5-2016

Managing Volume in Discovery Systems

Aaron Tay
Singapore Management University, aarontay@smu.edu.sg

Chapter 4: How much is too much?

Aaron Tay

The well-established measures of recall and precision are becoming increasingly relevant in integrated discovery systems. Given the way that most people search, which is the simple keyword box that searches all text anywhere in the record, integrated discovery systems will lead to increasingly large recall as we provide access to more linked items. Using the environment of academic libraries, this chapter will explore the following questions: Do we need to be careful about the sheer volume of items to which we can provide access via integrated discovery systems? Do we want these systems to become another Google, where precision of results is not always as accurate as we would like? Are we too obsessed with the notion of providing access to everything at the expense of the quality of the results?

4.1 Introduction – The promise of the one search

The Library with a Thousand Databases was the amazingly evocative title of a talk given by Matthew Reidsma as part of NISO's Virtual Conference on Web Scale Discovery in 2013 (Reidsma, 2013). Drawing on Joseph Campbell's work on myth, *The Hero with a Thousand Faces* (Campbell, 1968), Reidsma drew a parallel between Campbell's 'hero's journey,' where the aspiring hero ventures from the normal world (use of Google) to the special realm (use of specialized databases), and the experiences first time users face as they start a piece of academic research. A similar idea can be seen in *The Adventures of Sir Learnsalot* series of videos, where academic research is depicted as a hero's journey (Tolly, 2013).

The use of a library web scale discovery service – Summon - was then proposed by Reidsma as a partial solution to keep users in the 'normal world,' at least as much as possible. This is an interesting analogy, though I would argue that the challenge a first time user faces when starting an academic research project goes beyond just using databases, but there is no doubt that web scale discovery does make things easier from the point of view of searching.

Web scale discovery services like Summon (www.proquest.com/products-services/The-Summon-Service.html) and its competitors are generally designed to be easy to use and try to mimic web search engines in terms of usability. This was unlike OPAC systems that forced users to search in unnatural ways by inverting author names and had generally unappealing user interfaces that were designed more for expert users. While 'next generation catalogues' that preceded them, improved greatly on usability by include features like a one search box with relevancy ranking, facets and *did you mean* features (Wilson, 2007), they still did not resonate with users because they only searched for books and dvds from the local catalogue and lacked the ability to search for articles in subscribed databases in one search. Arguably the strongest feature that web scale discovery brings beyond enhanced usability is the promise to search through multiple silos of content in databases in one uniform search interface.

Unlike older generation of federated search systems, which had problems with speed and reliability (Helfer & Wakimoto, 2005), web scale discovery was able to side step this issue. By acquiring the content from content providers in advance, harmonizing the data, and putting them into one central unified single index for searching, web scale discovery accomplishes the task of a 'single search' at the cost of a slight time lag in the currency of results. Unlike federated search, web scale discovery search results are not retrieved in real-time from the databases and this can lead to results that can be several days if not weeks behind. For example, newspaper results retrieved from Web scale discovery services could be a couple of days old compared to a federated search but many would consider this a fair trade-off in return for a search that is more stable and has a faster response time.

While there was initial debate over whether centralized index searching alone was sufficient, or whether one should create a hybrid approach of centralized index and federated searching (Katzman, 2009), today most academic libraries have shifted towards using one of the four major discovery services – Summon, Primo (with Primo Central Index, www.exlibrisgroup.com/category/PrimoOverview), EBSCO Discovery Service (EDS, www.ebscohost.com/discovery), and WorldCat Discovery (www.oclc.org/worldcat-discovery.en.html), which are all based primarily, if not solely, on central index (Breeding, 2015).

Of course, for the user who is accustomed to Google, the idea that you need to painstakingly rerun the same search in different databases to search for what you want is absurd; how web scale discovery works is the normal expected state of affairs. Librarians acknowledge this, and in a 2012 survey of Association of Research Libraries (ARL) institutions that used Summon many chose to brand their search as 'OneSearch,' or variants such as 'searchall,' (Tay, 2012). The idea of a one search is very seductive and obvious to users, and the idea that one should stuff as much as possible into one search seems to be uncontroversial and yet, as we shall see, trying to put in as much content as possible in a web scale discovery search index has led to many problems, and most libraries choose instead to curate the content that is searchable in their web scale discovery service.

