

FROM Catalog TO Gateway

Briefings from the CFFC



Spring 2000
Volume 11, Number 1

Masthead



Aggregation or Aggravation? Optimizing Access to Full-Text Journals

Karen Calhoun and Bill Kara

Based on presentations at the CCS Catalog Management Discussion Group, January 1999, Philadelphia, Pennsylvania, and at the PCC Participants Meeting, June 1999, New Orleans, Louisiana, with updates.

Aggregations are collections of publications in electronic form. While aggregations come in all sizes and differ in content, all offer unprecedented access to online full-text journals for library users. Librarians must optimize end-user access to full text journals in aggregations, if they are to provide adequate service, and a catalog that fails to integrate print and electronic resources is no longer sufficient. The authors examine the benefits of aggregations, the challenges of describing aggregated full-text titles, and several alternatives for integrating print and electronic journal titles in library catalogs.

The Problem

“Does the library subscribe to *Academy of Management Executive*?” This is the sort of question that was once fairly straightforward, but no longer. At the time a library user posed this question at Cornell’s business school library, the online catalog indicated holdings from 1987 through 1989, and the user left the library shaking his head; he couldn’t believe the subscription wasn’t current. And guess what: the library **does** hold an unbroken run. Why wasn’t it plain to the user?—because the title was “buried” in an aggregator database, and the online catalog only reflected print holdings for this title.

This little scene is being played out in libraries all over the country. It represents a significant problem that must be solved if libraries are to provide adequate levels of service to their users. Many librarians are putting their best efforts into devising solutions, but, unfortunately, the solutions, while some have been ingenious, have for the most part been independent and uncoordinated.

The challenge of optimizing end-user access to aggregations of full-text electronic journals is turning librarians’ assumptions about the catalog’s scope and function upside down. Aggregations (such as SpringerLINK or Bell and Howell’s ProQuest

From the President

News & Events

Publications



ALCTS Links

Archive

Contents

- [The Problem](#)
- [What is an Aggregator Database?](#)
- [Why Aggregator Databases Are](#)

Direct) offer unprecedented opportunities to provide online access to large collections of full-text materials. At the same time, they demand that librarians decide how they will identify the journal titles that their libraries “hold.” Should each title within an aggregation be cataloged? If yes, can a library afford to catalog and maintain each title, when some aggregations contain thousands of titles? What are the options, what are the issues and challenges, who should do this work, and in what ways might libraries and vendors collaborate to minimize duplication of effort?

It is important for librarians to look at the issues surrounding aggregations and work together to resolve them, and there has been much interest and debate about the topic at library professional meetings. There are different solutions and pitfalls, but there certainly needs to be plentiful discussion, cooperation, and planning. Without such attention, each individual library may continue to tackle each aggregator database differently.

[Back to Table of Contents](#)

What Is an Aggregator Database?

In *Webster’s New Collegiate Dictionary*, aggregation is defined as:

1. a: the collecting of units or parts into a mass or whole
b: the condition of being so collected
2. a group, body, or mass composed of many distinct parts

This definition can apply to numerous items, including publications. Simply, when concerned with digital full-text materials, an aggregation is a collection of publications in electronic form (in this paper, the terms “aggregation” and “aggregator database” are used interchangeably). Aggregations come in all sizes and are remarkably different in content, not only in the titles contained in them, but how they’re organized and why they’ve been put together.

Some of the more common types of aggregations are those that are collections of titles by one publisher or based on a broad subject. Publisher-based aggregations are those in which all the journals in the collection are from one publisher. Subject-based aggregations are those that can include publications from numerous publishers but are related in that they share the same broad subject, whether business, medicine/health, law, literature, etc., or whether they’re meant to serve as a broad general periodicals collection. In addition, full-text publications can be accessed through vendors who have aggregated the journals of many publishers and make them available through their services. These might be the basic categories of aggregations, but there are variations, and each aggregation needs to be examined for its own content and structure.

