Library Discovery Directions

Lorcan Dempsey
VP, Membership and Research, OCLC

Please direct correspondence to:

OCLC Research
oclcresearch@oclc.org

Contents

Introduction

The variety of topics and perspectives represented in this collection is clear evidence of the diversified scope of library discovery. In only a few years the emphasis has shifted from consideration of the development of a particular library service (the evolution from catalog to metasearch to discovery layer) to a broader consideration of user behaviors and service development in a complex network ecosystem of offerings, from both libraries and third-party providers. A full appreciation of discovery and discoverability in a library environment now involves thinking about much more than the discovery layer.

(In writing this short piece I was interested to trace this evolution in two earlier pieces I have written about similar themes. (Dempsey, 2006; Dempsey, 2012) Considered alongside this one, the three pieces are written at approximately six-year intervals.)

For a sense of the diversity of the current ecosystem, consider the various current roles of Wikipedia, Google Scholar, reading lists, the library discovery layer, resource guides, Scopus and Web of Science, WorldCat, GoodReads, ResearchGate and Mendeley, Arxiv and PubMed Central. These and other resources are used by library users to find specific resources of interest to them, for serendipitous discovery, for exploration. They are used alongside library resources, and sometimes in combination with them. The transition from Google Books to a library catalog via WorldCat, or the transition from Google Scholar to a licensed article via a registered library knowledge base, or the serendipitous discovery of special collections in Google or Wikipedia, are all examples of such combinations.

In this environment, three broad related trends are of interest. First, "discovery often happens elsewhere." We know that library users now search for and find materials of interest in many places. These include the network level services that are now so much a part of our network lives (Wikipedia, Google, Amazon) and as well as more specialist resources (arXiv.org, for example). The discovered resources may be books or journals, or they may be software, research data, learning materials and so on. We also know that people may "discover" materials of interest in non-library services and then turn to the library to "locate" particular instances of those resources, whether these are on a shelf, licensed by the library, or potentially requestable from elsewhere. So, a user who spends time on ResearchGate, say, may turn to the library if a copy of a discovered article is not available there.

Second, while the library collection may be a part of the library's information universe, it is not necessarily central, nor the first used. This is a natural outcome of a shift from an environment of information scarcity (where the library acquired materials for local use and was the only place to get them) to one of information abundance (where the network environment is rich in relevant resources).

And third, we are observing a reconfiguration of workflows. In a print world, or a world of information scarcity, the learner or researcher had to build their workflow around the library. In a network world, many people will expect library resources to fit into a user's network workflows which may comprise multiple resources (citation management, network discovery services, and so on).

These changes are important, because historically library discovery and library collections went hand in hand. However, both our sense of discovery and our sense of collections are changing in different ways in the current network environment.

In this context, two important trajectories seem to me to lie just under the surface of much of the discussion in the following papers, and perhaps to provide an integrating pattern.

1. The peeling away of library discovery from the local collection.

2. The peeling away of library discovery from the local audience.

Collection and audience boundaries are both blurring, placing different requirements on discovery. Of course, each of these has very much coevolved with changing research and learning behaviors in a rich networked, discovery ecosystem.

This means that at the same time we can see a variety of other service and technology directions emerge, and I make notes about some salient areas in subsequent sections.

In this brief foreword, I can only say a little about such directions, which can then be traced with more detail and nuance in the contributions assembled here.

