11-9-2006

Collection development strategies for online aggregated databases

Susan McMullen
smcmullen@rwu.edu

Patricia B. M. Brennan

Joanna M. Burkhardt

Marla Wallace

Follow this and additional works at: http://docs.rwu.edu/librarypub

Collection Development Strategies for
Online Aggregated Databases
Susan McMullen, Patricia B.M. Brennan, Joanna Burkhardt,
and Marla Wallace

A Brief History of Electronic Collection Development

The 1989 Clip Note <u>#11, Collection Development Policies for College Libraries,</u> listed only one example of a college library with a Collection Development policy for "Special Formats," specifically computer software. Since that time the world has changed considerably and the amount of library material available in electronic format is staggering. Availability of library materials in electronic formats has increased exponentially since 1989. Collection Development managers have been required to address this increased availability in the Collection Development Policies for their institutions.

Selection of electronic products for the library has many of the hallmarks of the processes used for non-electronic materials. Clayton and Gorman (2001), Thornton (2000), Holleman (2000), and Metz (2000), all suggest that products considered in the selection process should:

- Meet user needs;

- Meet institutional goals, objectives and priorities;

- Comply with the institutional selection policy (which will address quality, relevance to the collection, and aesthetic aspects);

- Fit within the confines of the budget, balance the collection, and address resource sharing and cooperative development agreements.

Licensing, accessibility, hardware, archiving, format, copyright and delivery of electronic products are issues specific to electronic products. Collection Development policies must incorporate guidelines that speak to these product specific questions.

Collection Development Policies for Electronic Resources

There is general agreement among those who write on the subject that a detailed Collection Development/Acquisition policy for the institution in question is essential. Existing policies require updating on a regular basis and as technology and equipment change. The specific issues relating to electronic products can be incorporated into existing guidelines, added as a separate set of guidelines in the existing document, or an additional policy specific to electronic products could be created.

Many institutions have made their collection development policies accessible via the World Wide Web. A useful collection of institutions and the Web links to their Collection Development policies has been posted by Anne Oakerson at Yale University http://www.library.yale.edu/~okerson/ecd.html.

In examining some of these policies, a variety of approaches can be found. Some institutions (including University of Notre Dame, Monash University of Australia, and Dartmouth College) do not have a separate policy for electronic materials, perhaps indicating that this is an issue of format and does not impact on collection development per se. Some, such as Columbia University and Boston Public Library, list specific criteria for consideration when electronic format is an option, perhaps indicating that the format might have some impact on the collection development decision. Several institutions have a separate policy for collection of electronic materials which is invoked when consideration of electronic format for an acquisition is possible. University of

North Carolina at Asheville, University of Oregon, University of New Orleans, University of Minnesota, University of Wyoming, and University of Tennessee, Knoxville all have separate policies for electronic acquisitions. Yale is among the institutions that have one policy for materials purchased alone, and another which applies when purchasing electronic materials consortially.

In Patricia Bril's chapter on consortia collection development in *Collection Management in Academic Libraries* (Jenkins 1991), she indicated that patrons use the "law of least effort," meaning that they seek convenience of access before quality of the results. In 1991 when this book was published, Bril mentioned that consortial buying and sharing of electronic materials would have to wait until more standardization of computers had been accomplished. At that time most of the efforts at shared collection development between or among institutions had failed. However, in today's electronic environment, library users will opt for electronic access over a physical visit to the library. The standardized technology now exists to the extent that consortial purchase and use of electronic databases and other tools is not only possible, in some cases it is preferable.

The policy for selection and acquisition of electronic products should address the criteria listed above. However, the special features and requirements of electronic products demand additional guidelines. Electronic products come in many forms. Some are simply replicas of the paper version. These will be the exact equivalent of the paper in an electronic format. Some will have text plus added capabilities such as search, data manipulation or alerting. Still others will include text plus high tech features such as hypermedia, electronic analog models, motion and sound.

It is still very important to consider the capabilities of the electronic product in relation to the capabilities of the hardware available. If the product requires more computer power or memory than is currently available, the product will not be a useful addition to the collection.

Collection policies will also include consideration of single user vs. multi user access, speed and functionality of the product, content - including the way(s) in which full text is made available - archiving and long term access, links, the user interface, connection files, special features, authentication, remote vs. local use, bundling with other products or stand-alone, duplication of information in other formats, competing products, one time purchase vs. subscription, stability of the provider, technical support, and fluidity of contents. Many of these issues will be discussed below.

Aggregators - a semi-definition

What do we mean by an aggregator database? A sweep across the literature of librarianship, publishing, and various academic disciplines shows that the term can be and has been used in many contexts, including:

- Weblog/blog news summarizers such as Syndic8 and NewsIsFree

- Market research gateways such as MindBranch

- Permanent journal archive projects, or what Li and Leung (2001, 199) call "stable aggregators," such as JSTOR or Project MUSE

- Abstracting and indexing services that contain or are linked to full text material - or what Li and Leung call "unstable aggregators," such as databases from EBSCOhost and ProQuest

For the purposes of this overview, issues surrounding the creation and development of the "unstable aggregators" will receive the bulk of the attention. But first, a review of the history of aggregator databases is in order.