Relatively simple, straightforward collections do exist. These are the “vanilla” aggregations. In this category there are aggregations with a known list of journal titles (e.g., Project Muse), which are stable and well maintained over time. All the titles have some common element, whether they are from the same publisher or cover the same broad subject, or both. For each of the journal titles available in the collection there are complete issues (or relatively complete issues with only minor, known differences from their print counterparts), which are accessible by both journal title and specific issue. There is a commitment from the aggregator to maintain the list of titles by adding new journal issues as they’re published. It is easy to identify the parts of the collection, relatively easy to catalog (since they closely correspond to a print counterpart), and relatively easy to maintain over time

[Here to Stay](#)

- [Aggregations Break the Rule](#)
- [Assumptions a “Right” Answer](#)
- [Title Lists](#)
- [Single Record Technique](#)
- [Separate Record Technique](#)
- [Conclusion](#)
- [References and Notes](#)

[Karen Calhoun](#) is Head of Cataloging, Central Technical Services, Cornell University L. She may be contacted at may contacted at (607) 255-9915.

[Bill Kara](#) is Head of Technical Services, Albert R. Mann Libr. Cornell University. He may be reached at (607) 255-7952.

since the aggregator maintains a stable journal title list and would notify the subscriber of changes to the collection.

On the opposite end of the scale are the “tutti frutti surprise” aggregations. These can be quite large and amorphous. Titles available one month might not be available six months later, the user perhaps cannot select a specific journal title and issue, and the full issue of a journal might not be available, but only its articles pertaining to the broad subject category of the collection. Aggregation content might include monographs, reference books, and pamphlets (or parts, but not necessarily the complete full text, of these many publications). These aggregations serve very different purposes and pose very different challenges. If they’re successful in providing a unique or value-enhanced resource to the collection, their content, whether clearly defined or broadly based, can become an integral part of a library’s online collections and services.

Between the extremes of the “vanilla” and “tutti frutti surprise”, there are countless other variations. One important difference is the size. A “vanilla” aggregation of forty titles might seem relatively easy to catalog and maintain, although an aggregation of two hundred, six hundred, or even more titles, although well maintained by the aggregator over time, might pose different challenges for providing detailed access to individual titles in the collection.

[Back to Table of Contents](#)

Why Aggregator Databases Are Here to Stay

Although in some cases it is quite aggravating to identify effectively the precise content and to maintain this information over time, aggregations do have real benefits. Those benefits will keep aggregator databases part of libraries’ online collections. Before the advent of aggregations, individual full-text titles had to be selected, subscribed to, paid for, and maintained. Selecting and adding individual titles to the collection are time-consuming tasks, not only for collection development and acquisitions, but for cataloging as well. For each individually selected title one would need to review and sign the contract, pay for that title separately, and renew that subscription annually until cancellation. An aggregation of a hundred journal titles from the same publisher (many of which were already selected for the library’s print collections) would have only one contract and one invoice for those hundred titles. And, once ready for cataloging, they are sent through as a batch, potentially sharing many of the same bibliographic elements. The information technology staff also handles the individual titles in the aggregation similarly since each would share the same scripting and security requirements. Public services then deals with only one interface for those hundred titles, and searching across the whole collection is a benefit that would not be available if subscribing and linking to only individual titles. Once in the system they could be reviewed for renewal as a batch.

Yes, it could be argued that because some aggregators require a buy-all-or-nothing policy, a library would acquire (and pay for) online journals it would have not otherwise selected. But accessing and pricing aggregations are still evolving. Most vendor-based collections permit more tailoring to a library’s unique needs and pricing limitations. For those aggregations that do not permit much tailoring of their lists, the economy of scale is often a factor. It may be less expensive to subscribe to a full list of titles than to subscribe individually to a significant portion, but not all titles, on the list. As with other cataloging decisions, each aggregation needs to be evaluated on its own merits, features, and cost effectiveness. The evolution of aggregations, with the development of new products and services, is still underway, but clearly the trend has been to favor aggregations over individual subscriptions to

full-text titles. The challenge is for a library to identify the best collections and services for its users that are both affordable and could be maintained.

