including EBSCO Discovery Service, Primo, Summon, and WorldCat Discovery Services. We hope that the research discussed in this paper may serve as a starting point that inspires more libraries to do research on local collections in discovery services and motivates vendors to improve products.

Keywords: Discovery services, discovery tools, EDS, Primo, Summon, WorldCat Discovery Services.

INTRODUCTION

In 2014 when Rider University Libraries decided to deploy a discovery tool, one of the motivations was to boost dwindling circulation statistics and attract more users to use library collections. However, the circulation statistics for local collections did not increase visibly in subsequent years. Evidence indicates that "As a result of implementing a discovery service,

electronic journal usage went up, but print circulation went down" (Calvert, 2015). This claim was echoed by other studies (Greiner, 2014; Lawton, 2015; Levine-Clark, McDonals, & Price, 2014). One librarian wrote in his blog, "I have never come across a book from our catalogue in all my research using the discovery service… As a small library our book circulation stats are already pretty dead, I doubt the discovery service will put any extra nails in the coffin. However, it is still very much a possibility" (Regier, 2015).

The aftermath of implementing a discovery service may seem problematic and it was addressed by discovery vendors in their product documentation. EBSCO Discovery Service (EDS) reassured libraries that "…customers have the option of influencing the overall relevance weighting of their catalog and/or institutional repository. This optional setting enables all catalog and institutional repository records to appear higher (or lower) in the search results list relative to other content in the EDS profile" (EBSCO, 2017). Primo also claimed that customers "…can boost local items so that they appear higher on the result list than items coming from Primo Central, all other factors being equal" (Ex Libris, 2017). Summon and WorldCat Discovery Services (WorldCat) also made similar statements that the location of each item is analysed to increase the relevance of local collections and their products "Specify the search results to represent your library's holdings first…" (Ex Libris, 2014; OCLC, 2017).

In spite of the efforts by discovery service vendors, the local collections suffer increasingly low usage statistics. There is fear that as discovery service providers are also content providers, and they may favour journal articles in databases than library collections. A bias is suspected. "Discovery tool vendors may have better information about their own content, boosting certain articles higher in results" (Asher, 2013; Parry, 2014).

The following study was prompted by our curiosity about how local collections are being presented in major discovery services in design and search results. The findings may help us to understand more about the controversy.

LITERATURE REVIEW

A review of literature did not reveal any significant research on local collections in discovery services. There seems to be a gap on this subject. However, abundant research exists on other aspects of discovery services. Up to now most research on discovery tools has focused on usability, user behaviour, use pattern changes, and the impact of discovery services on reference and library instruction.

More relevant and closely related to our study is research on changes in use patterns after the implementation of a discovery service. An earlier study reported a 50% to 150% increase in students' use of full-text databases and online collections by Grant Valley State University after they had implemented Summon (Sean, 2010). Another study reported an increase of usage in e-journals and ebooks and a decrease in print collections in Manitoba University Libraries after the implementation of Summon (O'hara, 2012). Studies on EDS, Primo, and WorldCat Discovery Services produced similar findings that use for electronic materials had gone up and print collection down (Asher, 2013; Calvert, 2015; Greiner, 2014; Lawton, 2015; Levine-Clark, McDonalds & Price, 2014; Regier, 2015).

Especially noticeable is the finding that a discovery service use may influence researchers' experiences (Parry, 2014). According to a study by Bucknell University and Illinois

Wesleyan University, researchers tend to use what they are given by a search service. For instance, students who used EDS used more journal articles for their assignment. Students who used Google Scholar and the library catalogue used more books, while students who searched Summon used more newspaper and magazine articles (Asher, Duke, and Wilson, 2013). According to Andrew Asher, assessment librarian at Indiana University at Bloomington, "it's a logical impossibility to create a querying tool that doesn't have any form of bias" (Parry, 2014).

All the above cited studies and findings lay the background that led to the research project in this paper.

RESEARCH METHOD

Purpose and goal: Investigate how local collections are presented in discovery services in design and display of search results.

