Old Hopes, New Possibilities: Next-Generation Catalogues and the Centralization of Access

Joshua Barton and Lucas Mak

Abstract

Next-generation catalogues can be viewed as the latest manifestation of a tendency in library catalogue history to strive for centralization of access to collections—a single portal for the discovery of library resources. Due to an increasing volume of published materials and the explosion of online information resources during the Internet age, the library does not currently provide centralized access to its various information silos, nor does it provide a user-friendly search and retrieval experience for users whose expectations are shaped by Google and other major commercial Web sites. Searching across library resources is a complicated task, bearing high-attention “transaction costs” for the user, which discourage the use of library resources. Libraries need access systems that minimize complexity, easing discovery and delivery of resources for user populations. Here, the authors review past efforts of centralization of access, consider the potential of next-generation catalogues in the context of this historical tendency toward centralization of access, and describe what goals underlie that centralization.

Introduction

In the early days of library catalogues, entire local collections could be indexed. As information formats proliferated, especially journal publishing, libraries were unable to meet cataloguing demand. This work was taken on by abstracting and indexing (A&I) services, whose products supplemented the local catalogue. Library literature of the 1980s and 1990s analyzed the advent of the online public access catalogue (OPAC). There were expectations that comprehensive catalogues with centralized access to an entire library collection could be achieved, as it had been in some earlier catalogues.
The explosion of information availability with the Internet and the flourishing of silos of digital library collections through the 1990s and 2000s have challenged the early expectations of OPACs. Users’ expectations of OPACs have changed as well, owing to the wide use of online search engines like Google, through which “people expect to enter one or two words (or maybe a phrase) into a search box, click a button, get a list of relevancy ranked results returned, select an item from the results, and view/download the information” (Morgan, 2007, “Introduction,” §3). This is not an online searching experience that libraries have been able to provide yet. Searching for library resources is a complicated task, bearing high-attention “transaction costs” for the user, such as navigating across decentralized information silos. This complexity inhibits the use of library resources. Libraries need access systems that minimize complexity. They need true centralization of access that makes discovery and delivery of resources easier for a wide user population.

Next-generation catalogues and Web-scale discovery services (which will be referred to jointly throughout simply as next-generation catalogues) raise the prospect of centralizing access to today’s library collections through the provision of a single portal to the library’s resources, minimizing the need to search across decentralized silos. The features of next-generation catalogues, including faceted navigation and relevancy-ranked search results, are designed to make the search process more user friendly and more in line with user expectations honed by the World Wide Web, in addition to making the larger search scope of centralized access more navigable.

Next-generation catalogues can be viewed as the latest manifestation of a tendency in library catalogue history to strive for centralization of access to collections. The authors aim to consider the potential of next-generation catalogues in the context of that historical tendency toward the centralization of access, describe what goals underlie that centralization, and consider the capacity of next-generation catalogues to achieve those goals.

Centralizing Access—What Is It?
The centralization of access is the effective provision of a single portal for discovery of library resources, a single search environment that a user can usefully engage to search across all the information available to a library. Centralization, or something very close to it, has been achieved before but is currently disrupted by the fragmentation of library resources into separate information silos: the local catalogue, databases of aggregated journal articles, an institutional repository—each having to be searched separately, as will be discussed.

Centralization of access does not necessarily require that the underlying architecture of the library’s information resources be centralized. For example, libraries should not be expected to create or load MARC records
for every available item into their Integrated Library System (ILS)–based catalogue module. This is not a scalable approach, nor is it one for which existing OPACs are well-suited. The search should be experienced as centralized by the user, with materials being discoverable within a single, virtual environment.

**WHY CENTRALIZING ACCESS IS IMPORTANT**

Acknowledging that different libraries will have different needs in how they provide access to their resources, the authors believe that the prevailing expectations that users have of their search experiences make centralized access a general need. Library users have developed certain expectations of online searching activity from using sites like Google and Amazon. These and other Web-based commercial entities have a vested interest in making their sites intuitive and easy to use so users can be exposed to revenue-generating advertising or can be successfully guided toward a purchase opportunity. To meet these goals, commercial entities have developed incredibly sophisticated search tools for their large and diverse collections of information or products. Millions of products or billions of Web sites can be searched within a single interface with relatively dependable results of at least a few relevant hits for even very basic queries. A user study by Griffiths and Brophy (2005) suggests that these commercial search tools, especially Google, define the contemporary paradigm of the online search experience, at least for university students. Users come to the library’s Web presence expecting to be able to do the same kind of searching they can do on Google (Augustine & Greene, 2002).

Familiarity with the use of commercial search tools is so widespread that the library search tools employing different or comparatively outdated conventions are at risk of obsolescence. Users simply will not bother with them. To describe this, Lorcan Dempsey (Dempsey, 2006a, 2006b) invokes the language of Chris Anderson’s “Long Tail” argument, which purports that commercial success on the Web lies in providing supply to myriad online niche markets, or the “Long Tail” (referring to the long, descending tail of a probability curve) (Anderson, 2004). The Long Tail argument uses economic terms; thus, Dempsey describes libraries’ contemporary predicament as one of competition in the networked environment, where scarcity is no longer in physical space but in user attention. High transaction costs in library access systems—those interactions with the system that demand a lot of scarce user attention—inhibit use, and one can think of this as an economic articulation of Ranganathan’s (1964) “save the time of the user.” Within the networked environment, libraries are vulnerable because their catalogues have lagged behind in development compared to other online search tools and have come to provide access to only a portion of the library’s collections, creating a complicated, or costly, search process (Davidson, 1999; Morgan, 2007; Shadle, 2008).
Anderson’s Long Tail shows how the networked environment causes the “tyranny of physical space” to break down (Anderson, 2004). It allows for massive consolidations of information and products, like Google, Amazon, Netflix, and iTunes. The vast “inventories” of these entities were not possible in the physical environment—video rental shops cannot afford to stock the same volume of movies as Netflix. The Internet is a place where the Long Tail of low-use items, taken in aggregate, can see usage on par with the smaller number of high-use items. According to Dempsey (2006a), the relevance of the Long Tail for libraries is that it is “about how well supply and demand are matched in a network environment.” Success is found in how well one aggregates supply and aggregates demand, striving for low transaction costs.