Finally, and most importantly, library users increasingly want and expect access to full-text materials. Aggregator databases are an opportunity and a means of building these collections quickly and effectively.

[Back to Table of Contents](#)

Aggregations Break the Rules

Rules are for the obedience of fools and the guidance of wise men.—David Ogilvy¹

Aggregator databases are forcing librarians to color outside the lines. Unlike the large sets previously purchased by libraries, like microforms, an aggregation's full-text content typically includes high-demand, core titles that enable users to get information when they want it, where they are, with a few mouse clicks. For some library users, aggregator databases render library online catalogs just one more database among many available to users. At the same time, users want and expect the catalog to reflect all the resources to which they have access. To meet user expectations, provide good service, and make the best use of the funds being expended on aggregations, it is necessary to reexamine the catalog in *convergence* with the collection. An online catalog that is a self-contained information system providing access to a well-defined physical collection is no longer sufficient.

In 1997 Younger wrote, "the continuing evolution of Internet services has changed forever the landscape of document delivery and retrieval. . . . [The] functionality [of Internet services] highlights the interconnectedness of individual catalogs, databases, and search engines and, not surprisingly, many activities are directed toward creating a more coherent global system."² Similarly, Van de Sompel and Hochstenbach, writing in 1999 of the necessity of creating services that link related information entities—like library catalogs and full-text databases—noted "the expectations of a Net traveler are inspired by his hyperlinked Web experiences. To such a user, it is not comprehensible that secondary sources, catalogues and primary sources, that are logically related, are not functionally linked."³

Nevertheless, even if one is convinced that the catalog should point to the full-text journals in aggregator databases, knowing how to proceed is far from clear. First, some aggregations (like Lexis Nexis' Academic Universe) offer access to the articles in full-text journals, but they must be retrieved by subject searching; accessing the contents page of an individual issue is not possible. Questions arise in this case: Can the library be said to "hold" these journals? Are what such aggregators provide *journals* at all, as we currently define them? Second, the collection of full-text titles in an aggregation can be ephemeral. Subscriptions come and go, titles come and go, and on top of that, electronic locations can change, disappear, and be renamed. How will the library cope with the need for constant adjustment of links and records? Third, some libraries, faced with large collections of unarchived, licensed titles (previously available on CD-ROM, and now available remotely) have established policies that direct scarce cataloging resources only to what is actually part of the permanent collection. If that policy is changed, how can the effort be staffed? Fourth, the cataloging rules and practices for Web-accessible resources are evolving and far from stable. Catalogers cannot rely on a consensus of best practices, a stable body of rules, and many years of experience in doing this kind of work. Fifth, cataloging, especially serials cataloging, is time- and labor-intensive. Given the size of many aggregations (thousands of full-text titles) and the

speed with which links from the catalog are wanted (yesterday!), how can catalogers possibly accomplish the work quickly enough?

In summary, the presence of and growing demand for full-text aggregator databases are critical factors driving future cataloging decisions. And, instead of exerting pressure for incremental, evolutionary change in the nature of the catalog, aggregator databases demand in-your-face, right-now substantive changes in cataloging policy and practice.

[Back to Table of Contents](#)

Assumptions and “Right” Answers

Conventional cataloging could solve the problem of aggregations, but most of today’s resource-strapped cataloging departments do not have the means to do it. A good strategy, when a problem appears to have only one (impossible or unappealing) right answer, is to reformulate the problem. In doing so, a second, third, or fourth right answer can sometimes emerge. A first step in searching for alternative solutions is to establish a set of working assumptions to guide decision making. In the following analysis, the working assumptions are:

1. Records for full-text titles are surrogates that, functionally speaking, should permit a searcher to find, identify, select, and access the resources themselves.
2. Print and e-resources should be integrated in library catalogs.
3. Third, any solution to the problem of aggregator databases should be a solution that can be widely shared among libraries.
4. Fourth, the solution must be timely and affordable.