Scope: Only web-scale discovery services are included in the study including EBSCO Discovery Services (EDS), Primo, Summon, and WorldCat Discovery Services. They all use a central grand index for searching and coverage of resources. Web-scale is an Information Technology (IT) term denoting grand processing and high speed computing power. Web-scale discovery tools are top notch search engines for libraries.

Sample Libraries: To offset bias from local configurations, 40 libraries are randomly chosen from the customer list of the vendors or Internet with ten examples for each discovery tool selected We purposefully included English speaking but non-US libraries. The libraries in the sample comprised academic institutions with multiple disciplines (liberal arts and not specialized) with an English language discovery tool regardless their geographic locations.

Research Procedures: Local collections are defined as both physical items located in the library and catalog items which have a bibliographic record in the OPAC. The latter may include electronic collections as well as physical items. Most Primo and WorldCat libraries do not have a separate classic catalog. In that case we used catalogue searches within the discovery tool for catalogue data. We only examined the top 20 search results in discovery layers because past research shows that 87% of the users will not view search results beyond first 20 (Guan & Jia, 2016) and that most students stop reading beyond first page of search results (Hanneke & O'Brien, 2016).

Search Terms: One search was conducted using search term "pesticides AND environment" in each library's discovery service and catalogue. In case the library does not have a classic catalogue, we searched the catalogue or local records through the discovery service. Most discovery tools still maintain a separate catalogue presence in the system so we could isolate them for our research.

Data recording and analysis: We recorded the numbers of physical and catalogue items respectively in the top 20 hits in discovery tool search and their positions (1 to 20). Where possible, we would compare the items in local collections retrieved in discovery services to those found in the catalogue. When a library does not have a separate catalogue, we would locate the presence of the catalogues in discovery tools and separated local items and catalogue

items from the discovery interfaces. We ranked the top 20 items in a discovery retrieval based on authority, relevance, and timeliness. An Excel file was used to record data for analysis.

 RESULTS

The resulting sample includes libraries from Australia, Britain, Canada, Israel, Italy, New Zealand, Sweden, and United States (Figure 1). The following is a breakdown of the 40 libraries in the sample by their geographical locations. Most are from the US and 18 are from other parts of the world.

	AUSTRALIA	CANADA	UK	US	OTHERS	TOTAL
EDS	1	2	0	5	2 (ISRAEL & SWEDEN	10
PRIMO	1	0	2	7	0	10
SUMMON	2	1	1	5	1 (NEW ZEALAND)	10
WORLDCAT	0	3	1	5	1 (ITALY)	10
TOTAL	4	6	4	22	4	40

Figure 1. Geographic Locations of the 40 Libraries

Only 23 out of the 40 libraries in the sample maintain a separate catalogue in addition to a discovery tool. Most Primo and WorldCat libraries no longer maintain a separate classic catalogue, but one can still determine local collections in the discovery search results. EDS and most Summon libraries still have their legacy catalogue (see Figure 2). Only one Primo library and three WorldCat libraries still keep legacy catalogues.
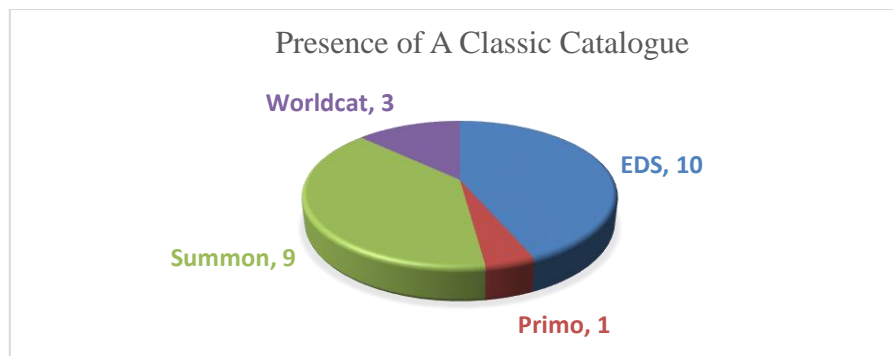