The Peeling Away of Discovery from the Local Collection
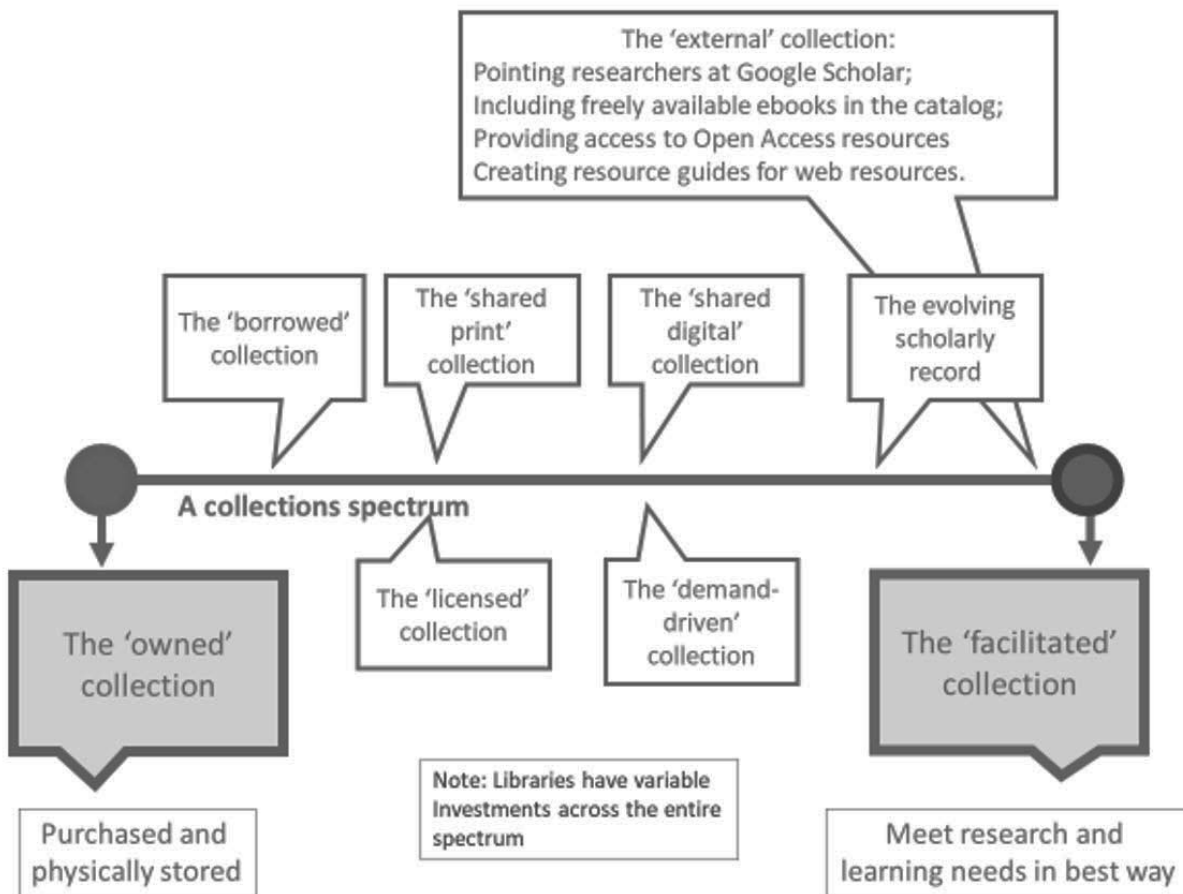
The facilitated collection and the collective collection

In a print world of scarce resources, assembling a collection close to the user made sense. And indeed, "goodness" was associated with the size of that local collection. This model extends into the electronic world, where resources were licensed for local use. Increasingly, we are seeing that the centrality of the acquired collection to the library service and identity is an artifact of a particular phase of library development, which we are now moving beyond. By "acquired" I mean the collections that the library has purchased (often in print) or has licensed (usually electronic) and makes available to a local audience.

But researchers and learners inhabit a network environment rich in information and workflow resources—for research, for communication, for archiving, for social sharing, and so on. These are general purpose (github, Wikipedia) or specialized (Amazon, arXiv, PubMed Central, Google Scholar, etc.). "Localness" is no longer a determining influence in information use. The acquired library collection is now actually only one resource among many of potential interest.

In this context, we see a progressive shift in library interests from the locally acquired just-in-case collection to the "collection as a service," facilitating access by researchers to resources of potential interest wherever they are. As indicated in this figure, a variety of services have been added successively over the years to the resources the libraries offer.

Figure 1: Evolution of library services

The 'external' collection:
Pointing researchers at Google Scholar;
Including freely available ebooks in the catalog;
Providing access to Open Access resources
Creating resource guides for web resources.

The 'borrowed' collection

The 'shared print' collection

The 'shared digital' collection

The evolving scholarly record

A collections spectrum

The 'licensed' collection

The 'demand-driven' collection

The 'owned' collection

The 'facilitated' collection

Purchased and physically stored

Note: Libraries have variable Investments across the entire spectrum

Meet research and learning needs in best way

Historically the library had a "purchased" or "owned" print collection. This was supplemented by the "borrowed" collection as libraries grouped together to share materials through inter library lending. Then the licensed collection emerged, first with abstracting and indexing services and then with the journal literature itself.