4.2 The rush to "Everything"

In the early days of web scale discovery systems, one of the selling points for these services was the amount of content included in the index. One of the most important considerations for

librarians when selecting a web scale discovery system was whether it included databases to which they subscribed (Hoeppner, 2012). In their early days, web scale discovery services owned by Proquest, EBSCO, and others, were announcing content agreements with content providers almost monthly, and as time went by, content providers such as Elsevier, Sage, and Taylor & Francis started to provide their content to these systems, to avoid risking reduced visibility, as results showed increased usage of content included in discovery services indexes (Way, 2010).

While it made sense for full text publishers to contribute content to discovery services, the value proposition for abstracting and indexing databases to contribute their metadata was not as clear. As noted in the NISO ODI survey report on libraries and content providers conducted in 2012, 'The majority of these respondents reported perceived risk to the value-added data available in their abstracting / indexing (A&I) databases and the need for identification of supplied content in the databases. Some content provider comments touched on tensions where the evolution of discovery services is seen by some as an alternative to traditional A&I services' (NISO ODI Working Group, 2013, p. 10). It is difficult, if not impossible, to quantify the value of contributing metadata from abstracting and indexing databases to discovery services. Web scale discovery services that merged metadata from various contributing sources, including full text publishers, aggregators and A&I databases, would often increase the retrieval of full text resources, but at the expense of hiding the contribution made by the value-added metadata contributed by the A&I source. For example, an item could be found solely due to a match in the value added metadata contributed by the A&I source to the discovery service, however upon clicking on the result, the user would be directed immediately to a full text source on another platform. This would be recorded as a full text download on the other platform, but usage of A&I source might actually plunge as the user never actually visits the A&I source. That said, as of today, two of the largest generalist abstracting and indexing databases, Web of Science and Scopus, are generally available in discovery services.

One of the greatest contentions was the role played by two web scale discovery services, Summon, and EBSCO Discovery Service. Unlike the other two discovery services WorldCat Discovery Service and Primo, they were both owned by aggregators of content (Proquest and EBSCO respectively), and played a role as a discovery service provider. The concerns were two fold. First was the fear that these discovery services would privilege their own content in the relevancy ranking, as compared to more 'content neutral' services. Second is the fact that unlike publishers such as Sage or Taylor & Francis, which were generally happy to provide their content (both full text and metadata) to all four major discovery services, there was concern that Proquest and EBSO would not provide their own content to other discovery services. A well-known dispute occurred among the Orbis Casade Alliance, Exlibris and EBSCO about releasing metadata from some EBSCO databases subscribed to by the Alliance, to be displayed in Exlibris's Primo (See series of letters in Orbis Cascade Alliance, 2013). This led librarians like Carl Grant to warn about libraries 'being locked into a content silo,' depending on the type of discovery service they selected (Grant, 2013). In recent years, this issue has eased to some extent with Proquest signing deals with Exlibris (Proquest, 2014) and EBSCO releasing a policy of sharing metadata (Quint, 2014), though problems still remain. Today, a typical academic library can provide access to a substantial majority of its subscribed content in a typical discovery service, with the actual amount depending on the fit between the subscribed content and the discovery service chosen.

The rush to content has created a mindset where some librarians would add and squeeze everything possible into the index for searching and let the relevancy ranking sort it out. This 'everything' includes

- Subscribed content from publishers and aggregators
- Subscribed content from A&I
- Local content from the library including, catalogues, and institutional and digital repositories, LibGuides, etc.
- Free and open access content

Putting as much of their content as possible into one discovery index achieves what Lorcan Dempsely calls 'full collection discovery,' where one can search for all or most of the library's collection in one search box (Dempsey, 2012). Other libraries have gone even further to achieve what is known as 'full library discovery,' by including content such as library webpages, FAQs, and even highlighting librarians and expertise in search results (Dempsey, 2012)

4.3 Google Scholar, the largest academic material index?

How large should our library discovery index aspire to be? A natural idea would be to try to benchmark against Google Scholar. Google Scholar is a major commercial competitor to our library discovery services, and various reports suggest that many of our users, including both faculty and post graduates, are increasingly turning towards Google Scholar for discovery purposes (Bosman & Kramer, 2015; Joint Information Systems Committee, 2012). As such, it might be instructive to consider the size of the Google Scholar index as a standard to compare against.