Figure 2. Presence of classic catalogues in sample libraries

Figure 3 shows the average number of physical items a library owns grouped by the discovery the library uses. The bar chart is a breakdown by discovery tools of library physical collections on "pesticides AND environment". It shows the average number of items in the physical collections of a library including print books and periodicals, CD/DVDs, and other physical items. Libraries that have EDS in the sample happened to have most physical items (more than 89 items) per library on average followed by WorldCat libraries (about 72), Summon Libraries (about 62), and Primo Libraries (about 56) with this query.
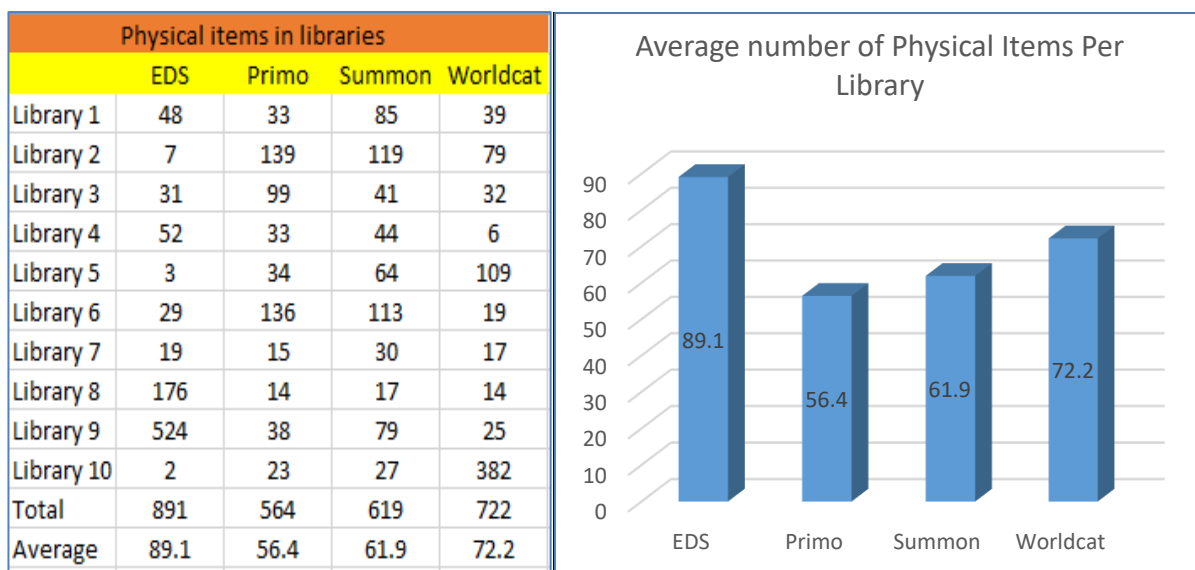
Physical items in libraries				
	EDS	Primo	Summon	Worldcat
Library 1	48	33	85	39
Library 2	7	139	119	79
Library 3	31	99	41	32
Library 4	52	33	44	6
Library 5	3	34	64	109
Library 6	29	136	113	19
Library 7	19	15	30	17
Library 8	176	14	17	14
Library 9	524	38	79	25
Library 10	2	23	27	382
Total	891	564	619	722
Average	89.1	56.4	61.9	72.2

Average number of Physical Items Per Library

EDS 89.1, Primo 56.4, Summon 61.9, Worldcat 72.2

Figure 3. Average number of physical items on "pesticides AND environment" per library in the sample