At first glance, one might think that supply here is equivalent to library resources and demand is equivalent to users, or use. After all, libraries have been stockpiling collections of resources for users for a long, long time. However, there is more to it than that. Aggregating supply, for Dempsey (2006a, §3), is “making it easier to find and obtain materials of interest wherever they are.” Aggregating demand is “extending the population to which a resource is accessible . . . [so that] resources have a better chance of finding interested users” (Dempsey, 2006a, “The Long Tail,” §13). One can see how these are closely related tasks, and one can imagine how they relate to so much of what libraries can and ought to be doing on the Web today, especially in our access systems, where we seek to capture the scarce resource of user attention.

The library’s mission to provide for its constituents’ information needs thrusts it into competition with commercial entities on the Web. In contrast with the streamlined and consolidated portals of Google and Amazon, libraries provide different portals to different segments of their collections that do not adequately aggregate the supply of information resources. In other words, library resources are difficult to find and obtain because they reside in disparate and fragmented silos: the OPAC, repositories of locally created digital content, specialized databases for subscribed electronic resources, etc. There is no one place for a user to begin his or her search, and there is a lack of integration between these various silos, stymying users’ ability to search comprehensively. When users reach dead ends, they can seek out help from librarians virtually or face to face, but not all users will. Thus, compared to other Web-based search tools, information transactions in the library’s fragmented discovery environment are attention-expensive. It is complex (i.e., high transaction cost) work to have to search one discovery interface after another. Because these high transaction costs inhibit use, they put libraries at risk of losing users to competing services, even users from libraries’ own institutions.

Nor do libraries adequately aggregate demand for their information resources. Apart from the argument that library resources are better-vet-
Centralization of Access

Ted and more reliable sources of high quality information, libraries do not have an effective means of drawing user attention away from the much more highly used Web-based search engines. Karen Coyle (2007a) points out that the resources of the open Web, those discoverable in Google and elsewhere, are not necessarily rivals to library resources. Open Web and library-vetted resources can be complementary. However, the competition to attract users to make use of these respective resources is very real. Coyle (2007a, 2007b) and others have suggested a reorientation to this problem that looks for ways to go where demand is already aggregated—to integrate library resources into the much more heavily used discovery environment of Google and the like, which likely entails a new conception of the role and purpose of the library catalogue within a very broad discovery framework of the future. For now, however, within the context of a single institution or constituency, libraries should still be mindful of their visibility and capacity to attract users to their existing access systems, that is, their capacity to aggregate demand among their own constituents.

Two Problems

To summarize the current need, libraries require catalogues that do a better job of aggregating supply and—one way or another—aggregating demand in order to achieve and maintain low information-transaction costs. The lower the transaction costs, the less inhibited use will be. In working toward these goals, two key tasks are to

- facilitate centralized, single-portal searching across all library resources;
- make the search experience as intuitive and user-friendly as possible.

Both of these need to happen, not just one or the other. A usable system that does not aggregate supply will continue to have high transaction costs, so it will not aggregate demand either. Libraries in today’s Web-dominated environment need centralized access to their collections; they need a single search portal equipped with the tools and features that will make its vast, diverse contents navigable.

Centralizing Access in Catalogue History

As urgent as the need to centralize access manifests today, it is not without precedent in the library world. The profession has been trying to consolidate access to intellectual content for a very long time, revealing a recurrent interest in bringing efficiency to the search process, but with slightly different goals at different times.

Early on, access was centralized within the library catalogue. Most early library catalogues served as inventories to relatively small physical collections, but in the eighteenth century, as libraries grew in size and scale, collections came to require finding aids rather than lists of items. The purpose of the catalogue shifted from inventory to indexing (Hanson &
Daily, 2009; Strout, 1956; Tyckoson, 1991; Weber, 1964). Certain libraries came close to collecting and cataloguing the published knowledge of the West. The Bibliothèque Nationale de France, comprised of confiscated royal and aristocratic libraries, was one of these. Its achievement made a comprehensive catalogue of the West's knowledge (though perhaps taken for granted at the time as the “World’s” knowledge) a conceivable prospect, not dissimilar from the idealized ambition of summarizing the “World’s” knowledge in the Encyclopedia of Diderot and d’Alembert (Tyckoson, 1991). Whether such lofty aspirations were conceivable or not, as the volume of publication increased, the prospect of a complete catalogue was rendered impossible. The growing number of publications forced libraries to collect selectively, but even then cataloguing resources were not sufficient to catalogue every item in a collection. Journal publications became more common, so libraries were no longer collecting only books and had to make decisions about whether to include entries for journal articles in their catalogues. Some libraries, like the Boston Athenaeum under the direction of Charles Cutter, did allow for analytical catalogue entries for periodical articles and contributions to proceedings well into the nineteenth century, but such practice soon became unsustainable. Libraries came to depend on A&I services from scholarly societies and private companies for article-level, intellectual access to journal content while the library catalogue would provide information on the local availability of journals at the title level (Tyckoson, 1989; Weber, 1964). From that point on, obtaining an article held a higher transaction cost for user attention. It was a two-step process of locating an article in an A&I service and checking for availability and shelf location in the library catalogue.

Aside from the issue of scalability, libraries were likely to consider journal content as a secondary concern. Library cataloguing was traditionally linked to the humanities, and thus to the monograph, while periodical literature had emerged from the sciences. As such, periodicals were not equitably accounted for in early cataloguing codes but rather left to the specialized treatment of bibliographers (Hanson & Daily, 2009).