More recently we have seen the demand-driven collections emerge. Narrowly, this relates to demand-driven acquisition. More broadly, we will see more reliance on data-driven decision-making around acquisition choices. In this context, there is an interesting flip. Previously, collections drove discovery; in a more demand- or data-driven environment, discovery may drive collections.

As research and scholarly network resources have proliferated, we now see greater reliance on external, freely available resources. A library may proxy access to Google Scholar, may add metadata for free ebook collections or open access collections to the discovery layer or knowledge base, and is developing resource guides based on broad user interests rather than on what is solely available in the collection.

One aspect of this facilitation is the "collective collection" where collections are managed in some way at a shared level above the institution. This is becoming more common, as libraries group together for resource sharing, shared print initiatives, a shared library system, or some other group activity. In the US resource sharing consortia are common; in other countries there may be national or regional groups (e.g., the Library Hub services in the UK, Sudoc from Abes in France, or Libraries Australia). The level of

coordination may vary. As an example of strong coordination, ReCAP manages a physically consolidated print collection on behalf of its research library members. (This is also interesting from a discovery point of view, as members add an index of the ReCAP shared collection to their discovery layers, which has resulted in significantly increased use of the shared collection—it would be interesting to explore how scalable an approach this would be across more libraries.)

There are now also collective collections comprising shared digital collections (such as those made available by Europeana, for example), or shared scholarly materials (such as the Australian Research Data Commons, for example).

What has this meant for discovery?

Well, the library interest in discovery has broadened. However, the extent of the shift may be partly obscured by organizational issues. Discovery is in fact supported across multiple library divisions or specialties.

So, a discovery layer may be managed to provide access to the acquired collection, and this tends to be seen still as the main "discovery" support. Typically, it will provide access to the catalog (and maybe to WorldCat or another union catalog), to licensed article data (and will work with resolver/knowledge base infrastructure), and maybe to unique local digital materials (from a local repository or repositories). And as noted above, various additional freely available resources may be added. In fact, management of the discovery layer has become more complex, as access to resource sharing groups is added, and as it is supported from a variety of metadata streams.

Direct access may also be provided to Google Scholar, WorldCat, PubMed Central, Scopus or Web of Science, and core disciplinary databases (PsychInfo, for example), among others. A knowledge base may be configured in Google Scholar to ensure transition from that discovery environment to resources licensed by the library. As noted, a discovery layer may be articulated in some way with a union catalogue for access to a collective collection of which the library is a part (as is the case with our neighbours here in Ohio, for example, who have access to the shared resource of OhioLINK alongside their local collections).

A major outcome of this shift to facilitated collections is the rapid development of resource guides, which guide users to resources of potential interest to them, inside and outside the local collection. To meet the strong interest in open access, services have emerged that may be integrated with a discovery layer, a knowledge base, resource sharing, or deployed as browser plugins. And new areas of interest continue to emerge, research data management and open educational resources in recent years for example, where again, the library may wish to provide discovery options.

In this way, we can see the library is guiding users to resources in multiple ways. And, of course, as discussed below librarians may be advising their users on finding and using resources in a much broader discovery ecosystem.

Full library discovery

In current services, a library user is usually presented with a collection search, and a range of web pages about other library services or expertise. The two are separated in the website experience but may not actually be separate in the user expectation of the library. If I am interested in demographics, for example, why not find a curated resource guide or the name of a subject specialist as well as relevant articles or books when I do a search?

And, indeed, we have seen discovery environments which provide a layer over not only the collections, but also potentially over the library website, library staff and expertise, resource guides, and so on. In other words, we are seeing approaches to "full library discovery" emerge.

The steps that have been taken include a search over the website generally; over events, exhibitions or specialist services; over people and expertise (sometimes associated with resource guides); and over resource guides (as discussed above). Indeed, in the absence of dedicated services (for events, for example), resource guides are often used as a simple content management framework for various types of information about the library, and not only for lists of information resources.

Of course, collections are still key for libraries, but as library services continue to evolve beyond the collection and as there is a focus on deeper engagement with user communities it makes sense for discovery services to represent more of what the library does and can provide.

This is of interest in academic libraries and is also of special interest in public libraries where event and meeting management are an important part of outreach.

We can see some systems development alongside this. One is the use of Drupal, Blacklight and other frameworks to deliver unified results across several resources including website, and other elements of library operation. In public libraries we have seen the use of events management software, customer relationship management systems and other systems for engagement. Bibliocommons is interesting here. OCLC's Wise system puts community engagement at the center of its operations, rather than the collection.