Likewise Figure 4 shows the average number of catalogue items a library has grouped by what discovery service they use. The bar chart is a breakdown by discovery tools of catalogue items on "pesticides AND environment" in a library. Catalogue items include all items that have a bibliographic record in the local catalogue or database. Catalogue items are supplied by libraries for indexing in the central index of a discovery tool and they may include electronic collections. Please note that there is missing data from one Summon library. In this particular example, we could not isolate catalogue items from discovery data which makes it impossible to compare the totals between nine and ten libraries. Therefore we used averages as the basis for our calculation in this paper. This way missing data will present less a problem for data analysis.
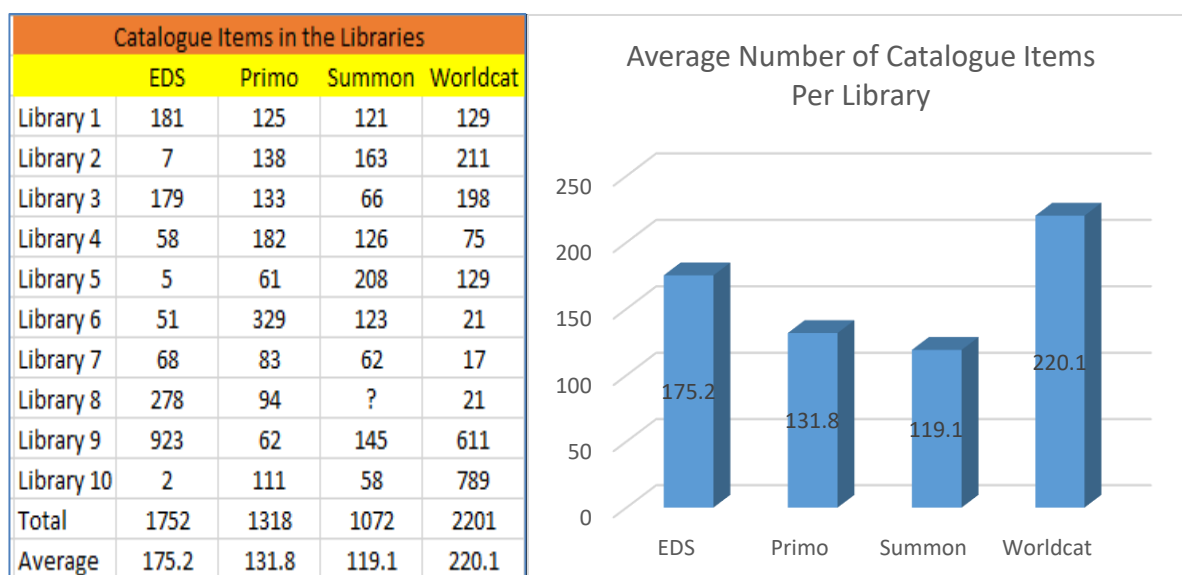
Catalogue Items in the Libraries				
	EDS	Primo	Summon	Worldcat
Library 1	181	125	121	129
Library 2	7	138	163	211
Library 3	179	133	66	198
Library 4	58	182	126	75
Library 5	5	61	208	129
Library 6	51	329	123	21
Library 7	68	83	62	17
Library 8	278	94	?	21
Library 9	923	62	145	611
Library 10	2	111	58	789
Total	1752	1318	1072	2201
Average	175.2	131.8	119.1	220.1

Average Number of Catalogue Items Per Library

EDS 175.2, Primo 131.8, Summon 119.1, Worldcat 220.1

Figure 4. Average number of catalogue items on "pesticides AND environment" per library in the sample

Figure 5 shows the composition of the top 20 hits in discovery tools and how much local collections get into the prominent display. The bar chart shows the percentage of local and catalogue items in the top 20 discovery search results. In the following bar chart, Summon's top 20 search results contains most items from local physical collections in comparison, about 31.30%. Primo and WorldCat's top 20 hits contains 48.50% from the local catalogue. EDS displays the least local collections, both physical and catalogue at 2.5 percent and 3%.
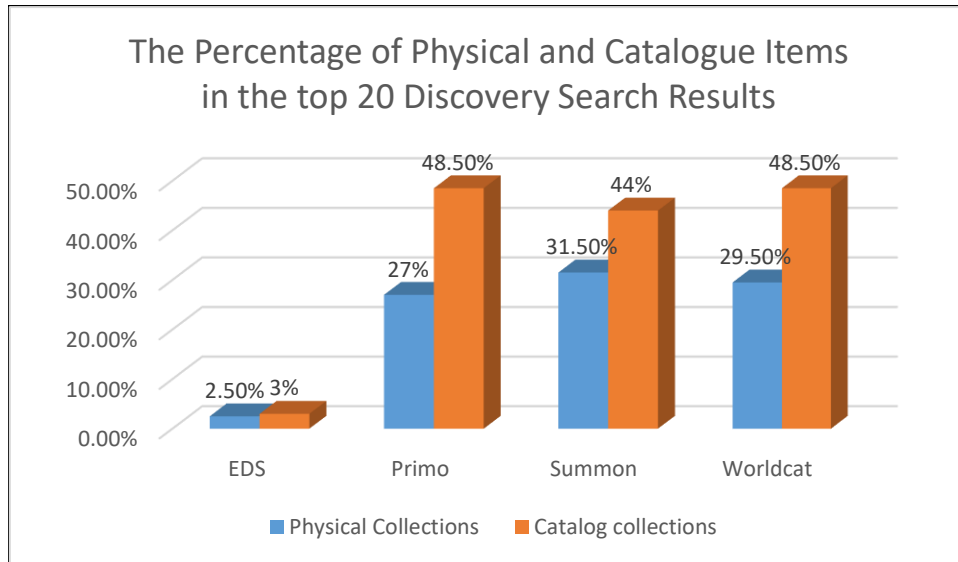