The displacement of journal articles from the catalogue is a practical compromise from one vantage point, but such outsourcing does not mean that libraries have given up on providing a central point for accessing library materials. Library catalogues of the nineteenth century included nonbook items like maps, charts, and other materials that diversified the formats discoverable by users. As material types have proliferated, cataloguing rules have tried to accommodate emerging formats, including the remote electronic resource formats of the present day. Other than cataloguing resources in AACR/MARC upfront, libraries also have been trying to reincorporate digital resources already described in other metadata standards back into the catalogue through metadata cross-walking, batch
ingest, or manual recataloguing (Deng & Reese, 2009; Taber & Conger, 2010).

The impact of decentralized access to information silos was initially not very pronounced. Before widespread implementation of online catalogues, library resources were generally experienced as centrally accessible inside a physical space through the three-part infrastructure of the catalogue, indexes, and reference librarians, who were available as guides and mediators for moving between these tools. In a pioneering article on catalogue automation, Swanson (1964) acknowledges how these three parts form a holistic point of contact for users within the library. Swanson describes the form and function of a hypothetical automated catalogue, which he calls a “console.” This console, housed inside the library building, would provide immediate, unmediated service for information needs. “Printed products,” bibliographies, indexes, and so on, would be consulted where the console might be lacking. Finally, reference librarians were to be available as a means of support. “All three forms of interaction are necessary” (Swanson, 1964, p. 113). One can presume that Swanson’s sense of the necessity and interrelation of these three parts extends from experiencing an analogous dynamic between reference librarians, bibliographies and indexes, and a print card catalogue, all inside a library building. Having these tools and resources housed together in a single physical environment facilitated the experience of centralized access, with librarians as the bridge between the different discovery components, ready and able to guide users or notice and intervene when users’ search experiences were challenged. In Dempsey’s terminology, one can think of the physical library environment facilitating various transaction cost controls insofar as it affords librarian interventions in the search experience. Nevertheless, Swanson foresaw the utility of having a console that could subsume some of the content that required librarian mediation. He envisioned, among many features that would come to be a part of online catalogues in the next three decades, a merger of information silos in which a user “may choose whether he wants to see an entire book, a specific page or chapter, a journal article, a reference work, an abstract, or a review or critique” (Swanson, 1964, p. 119).

Indeed, Swanson and others saw utility in this level of automation largely because the literal legwork of moving between library discovery tools was arduous for users, and the work to maintain large print card catalogues was demanding for libraries. In this he shared the same premonition of the benefit of automated access to information as Vannevar Bush in his famous conception of the Memex (Bush, 1945), though Swanson did not cite Bush in 1964. As the technology to automate library operations became more affordable and widely available, many libraries were keen to seize upon an automated catalogue as an alternative. By the
1970s, the MARC format had already been adopted to facilitate electronic transmission of catalogue records and realize efficiencies in printing catalogue cards. Automated catalogues were able to store and display existing MARC metadata in computer terminals inside the library. As these new catalogues developed and were adopted by users, libraries could eventually eliminate their card catalogues, alleviating problems of space consumption and labor-intensive physical maintenance (Saffady, 1989).

In 1964, Swanson devised his console to illustrate fundamental requirements that he thought future automated catalogues should have. He had four major functions in mind: finding a specific work, selection based on subject, selection based on similarity to other works, and selection based on one’s own or an expert’s previous use of a work. Revisiting Swanson’s conceptual online catalogue thirty years later, Su said, “While much of today’s jargon had not been invented some thirty years ago, many of Swanson’s ideas can actually be characterized . . . as ‘user-friendly.’” (1994, p. 151).

However, the motivation for the actual implementation of early automated catalogues was efficiency, with the impact on service to users as a secondary concern (Borgman, 1997; Hildreth, 1984; Saffady, 1989). According to Hildreth’s framework for the three generations of online catalogue development, these were first-generation online catalogues, primarily suited for known-item searching by author, title, or control number. They had fairly restrictive search options, requiring exact phrases within a single interface. Despite these limitations for use, many libraries pursued these first online catalogues for the efficiency they would yield.

Second-generation online catalogues brought enhancements such as controlled vocabulary subject searching, keyword searching, and the ability to apply Boolean operators to searches, to limit and refine search results. Many of these catalogues offered the option to search in a basic or an advanced interface. These provided powerful search options for users but had a steep learning curve (Hildreth, 1984). The urgency to attend to catalogue usability was stressed in some literature of the time, with an emphasis that user-behavior analysis be harnessed in the design of the catalogue (Borgman, 1986; Hildreth, 1984, 1987). At the same time, many A&I services were becoming available as electronic databases that could be searched on site and, later, remotely. There were calls for a merger of the catalogue and the electronic A&I databases of periodical literature, either by entering publisher-created A&I records directly into the catalogue (Hildreth, 1987) or by offering separate catalogue and periodical search portals (Hildreth, 1987; Tyckoson, 1989), though Hildreth said the “better approach would be to process an author, title, or subject search without regard initially to the form of publication which is indexed.” (1987, p. 663). Despite the calls for development, Hildreth feared that “these commercial suppliers of online catalogue systems will become stuck on
the plateau of second generation developments” (1987, p. 649). Aside from small, iterative progress, Hildreth had apprehension that catalogues would not move into the third generation of his framework of online catalogue development. By the time of Su’s revisitation of Swanson’s work in 1994, the expansion of the catalogue to include records for nontraditional items like journal articles had still not been accomplished.

With the coming of the World Wide Web it became possible for more users to conduct searches off-site. Integrated Library System (ILS) vendors, many of whom already offered an online catalogue module as part of their system, were now under pressure to optimize the online catalogue for Internet use. By this time, the ILS-bound online catalogue module was widely referred to as the OPAC, and all were still second-generation catalogues (Butterfield, 2009).