History of Aggregator Databases

In the glory days of the CD-ROM revolution, before the rapid rise of the public Internet, and long before the creation of the Web, University Microfilms, Inc. (UMI) was the dominant, one might almost say exclusive, creator and distributor of periodical and newspaper backfiles for libraries. UMI realized that it was sitting on a potential goldmine. Periodical indices on CD-ROM had gained some level of acceptance in libraries of all types. What if actual article images could be linked to a powerful aggregation of indexes to give users access to, instead of mere awareness of, the contents of hundreds of periodical publications? Enter ProQuest. According to the announcement following its debut at the American Library Association meeting in June 1990, "'ProQuest' is the new name for UMI's family of CD-ROM products. It will apply to UMI's current line of [CD-ROM] databases, and any subsequent database and hardware offerings." (UMI 1990:22) Little did anyone realize at the time what momentous changes that first aggregator "banner" would weather. Within six months, vendors were offering CD-ROM minichangers - or towers - with almost 4 gigabytes of online storage to accommodate the increasing amount of direct content offered with CD-ROM-based periodical indexing. By late 1992, *Library Journal* published a comparative review of four periodical indices with accompanying article content on CD-ROM: EBSCO Magazine Article Summaries Full Text Elite, InfoTrac Magazines ASAP Plus, ProQuest Magazine Express, and Wilson Readers' Guide Abstracts and Readers' Guide Select

Edition. The Big Four aggregators went head to head in 1992 and they continue to do so today.

But notice that these products focus coverage on "magazines." Generalist periodical literature of the type librarians have always associated with Wilson's Reader's Guide to Periodical Literature made these products extremely attractive to large, well-funded public libraries, but only moderately interesting to academic libraries. At the next level of sophistication, aggregations of business periodicals appeared on CD-ROM almost simultaneously with their more generalist cousins - a bit more interesting to college and university libraries with established curricula in these areas. But the vendors were not yet grappling with the body of scholarly journals and periodicals that are the life-blood of the academic library. Nor were they yet prepared to address many of the issues that flow from the academic mission to support permanent, complete access to particular publications or bodies of literature.

Academic Libraries and Electronic Journals

The purposes and aims of popular, magazine publishing and scholarly, journal publishing are fundamentally different. Magazines exist to make money for their publishers by filling a market niche, providing a service, advancing a cause, or publicizing a socio-political point of view. Christopher L. Tomlins, (editor *of Law and History Review*, the journal of the American Society for Legal History), in a wonderfully cogent paper, republished by the American Council of Learned Societies, writes: "...scholarly journals...are created to be disseminators of authoritative scholarship - authoritative in three senses. First, authoritative in the sense that the scholarship the journal publishes is certified by its editorial practices to be as reliable as the collective expertise of the group

of professionals who have produced and judged the scholarship can make it (a group that encompasses originating author, editor, referees, copy editor, managing editor, typesetter, production manager, et al.). Second, authoritative in that the journal represents authored knowledge, knowledge that is definitively attributable, that has achieved a unique representation for which an author and an editorial process is answerable and is identifiably responsible, knowledge that is substantively stable and will not alter without becoming something identifiably different, as in differently authored, or differently attributable-and if thus altered, then altered in accordance with defensible and acceptable and reasonable professional canons of what difference constitutes. Third, authoritative in the sense of authorized for inclusion in an archive, for that is what a scholarly journal also is." (Tomlins 1998, 3) Having brought the perspective of the academy particularly as it applies to the humanities and social sciences, as well as the point of view of the editor of a well-regarded but small academic journal to this view of the scholarly journal, Tomlins goes on to capture the counterbalanced threats and opportunities of electronic publishing for scholarly journals at the cusp of the twenty-first century. Outside the academy, many of those involved in the actual electronic dissemination or distribution of scholarly journal contents -- and therefore involved in determining their future in the online environment --view academic journals as "bundles of resources, mobilized in the rapid commodification of information that dominates [our] culture." Proponents of this view include "publishers, aggregators, information services, and their institutional clients." (D'Arms 1998, np)

Aggregators' Impact on Collection Development Decisions

By the early to mid-1990s, the increasing sophistication of the public Internet, followed by the advent of the World Wide Web, spread both the benefits and the pitfalls of electronic publishing. The need to extend the reach of the useful and accurate as a counterbalance "between the traditionally sanctioned and well-funded ...the unfunded and unsanctioned ...or lone lunatic" (O'Malley and Rosenzweig 1997, 136) propelled many publishers of academic journals to find ways to make their material accessible electronically. Many avenues were explored and many of those avenues remain well-traveled, from free access to the same material published in paper (Proceedings of the National Academy of Sciences - PNAS), to cover-to-cover reproduction and archiving of specific groups of titles (JSTOR and Project MUSE), to direct access by subscription to a publisher-maintained Web site, to loading of content in "unstable aggregators." But aggregation has frequently interfered with, or failed to maintain, the senses of authority so valued by the college and university. Instead, it emphasized the commodity value of individual pieces of information to individual users. The bundling, or aggregation, of large numbers of journal titles, based principally on current availability of content by publisher contract, often led to the unwitting "acquisition" - however temporary - of titles which a library would not have considered acquiring by individual subscription. The enthusiasm of the library community for the astounding availability of so much digital content - from a gross of aggregated titles on CD-ROM fifteen years ago, to a few hundred aggregated titles on CD-ROM in the early 90s, to hundreds of aggregated titles via telnet on the Internet, to thousands of aggregated titles via the World Wide Web - led to some instances of uncritical collection development decisions. Assuming that once content was included, access to it would remain stable, some libraries withdrew or

deaccessed titles included in aggregator databases only to discover that publisher

contracts were fluid and that content could be added or withdrawn by the publisher or

aggregator at any time. The library was renting - not buying.