Figure 5. Percentage of physical and catalogue items in the top 20 discovery search

Figure 6 describes a comparison between what a library owes and how much each discovery layer displays in the top 20 search results from the library holdings. It shows the percentage of library physical collections that is displayed in discovery tool top 20 hits when searching "pesticides AND environment". Even though EDS libraries have the largest local physical collections, an average of 89.1 per library, it only presents 0.99% of a library's physical collections to users. In other words, if a library has 89 physical items on "pesticides AND environment", EDS may display about one percent of this number. WorldCat, Primo, and Summon display between 11 to 15 percent local physical collections respectively in their top 20 search results, with WorldCat displaying most physical local collections at 15% in our search.
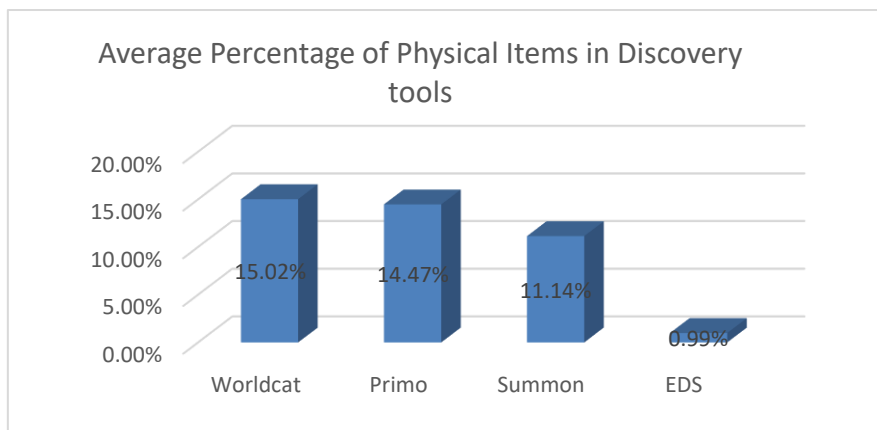

Figure 6. Average percentage of physical collections represented in Discovery tools

Likewise Figure 7 is a comparison of what a library has in its catalogue and the percentage of the catalogue items displayed in the top 20 search results in discovery tools. In this case WorldCat displays about 13% of catalogue items on "pesticides AND environment" and EDS displays 0.68%.
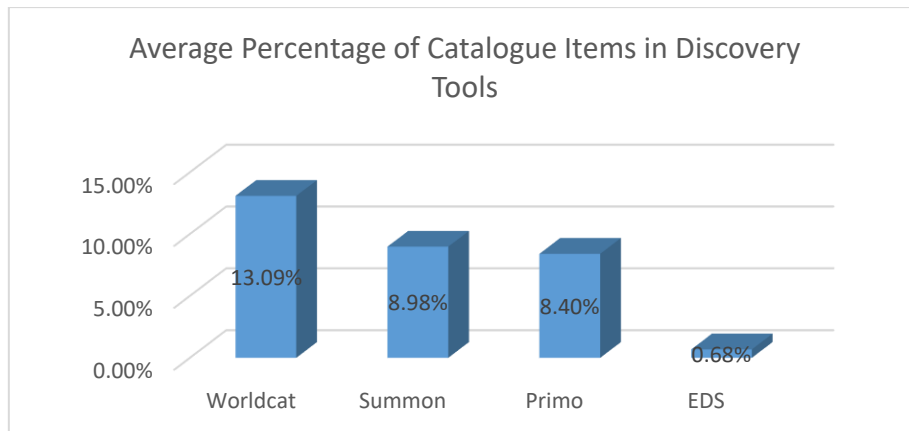