The progression from first- to second-generation online catalogues, and the movement of these onto the Web, yielded an increasing amount of unmediated searching by users, inside the library at first, but then moving off-site. When the catalogue was still a part of the library’s physical space, it could still be situated in the three-part point of contact acknowledged by Swanson. As off-site catalogue searching became more common, with users searching in their offices or homes, centralized access to the library’s resources was disrupted. Users had less recourse to the physical proximity of reference librarians to guide them in and between catalogue and noncatalogue discovery tools. With the user’s search process extracted from the library building, there was less coordination of services and resources for users at their point of need. Without this coordination, the library search experience became fragmented and decentralized, with users navigating the online catalogue and an increasing number of remote electronic databases, stitched together with varying success in the emerging and constantly changing display conventions of the Web. Further exaggerating the troubles of online catalogues was wide acceptance of online search engines and commercial Web sites, which, as already discussed, came to define user expectations for online search experiences, both in their graphical, visual conventions and in the actual mechanics of the search process. User expectations continue to closely align with the features and ease of use of commercial Web sites, so there continues to be frustration with the lack of centralized access to library resources and having to search in multiple places.

Another factor to bear in mind here is the role that union catalogues have come to play in libraries as yet another information silo of “available” materials, even though they are located in another library. Libraries have been expanding the scope of their catalogues through union databases and breaking away from the “physical” assumption of a catalogue. From day one, the library catalogue’s main function has been to “describe and index the holdings of a particular library” (Lewis, 1999, p. 263). This used
to mean a unique local physical library collection. For universities with multiple campuses or public libraries with a number of branches, there was a need to provide a centralized catalogue to account for materials held in all constituting library branches. This kind of intramural catalogue expanded the scope to collections located beyond a physical building, but “the local catalogue was [still] clearly ‘our’ product, ‘our’ service to ‘our’ patrons, the only gateway to ‘our’ materials” (Turner, 2010, p. 272). Needs for resource sharing between libraries gave an economic reason for developing interinstitutional union catalogues or utilizing existing union databases for interlibrary loan, in addition to cataloguing. Considering North American examples, the coverage of these union catalogues could be general (e.g., National Union Catalogue [NUC], OCLC), genre-specific (e.g., National Union Catalogue of Manuscript Collection [NUC-MC]), or subject-specific (e.g., TALON Union Catalogue of Monographs, Midwest Medical Union Catalogue) (Binkley, 1965; Hall, 2004; Kronick & Bowden, 1978; Su, 1994). They could also be regional, national, or even international in scope. Nonetheless, the concept of an interinstitutional union catalogue as a central access point to provide access to materials held and not held by a participating institution is a clear breakaway from the traditional “local” assumption of the library catalogue.

Library catalogues also went from local to beyond local when they started providing links to full-text content licensed from third-party vendors or stored remotely. From this perspective, library catalogues have marked a shift from providing access to locally and physically held collections to materials potentially available to users from anywhere, creating a much broader discovery environment. The problem, again, is that access to this discovery environment is not centralized; it is fragmented across many silos that include the union catalogues and full-text licensed content just mentioned.

The State of Centralization Now

Failure of a Monolithic System
The library catalogue, in either print or online form, has provided varying levels of centralized access to materials collected by libraries at different points over time. However, centralization of access through the traditional catalogue infrastructure alone is no longer scalable, sustainable, or appropriate. In order to accommodate emerging formats, the cataloguing community has revised cataloguing rules or provided new guidelines to existing rules. They have expanded the MARC format by adding new fields and redefining or adding newly approved values to existing ones. Cataloguing rules and the MARC format have been stretched repeatedly to try to make everything fit into the existing catalogue infrastructure (Seys, 1999).
Given the rise of digital resources, descriptive metadata alone cannot sufficiently present the complexity of nor guarantee perpetual access to a digital object. The shift from print to digital has presented a challenge that the traditional catalogue infrastructure is not designed to cope with. According to the DLF ILS Discovery Interface Task Group (2008, p. 8), “current OPACs are limited in their support for multiple metadata standards.” Administrative, technical, rights, structural, and preservation metadata are at odds with cataloguing rules and the MARC standard. Consequently, libraries have established separate repositories to store digital objects and their metadata for purposes of delivery, access control, display, and preservation. More and more bibliographic data, created under rules other than AACR2 and encoded in standards other than MARC, are living outside library catalogues. The library catalogue, therefore, is just one of the many fragmented information silos maintained, and it accounts for a smaller and smaller percentage of the resources to which libraries provide access.

Faced with growing heterogeneous metadata standards, the library catalogue remains monolithic in terms of its data structure. Lewis (1999) advocates for incorporating other data structures into the system and adapting the catalogue according to the changing nature of the collection. However, in the past fifteen to twenty years, the cataloguing community has demonstrated a tendency to force as much as possible into the traditional AACR2/MARC-based catalogue infrastructure by such means as converting non-MARC bibliographic data into MARC, to be ingested into the catalogue. Due to the sheer volume of items to be ingested this way, “[the] belief that all library-selected resources should receive MARC/AACR cataloguing is a pipe dream” (Medeiros, 1999, p. 303). The inherent monolithic structure of the ILS-bound, MARC-based OPAC is not compatible with the size and heterogeneity of the current information environment.

The User-Unfriendly OPAC
Besides the gradual failure to bring everything into the traditional catalogue infrastructure, the catalogue search experience itself is riddled with high transaction costs—a turnoff to library users. A survey done by the DLF ILS Discovery Interface Task Group (2008, p. 8) highlights several recurring concerns of library professionals regarding the OPAC:

- Current OPACs are limited in their support for multiple metadata standards and lack support for Functional Requirements for Bibliographic Records (FRBR).
- The OPAC is limited in that it searches only items owned by the subscribing institution.
- The OPAC interface is difficult to use and is not intuitive compared to other search tools (particularly search engines and e-commerce sites).
The more powerful features of the catalogue search are mostly hidden or exposed in such a way as to confuse the users.