We may not have thought of this as "discovery" in the past, but it is an important part of the "discoverability" of the library's capacity and potential value. And this is increasingly important for libraries of all types.

The Peeling Away of Discovery from the Local Audience

The inside-out collection

Much of the discussion of discovery has been about the acquired collection, the materials bought and licensed and assembled for local use. This is an "outside in" collection – the materials originate outside the institution.

However, in a network environment, institutions themselves are important producers of cultural and scholarly resources of potentially broad interest to different external audiences.

In this context, think of the scholarly workflow. Historically, the final product of research, the book or article, was published and it was the main research output. However, as workflows are digital and as they are supported by various workflow tools, there is more interest in other outputs – software, methods, preprints, research data. This is encouraged by policy mandates (at national, funder, and institutional levels), by changing norms of science and research, by an interest in reproducibility, and so on. Research outputs appear across the research life cycle.

A related scholarly activity involves the creation of expertise and research profiles. Institutions may have their own systems for this, which pull together data about researchers, including bibliographic data. One aim is to improve the discoverability of local faculty for various reasons. And of course, many researchers themselves are very interested in managing reputation and visibility through Google Scholar,

ResearchGate or other services. The library may advise here or encourage the use of ORCID iDs, which again support discoverability in the scholarly ecosystem.

Or think also of special collections and archives. These tend to be assets that are unique to the institution or are rare, which have reputational value, and where the institution may accept a preservation responsibility as part of the broader scholarly or cultural record. There is growing interest in digitizing distinctive materials and making them more broadly available. This is to integrate them more effectively in local learning and research activities. However, it is also importantly to disclose them to a wider audience outside the institution, or to allow them to be placed alongside related materials from elsewhere.

In each of these cases, there is a growing interest in "discoverability" by an audience outside the institution (as well as inside). In this sense, these are "inside out" collections (Dahl, 2018). The discovery dynamic is very different. While it makes sense to make sure they are represented in local discovery systems, it may be more important that they are effectively represented in external discovery systems used by potential audiences. So, care may be taken that they are indexed effectively by Google, or metadata for them may be marked up with additional links to improve chances of crawling, indexing or ranking. Metadata may be provided to aggregators of cultural, scholarly or open access materials. Links may be added to Wikipedia. And so on.

The library here has a role in improving discoverability for audiences outside the institution, to enhance reputation, to increase the impact of local scholarship and research, and to share awareness of distinctive scholarly or cultural materials.

Attention here is diffuse: it has no single organizational focus within the library. It is not entirely clear how much effort libraries are putting into this area, and this will certainly vary across types and scale of library. But supporting discoverability more directly in this way clearly represents a different and important orientation.

Some Technology Directions

Several technologies and techniques are important and are discussed in later contributions. Here are some interesting directions.

- **Linked data**. Important intellectual work has been done by libraries on describing people, subjects, works and other entities and an extensive apparatus of authority files exists. That work is now being mobilized in a new environment. Our bibliographic infrastructure is evolving towards a more entity-based approach, as we think about modeling and exposing data about structured entities of interest (graphs of works, authors, places, for example), rather than only shipping around bundles of data about titles (records). Work on data modeling, entity backbones, and data augmentation is being carried out by multiple agencies, including national libraries, publishers, and individual libraries, as well as OCLC. This is part of a more general trend, of which Google's knowledge graph and WikiData are important examples. As the volume and variety of links grows, and as those links resolve to metadata about "things," it becomes possible to match and merge data more easily at scale, to build greater navigation and context into interfaces, and to build relationships across the web. This work is at early stages but shows promise in improving discovery and discoverability.

- **Data science and machine learning**. Discovery is a data intensive field and will benefit from advances in data science and machine learning. For example, it is likely that discovery in the journal literature will be facilitated by programmatic analysis of bodies of literature assembled by

publishers, researchers and others. The assembly of large amounts of data in this way and the collection of transaction or "intentional" data from users mean that machine learning will increasingly be mobilized to develop scalable approaches to extract insights. There is an interest in entity recognition, topic modelling, summarization, plagiarism detection, recommendation, ranking, personalization, and in identifying patterns that may support new findings or directions. Of course, this also raises issues about privacy, algorithmic transparency, authority, and appropriate use, and libraries are stepping up to a critical role as advocates for the interests of users and researchers.