Figure 7. Average percentage of physical collections represented in Discovery tools

Figure 8 is the combined average percent of physical and catalogue displays in discovery tools. For our particular search, WorldCat displays most local collections (14.06%) among the four systems while EDS displays the least (0.84%). It remains unknown if 10.06% is the ideal amount of local collections in the top 20 hits in discovery search. So far there is no benchmark or clear standard what percentage local collections libraries would like the discovery tools to display.
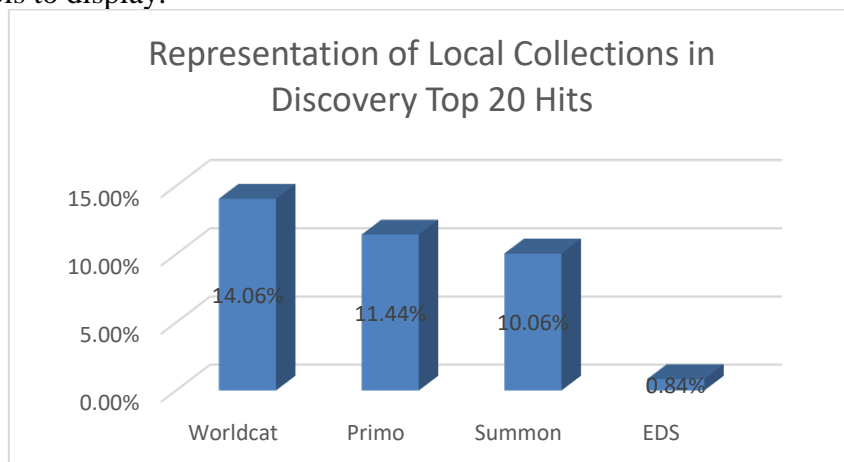

Figure 8. Average percentage of physical and catalogue collections in top 20 discovery search results

Figure 9 describes retrieval of the same physical items in discovery tool and catalogue with the same search "pesticides AND environment". "Y" means all the physical items retrieved in discovery tools are also among the top 20 search results in the catalogue search. 'Y/N" denotes a situation where items retrieved in the top 20 discovery layer are partially found in the catalogue. "N" means items retrieved in the top 20 discovery search results are not in the catalogue top 20 search results. When "N" happens, users are retrieving totally different results when search "pesticides AND environment" in discovery layers and catalogues. "N/A" are cases where the top 20 hits do not have the presence of local collections.

Search Pesticides and Environment Duplication in Discovery and Catalogue				
	EDS	Primo	Summon	Worldcat
Library 1	Y	Y	Y	Y
Library 2	N/A	Y	Y/N	Y
Library 3	N/A	Y	Y/N	Y
Library 4	N/A	Y	Y/N	Y/N
Library 5	Y/N	Y	Y/N	N
Library 6	N/A	Y/N	Y	N
Library 7	N/A	Y	Y/N	Y/N
Library 8	N/A	Y	Y/N	Y
Library 9	N/A	Y	Y/N	Y/N
Library 10	N/A	Y	Y	N

Figure 9. Duplicate items in top 20 search results in catalogue and discovery tools

We analyzed the quality of 800 items from the top 20 hits from 40 libraries based on a rubric with criteria including authority, relevancy, and timeliness. We also recorded the positions of local items on a scale of 1-20 to see where they stand in the top 20 list. However, we did not find significant differences in quality regarding local items vs. non-local items in the top 20 discovery search results.