- Exploratory searching is difficult, and OPACs often lack basic features like spell-checking and good relevance algorithms. Functionality does not encourage browsability or serendipity.
- Searching for known items can also be problematic, if users do not know exact titles or filing rules.

The growth and diversification of the catalogue throughout the years lead to the need for new tools to navigate this larger, more diverse environment. To correct issues like those highlighted in the DLF ILS Discovery Interface Task Group survey, an even more fundamental need is interfaces designed with user information-seeking behavior in mind. Many libraries today continue to use OPACs that do not benefit from such study or that have built-in, faulty assumptions about user search behavior.

The catalogue “objects” of Charles Cutter have represented the ideal for catalogue functionality for over a hundred years. Even the latest conceptual model for catalogue functionality, the FRBR, and the new cataloguing code based on FRBR, Resource Description and Access (RDA), are descended from Cutter’s objects (Denton, 2007; Tillett, 2004, p. 5). FRBR and RDA account for relationships between works, expressions, manifestations, and items (in the parlance of FRBR), and the relationships of persons, families, and corporate bodies to these works, expressions, manifestations, and items. Furthermore, FRBR expands Cutter’s original objects into the “user tasks” of: find, identify, select, and obtain.

The details of FRBR and RDA and their prospect for the catalogue have been discussed at length in many studies and summaries (Carlyle, 2006; Dickey, 2008; IFLA Study Group on the Functional Requirements for Bibliographic Records & International Federation of Library Associations and Institutions, Section on Cataloguing, Standing Committee, 1998; Taylor, 2007; Tillett, 2004). The FRBR model provides a conceptual framework to understand bibliographic entities, their attributes, and the relationships between them, and to support information organization tools (Dickey, 2008). Using FRBR as its conceptual basis, RDA reorients description practice toward a focus on relationships between bibliographic entities. By bringing out these relationships, FRBR and RDA together provide a foundation that can enhance browsability and navigation in emerging library catalogue systems (Dickey, 2008).

Despite the possibilities of FRBR and RDA, it has been said that the catalogue has never lived up to Cutter’s original ideal (Lewis, 1999, pp. 263–264). There is also criticism to be leveled at Cutter’s objects themselves, directed at the assumptions about user behavior that Cutter’s objects have engrained into the catalogue. For instance, Borgman (1996) describes how Cutter’s objects do not rely on any empirical data on actual
user behavior but rather assume that users approach the catalogue knowing at least one of three possible access points: title, author, or subject. This assumption was reflected in the arrangement of items in the card catalogue. Reconciling the navigation of these access points with a user’s information need was partially supported with cross-references in catalogue cards but largely dependent on the mediation of reference librarians and user development of a gradual familiarity with catalogue intricacies. Users could retrieve results in a faster manner and even far away from the library after OPACs were made available via the Internet. However, this improvement did not alter the demand that users learn how to work with the system. The OPAC, in its early stages, was criticized as “nothing more than a mechanized card catalogue” (Su, 1994, p. 137). Second-generation functionalities like keyword searching and Boolean operators did not change the inflexible demands on users to avoid typos and input queries in the preferred language of the system. OPACs continued to render record data in visual displays similar or identical to that of physical catalogue cards.

Given the superior usability and heavy user reliance on commercial search tools, the disappointment of user expectations and the unfriendly assumption that “the user needs to learn to master the bibliographic complexities of the catalogue” have isolated the OPAC in the online world (Lewis, 1999, p. 266). For example, when analyzing known item searching, Kress, Del Bosque, and Ipri (2011, p. 164) note that the advanced search interface of the OPAC requires users to “figure out the elements of the citation to enter the elements into the right fields.” Craven, Johnson, and Butters (2010, p. 75) found that the majority of users’ issues are related to “ease of use” or, rather, a lack thereof. Users expect the OPAC to function like Web search engines with a single search box, relevancy ranking, and interactive search features. The OPAC falls short of these expectations. The persistent situation has made criticism of the OPAC increasingly harsh. Not only is it “hard to use” (Borgman, 1996) but it “is in danger of becoming invisible and even inconsequential” (Davidson, 1999, p. 283), it “sucks” (Schneider, 2006a, 2006b, 2006c), or it is already “dead” (Shadle, 2008). These criticisms carry weight along with the evidence that, in 2010, 84 percent of information consumers began their information search on a search engine, 0 percent with libraries (De Rosa, 2010).

The unpleasant search experience does not end within the local OPAC. The multiple silos that users must comb through to do complete searches have their own varying levels of usability. In 1991, Tyckoson (p. 15) criticized the outsourcing of journal indexing that resulted in “lack of consistency and inconvenience to our users”—an inconsistency due to various forms and completeness of metadata and the various presentations and user interfaces of third-party indexing services. As a result, users had to negotiate between different thesauri and descriptive standards adopted by A&I services, library catalogues, and digital repositories (Tyckoson,
This fragmented information-seeking environment continues to confuse users today, as demonstrated by a usability study by Kress, Basque, & Ipri (2011) in which participants did not know where to begin their search. The complexity in coverage and holding information of resources accessible through the library’s different silos hampered efforts to even begin the search process.

The fragmentation of information silos has other direct impacts on the usability of the library’s entire discovery environment. The traditional OPAC still forces users to search against locally held and subscribed resources before bringing users to resources held in other libraries, despite the fact that union and consortial catalogues could expose users to a wider range of available resources. This obscures the availability of resources that are requestable or accessible from other libraries. The disconnect between the library catalogue and other local or licensed silos complicates use even further. Before the digital age, the library catalogue was a dead end that allowed users to search and locate library resources, but delivery of content was outside of the catalogue’s scope. Swanson's vision in 1964 of delivering content beyond bibliographic data to users through a console finally became a reality when full-text access became available through locally mounted databases in the late 1980s and through the Internet in the 1990s (Borgman, 1997; Su, 1994). For delivery, the catalogue changed from a dead end into a through road. However, catalogues still fall short of providing seamless search, delivery, and requesting for users, since each of these functions is done in a separate interface. With better integration between systems like interlibrary loan and identity management, users would be able to complete more transactions from search to delivery within a single interface, but the centralization of access to all resources into one search portal remains the paramount need.