The effects of these collection development decisions are highly dependent on the

importance attached to particular periodicals by individual academic communities. One

institution's vital publication is another's secondary purchase or irrelevant title. So

although some libraries found themselves scrambling to reassemble in physical form or

through direct electronic subscription materials that they had previously withdrawn, most

had cautiously continued to hold the most valued titles in physical form as the electronic

publishing landscape began to offer more stable choices for archived electronic versions

of those journals. Yet as recently as 2000, Reich and Rosenthal, writing for Highwire

Press at Stanford, noted: "Librarians have an equally well-founded skepticism that they

can provide their readers with long-term access to material published on the

Web. Important content can be changed or removed at the publisher's whim. It is leased

to subscribers; they don't own it. There is no local 'collection' for a library to develop or

manage. The publisher's promise of 'perpetual access' is empty, there is no enforcement

mechanism behind it. The result is that libraries are reluctant to subscribe to Web

journals (Reich and Rosenthal in Curtis, Scheschy and Tarango 2000, 19). This

skepticism was born out as recently as 2003, when Sage Publications withdrew its

prestigious social sciences journal content not only from various aggregators, but from

OCLC's Electronic Collections Online - whose mission it is to permanently archive

content for institutional subscribers.

Although some parts of our user community - typically undergraduate students - would like to see the electronic format replace physical access as quickly as possible, the other parts of the academic community will continue to rely on our prudence, caution, thorough knowledge of the library's user community and its curricular emphases as well as forward-thinking to inform our decisions about the library's periodical "collection." "The online environment is not the wave of the future, it is the wave of the present (Tomlins 1998, 6)...in an information environment that is already highly commodified and becoming daily ever more so...[there is] no doubt that to continue to play its first, preferred role of scholarly authorizer, the journal will have to prove itself attractive as a player in the second incarnation too, as a supplier fit to be included in the commodification of information...(Tomlins 1998, 5). To that end, academic publishers continue to offer their content to aggregator databases. And academic librarians have continued to influence aggregators to become cognizant of and attempt to address those issues of continuing importance to the academic community - continuity of access, completeness of coverage, and stability of the backfile.

Evaluating Database Content and Features

In choosing an aggregator database, libraries need to decide which features are most crucial to their users and develop a methodology for evaluating these features. In developing selection criteria, it is useful to develop a list of guidelines concerning database functionality. Which functions or features are of primary importance to your library, which are marginal, and which features can provide value-added services to your patrons? Areas which require particular analysis include:

- Breadth and depth of coverage

- Quality of indexing

- Usability, which includes screen design and ease of searching

- Ability to customize the database

- Delivery options

- Accessibility

- Availability of statistical usage data

- Quality of support.

Depth and Breadth of Coverage

What is the content of the database? Evaluating database content goes well beyond such preliminary indicators as quantity of titles indexed and numbers of full-text journals included. The value of a full-text database is greatly dependent on its ability to offer additional quality sources to the library's collection that otherwise may be beyond the reach of its serials acquisitions budget. Consequently, librarians should be examining the full-text titles contained within a database to determine how many are peer-reviewed. Of the peer-reviewed journals included, how many are unique or exclusive to a particular database? Another indication of database content value is how the database is evolving over time. Over the past year, how many scholarly journals have been added to the database – as compared with how many scholarly titles have been halted or removed? Look at the same statistics over a two-year and three-year time frame as well. (Brooks, 2001) Can you discern any trends in the direction of database content? What are the dates of coverage for each title? How is the vendor handling cessations? If a publication is halted or removed, how is the customer notified? Is it easy to tell which titles have

been ceased? Are archival issues still available within the database or have they been

eliminated? Is indexing to these previously full-text journals still provided?

In their study, Brier and Lebbin stressed the importance of evaluating title coverage of

full-text databases in relation to the library's print collection. The database that provides

the greatest number of titles may not necessarily provide the highest quality journals or

the most appropriate titles for a particular library's collection needs or clientele. (Brier,

1999) Since the goal for an academic library is to provide more scholarly resources to

the library's collection, a subject content analysis of peer-reviewed title coverage may

also help librarians determine if a particular vendor's database content may add value to

their subject journal collections. Additionally, one might consider how many of the

library's print and electronic subscriptions are indexed by the database, thus ensuring

access to library-owned collections. Though this criterion is definitely a convenience, it

is not an indication of the quality of titles indexed that would be available from other

libraries through interlibrary loan.

Due to the fluid and ever changing nature of publisher licensing for journal content in

aggregator databases, librarians have to be concerned about the holes and gaps that may

occur in their collections when key publications are dropped from an aggregator's

database. In a volatile marketplace, it is often risky business to drop a print subscription

based solely on its inclusion in an aggregator database. A perfect example of this type of

volatility occurred when Sage Publications decided to remove the full-text content from

Sage journals which had been previously available in aggregator databases, such as

EbscoHost and Proquest, and begin selling their own full-text content through separate

"Sage Full-Text Collections." Each collection is currently licensed separately through Cambridge Scientific Abstracts.

Embargo periods lasting a year or longer have also made it more problematic for libraries to drop print subscriptions to journals when current content is essential for research. An embargo is a publisher-imposed restriction that does not allow vendors to display full-text content from their journal titles for a defined period of time. An embargo period is meant to prevent libraries from canceling their print subscriptions just because they get the content electronically through an aggregator database.

Questions to ask vendors concerning embargo periods include:

- What percentage of their full text journals have embargo periods?

- What are the specific titles of the journals with embargoes -- are they titles for which you need the latest twelve months of coverage – do you have print back-up available?

- How easy is it to determine which journals have embargoes and what is their embargo period?

- Are embargo periods clearly stated on title lists and are those title lists easily accessible online?