Although Google Scholar is often seen as the largest index of academic material, its exact size has not been disclosed by Google, which has not stopped many from trying to quantify its size. Using the catch-recapture method, the size of Google Scholar was estimated to be 100 million records, as of May 2014 (Khabsa & Giles, 2014). Using a range of methods, Orduña-Malea, Ayllón, Martín-Martín, and López-Cózar (2014) estimated the size of Google Scholar to be between 126.3 million to 176.8 million records.

How large are web scale discovery indexes in comparison to Google Scholar? The only data I can find is for Summon, though the other three discovery services should be in the range. In 2014, Summon contained 2.1 billion deduped records but, more importantly, approximately 142.8 million items in the index are 'from commercial and open access resources that are available for clients to access based on their individual subscriptions' (Proquest, 2015). This 142.8 million total is roughly comparable to the size of the estimated Google Scholar index found by the stated studies. For someone new to web scale discovery, this seems to be good news, as our web scale discovery indexes seem to be the same size as our commercial rival, and to be competitive, we should try to make searchable everything possible in the index,including

1. Items in the library local catalogue/ institutional repository,
2. Articles in the index to which the library subscribes,
3. Open access articles in the index, and
4. Items to which the library does not subscribe.

A naive approach would adopt a 'put them all in the index and let the relevancy ranking sort it out' approach, and make it all searchable. Yet in practice, you will find few libraries making

items in the last two categories searchable, and some libraries even refuse to do a blended search of catalogue items and subscribed articles, and separate the search results from the first two in different search boxes, because a single blended list can bury relevant results.

4.4 The collection spectrum in discovery

Libraries today have more fluid boundaries on what counts as their collection; or example the library collection can include

- Purchased and physically stored collection
- Licensed online collection
- Demand driven acquisition collections (print or online)
- Shared print collections (available via consortium agreements)
- Open access and free online collections (Dempsey, Malpas, & Lavoie, 2014)

Which of the above sets of information should be shown when searching? While some libraries do include content to which they have no access (MIT Libraries for instance), most do not. Part of the reason is because a large target of such services is undergraduates, who expect their library discovery service to show only items to which they have immediate access. Faculty members that have outsourced their discovery needs elsewhere such as Google, Google Scholar and Google books and use discovery services only to check if an item is available, also expect discovery services to list only what is immediately available, so most libraries choose not to show items that are not available.

What about open access and free items? While making all available open access resources searchable seems to be a good idea, we find that in practice, many libraries shun away from adding available open access collections. There are many reasons for this, such as the poor quality of some of these collections, many broken links, and links to non-free full text resources (Renaville, 2015). Finally, many academic libraries are beginning to argue against the obvious 'blended' approach of including items normally found in a catalogue together with articles from bibliographic database in one result list. Some argue that a blended model that mixes different classes of content in one list confuses less experienced users, who are unaware of the differences between book and article results (Rochkind, 2012; Tay & Yikang, 2015). Notably, even Google Scholar generally shows only articles, and rarely shows books. This has led to the rise of the idea that a 'bento style' approach might be better, separating the display of results from different silos in different containers on the same page, rather than showing only one blended list of results. Libraries such as North Carolina State University, , Duke, and Stanford, now employ this type of search (Lown, Sierra, & Boyer, 2013).

4.5 Why relevancy ranking with everything is hard

Libraries have found that an approach that throws **everything** into the index and let the relevancy ranking sort it out does not work. The weakness in relevancy ranking has manifested in two ways:

1. Known item search, where the wanted item does not appear on the first page, and
2. Topic search, where the precision of the results is often very poor.