**Next-Generation Catalogues**

Catalogue history and literature from the earliest conceptions of catalogue automation to the present assessments of the OPAC have demonstrated a preference to centralize major information silos into one portal for discovery, and the need for change in this direction is evident. Originally, appeals to merge journal literature into the catalogue were based on an assumption that, given the labor or technological capacity, this would be a useful thing to do. Cutter’s early attempts to include analytical entries for articles, and all the revisitations of the prospect since, suggest an inclination by librarians across time that users should not have to look in multiple places for one search. Now, however, it is no longer a matter of just utility, but also relevance. Libraries are no longer the default destination for users with information needs. Libraries’ fragmented search experience is complex and costly for user attention. If libraries do not centralize access to their resources and if they continue to have high trans-
action costs for users, they will lose those users to competing information providers. So how is centralization to be accomplished for today? The host of next-generation catalogue products that have emerged in recent years have finally pushed the maturity of available catalogues into a new, or “next-generation,” phase and raise the prospect of achieving centralized access to library resources in today’s information environment. The discussion that follows examines some of the categories and features of next-generation catalogues and considers the extent to which they might help libraries aggregate supply and demand for the present and future.

**Terminology**

**Next-Generation Catalogues.** There seems to be varying definitions of the next-generation catalogue, but it is related to what Charles Hildreth referred to as a third-generation catalogue in his framework for three generations of online catalogue development (Hildreth, 1984). Besides the enhancements already widely deployed in second-generation catalogues, third-generation catalogues were to provide more adaptive search assistance, support for natural-language keyword searching, an expanded scope to the catalogue linked across library databases, and context-based support for typos and other errors. The term “next-generation catalogue” came to refer to the hypothetical catalogue that would embody these and other enhancements as expectations for the catalogue continued to evolve, notably in principles outlined by Eric Lease Morgan (2007). With prominent implementations, like that of North Carolina State University’s Endeca catalogue in 2006, the term came to refer to a tool in the here-and-now, extant but in development, despite the fact that NCSU’s Endeca catalogue lacked some features of Hildreth’s third-generation definition (Antelman, Lynema, & Pace, 2006, p. 137).

Using the term “next-generation catalogue” may require the caveat that “catalogue” as it is used here “doesn’t do justice to the new, expanded vision of the library’s search environment” (Breeding, 2007, p. 5). Indeed, one of Morgan’s assumptions about the next-generation catalogue is that “it is not a catalogue” (2007). Many of these tools merge traditional catalogue data with digital collections, institutional repository content, and, to varying degrees, journal article indexing. This centralizes access to a great extent, with wider and deeper access than the term “catalogue” may portend. Different terms have come into use because of this, with some defining next-generation catalogues narrowly to distinguish them from the more recent Web-scale discovery services (Nagy, 2011; Vaughan, 2011). However, Breeding (2007, 2010) and others (Yang & Hofmann, 2011; Yang & Wagner, 2010) have still used “next-generation catalogue” to refer to all tools in the new, broader search environment, and the authors follow suit here.

**Web-Scale Discovery Services.** Vaughan (2011) identifies Web-scale dis-
covery services as those next-generation tools that employ an aggregated search technique across local and nonlocal information silos. Aggregated search is facilitated by the compilation of a single, preharvested index of metadata from across the library’s owned and subscribed collections. The Web-scale discovery service harvests metadata from library resources, like the local catalogue and institutional repository, as well as A&I or online journal content from vendors and publishers through negotiated arrangements. These metadata are then compiled into a single index, which serves as the basis for searching in the discovery layer. Web-scale discovery service providers may conduct normalization of metadata after it has been harvested to alleviate some of the inconsistency of metadata from numerous sources. With a single index, these tools tend to have fast performance.

These are distinct from other, older services that rely solely on a more distributed model of federated searching. Tools using the distributed model search library databases one-by-one and compile search results on the fly into a single display, a technique that yields slower performance. The scalability of the distributed search model eventually came into question, with Breeding declaring that “the current strategy of metasearch that depends on live connections casting queries to multiple remote information sources cannot stand up to search systems based on centralized indexes that were created in advance based on harvested content” (Breeding, 2005, p. 27). Nevertheless, some Web-scale discovery services and other next-generation tools rely on a combination of the aggregated and distributed search models.

Other Terms. Both next-generation catalogues and Web-scale discovery services may be referred to as “discovery layers” or “discovery interfaces.” Both facilitate a broader environment for searching as well as features to support it, and both function as separate catalogue layers, detached and independent from the library’s ILS.

Next-Generation Catalogue Features
In order to establish a framework for consideration of next-generation catalogues’ prospects for centralizing access, a general description of next-generation features is helpful, without reference to specific products or implementations. Several recent reviews of next-generation products already exist (Breeding, 2010; Vaughan, 2011; Way, 2010; Yang & Hofmann, 2011; Yang & Wagner, 2010), undoubtedly with more to come as these products undergo rapid change and development. Breeding’s (2007, 2010) list of next-generation catalogue features fully describes the general state-of-the-art. However, listed here are only the features that seem most pertinent to the goals of centralizing access and aggregating supply and demand.

Single Point of Entry. The first and foremost feature of the ideal next-generation catalogue is “a single point of entry to all of the library’s in-
formation” (Breeding, 2007, p. 10) where one search can yield relevant results from across all available library materials. A single portal for searching is the “holy grail” of centralizing access (Yang & Wagner, 2010, p. 698), because without it, searching continues to be fragmented across information silos, user expectations fail to be met, and search transactions continue to be of high cost. The single-portal search is the core of what the authors consider to be centralized access.