Another indicator of depth and breadth of coverage is the number of scholarly or peer reviewed journals found within search results. Search results that contain articles in the greatest number of scholarly journals may not always be found in databases with the largest journal count (Brennan, et al., 1999).

Quality of Indexing

In our study completed in 1999, librarians from HELIN, the Rhode Island academic library consortium, wanted to find out what electronic full-text really meant. What exactly are the indexing criteria used by the major database vendors? Librarians compared actual print issues of several journal titles to the full text content available within the database to determine how closely the online full-text or full-image content matched that of the print journal. "What the vendor calls full-text access may not mean cover-to-cover full text access."(Brennan, et al., 1999) What type of content is provided? Substantial articles? Brief articles (less than a column in length)? Editorials? Commentary? Special issues? To do this type of evaluation, it is necessary to display online the contents of an individual issue. Being able to view the table of contents for a particular journal issue is also vitally important today as a means of tracking articles that have been cross-referenced from other databases or from the library's online catalog. When replacing paper subscriptions with online full-text access, librarians should be aware that they not only risk losing access to the entire title if the publisher pulls its content from the database, but because of variations in indexing criteria and quality, everything that is contained in the paper copy may not be included or accessible by the aggregator.

The subject authority used to construct the database will have a clear impact on its search and retrieval functionality. Does the database use LCSH or another standard thesaurus or its own controlled vocabulary? What record fields are indexed and searchable? What fields are automatically searched when performing a basic keyword search? Does the database support subject as well as keyword searching? Can a user browse by subject or view an online thesaurus?   The availability of subheadings and related terms within a

subject browse can be especially useful for the user. Linked subject descriptors for each record are essential. Explore these links to see if you can determine the depth and breadth of subject indexing. Is the subject linking well developed or does it capture only a portion of the articles that may be available on a subject in the database?

Another indexing consideration is the availability of information on title changes. Journals frequently change titles, often leaving users and librarians confused about which title is a continuation for which journal. Are there cross-references to the older title, or are the old and new titles handled as separate entries with no cross references? If a Browse by Publication feature is available, a link giving a detailed history of the title and its variations with dates is very helpful.

Finally, librarians should be concerned about the currency of the indexing. How readily are citations loaded after receipt of publications? Is there a time lag between the loading of citations and the availability of full text? Are there restrictions, such as embargo periods, that govern how soon the full text is made available after the citations are loaded? If so, does the vendor provide this information within the product in a publication list or elsewhere? (Brennan et al., 1999)

Usability: Screen Design and Ease of Searching

Ideally, vendors perform usability testing on user interfaces to ensure that their database is easy to search and navigate. Librarians should be particularly concerned about the intuitiveness of the search interface. To test the usability of the interface, librarians simply have to ask a few users to perform selected searches to see how successful they are, or observe users in their own search process. Note how search terms are entered and if satisfactory results are retrieved. No matter what the content of the database is, if a

user cannot successfully navigate his way through the search and retrieval process to find satisfactory results to his query, its viability is compromised. (Still and Kassabian, 1999). Although there is no substitute for being an informed database searcher, users should receive some clues about how to search and have ready access to context-sensitive help screens. Search options should be readily apparent and searching should be self-explanatory. For example, users might receive clues from the availability of drop down boxes containing search options or selection choices for searching all of the words, any of the words, or the exact phrase.

Usability/ ease of searching questions to ask include:

- Is terminology clear and consistent throughout the search and retrieval process?

- Are icons and buttons labeled with words?

- Is the navigational scheme clearly understood? Are links to the search screen, results lists, marked items, etc. clearly recognizable?

- Does the results screen offer users options for improving their search results?

- Is it clear to users how to limit or expand their search results? Which limiters are supported?

- Are advanced search options readily available and easy to understand?

- How are search idiosyncrasies handled, such as hyphenated names and Spanish surnames?

- If zero hits are retrieved, are users redirected or given the proper syntax for improving their search? What automated error checks or redirects are available?

- What delivery options are available and are they clearly recognizable and easy to use?

These questions are reflective of what usability experts such as Jakob Nielsen have established as usability heuristics for Web design. Nielsen's ten usability heuristics provide a standard for measuring the usability of Web-based databases.

1. ***Visibility of system status (where am I and where can I go next?)*** -- Is it readily apparent to the user what to do upon entering the database? Are the options for searching readily available? Once citations are displayed, is it obvious how to obtain full-text? Is it clear to the user what is happening when certain display options open multiple windows?

2. ***Match the system and the real world*** – Is terminology clear and understandable to the user?

3. ***User control and freedom*** – Is the user in control of the search process? Can they easily return to the search screen? Are there options available for refining searches available from the results screen?

4. ***Consistency and standards*** – Does the database follow standard conventions that are recognized across a variety of databases? For example, is an asterisk * used to represent truncation? Are search buttons labeled and located where the user would expect them to be?

5. ***Error prevention*** – What types of hints does the user get for entering proper syntax for retrieving desired results?

6. ***Recognition rather than recall*** – Is the design of the interface intuitive enough for the user to know how to proceed without having to remember specific usage instructions?

7. ***Flexibility and efficiency of use*** – Is it easy for the user to move among search, display, and result screens? How easy is it to manipulate results? Can users download them easily into bibliographic management software?

8. ***Aesthetic and minimalist design*** –The interface should be clean, and it should be easy for users to understand search options. Instruction should be minimal and written for scanability.

9. ***Help users recognize, diagnose, and recover from errors*** – If a user does not retrieve expected results, is it clear to the user what has happened? What types of error checks are in place for improper syntax or zero hits?