As noted by various experts like Marshall Breeding and Roger Schonfeld, one of the major problems with web scale discovery is known item search (Breeding, 2015; Schonfeld, 2014). In addition to anecdotal stories from librarians who have implemented web scale discovery, some

studies have found serious problems with known item searching in web scale discovery systems. In a random sample test of known items queries using Summon, Namei and Young (2015) found that only 70% of queries succeeded in ranking relevant items in the top 10 results. As known item searches make up an estimated 44% to 50% of search queries in a typical library discovery system, (Chapman et al., 2013; Schlembach, Mischo, & Bishoff, 2013), a 30% failure rate in what is considered a simple basic task, is extremely serious. Similarly, Singley (2014) found that testing with single word titles, titles with stop words, title/author keyword, and citations, led to failures for many library discovery services, ranging from only 20% success rates to slightly over 90%. While some of the problems, such as searching with complete citations have since being remedied, not all problems have been solved.

Why have library systems regressed to the point where simple known item searching has become difficult? While it is true such systems do not browse titles, a keyword search is generally sufficient for known item searches in next generation catalogues that include only local items. Obviously, it is the greater mass of possible matches due to the inclusion of millions of articles in the unified index that makes matching not so clear cut, particularly for local material from the catalogue, such as hard copy books, microforms, and so forth. This is also one reason why the bento search approaches mentioned above have become popular. Always showing the first 5 or 10 results of every category of items sidesteps this problem, although other approaches, such as using search assistance to detect known item searches and recommend links might help. Such an approach is employed by the University of Illinois Urbana-Champaign Library (Mischo, Schlembach, & Norman, 2013)

4.6 What impacts discoverability?

It is not just known item searching that suffers in web scale discovery systems, but also subject searching. Web scale discovery systems struggle with relevancy ranking because of the diversity of content types (in terms of length and type of metadata) they needs to rank. Unlike Google Scholar, or almost any type of library database, that ranks the relevancy of only one type of material, web scale discovery systems attempt to rank the relevancy of a variety items, including books, book chapters, articles, DVDs, online videos, microforms, music CDs, and more. This is very problematic, because each type of material differs greatly in terms of whether full text is available for indexing, as well as the amount of metadata available.

As web scale discovery systems can only work with the metadata and or full text they are provided by content providers, they get content that ranges from the most minimal metadata, for example, e-book records with only title, author and a broad subject, to fully described 'thick metadata,' including fully described controlled subject headings, tables of contents, and reviews found in abstract and indexing databases. Full text may or may not be available for indexing either, depending on the content provider, or the type of content indexed; for example, discovery services may include many non-text material, such as streaming videos, and print books with no full text available from electronic copies). Clearly, relevancy ranking becomes extremely difficult due to the diversity and range of full text and metadata available. Figure 1 below shows one possible way to visualize the diversity of content that the relevancy ranking must work with.