The next sections of this paper are an evaluation, based on the four assumptions laid out here, of several alternatives to providing access to full-text titles in aggregator databases.

[Back to Table of Contents](#)

Title Lists

Many libraries mount lists of online journal and newspaper titles on their Web sites to identify the journals to which they provide networked access. The links on such lists serve as surrogates, as defined above. Certainly, using a list and cataloging full-text titles for the catalog aren’t mutually exclusive; some libraries do both. Usually, Web title lists are “static”—that is, the contents of the list aren’t dynamically generated from the real-time contents of a database. Instead, a person prepares the list from what is known at a certain point in time. Lists can be searchable, browsable, or both.

From the perspective of the first assumption—that surrogates should permit users to find and access the resources themselves—title lists can do the job, provided the searcher knows to look for them. In practice, lists alone are ineffective discovery and retrieval tools, especially if the library provides access to more than a few titles. Title lists fail to satisfy the conditions of the second assumption—that print and e-resources should be integrated in the catalog. They also fail to meet the third requirement—lists are difficult to share with other libraries. From the perspective of the fourth assumption—that the solution be affordable and timely—lists are

relatively fast and easy to put up, although maintaining them requires constant effort.

[Back to Table of Contents](#)

Single Record Technique

The single record technique, an option endorsed by CONSER and described in Module 31

(<http://web.archive.org/web/20010620162004/http://lcweb.loc.gov/acq/conser/mod31pt1.html#interim>)

of the *CONSER Cataloging Manual (CCM)*, gathers information about a journal title and the library's holdings together on a single record, if the library owns a hard copy version and provides access to an electronic version. Many have argued that this approach is the best for providing index and record displays that users can readily understand. It certainly reduces the number of records needed in the online catalog for a given journal title. For an example of the single record technique, examine the record for *ARC News* (Redlands, Calif.) ([example 31.20.4](#) in part 2 of *CCM* Module 31).

From the perspective of the first assumption, the single record technique provides a useful means for the user to find, identify, select, and access full text online, although the result of the technique can be an emphasis, in catalog and index displays, on the details of the print version and holdings. As for the second assumption, the single record technique nicely integrates print and e-resource information in the catalog.

From the perspective of the second two assumptions, the single record approach is not as successful. Measured for its effectiveness as a cooperative (shared) solution, each library's editing its own records, in its own local library system, is duplicative and wasteful. Even if the serial records are edited centrally, in the CONSER database, not all libraries hold print equivalents of the journals they can access via aggregations, and libraries subscribe to different aggregations. Therefore, creating sets of records for wide sharing and distribution, using the single record technique, is problematic.

From the vantage point of timeliness and affordability, multiple version records add to the complexity of cataloger decision making. Based on Cornell's experience applying the single record technique to the nearly two thousand full-text titles in ProQuest Direct, the person applying the single record technique to a group of titles needs to have a good grasp of serials cataloging. It was necessary to conduct extensive quality reviews on the work of students and some clerical staff and to allocate considerable time from highly trained staff to manage the many exceptions to standard processing. In sum, the need for highly trained staff was costly, and resolving the many exceptions slowed down the work. In addition, now that the information about multiple versions is woven together in single records, a record-by-record approach to maintenance is a necessity. If an automated batch solution to maintenance of the ProQuest Direct titles can be found at all, the programming is likely to be complex.

[Back to Table of Contents](#)

Separate Record Technique

CONSER guidelines also provide the option of the separate record technique for

cataloging electronic journals with print equivalents. When e-versions of print journals first began appearing on the cataloging scene some years ago, almost no cataloging copy for e-versions cataloged as separate records existed in the utilities; thus a choice to create separate records equaled a choice to do original cataloging, and as a result many libraries opted for the single record technique. Over the last few years, many more separate records have appeared in OCLC, thanks in part to OhioLink's contracting with OCLC's contract cataloging service to prepare records for the titles in several aggregator databases licensed by OhioLink. More are beginning to appear now, as OCLC solicits libraries to prepare sets of separate records for its WorldCat Collections Sets (for more information, see <http://web.archive.org/web/20010620162004/http://www.oclc.org/oclc/colset/index.htm>). An example of a recent project, being undertaken by the University of Illinois at Chicago, is for H. W. Wilson Select. [Figure 1](#) provides an example of a project record for the H. W. Wilson Select title *Accounting Horizons*.