10. ***Help and documentation*** – Is help readily available? Is it context sensitive? Is there a searchable index of help topics? Does the database have its own tutorial?

Search Capabilities

Under the category of search capabilities, evaluation criteria may include whether or not Boolean logic is supported, if both subject and keyword searching is available, and what advanced or complex searching techniques are provided. Users should have the ability to perform field specific searching, use Boolean connectors and proximity operators, and perform ordered searching with parentheses. Beyond keyword searching, what additional types of searching are available? Is natural language searching supported and if so, does it produce reliable results? Can users browse contents by publication? This is often an important capability for users wishing to see what is available within a particular journal title. What types of search limiters and expanders are provided – can results be limited to peer-reviewed journals, by date, or by publication? Is limiting provided from within a search results screen or must the user set limits before executing a search? Providing

additional search features within results can also prove very beneficial to a user. The ability to "find articles like this one" helps users unfamiliar with linked subject headings discover similar articles on a particular topic once they have found one that may be especially relevant to their research needs. By cross-referencing and linking the author field, users can also find additional articles by the same author.

Results and Delivery Options

When evaluating results and their delivery options, the biggest concerns are functionality and ease of use. How easy is it for the user to view and manipulate their results? The ability to select individual records to export into a marked records list should be clearly indicated and easy to perform. How long does the database hold onto the marked records before they are cleared? Do users have enough time to evaluate and mark long lists of citations without fear of losing them before they are able to finish? When users select marked lists for download or emailing, is it clear whether the user will receive a list of citations or will full-text automatically be included when available? Do they have a choice to download or email just the citation records? Can they download results directly into a bibliographic management program?

Printing, downloading, and emailing of results should be standard features in any vendor's product. Most aggregators deliver full-image content in PDF format requiring the simple download of *Adobe Acrobat Reader*. Specialized database content providers may require the downloading of their own reader software. This can sometimes prove cumbersome for the occasional user of the product. Even when using standard PDF, users may experience difficulties interfacing with different software platforms. What version of *Acrobat Reader* is needed to open the PDF content within the database? Does

it work equally well on Macs and PCs? Are special print drivers required to print non-standard fonts? Does full-image content display in the same window or does it require opening a separate window?   To email full-image content, vendors should provide users with an automatic means to send the article as an attachment. When only full-image content is available, delivery options should default to this attachment option.

Customization

Recognizing that libraries have unique collections and patrons, a vendor can greatly enhance a database's viability to individual constituencies by allowing for customization. An adaptable database allows librarians to tailor certain features and default screens to the skills and preferences of its clientele. For example, many librarians may wish to make a database's advanced search screen the default search interface for their users. By offering users a visual presentation of separate search boxes with options for Boolean connectors, many librarians feel that patrons will achieve better search results. Usability studies at other institutions may indicate that the advanced search screen is too confusing and those libraries may prefer a more simplistic, single search box interact that defaults to an "and" search.

Other search customization options may include:

- How much of the backfile to search automatically – all of it? The current year? Past three years?

- The ability to search all vendor subscribed databases simultaneously, individually, or by subject area.

- The ability to choose search limiters and expanders.

- The decision to include links to vendor evaluated Web resources in the search results.

Result display screens should also offer several options for customization. At the most basic level, options for displaying results might include how many citations to include on a screen or how to sort results (i.e. by date or relevance). When displaying full-image content in PDF, is there an option for displaying the article in the same window or opening up a separate window? Results may be further customized to include a link to the library's OPAC or a note about local holdings. Is cross-linking to other full-text databases and e-journal collections available? If so, should links be made available to these other collections? Librarians may wish to ask themselves how useful these links may be if they need to be unilaterally turned on to appear in all title results. Is there a way to customize which links to other e-journal collections appear with which journal titles?

Branding the database with the library logo and with a link back to its home page can also be utilized as a marketing tool. This type of branding may send a message to users that the contents of the database are being offered by the library and are customized for their individual use.

Value-Added Features

As mentioned above, cross linking amongst databases can be a very powerful tool in helping users find full-text content from outside of the database they are searching. How sophisticated is the level of cross-linking within the database? At its most basic, the database should link to the library's OPAC. The OPAC should then in turn provide information on the library's holdings as well as links to online full-text title content from

a variety of vendors. To provide this type of linking, vendors need to supply libraries with MARC formatted title records to be integrated into the library's catalog. Consequently, other criteria for evaluation would be to determine if MARC records are available and how extensive the records are. What other types of cross-linking between database products are provided? Are database cross-links available only to some vendor subscribed databases?

Other value added features to consider:

- Can users set up subject or search profiles so that they can be notified when new content becomes available in the database based on their subject interest or on a particular saved search? (Barber 2000, 19)

- Does the vendor supply persistent URLs which can be used to include on library produced course guides, instructor syllabi, or other Web-enabled instruction venues?

- Can results be downloaded into a bibliographic management databases such as Endnote or Reworks? How cumbersome is this download process?

Accessibility and Customer Support

More than ever before, library users are accessing database content from remote locations. Librarians need to be concerned with ease in authenticating their users for access to these expensive database products. How are users authenticated? Does your library proxy server interface seamlessly with the database? How many user logins does your license allow? For general aggregator databases, a site license with unlimited logins is the standard.

With so many users accessing databases from outside of the library building, the need for enhanced Help screens and other forms of customer support also becomes more essential. Is Help context sensitive? How usable and complete is the listing of search Help contents – and are the contents searchable?