Still, there are questions to be answered about how single-portal search will be facilitated. Preaggregating metadata from across libraries’ collections into a single index is a huge step forward but not necessarily sufficient to centralize access in a way that is impervious to error and user frustration. Other features of the next-generation catalogue described below will have an influence on the usability of the next-generation catalogue, but one aspect that deserves special attention is the interoperability of diverse metadata within the preaggregated index itself. Cataloguers have long prided themselves on the value-added work of controlled vocabulary and authority data that underlie the smooth functioning of library catalogues. Some next-generation features, like faceted navigation, stand to benefit a great deal from this cataloguing work when it is well managed and mostly error free. Naun (2010, p. 340) sees the advantage afforded to next-generation catalogues by traditional cataloguing, including subject headings and classification, as “both a vindication and a challenge” to cataloguing practice.

So metadata from the local catalogue may perform well in the next-generation interface, but what happens when metadata from resources other than the catalogue is dumped into the same index? As Breeding (2010, p. 80) states, “To the extent that the new discovery environment combines resources that were previously separate, the library will need to enforce consistent metadata practices across its collections to the largest extent possible.” Metadata from different sources can vary in its data structure (e.g., MARC versus Dublin Core) as well as value standards in the records (e.g., Library of Congress Subject Headings versus Art and Architecture Thesaurus). Published crosswalks between popular data structures are already in existence, and some next-generation catalogues like Summon and WorldCat Local have been mapping metadata from different sources into a common data structure for indexing (Vaughan, 2011). However, mapping between various controlled vocabulary systems remains a daunting task. Libraries may be able to have some influence on the value standards employed and the overall quality of metadata from local digital collections or institutional repositories, but that may involve labor-intensive new work. Furthermore, when metadata is included from the hundreds or thousands of the library’s licensed databases, there is no telling what diversity of metadata might come to coexist. Negotiating access to metadata from these vendors and publishers ahead of time is a
huge task, so most libraries will not have the option of ensuring quality or agreement with whatever metadata value standards may be preferred for use in the next-generation interface. The resultant mix of metadata in the preaggregated index may be fairly inconsistent, which, if unresolved, could be counterproductive to the centralization effort. Next-generation catalogue vendors may be able to harmonize this mix of metadata, but it is yet to be seen how or to what extent this can be done. It could involve centralizing all metadata value standards into one preferred standard or developing crosswalks between the most common standards, but both options entail the rigorous work of mapping between semantic values, which is what metadata value standards, such as controlled vocabularies, comprise.

*State-of-the-Art Web Interface.* Though “state-of-the-art” here is difficult to objectively define, the interface should be designed to meet user expectations. It should be intuitive and easy to use, taking the successful interfaces of open Web search engines as a point of reference. Ideally, the interface design should also take study of actual user search behavior into account. Meeting this criterion relates to other features identified by Breeding, including faceted navigation, relevance ranking, visually enriched displays, search-term recommendations, user-contributed content like tags and reviews, personalization features, “more-like-this” recommendations, and other innovations in usability that we have yet to see.

What makes a usable interface critical to centralizing access is the greater need for intuitive navigation that is introduced by the expanded scope of the catalogue. If a library were to centralize access to their collections with an interface that was unfriendly to users or just inherently difficult to use, users simply would not use it, and the effort would be wasted. Usability plays into information-transaction costs just as significantly as do fragmented silos.

*Connections with External Applications.* Library catalogues have little gravitational pull on their users, who by far prefer Web-based search engines (De Rosa, 2010). For next-generation catalogues to aggregate demand, they would do well to find ways to integrate into discovery platforms that are already heavily used, like Google. In an academic context, the next-generation catalogue should be able to integrate itself into the institution’s course-management system. Going where users already are increases the likelihood of drawing them to library resources. This will also require reliable interaction with the institution’s identity management and authentication processes, so that affiliated users can be readily authenticated in a seamless transaction.

Similarly, integration with interlibrary loan applications is important. Users should be able to search for, locate, and request materials available from other libraries and, ideally, have those materials delivered as well, all within the same interface.
**Will Next-Generation Catalogues Deliver?**

A study by Yang and Hoffmann (2011) examines the features of a sample of next-generation catalogues deployed in academic libraries. Less than half of the sampled next-generation catalogues provide true one-stop searching, the most important feature with respect to centralizing access. Some services provide a true single point of entry for all resources, while others split article searching and “other resources” searching into two different steps. The biggest hurdle for providing a one-stop search in any discovery service is acquiring metadata from publishers and other third parties. In 2005, Breeding (p. 29) said that “having the publishers of content resources expose their entire collections for metadata harvesting and document indexing just hasn’t been practical from a technical or a business perspective.” However, the tide has turned since OCLC rolled out WorldCat Local in late 2007, providing article citations from databases and subsequently entering into partnership agreements for article content with more publishers in 2009. Other major vendors in the discovery service business have also entered into similar partnerships with database vendors, aggregators, and publishers, and adopted aggregated search as a backbone technology, with or without the federated searching as a supplement (Vaughan, 2011). Among these are Summon and EBSCO Discovery Service. As these service providers continue to compile large bodies of metadata and full-text content from publishers, they do the work of aggregating supply for the libraries that partner with them. Success in this market may well be determined by who aggregates the most data.

However, as discussed earlier, aggregating metadata from different sources without paying attention to its consistency and quality may be counterproductive to the centralization effort. Next-generation catalogue vendors may be able to map metadata into a single data structure to facilitate searching, but variations in vocabulary authorities used in different sources may pose a challenge in providing an optimized retrieval experience. Although keywords drawn from full-text harvesting by some next-generation catalogues like Summon may be able to mitigate part of the negative effect of this inconsistency, faceted navigation and other functions that rely on consistent metadata may suffer.