DISCUSSIONS

Based on the above findings, it seems that ddifferences do exist among discovery services as to how they present local collections (both physical and catalog items). Some discovery tools may be more local collection friendly and show less bias than others. Visibly WorldCat, Primo, and Summon display local collections significantly more than EDS, with WorldCat representing 14.06% of local collections versus EDS's 0.84% in their prominent display. In comparison, Summon displays the most physical collection in its top 20 hits, about 31.50%. About 48.50% of Primo and WorldCat's top 20 hits comprise catalogue items. Libraries are happy to have the local holdings high in the search results. However, it is difficult to determine what those number denote or if a bias exists as there is no established benchmark or standard as to how much a discovery service should display local holdings on the top 20 hits. If 50% is a fair dividing line, then Primo and WorldCat are close. In the process of data collection, we discovered a few libraries that filled the top 10 search results with local collections followed by journal articles and other electronic collections from the 11th item down. This may be a viable option that libraries could take to configure their discovery services, but not the discovery services show such capability.

Besides a possible bias, many other factors may contribute to the composition of the top 20 search result display. The subjects may affect how much local collections get to the top of discovery tool search results. The subject of pesticides and their impact on environment may favor more current data in journal articles than print books. Meta data is another factor that may affect the discovery tool displays. For instance, journal articles have fuller metadata than MARC records providing extra fields such as author supplied keywords, abstracts, major/minor subject headings, thesaurus, full text indexing and more (see Figure 10). Naturally they are given more weight, increasing in electronic resources and decreasing in print materials and impacting the discovery tool displays. By 2020 about 80% of the library

budget will be on electronic resources (OCLC, 2013), making local print resources less relevant.

Discovery tool	Metadata	Academic Score	Local Collections	Other
EDS	Subject, title, author supplied keywords, abstract, authors, full text	Peer-reviewed, pub date, citations, material type	Option of influencing overall weighing of local collections	Adjacency bias
Primo	Author, title, subject, and other fields	Peer-reviewed, pub date, citations, material type	Personalized ranking to boost local collections	Query type
Summon	Title, subtitle, subject, author, abstract, full text	Peer-reviewed, pub date, citations, material type	Boost local collections (did not say how)	
Worldcat Discovery	Author, title, and rest of the record, term frequency and proximity	Date and number of holding libraries	Local collections come first by default	

Figure 10. Relevancy ranking by vendors

What we found shocking is the fact that a search of 'pesticides AND environment" could retrieve different set of items in the top 20 hits of discovery interface and the catalog. Even in a situation where no separate catalog is maintained and the catalog data could only be searched within a discovery surface, the retrieved items could be very different, but highly relevant at the same time. That means that users may get different results by conducting the same search. This could be confusing.

When librarians started discussion on next generation catalog, they were hoping for a single user interface that would replace the catalog and search across all library resources. In 2012 and 2013 most libraries still had a catalog and discovery tool side by side. Now things are changing. We see a trend to move to one single user interface by merging catalogue and discovery layer. Primo and WorldCat can serve as both, but Summon and EDS seem not able to fulfil this purpose yet. In our sample, most Primo and Wordcat libraries maintain one single user interface. A few libraries totally merged catalog into discovery layers. For instance, WorldCat labels all the items in the grant index as local holdings and those include journal articles in remote databases and ebook subscriptions. This raised a question about the catalogue both as a concept and physical entity and if we still need it. Is the catalogue becoming irrelevant and obsolete? What is the justification for its existence?

According to a recent study that involved a survey of over 2000 faculty members across disciplines, users may prefer journal articles and electronic resources more than print materials. Faculty often resort to journal articles first in their research (Angel & Anglada, 2016). Working at reference desk for many years, we know this may not be always true though. However, this raised the question if it is the right thing to do by promoting local holdings in discovery services, while users prefer journal articles and ebooks.

CONCLUSION

The findings from this study definitely detected a difference among four major discovery tools as to how they display local collections. Some have less bias and are more local collection friendly than others. Libraries have yet to decide for themselves if they are satisfied with how web-scale discovery tools handle local collections the way they do now. There are pros and cons associated with each discovery tool. Libraries should also consider how they want to take advantage of discovery service in promoting local holdings and what benchmark they want to establish to reach.

This study intends to serve as a starting point to spark more interest in local collections in discovery services, and additional queries in different disciplines will be explored as our project continues. So far we only get a glimpse of local collections in discovery tools. The more we understand how those services work and behave, the better libraries are to take advantage of them for good purpose. By doing so we are also helping vendors to improve products to best serve users.