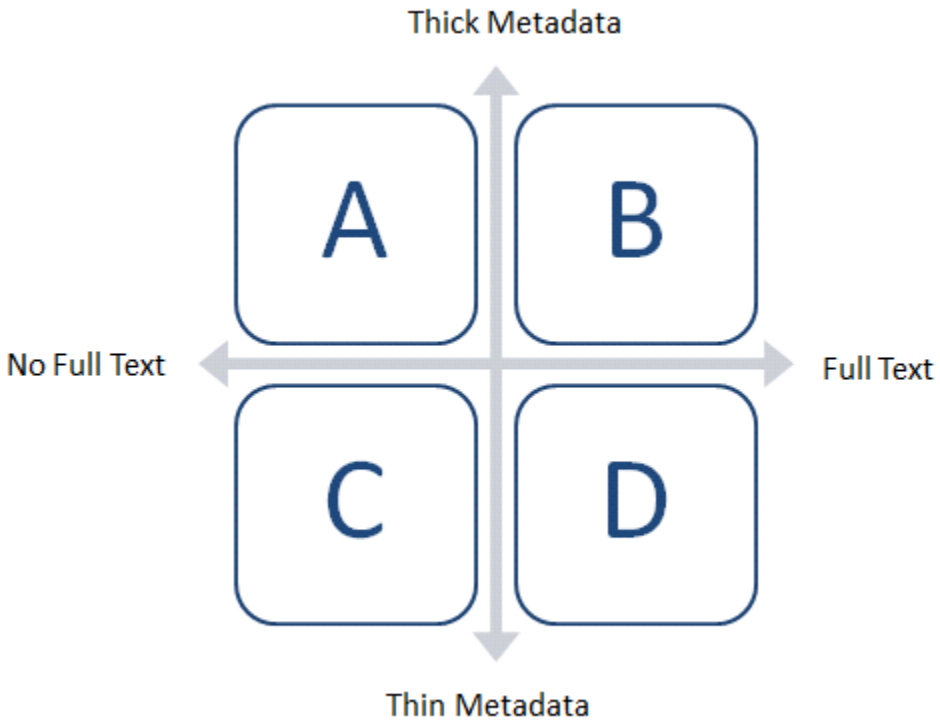

Figure 1. Two dimensions of content discovery

On the horizontal axis, we have content that ranges from having full text to those without full text. On the vertical axis, we have content that ranges from having full metadata to those with thin metadata. Some examples of the type of content in the 4 quadrants:

A. Thick Metadata, No Full text - e.g., abstracting & indexing (A&I) databases, such as Scopus, Web of Science, APA Psycinfo

B. Thick Metadata, Full text - e.g., EBSCO databases in the EBSCO Discovery Service, combined super-records in Summon that include metadata from A&I databases such as Scopus, and full text from publishers

C. Thin metadata, No Full text - e.g., publisher provided metadata with no full text, online video collections, and institutional repository records.

D. Thin metadata, Full text - e.g., many publishers provide content to Summon/Primo, and so forth.

Obviously items in quadrant C with minimal metadata and no full text are often difficult to discover if the web scale discovery service matches on both metadata and full text.

4.7 The problems with low precision and high recall

We can now understand why it can be difficult for known item searches for certain items (e.g catalogue items with short generic titles), to produce relevant results. A web scale discovery service that matches on both full text and metadata will have extremely high recall like Google or Google Scholar, but at the expense of precision, which can often be fatal for known item searches, which have an expectation of high precision. In such searches, users expect their results to appear near the top of the page or, at most, on the first page (typically 10 results per page). In such a situation, overall recall is almost irrelevant, as the expectation of the searcher is that recall for a known item will be 100%, assuming the item exists. A result that ranks the desired item in the 100th position while still achieving 100% recall, for example, would be useless. Unless carefully managed, the results of known item searches of local catalogue items with minimal metadata will often be buried, particularly if the title is short and generic. Non-text items such as music CDs, microforms or even books with no full text indexed suffer particularly if they have short common titles. Known item searches are not the only type of search that suffers from low precision: subject searches can easily be affected for the same reason.

As the most relevant items might not have the most complete metadata or even full-text, it is not necessarily the case that they will be ranked highly. While the relevancy ranking can be fine-tuned to a point, in general, the larger the index searched, the greater the recall but at the price of precision; beyond a certain point, it can get problematic. Once again, it is perhaps instructive to review Google Scholar and the experience medical librarians had with using Google Scholar for systematic reviews. In a study of the use of Google Scholar for systematic review, Gehanno, Rolling, and Darmoni (2013) concluded that 100% of the 738 studies retrieved for systematic reviews could be found in Google Scholar, leading to hopes that one could just use Google Scholar a to find such studies and forgo the tedious need to set up search strategies for multiple databases. Unfortunately, hopes were dashed when it was found that while Google Scholar theoretically had extremely high recalls, the precision of searches was so low it was not practical to use Google Scholar for systematic reviews (Boeker, Vach, & Motschall, 2013). The study estimated that due to the much lower precision of the results, users had to 'check about 20 times more references on relevance compared to the standard approach using multiple searches in traditional literature databases.' Boeker, Vach, and Motschall (2013) suggested that Google scholar queries inherently had low precision compared to databases, due to the lack of advanced search features. For instance, Google Scholar lacks deep nesting and truncation operations, and has a 256 character search limit, which limited the ability to create complicated precise searches. This, together with the auto stemming capabilities of Google Scholar, results in searches with extremely low precision. While most library web scale discovery services such as Summon, Primo or EBSCO Discovery Service support some of the missing features in Google Scholar, such as truncation, they are also generally not used for systematic reviews alone because of low precision caused by the large size of the index and matching of full text.

4.8 Practical advice for handling content in web scale discovery services

Think carefully about the type of users you have, and the role your web scale discovery service should play in the repertoire of search tools available to your users. Should the discovery service strive to be the broadest possible search tool when tools like Google Scholar already play this role? Or should it aim for the sweet spot of being broader than individual library databases, yet better curated and more targeted than Google Scholar?