From the perspective of the first assumption, separate records permit searchers to find, identify, select, and access full-text titles in aggregator databases. Arguably, they do this as well as single (multiple version) records. Since they describe the e-version separately, they allow more information about the e-version to be displayed in the body of the record and at the index level of the OPAC. For example, if the single record technique had been applied to *Accounting Horizons*, the user would not see the general material designation "computer file" at the title index level of the catalog or in the individual record display.

Nevertheless, when separate records are prepared by catalogers, title-by-title, according to current bibliographic input standards, they tend to be very full records, and expensive to create. The input standards have developed in a world in which the surrogate (i.e., the catalog record) must provide enough information to allow the searcher to decide to take the further step of examining the work itself. One must ask, given that the electronic resource being described is usually one click away, is it really necessary to record all the attributes and relationships that are usually recorded in a catalog record? How much is enough?

Some librarians have created computer programs to derive brief separate records from other available data about the full-text journals available in several aggregator databases. At a panel session during the 1999 ACRL national conference, David Atkins and Flora Shrode reported on a project at University of Tennessee at Knoxville to automatically create, import, and maintain full-text journal title records in the UTK online catalog.⁴ [Figure 2](#) is an example of a record created by the UTK library's Perl scripts and the utility MarcMakr (available for download at <http://web.archive.org/web/20010620162004/http://lcweb.loc.gov/marc/marcsoft.html>). While the UTK records do not conform to current bibliographic input standards for manually crafted records, when one of the authors questioned Atkins at the ACRL conference, Atkins reported a high level of satisfaction among library users with the solution UTK has provided.

The PCC Task Force on Journals in Aggregator Databases' 1999 interim report provides details about its collaboration with EBSCO to define and implement specifications for machine-derived separate records.⁵ A later report in CONSERline provides an update.⁶ The project focused on the titles in EBSCO's Academic Search Elite aggregation. Separate records for the full-text journals in the aggregation were created programmatically from existing records for the print journals. Last summer, the set of about 1,100 brief MARC records was made available at no charge to EBSCO subscribers. [Figure 3](#) is an example of a test record created in the project. The task force is now working with other vendors to encourage them to produce MARC record sets for their aggregations and make them available to their subscribers.

Separate records satisfy the conditions of the second assumption, that print and e-

resources be integrated in library catalogs. One disadvantage of using the separate record technique, however, is the resulting complexity of index displays (at the index level, the user is faced with multiple choices for the same title). This situation could become quite unwieldy in a large research library, which typically licenses multiple aggregations, so that the index could display a confusing array of print and e-versions of the same title to the user.

From the perspective of the third and fourth assumptions—that solutions be shared, timely, and affordable—the separate record technique is a good fit. As is being demonstrated through the projects mentioned in this paper and elsewhere, under the right conditions, records can be economically created and made widely available as sets. However, it's important to note that given the volatility of aggregator databases, such sets need to be constantly and consistently maintained, and it makes more sense for maintenance to be centralized or shared, rather than repeated by every library holding a copy of the set.

[Back to Table of Contents](#)

Conclusion

Great progress has been made in the past year toward optimizing access to full-text journals in aggregations. Conversations about aggregator databases among selectors, acquisitions staff, catalogers, and information technology and reference staff have yielded a broad-based perspective on user expectations, what constitutes good service, and the function of the catalog in 21st century libraries. A spirited public dialogue among catalogers at numerous professional meetings has resulted in a greater understanding of the issues, if not consensus. Catalogers' efforts have advanced online catalogs closer to being "one-stop shops" for both print and aggregator-provided full-text journals. Librarians are devising, engineering and negotiating collaborative, cost-effective and timely solutions that promise to enhance online catalogs and to make the most of the millions of dollars that libraries invest each year in aggregator databases.