Usage Statistics

Though sometimes overlooked, the availability and accessibility of usage statistics is a very important component of the database evaluation process. The reporting of usage statistics from database vendors is essential for library decision-making and for validating the continued budgetary commitment for online database products. Usage statistics should provide librarians with enough information for accurate cost analysis, justifying expenditures, and determining usage trends over time. Usage indicators may include a count of sessions, time per session, a count of searches, and count of full-text downloads. "In addition, usage statistics can show a variety of information, including success or failure of user access (e.g., turn-aways per time period per specific database), user access methods (e.g., telnet vs. browsers), access levels at one institution compared against peer institutions, cost of access (e.g. cost per downloaded item) and other items pertaining to user behaviors." (Shim and McClure, 2002) When looking specifically at aggregator databases which supply the full text of many journal titles, it would be useful to know the number of full-text article downloads provided from each journal title in the database. This is particularly useful for cost analysis, allowing comparison of the actual cost per article download against the cost of the print subscription.

When evaluating networked databases, librarians should be sure that usage statistics will be provided to them on a regular and consistent basis. Are usage statistics continually

available from a vendor's Web site or does one have to wait for monthly, quarterly or even annual reports? How will the data be distributed – by email, fax, U.S. mail, or via the vendor's administrative Web site? Is the data automatically distributed or does it have to be requested? What about statistics in a consortial arrangement – can vendors successfully sort usage statistics to provide individualized reports to each institution in the consortium? When viewing statistics from a vendor's Web site, is there a choice of viewing formats? Some of the most common formats include Excel, PDF, CSV, ZIP, and HTML. Librarians should look for usage reports that are easy to retrieve, manipulate, and interpret. Think about what is most convenient for your institution. Text format allows librarians to easily input data into spreadsheet format; PDF does not allow data manipulatation. Does your institution have the ability and staff to use raw data for interpretation and manipulation? Most importantly, once the data is retrieved, is it understandable? Is there documentation available that will provide enough details to enable you to interpret and understand the statistics provided? (Shim and McClure, 2002). The methodology employed by the vendor for collecting usage data is also an important consideration for understanding how the database is being used. Can you retrieve information from the vendor's Web site about terminology and process – how the statistics are collected? Is it consistent with other vendors? Reports from different vendors are often inconsistent making it difficult to make comparisons.

To make informed decisions, measures of usage data should be trustworthy, consistent, and comparable with other database products. In the past few years, several initiatives have been investigating the standardization of electronic measures in order to provide

guidelines for assessing networked resources. In evaluating database products, it would
be useful to see if the vendor adheres to the guidelines outlined by these initiatives:

- The E-Metrics Project (http://www.arl.org/stats/newmeas/emetrics/index.html)
  was an effort by Association of Research Libraries (ARL) to identify new
  standardized measures for evaluating usage of electronic resources and to
  promote collaboration between database vendors and research libraries.

- The International Coalition of Library Consortia. Their December 2001 revision
  of *Guidelines for Statistical Measures of Usage of Web-based Information
  Sources* (http://www.library.yale.edu/consortia/2001webstats.htm) defines
  mandatory data elements and provide parameters for confidentiality, access, and
  delivery of usage measures.

- *Counter*, an international group supported by publishers and library
  organizations, released an International *Code of Practice* in December 2002 that
  "specifies the data elements to be measured, definitions of these data elements;
  usage report content, format, frequency and methods of delivery."
  (http://www.projectcounter.org/)

Methodology

The methodology used for a library's evaluation of aggregator databases will necessarily
need to conform to the specific goals of the evaluation project.

While the specific goals of an evaluation project may differ, each analysis is likely to
consist of three phases: analysis of user needs, analysis of vendor policies and standards,
and determining the extent and quality of full-text coverage provided by the database. At

each phase of the evaluation, it is helpful to create worksheets and check lists to track the data you are compiling.

This section suggests issues to consider before designing the project, and describes the methodology used in two database comparison projects.

Analysis of user needs. This phase lays the groundwork for the evaluation. Questions to consider in this phase include:

- Is the goal to select a multidisciplinary aggregator database, or a discipline-specific one? Are you considering replacing your existing product, or purchasing a new one?

- Is the plan to replace current hard copy subscriptions with electronic access, or to increase the depth of the collection by creating electronic access to new titles?

For a quantitative analysis of a database, compare the vendor's full-text list to a standard that makes sense for your needs. If the database is multidisciplinary, compare the vendor's list to selected titles from *Magazines for Libraries*, for instance. Or, to assess the database's strength in business periodicals, compare the titles in a standard such as the *Harvard Business School Core Collection* to the vendor's list. To determine how many additional titles you will gain, compare the vendor's full-text list to your library's subscription list.

Brier and Lebbin (1999) describe the title evaluation coverage project at the University of Hawaii-Monoa (UHM). The goal was to compare the title coverage offered by three periodical database vendors to the library's print collection, and to evaluate the findings according to collection development principles.

The evaluators retrieved title lists from EBSCOhost Academic Search, Information Access Company's Expanded Academic Index, and UMI's Periodical Abstracts Research

II. Using *Magazines for Libraries* as the standard for an undergraduate collection, the evaluators aimed for a qualitative as well as a quantitative analysis. For the purposes of pretesting, they created titles lists from the three vendors, *Magazines for Libraries*, and UHM serial titles, then extracted only those titles beginning with the letter "A." At the end of the evaluation, they were able to assign three discrete values to each database:

- The *full-text value* reflected the number of number of full-text titles new to the UHM libraries – that is, not available in the libraries' print collection - that were also listed in *Magazines for Libraries*. Academic Search proved to be the database of choice if the goal was to provide maximum access to new titles. Periodical Abstracts Research II, on the other hand, had a higher number of titles that were both new to the UHM Libraries and appeared in *Magazines for Libraries*, and therefore was assigned a higher quality ratio.