What type of searchable content should you add to the index? If your discovery service allows you to display items not owned or licensed by your institution, should you set that to be default? The answer to that would depend on the sophistication of your users. If, like most academic libraries, you decide on a more restrained approach of only including items owned or licensed by your institution, should you add every available free or open access collection possible into the index for searching? Adding content to be searchable in your discovery service just because you can, and 'just in case' is often a bad idea, particularly in areas or subjects you know your users are unlikely to want to use, as precision is likely to suffer. Adding the Pubmed Central collection, or free foreign language collections when your users do not generally look for such material might be a bad idea, because while these additions might improve recall slightly, the precision of the results will be more adversely affected. Even subscribed content which presumably matches the type of content that users might want to see might not be an automatic inclusion. For example, you may consider not adding newspaper collections from providers like Factiva, even if the items are discoverable in the index. This is because newspaper collections tend to have faulty links, in addition to the sheer mass of newspaper articles that can often bury other results.

Be familiar with the options your discovery service offers. If your discovery service provides the option of metadata only search while excluding full text as a default, you may want to test with a series of actual known item searches as well as typical topical search to see if the relevancy of results improves using metadata only but no full text search, or whether you can get away with including full text search. Also, check if your discovery service provides the option of adjusting the relevancy ranking: some systems might allow you to broadly increase the weight of local holdings, which may help with known item search of such items, and help prevent them from being buried by the articles in the index, while others may allow more specific changing of field weights. Above all, test with actual popular queries made by users, with an extra focus on items that have no full-text and poor metadata. Depending on your source of data, e-books or other material might have very poor metadata as well.

Investigate if your discovery system provides or allows the possibility of a recommender system for search assistance that can be customized, this can help with searches that have poor relevancy.

4.9 Conclusion and future of web scale discovery

While implementing and maintaining web scale discovery has become an expected part of academic librarianship, there is a brewing debate over whether libraries should continue to fight for a central role in discovery or to cede to commercial and other services that operate at the network layer such as Google, Google Scholar, Mendeley. Utrecht University (Kortekaas, 2012) for example, is well known for announcing they have decided to focus on delivery and surrendered discovery (Kortekaas, 2012). This university has reason to believe that discovery happens elsewhere, as evidence suggests that the library website is not the starting place for discovery, for not only undergraduates (Perruso, 2015). but also researchers (Housewright, Schonfeld & Wulfson , 2013). As the popularity of tools like Google Scholar as a general search system continues to skyrocket, is the best move to bow out of the discovery business?

If that is the case, Utrecht University asks, why fight the change? Instead one should focus on supporting delivery wherever the user is, whether it is via the Google Scholar library links program or via OCLC or other service providers. More recently, Utrecht University designed a Chrome extension that would alert off-campus users whether any domains they are using can be proxied for nearly seamless access. The ultimate aim would be to focus on supporting researchers to use Google Scholar and library subject databases, and do away with discovery services or even catalogues (Utrecht University 2015). A similar idea would be to focus on supporting mainly known item search only in library systems and leave discovery to other systems.

Not everyone agrees with such a view, of course. Ken Varnum believes that individual institutions can have an edge over global systems like Google Scholar by creating specially tuned discovery services that include features, in particular scopes, which appeal specifically to their unique communities that cannot be easily duplicated by Google Scholar (Varnum, 2014).

In the example he gave the University of Michigan library built an experimental tool that created collections of journals based on subject librarian's categorization of journal titles by subject and by one of three categories - novice, expert, or both. Novice journal titles had articles that were more accessible and popular, while expert journal titles had articles that was more narrowly focused and specialized.

A even more granular approach would be to build collections would be for librarians to scope collections around various subject terms (both controlled and uncontrolled).

When the user logged into the tool the system would be able to identify the course enrolled by the user and offer an option to only search within resources identified by the librarian to be suitable for him. So for example when someone was enrolled in a basic Economics 101 course and searched Depression, he could get results from journals identified by the Economics subject librarian suitable for that level of study.

Such an approach would require even more careful curation of content and metadata of course.

In conclusion, faced with competition from Google, Google Scholar , Mendeley and other non-library discovery services, academic libraries are at the cross roads.