References

Asher, Andrew D., Lynda M., Duke, and Suzanne Wilson (2013). "Paths of Discovery: Comparing the Search Effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and Conventional Library Resources." *College & Research Libraries* 74, no. 5: 464-488.

Calvert, Kristin (2015). "Maximizing Academic Library Collections: Measuring Changes in Use Patterns Owing to EBSCO Discovery Service." *College & Research Libraries* 76, no. 1: 81-99.

Dempsey, Megan, and Alyssa M. Valenti. (2016). "Student Use of Keywords and Limiters in Web-scale Discovery Searching." *The Journal of Academic Librarianship* 42: 200-206.

EBSCO (2017). "How Is Relevance Ranking Determined in EBSCO Discovery Service (EDS)?" EBSCO Help. Accessed June 27, 2017. https://help.ebsco.com/interfaces/EBSCO_Discovery_Service/EDS_FAQs/relevance_ranking_determined_in_EDS.

Ex Libris (2014). "Summon: Relevance Ranking." Ex Libris Knowledge Center. AccesJune 27, 2017. https://knowledge.exlibrisgroup.com/Summon/Product_Documentation/Searching_in_The_Summon_Service/Search_Results/Summon%3A_Relevance_Ranking.
Ex Libris (2017). "Relevance Ranking in Primo." Ex Libris Primo. Accessed June 27, 2017. http://www.exlibrisgroup.com/category/Relevance-Ranking.

Greiner, Tony (2011). "How Does Switching to a Discovery Tool Affect Circulation?" *Declaration of Interdependence: The Proceedings of the ACRL 2011 Conference* http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/national/2011/papers/how_does_switching.pdf.

Guan, Jingjng, and Guizhi Jia (2016). "A Comparison of Discovery Tools in Chinese Academic Libraries." *National Library Journal* 108, no. 6 : 71-79.

Hanneke, Rosie, and Kelly K. O'Brien (2016). "Comparison of three web-scale discovery services for health sciences research." *Journal of the Medical Library Association* 104, no. 2: 109-117.

Lawton, Aoife (2015). "Use of ESBCO Discovery Tool at One University Reveals Increased Use of Electronic Collections but Decreased Use in Circulation of Print Collections." *Evidence Based Library & Information Practice* 10, no. 4: 244-246.

Levine-Clark, Michael, John McDonald, and Jason Price (2014). "The Effect of Discovery Systems on Online Journal Usage: A Longitudinal Study." *Insights* 27, no. 3 : 249-56.

OCLC (2017). "How Does Relevance Ranking Work in WorldCat Local?" OCLC Support & Training. Accessed June 27, 2017. https://www.oclc.org/support/services/WorldCat-local/faq/search.en.html.

Nelson, David, and Linda Turney (2015). "What's In A Word? Rethinking Facet Headings in a Discovery Service." *Information Technology & Libraries* 34, no. 2: 76-91.

OCLC (2013). "Meeting the E-Resources Challenge", accessed July 31, 2017. http://oclc.org/content/dam/oclc/reports/pdfs/OCLC-E-Resources-Report-UK.pdf.

O'Hara, Lisa (2012). "Collection Usage Pre- and Post-Summon Implementation at the University of Manitoba." *Evidence Based Library and Information Practice,* 7,no.4: 25-34.

Parry, Marc (April 21, 2014). "As Researchers Turn to Google, Libraries Navigate the Messy World of Discovery Tools." The Chronicle of Higher Education. Accessed June 29, 2017. http://www.chronicle.com/article/As-Researchers-Turn-to-Google/146081.

Regier, Ryan (June 24, 2015). "Relevancy Ranking in Discovery Services." *A Way of Happening: Library Research Blog*. Accessed June 28, 2017. https://awayofhappening.wordpress.com/2015/06/24/relevancy-ranking-in-discovery-services/.

Sean, Fitzpatick (2010). "Summon Summons 100th Customer." *American Libraries* no. 10: 24.

Stuart, Kenyon, Ken Varnum, and Judith Ahronheim (2015). "Measuring Journal Linking Success from a Discovery Service." *Information Technology & Libraries* 34, no. 1: 52-76.