For all their ingenuity, however, the recent efforts to produce and share sets of separate records for journals in aggregations represent a pragmatic response to an immediate problem, a kind of finger in the dike. In the broader context of organizing Internet resources, publishers, vendors, researchers, and libraries are all striving to find ways to link related information entities of all types—links from citation databases to catalogs and full text, links from citations in scholarly articles to other scholarly articles, links from catalog records to finding aids and primary sources, links to book reviews and images, and more. Library and information science researchers are working on more elegant solutions to "reference linking," or a system of links from one information object to another.

In summer 1999, Caplan and Arms reported on the progress of reference linking during the past year.⁷ The focus of reference linking is often on linking citations to journal articles, but it can be extended to any context in which metadata points to an object stored elsewhere. An exciting development was announced in fall 1999; twelve publishers are collaborating on a reference linking initiative, expected to launch during the first quarter of 2000.⁸

Long-term efforts to provide access to the full-text titles in aggregator databases promise to integrate today's somewhat ad hoc, pragmatic solutions with more robust, standards-based, interoperable systems. For that to happen, it is essential that librarians proactively collaborate with information technologists, publishers, vendors, computer scientists, and digital library specialists to share what they know,

in order that everyone involved understands each other's concerns and methods, and develops shared solutions.

[Back to Table of Contents](#)

References and Notes

1. Ogilvy, David. *Confessions of an Advertising Man* (New York: Ballantine), 1971). Quoted in Roger Von Oech, *A Whack on the Side of the Head* (New York: Warner Books, 1983), 73.
[back to article](#)
2. Younger, Jennifer A. "Resources Description in the Digital Age." *Library Trends* 45, no. 3 (Winter 1997): 462-87.
[back to article](#)
3. Van de Sompel, Herbert and Patrick Hockstenbach. "Reference Linking in a Hybrid Library Environment: Part 1, Frameworks for Linking." *D-Lib Magazine* 5, no. 4 (April 1999). Available:
www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt1.html
[back to article](#)
4. Atkins, David. "Expanding Periodical Holdings: University of Tennessee's Case Study in Breadth." Presentation at the panel session, "The Bleeding Edge of Access to Full-text Electronic Information: Three Case Studies in Science and Technology." ACRL Ninth National Conference, 1999.
[back to article](#)
5. Calhoun, Karen and John Riemer. "Interim Report: PCC Standing Committee on Automation (SCA), Task Group on Journals in Aggregator Databases." May 1999. Available:
<http://web.archive.org/web/20010620162004/http://lcweb.loc.gov/catdir/pcc/aggupdrpt.html>
[back to article](#)
6. "Progress Continues for Aggregators Task Group." *CONSERline* 14 (Summer 1999). Available:
<http://web.archive.org/web/20010620162004/http://lcweb.loc.gov/acq/conser/consln14.html#aggregator>
[back to article](#)
7. Caplan, Priscilla and William Y. Arms. "Reference Linking for Journal Articles." *D-Lib Magazine* 5, no. 7/8 (July/August 1999). Available:
<http://web.archive.org/web/20010620162004/http://www.dlib.org/dlib/july99/caplan/07caplan.html>
[back to article](#)
8. "Reference Linking Service to Aid Scientists Conducting Online Research." Press release from Elsevier Science. *Newsletter on Serials Pricing Issues*. 240 (November 29, 1999) [electronic discussion list]. Back issues available:
<http://web.archive.org/web/20010620162004/http://www.lib.unc.edu/prices>
(*Editor's note: Issue no. 240 of the Newsletter on Serials Pricing Issues was still too recent to be posted at the URL cited above at the time this issue of ALCTS Newsletter went online. If you do not find 240 posted with back issues at this site, please be aware that it will appear eventually.*)

[*back to article*](#)

[Back to Table of Contents](#)