In one future path, they could decide that it was inevitable that academic libraries will lose their central position as a discovery source and maintain a minimum level of investments on library discovery services in preparation for obsolescence.

Such an approach is not without risks, as the long term sustainability of commercial tools like Google Scholar cannot be taken for granted. Also as already mentioned above, as good as tools like Google Scholar are for exploratory searches they still lack the precision we get from focused subject databases.

In another future path, they could decide to focus on differentiating from such tools by creating custom features in discovery services that are the best fit for their individual communities as

suggested by Varnum. Perhaps by combining librarian expertise (both of the subject and with their community) with algorithms, a more precise yet broad tool can be created.

References

Boeker, M., Vach, W. and Motschall, E. (2013) Google Scholar as Replacement for Systematic Literature Searches: good relative recall and precision are not enough, *BMC medical research methodology*, **13**(1), 131.

Bosman, J. and Kramer, B. (2015, 23 June) *First 1000 Responses – most popular tools per research activity*, (23 June), https://101innovations.wordpress.com/2015/06/23/first-1000-responses-most-popular-tools-per-research-activity

Breeding, M. (2015) *The Future of Library Resource Discovery: white paper*, National Information Standards Organization.

Campbell, J. (1968) *The Hero with a Thousand Faces*, 2nd edn, Princeton University Press.

Chapman, S., Desai, S., Hagedorn, K., Varnum, K., Mishra, S. and Piacentine, J. (2013) Manually Classifying User Search Queries on an Academic Library Web Site, *Journal of Web Librarianship*, **7**(4), 401-421.

Dempsey, L. (2012) Thirteen Ways of Looking at Libraries, Discovery, and the Catalog: scale, workflow, attention, *Educause review online*, **88**, http://er.educause.edu/articles/2012/12/thirteen-ways-of-looking-at-libraries-discovery-and-the-catalog-scale-workflow-attention

Dempsey, L., Malpas, C.and Lavoie, B. (2014) Collection Directions: the evolution of library collections and collecting, *Libraries and the Academy*, **14**(3), 392-423.

Gehanno, J.-F., Rollin, L. and Darmoni, S. (2013) Is the Coverage of Google Scholar Enough to be Used Alone for Systematic Reviews? *BMC Medical Informatics and Decision Making*, **13**(1), 7.

Grant, C. (2013) *Do-they-or-don't-they? Ex Libris & EBSCO Information Services: content-neutrality & content silos. A living example of why librarians should pay attention when warned of the dangers of content silos*, (13 June), http://thoughts.care-affiliates.com/2013/06/do-they-or-dont-they-ex-libris-ebsco_13.html

Helfer, D., and Wakimoto, J. C. (2005) Metasearching: the good, the bad, and the ugly of making it work in your library, *Searcher*, **13**(2), 40

Hoeppner, A. (2012) The Ins and Outs of Evaluating Web-Scale Discovery Services, *Computers in libraries*, **32**(3), 6.

Housewright, Ross, Roger C. Schonfeld, and Kate Wulfson. "Ithaka S+ R US faculty survey 2012." New York: Ithaka S+ R. Retrieved April 30 (2013), http://lgdata.s3-website-useast1.amazonaws.com/docs/923/721668/Ithaka_S_R_US_Faculty_Survey_2012_FIN AL.pdf

Joint Information Systems Committee. (2012) Researchers of Tomorrow: the research behaviour of Generation Y doctoral students, www.webarchive.org.uk/wayback/archive/20140614040703/http://www.jisc.ac.uk/publications/reports/2012/researchers-of-tomorrow.aspx

Katzman, D. (2009) Is Speed Worth It? Federated search vs. unified index, (30 October), http://deepwebtechblog.com/is-speed-worth-it-federated-search-vs-unified-index

Khabsa, M. and Giles, C. L. (2014) *The Number of Scholarly Documents on the Public Web*, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0093949

Kortekaas, S. (2012) *Thinking the Unthinkable: a library without a catalogue. Reconsidering the future of discovery tools for Utrecht University*, (4 September), http://libereurope.eu/blog/2012/09/04/thinking-the-unthinkable-a-library-without-a-catalogue-reconsidering-the-future-of-discovery-tools-for-utrecht-university-library/

Lown, C., Sierra, T. and Boyer, J. (2013) How Users Search the Library from a Single Search Box, *College & Research Libraries*, **74**(3), 227-241.