- **Presentation**. When the discovery layer was introduced into libraries, there was a lot of discussion of a "Google like experience" and the "simple search box." However, not even Google does a "Google like experience" anymore. There is a simple search box, but a huge amount of work goes into presenting the results. There is no longer a single ranked flow. There is the now-familiar knowledge card summarizing what Google knows about an entity. It may pull out images, news items, scholarly articles or other elements for highlighted presentation. Different entities will have different elements presented. A literary work will have book covers of various editions, various fulfilment options, maybe works in the same genre. For organizations, a map may be presented. In some cases, a carousel of similar or related items is presented at the top of the page. And so on. Libraries rely on a small number of suppliers for their discovery layer systems (one of which is OCLC). Some then layer different presentation environments over them (e.g. Blacklight). There has also been some experiment with the so-called bento box display, where results from different sources (e.g. resource guides, article index, catalog) are presented in different sections of the screen. As discussed in later contributions, there is some discussion how best to present results given these options. However, given the full range of the discovery experience I have described here, it seems to me that we are at an interesting point in the evolution of library discovery where richer options for coordinated presentation of discovery results will evolve. I have focused on presentation here, what is generally presented for direct interaction. Of course, we may also see a renewed interest in more push or alerting approaches, views of the facilitated collection customized to particular group or individual perspectives.

Some Service Directions

The evolving library consultative role

The shifts I have spoken about are mirrored in the consultative role of the library. It is evident in the evolution from subject to a wider liaison role, and in the progressive broadening of the "literacy" role – from bibliographic instruction, to digital literacies, to a more reflective consultation around the complexities of the emerging information environment. Barbara Fister writes nicely about the transition (Fister, 2019).

One could think about two aspects of this, echoing the discussion above. The first is an outside-in one, where consultation about the use of discovery and other information resources in a complicated environment becomes more important. This encompasses thinking critically about a complex evolving network information environment, understanding the structure of disciplines, thinking about relevant network resources, as well as advice abound surveillance, algorithmic retrieval, and #fake. Again, this parallels the move from use of a defined local collection to more nuanced facilitation of information discovery and use in a network ecosystem.

The second is an inside-out one, where it becomes increasingly important to advise researchers and others about their informational lives as creators. One important aspect of this is to understand more about the discoverability of people and their work and how to optimize this in a network world, especially where systems and services increasingly measure, rank, recommend and cluster work algorithmically.

Advice about the assignment of persistent identifiers such as the ORCID iD for researchers, or DataCite for data sets, has become interesting here, as has advising about use of consistent names for organizations and people to improve discoverability and matching. Areas of interest include researcher profiles, copyright, publishing choices and OA, data consultancy, advice about mandates and best practices, and so on.

Of course, these twin aspects are connected in a network environment as we are both creators and consumers of resources. Indeed, for some, the boundaries between workflow, content, and online identity have become blurred (think of the use of ResearchGate or Mendeley for example).

In this way, advice about discovery and discoverability includes the local collection and discovery system, but now ranges across research and learning behaviors in a rich network environment.

Discovery to fulfilment

We know that network users prize convenience and predictability. And libraries are very focused on improving the user experience so as to encourage use and satisfy demand. A major part of this is ensuring that the path from discovery to fulfilment is as efficient as possible and fragmentation or delay in the user experience is minimized. This becomes more important as more discovery and fulfilment options are stitched together. Discovery without fulfilment can be frustrating.

One outcome of this is that we will see greater convergence between discovery, resource sharing, collection development and acquisitions. We can expect to see the links between these service areas become more automated and data driven. So, in a simple example, acquisition of articles or books may be triggered at some level of demand as indicated by discovery or resource sharing patterns of use.

The links between discovery and fulfilment services are also more important. Libraries are interested, for example, in ensuring the connection between local discovery and a consortial discovery and requesting system works well. Consortial borrowing systems emphasize speed and predictability of delivery. Facilitating integrated access to open access materials improves delivery.

That said, the library discovery to fulfilment environment can be difficult or cumbersome to navigate. A major cause of this is precisely the fragmentation across multiple service and system boundaries. It is difficult to achieve the gravitational attraction of, say, Google Scholar, when library services are built on top of a patchwork of system boundaries. In some ways the discovery layer was a response to the fragmentation and inefficiency of approaches based on metasearch. For some libraries, there has also been some consolidation at the group level, with shared discovery environments. However, given the nature of the environment in which they work, where they are bringing together diverse services, it is inevitable that there are integration costs – in terms of both systems work and user experience.