- The *abstract value* measured the number of titles owned by the UHM Libraries that were abstracted in the database but not available in full text. Again, Academic Search rated highest if the goal was providing the largest number of abstracted titles. However, Expanded Academic Index had a lower number of abstracted titles that were not owned by UHM. Expanded Academic Index, then, was assigned a higher qualitative value for indexing and abstracting a larger number of titles owned by UHM, "minimizing the frustration experienced by users attempting to retrieve articles not owned by UHM Libraries." (Brier 1999, 475)

- The *Interlibrary Loan Value* measured the number of titles abstracted in the database and listed in *Magazines for Libraries*, but not owned by UHM Libraries. The

database with the highest number of titles meeting these criteria was considered the most desirable for providing references to high-quality interlibrary loan material. While the evaluators did compile valuable data from this project, they also note that the values alone did not indicate a clear winner among the three databases. For example, "Periodical Abstracts Research II had the best full-text value, Expanded Academic Index the best abstract value, and both Periodical Abstracts Research II and Expanded Academic Index had the best ILL value for the UHM Libraries collection." (Brier 1999, 477) In addition, the authors note that other factors need consideration in deciding on a full text database, including the library's mission and the needs of the user community. While abstract value may be a useful measure, for example, it needs to be balanced against the frustration many users experience when they have become accustomed to retrieving the full text of an article.

Analysis of vendor policies and standards. Create a checklist of questions for vendors to determine how closely their product meets your needs. Questions should cover, but are not limited to, the topics listed below. Refer to "Evaluating Database Content and Features" for a comprehensive discussion of desirable features.

- Indexing standards and practices

- Subject authority

- Search and limiting features

- Extent and quality of online Help

- Customer service and technical support

- Patron authentication and remote access

- Pricing

In preparing for the database evaluation project described in Brennan (1999), reference librarians representing all of the libraries in the HELIN consortium held brainstorming sessions to compile a list of questions for vendors. The list was based on features considered necessary for a database, and also drew on areas of dissatisfaction with the existing database.

The list became the basis of Figure X, Vendor Responses, a checklist used to compare the features offered by the four vendors we were evaluating.

Figure X: Checklist of Vendor Responses

	Vendor A	Vendor B	Vendor C	Vendor D
Sources indexed (number of refereed and general titles)				
Journal list available?				
Start date of FT coverage				
Indexing staff and practices				
Subject authority				
Fields indexed				
Searchable fields				
Subject heading/descriptors displayed?				
Indexing criteria				
How soon are citations and full text loaded after receipt of issue?				
Limiting features				
Fields searched w/ keyword search				
Mark or tag feature?				
Context-sensitive Help?				
Email capability?				
Tech/customer support (email, toll-free number, 24/7, standard response time?)				
Support for remote access? IP? Proxy?				
Pricing (consortium, FTE)				

Extent of full-text coverage. The database evaluation project described in Brennan (1999) examined how closely the online version of a periodical title mirrored the hard copy. It was discovered that none of the four vendors selected for the evaluation provided cover-to-cover content in the electronic format. Letters to the editor, book reviews, and brief columns were among the components most frequently missing from the electronic version. Publisher embargoes and deleted articles from freelance authors, a result of the Tasini decision, can further constrain the amount of full text available online. (The Supreme Court ruling in *The New York Times v. Tasini* affirmed the copyright privileges of freelance authors whose works were originally published in periodicals and then licensed to commercial databases. For an overview of how this decision affects libraries, refer to http://www.arl.org/newsltr/217/tasini.html). A comparison of the online version to the hard copy is particularly critical if you are considering canceling the hard copy subscription when the electronic version becomes available.

The librarians who designed the 1999 database comparison project decided that a page-by-page comparison of a hard copy issue and its electronic counterpart was the only way to determine the extent of the database's full-text coverage. Each of the six libraries in the HELIN consortium selected five journals that their library subscribed to and were also available in full text in each of the vendor's databases. We designed a one-page worksheet, Paper vs. Online Journal Comparison (Figure X), to guide library staff in their comparisons. The questions on the worksheet combined objective ("What is the date of the most recent article found in this journal") and subjective ("If you were to stop receiving the paper copy of this journal, would this database adequately replace this journal?") analysis.

Figure X:
Database Review
Paper vs. Online Journal Comparison

I. Date/time of session _____

II. Vendor's Database used (circle one): Vendor A Vendor B Vendor C Vendor D

III. Title, date, and issue of journal (use current issue)

_____

IV. Questions to consider:

1. Does the online version have a Table of Contents (TOC)? Yes ___ No____

2. If Yes, how does the TOC compare to the paper version? Same? Major
    articles only?

    Remarks: _____

3. What is the date of the most recent online article? _____

Note: Can you determine the time elapsed between the time the paper articles are
published and they first appear online? _____

4. How is the journal organized and/or indexed? (i.e., by issue; by subject)

_____

5. Does the online version contain the "special features," such as editorials,
columns, or recurring sections that are found in the paper copy? _____

_____

6. Consider how the online version deals with full text articles:
    a) What are the Start dates for the following features:
        1) Citations/abstracts _____
        2) Full text _____
        3) Full text w/graphics _____
    Remarks: _____

    b) Are the online articles really the entire (full) text? Yes ___No__

    c) How are pictures/tables/graphics treated? _____

d) When compared to the paper copy, what percentage of the articles are:

Full text _____ %   Abstract _____ %   Not covered _____ %

V.   How easy was it to search for the specific journal and/or articles? _____

_____

VI.   Comment on the ease of use for each of the following:

Printing _____

Downloading _____

Emailing _____

VII.   Overall evaluation:

If you were to stop receiving the paper copy of this journal, would the online version be an adequate replacement? Specify Yes or No, and include comments as needed.