Some services do not limit the scope of user searches to just locally held or subscribed content. For example, libraries using WorldCat Local can have their service configured to show local-, consortial-, and worldwide-level holdings, given that the service is based on OCLC’s vast WorldCat database. Summon also allows users, if they so wish, to search digital collections originated from other Summon institutions as well as journal content not currently subscribed to by the local institution. The next-generation catalogues that are providing a search scope that is beyond local are maximizing the width and depth of access to intellectual content.

According to a recent study by Yang and Wagner (2010), no open
source or commercially available next-generation catalogue provides all of the features indicated by Breeding. Nonetheless, most of the features that are available on existing next-generation catalogues are there for the purposes of enhancing usability and navigation in the search experience. The usability of these new interfaces is under ongoing study, including in recent work by Majors (2011). Preliminary results by Majors suggest that these interfaces are improving the search experience. Even with cases of the next-generation catalogue succeeding in facilitating search across silos and improving the overall search experience, it has not superseded individual silos in the library information environment. These are still available as separately searchable portals, able to perform within their own respective strengths for specific functions or subject domains. Librarians and users may continue to prefer using individual silos in certain cases, including the OPAC, if their next-generation catalogue is insufficient in a particular function or feature. For example, academic faculty and subject experts who might prefer the OPAC-style precision of pre-coordinated author and title field searching will still have recourse to it, even as next-generation interfaces become increasingly robust in known-item searching.

Discovery is not the only goal of next-generation catalogues; delivery also needs attention. As Breeding (2007, p. 8) puts it, “the second half [of the information-seeking process] involves putting the actual content in front of the user through the online viewing of electronic content or services related to providing physical materials to the user.” This hearkens back to the need for connections to external applications. While link-resolver technology for instant electronic delivery of online content subscribed to by the institution is a common feature that can be connected to next-generation catalogues, there is still work to be done for most systems to truly streamline request and delivery services, especially for content that is not subscribed to or held by the institution. Libraries, library cooperatives, publishers, and next-generation catalogue vendors will need to do more work to achieve seamless authentication for accessing licensed materials and seamless integration of consortial requesting and interlibrary loan services into their discovery services. This will significantly improve user experience by lessening information transaction costs. Instead of asking users to log-in and identify their affiliation when requesting materials every time, the next-generation catalogue should be made to communicate with authentication identity management systems and document requesting systems in the background and seamlessly to trigger request and delivery with a single click. Without seamless integration, discovery and delivery are still two related but disconnected steps in the whole information seeking process, raising each user’s information-transaction costs.

Connections to applications on the open Web must also be established to achieve the maximum level of exposure for library resources. This is chiefly about taking advantage of an already aggregated demand. “Think
Centralization of access: they aggregate demand by mobilizing large network audiences for resources” (Dempsey, 2006b, “Matching Supply and Demand,” §1). “Rather than waiting in the catalogue for users to search, library data should ‘leak out’ into the information space, should be indexed by search engines, and should be formatted such that entries can be included in other databases or resources” (Coyle, 2007b, p. 414).

Next-generation catalogues have yet to see much development in this regard. Until libraries have tools to facilitate this “leaking out” into other locations with high aggregated demand, they will be limited to the domain of their own constituent users. Obviously, this is an incredibly important domain, and the one that most libraries are mission-bound to serve, but discovery and use beyond the library’s own institution is still valuable insofar as it buttresses the relevance of libraries with demonstrable outcomes. It is a step in the direction of library service to the world, throughout the world.

So, next-generation catalogues are making long-awaited progress in centralizing access, aggregating supply, and, on an institutional level, aggregating demand. These tools represent the most concrete prospect libraries have in the here-and-now for lessening their users’ information-transaction costs. Next-generation catalogue service providers that are building stockpiles of aggregated data (those referred to as Web-scale discovery services by Vaughan and Nagy) appear to offer libraries the highest level of centralized access, as their single search portal will be more likely to index all of the materials that the library owns or licenses. Those with the greatest pools of aggregated data set the pace for other service providers. Though the single search portal to a vast data store is the “holy grail” of centralized access, the stockpile of data must be coupled with effective features for usability. Study and development of usability features should continue and should inform the design processes of vendors to prevent libraries’ access systems from lagging behind those of open Web search engines. Aggregated data and usability should be pursued in tandem to maintain low information-transaction costs for users. Future study and development should lower transaction costs further, as could be accomplished by seamlessly integrating discovery and delivery in the next-generation catalogue and letting library resources “leak out” into the open Web.

Conclusion
Next-generation catalogues have reintroduced the prospect of centralized access. Although library resources reside in separate silos, next-generation catalogues endeavor to pull them together and provide a single interface for searching. Centralizing access in this way minimizes the user’s transaction costs, that is, it minimizes some complexity that may inhibit use. Additions of faceted navigation, relevancy ranking, connections to external
services, and other features tailored to enhance usability respond to new user expectations and work in tandem with the single search portal to reduce complexity. Instead of the libraries’ resources being accessed across a fragmented architecture, the ideal next-generation catalogue will effectively weave silos and subsystems together into one centralized, usable portal for access. The authors echo others who cite Web-scale discovery services as the tools that bear the most promise in this regard.

REFERENCES


Joshua Barton is head of copy cataloguing at Michigan State University Libraries, East Lansing, Michigan, a position held since 2011. From 2007 to 2011, he served as serials catalogue librarian and philosophy librarian at MSU. Joshua holds a MS in information with specializations in libraries and archives from the University of Michigan. He is interested in the future of catalogues and cataloguing, hidden collections issues and alternative literature in libraries. He has previously published and presented on next-generation catalogues and zines in libraries.

Lucas Mak is the metadata and catalogue librarian at Michigan State University Libraries, East Lansing, Michigan, a position he has held since 2006. He holds a MS in library and information science from the University of Illinois at Urbana-Champaign. He is interested in authority control, metadata cross-walking, and cataloguing workflow efficiency improvement through batch processing.