Mischo, W. H., Schlembach, M. C. and Norman, M. A. (2013) In *JDCL'13: proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries*, ACM.

Namei, E., & Young, C. A. (2015). Measuring Our Relevancy: Comparing Results in a Web-Scale Discovery Tool, Google & Google Scholar.

NISO ODI Working Group. (2013) *ODI Survey Report Reflections and Perspectives on Discovery Services*, www.niso.org/apps/group_public/document.php?document_id=9977

Orbis Cascade Alliance. (2013) *Ebsco and Exlibris*, www.orbiscascade.org/ebsco-ex-libris

Orduña-Malea, E., Ayllón, J. M., Martín-Martín, A. and López-Cózar, E. D. (2014) About The Size of Google Scholar: playing the numbers, *Scientometrics*, **104**(3), pp 931-949.

Perruso, C. (2015). Undergraduates' Use of Google vs. Library Resources: A Four-Year Cohort Study. College & Research Libraries, crl15-826.

Proquest. (2014) *ProQuest Collections Are Now Indexed and Discoverable through the Central Index of Ex Libris Primo*, www.proquest.com/about/news/2014/ProQuest-Collections-Are-Now-Indexed-and-Discoverable-through-the-Central-Index-of-Ex-Libris-Primo.html

Proquest. (2015) *Summon Index Growth: a review of 2014*, (2 February), www.proquest.com/blog/pqblog/2015/WFS2015-Summon-Index-Growth.html

Quint, B. (2014) *EBSCO Opens Masses of Content to Third-Party Discovery Services*, (May 6), http://newsbreaks.infotoday.com/NewsBreaks/EBSCO-Opens-Masses-of-Content-to-ThirdParty-Discovery-Services-96783.asp

Reidsma, M. (2013) *The Library with a Thousand Databases*, (26 November), http://matthew.reidsrow.com/articles/58

Renaville, F. (2015) *Open Access and Discovery Tools: how do Primo libraries manage green open access collections*? http://arxiv.org/abs/1509.04524

Rochkind, J. (2012) *Article Search Improvement Strategy*, (2 October), https://bibwild.wordpress.com/2012/10/02/article-search-improvement-strategy/

Schlembach, M. C., Mischo, W. H. and Bishoff, J. (2013) The Use of Transaction Logs to Model User Searching Behaviours, *Qualitative and Quantitative Methods in Libraries*, **4**, 365 – 369.

Schonfeld, R. C. (2014) *Does Discovery Still Happen in the Library?* (24 September), www.sr.ithaka.org/blog/does-discovery-still-happen-in-the-library-roles-and-strategies-for-a-shifting-reality/

Singley, E. (2014) *Discovery Systems – testing known item searching*, (8 March), http://emilysingley.net/discovery-systems-testing-known-item-searching

Tay, A. (2012) *Branding Library Discovery Services - what are libraries doing*? (17 June), http://musingsaboutlibrarianship.blogspot.sg/2012/06/branding-library-discovery-services.html

Tay, A., & Yikang, F. (2015) Implementing a Bento-Style Search in LibGuides v2. *Code4Lib Journal*, **29**, http://journal.code4lib.org/articles/10709

Tolly, L. (2013) The Information Quest: mapping the information adventure to 'The Hero's Journey' of Joseph Campbell's Monomyth. In Walsh, A and Coonan, E. (eds), *Only Connect... discovery pathways, library explorations, and the information adventure*, Lulu.

Utrecht University (2015) UU Easy Access - browser extension, www.uu.nl/en/university-library/about-the-library/uu-easy-access-browser-extension-beta

Varnum, K. (2014). Library Discovery: from ponds to streams. In K. J. Varnum, K. J. (ed), *The Top Technologies Every Librarian Needs to Know: a LITA guide*,ALA TechSource.

Way, D. (2010) The Impact of Web-Scale Ddiscovery on the Use of a Library Collection, *Serials Review*, **36**(4), 214-220.

Wilson, K. (2007) *OPAC 2.0: next generation online library catalogues ride the Web 2.0 wave*! *Online Currents*, **21**(10), 406.