The discovery layer

The discovery layer has dominated discussions about library discovery in recent years. It is an important focus for library attention, often seen as a "shop window." However, we know that much discovery happens elsewhere.

It is interesting to consider the three categories of licensed, purchased, and institutional materials in relation to the discovery layer. Article discovery is an evolving and fluid area, with the migration to open access creating new opportunity and uncertainty. We have seen the emergence of new network level discovery services alongside Scholar (e.g. Meta, Microsoft's Academic Search, Dimensions from Digital Science, and others), and while the long-term sustainability of some of these may be unclear, they provide additional options. ResearchGate, Mendeley and other research networks are important venues.

Preprint archives and other open resources continue to appear. Large publishers and others will likely have discovery offerings, but also new mining and interpretive services on top of large bodies of literature. At the same time, the range of research outputs to be discovered is growing (including research data, methods, and software).

Turning to purchased materials (books, maps, etc.), these typically enter the discovery layer through the cataloging stream. There has certainly been some move to group level for discovery of these items and this is likely to get more common as more shared approaches to managing print collections emerge.

Finally I have discussed how libraries may provide access to institutional materials (research outputs, special collections, …) locally, but that there is a strong interest in pushing these out to audiences outside the institution in various ways, so the discovery layer is only one place where this might happen.

At the same time, I have discussed how the scope of discovery itself has diversified, as the library addresses needs in a rich network environment.

Taking all of these factors together, this means that the current library discovery layer addresses a part of the discovery need and that the services it offers are situated in a fluid and evolving environment. It seems very much like a service category in transition.

However, few libraries have taken the step that Utrecht University Library has, in dispensing with the now conventional local library discovery layer altogether. In its place, they now rely on several network level resources as the primary entry point (Utrecht, n.d.).

The university library finds that researchers and students often like to start with a wide search when looking for scientific literature, for instance by visiting Google Scholar first. That is why we offer support in major, worldwide scientific search engines.

The library then offers a list of network level discovery resources, some openly available, some licensed (Scopus, Web of Science). Secondarily, a search is offered in the library's collection. This is through the Dutch union catalog (provided through WorldCat) and access to the resolver for journal title searches. Interestingly, they then provide a list of services for discovery of research data.

While this model may not be widely adopted it is an illuminating response to some of the issues discussed here. It shows a particular response to the scoping, scaling, and sourcing choices that face libraries as they design for changing needs. A scoping choice was made here that the discovery layer was not necessarily the best investment of library time and resources, given changing patterns of information use. The context established in the quote above is that searching behavior starts at webscale, with a "wide search." Discovery, they suggest, often begins at network scale, rather than at library scale, but does require links back to the library. Given that, they are sourcing their solution through a mix of network level resources and library resources. They are not sourcing exclusively through a library level discovery layer.

In the discovery environment, libraries are looking at what they should do (scoping), what the relationship between library and network level external services should be (scaling), and what combination of systems and services best needs their users' needs (sourcing).

In Conclusion

Discovery has been core to library operations, closely coupled with a local collection for most of its history. In a network world the relationship between discovery, collections and behaviors is changing in

important ways. My interest here has been to provide some background context. As the papers collected here show, there is much to explore.

Bibliography

Dahl, Mark, 2018. "Inside-out Library Services', Challenging the "Jacks of All Trades but Masters of None" Librarian Syndrome (Advances in Library Administration and Organization, Volume 39)." (2018): 15-34.

Dempsey, Lorcan, 2006. "The Library Catalogue in the New Discovery Environment: Some Thoughts." *Ariadne*, 30 July 2006, http://www.ariadne.ac.uk/issue/48/dempsey/.

Dempsey, Lorcan, 2012. "Thirteen Ways of Looking at Libraries, Discovery, and the Catalog: Scale, Workflow, Attention." *EDUCAUSE Review*, Dec. 2012, https://er.educause.edu/articles/2012/12/thirteen-ways-of-looking-at-libraries-discovery-and-the-catalog-scale-workflow-attention.

Fister, Barbara, 2019. "Information Literacy's Third Wave." *Inside Higher Ed*., 2019. https://www.insidehighered.com/blogs/library-babel-fish/information-literacy%E2%80%99s-third-wave.

Utrecht, n.d. "*Searching for literature*." Utrecht University. Retrieved 5 September 2019, from https://www.uu.nl/en/university-library/searching-for-literature#searchengines.