_____

_____

At the end of the trial period, the information gathered from the vendor responses, our comparison of the vendors, and feedback from the Journal Comparison worksheets were compiled into a Pro and Con table for each vendor.

Like Brier and Lebbin, we came to the conclusion that the hard data we had assembled was only a starting point for the complicated decision of selecting a database. Questions and considerations remained, such as How much full text is 'enough'? Is it too risky to drop a paper subscription and rely on an electronic backfile? If an otherwise quality database lacks a significant feature, such as a thesaurus or subject search capability, do we disqualify it? While a well-planned methodology yields invaluable information,

librarians need to consider usage patterns, patron preferences, accreditation standards,

and other intangibles before deciding on a product to purchase.

Conclusion

The advent of library materials in electronic format has required Collection Development Departments to create new policies to address the specifics of both the format and the content of electronic materials. In some cases guidelines applied to more traditional library materials may also apply to electronic materials. However, aggregator databases are a new breed of electronic materials, and policies for their collection must, necessarily differ to some extent. While collection policies may ask the same types of questions for aggregator databases, the means of acquiring answers to those questions can be quite different.   As technology continues to change and the library materials take on new formats, collection development personnel must continue to adapt their practices, forms, and procedures to address those changes. A static checklist of questions may no longer fit the bill for adequate evaluation of the library materials of the future.

References

*Collection Management in Academic Libraries.*1991. Edited by Jenkins, Clare and Morley, Mary. Burlington, VT: Gower.

Allison, Dee A., Beth McNeil, and Signe Swanson. 2000. Database selection: one size does not fit all. experience of the University of Nebraska. *College & Research Libraries* 61, no. 1:56-63.

American Library Association. Association of College and Research Libraries. College Libraries Section. 1989. *Collection Development Policies for College Libraries.* Vol. #11. Chicago: American Library Association.

Barber, David. 2000. Full-text Periodical Databases and Journal Archives: The User Experience. *Library Technology Reports* 36, no. 5:11.

Bates, Mary E. Now they sell it, now they don't. *EContent* 26, no. 4:46.

Brennan, Patricia B.M., Joanna M. Burkhardt, and Susan McMullen. 1999. What does electronic full-text really mean? A comparison of database vendors and what they deliver. experience of Rhode Island's Higher Education Library Information Network. *Reference Services Review* 27, no. 2:113-126.

Brier, David J., and Vickery K. Lebbin. 1999. Evaluating title coverage of full-text periodical databases. *The Journal of Academic Librarianship* 25, no. 6:473-478.

Brooks, Sam. July 2001. Integration of Information Resources and Collection Development Strategy. *Journal of Academic Librarianship* 27, no. 4:316 - 319.

Clayton, Peter and Gorman, G.E. 2001. *Managing Information Resources in LIbraries.* London: Library Association Publishing.

D'Arms, John H. 1998. Introduction. In *Wave of the present: the scholarly journal on the edge of the internet*   American Council of Learned Societies.

Fulton, Marsha L. 2003. Market research aggregators. *Searcher* 11, no. 2:10.

Gregory, Vicki. 2000. *Selecting and managing electronic resources: a how to do it manual.* New York: Neal-Schumann Publishers.

Holleman, Curt. 2000. Electronic resources: are basic criteria for the selectionn of materials changing? *Library Trends* 48, no. 4:694-710.

Kidd, Tony and Lyndsay Rees-Jones. 2000. *The Serials Management Handbook: A practical guide to print and electronic serials management.* London: Library Association Publishing.

Kovacs, Diane. 2000. *Building Electronic Library Collections: the essential guide to selection criteria and core subject collections.* New York: Neal-Schumann Publishers.

Lee, Stuart D. 2002. *Electronic Collection Development: a practical guide.* New York: Neal-Schumann Publishers.

Li, Yiu-On and Shirley Leung. 2001. Computer cataloguing of electronic journals in unstable aggregator databases: the Hong Kong Baptist University Library experience. *Library Resources and Technical Services* 45, no. 4:198 - 211.

Metz, Paul. 2000. Principles of selection for electronic resources. *Library Trends* 48, no. 4:711 - 728.

Nielsen, Jakob. Ten Usability Heuristics. 1994 [cited 2003, 10/28 Available from http://www.useit.com/papers/heuristic/heuristic_list.html.

Norman, O.G. Impact of electronic information sources on collection development: a survey of current practice. *Library Hi Tech* 5, no. 1-2:123 -132.

Notess, Greg R. RSS, aggregators, and reading the blog fantastic. *Online* 26, no. 6:52.

O'Malley, Michael and Roy Rosenzweig. 1997. Brave new world or blind alley? American history on the world wide web. *Journal of American History* 84, no. 1:132-155.

Reich, Vicky and David Rosenthal. 2000. *Making web publishing irreversible.* Edited by Donnelyn Curtis, Virginia M. Scheschy and and Tarango, Adolfo R. Stanford University ed. New York: Neal-Schuman.

Still, Julie, and Vibiana Kassabian. 1999. Selecting full-text undergraduate periodicals databases. *EContent* 22, no. 6:57.

Thornton, Glenda A. 2000. Impact of electronic resources on collection development, the roles of librarians and library consortia. *Library Trends* 48, no. 4:842 - 856.