September 2002

Aggregation, Integration and the Library Without Walls?

Adam Hodgkin

xrefer

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

Part of the [Library and Information Science Commons](#)

Recommended Citation

Hodgkin, Adam (2002) "Aggregation, Integration and the Library Without Walls?," *Against the Grain*: Vol. 14: Iss. 4, Article 8.
DOI: <https://doi.org/10.7771/2380-176X.5452>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

of developing a new budget model for our collections. In the past, the reference budget included many subject resources: indexes and abstracts, specialized dictionaries; statistical sources. We are trying to move those subject resources out of the reference budget and into their appropriate subject area, to get a better handle on how well we are supporting those disciplines. Certainly, if one looked at indexing and abstracting sources, I would estimate more than 85% are now online only. More general reference sources online would be a much smaller percentage of the budget."

Linda Keiter and Margaret Landesman, University of Utah: "We are now spending substantially more on electronic than on print reference sources. Because the electronic materials are funded on different accounts and reported in different categories, we don't have very accurate comparisons with print expenditures. But we would note that our library is paying more for just **FirstSearch** and **Academic Universe** than we are spending for print reference titles.

"Can we afford to purchase both? This isn't a new dilemma. Research libraries have always resisted buying the same information in more than one format. And we've always done it too. The old maxim of never buy more than one copy or more than one format maximized the number of titles available to users. If it's really expensive to buy both, we would push ourselves to make some kind of decision. But if the price is affordable, and we believe that we have users who want the print and also users who want the electronic, then we probably would buy both. What we can't afford to do is to buy everything we bought in print both in print and electronically. What we can't afford to do is to buy things people don't use."

Rumors from page 16

in Korea — sign up. Today there are thousands of users in more than 50 countries, across a diverse range of markets." Remember the article — "The Development of the First PDA Library Portal for Academic Libraries" (see *ATG*, June, 2002 p. 65) by **Michael Binder** (Dean of Libraries, **Western Kentucky University**) discussed **AvantGo** software. Apparently the same free software is used with **MobileEditions**. <http://www.wiley.com/>.

And speaking of articles in *Against the Grain*, did you read the one by **Niko Pfund** — "Technology and Change in the Library Marketplace," November 2001 (v.13#5, pp.22, 24). Anyway, **Niko** has another very incisive article in the June 18, 2002, *Chronicle of Higher Education*, "University Presses Aren't Endangered..." **Peter Givler** from the **American Association of University Presses** will speak in **Charleston** (thanks to **Lyman Newlin**)

continued on page 38

Aggregation, Integration and the Library Without Walls?

by **Adam Hodgkin** (xrefer) www.xreferplus.com

Content aggregators are playing an increasingly prominent role in delivering electronic subscription resources to libraries. It is timely to consider some of the advantages that can be conferred by aggregation but also some of the potential limits in the aggregator role.

The typical aggregation service will collect together in one on-line electronic resource copyrights licensed from a variety of publishers or authors. But aggregators do more than collect content from a variety of sources; these resources will be delivered to end-users through a common interface, they will usually be supported by a single search service and they will be part of a single delivery mechanism. There is a surprising degree of variety in the commercial terms offered in different areas, but increasingly aggregators are moving towards forms of subscription publishing. I work for **xrefer** which is an aggregator of "reference resources" so I will tend to draw examples from this field; but aggregators are operating in a number of different areas of the market: news stories, financial data, eBooks, eJournals etc. There are plenty of areas in which content from disparate sources needs to be aggregated if it is to be well used on the Web.

From the librarians' or the end users' point of view, I would suggest that there are three main advantages in having an aggregator deliver content resources into the library:

- All the resources are presented through a common user interface and style of presentation (this might be as simple as putting all the documents into **Adobe Acrobat** format).
- The resources share a common location (which for the end user means finding a large variety of similar resources within one service).
- The resources are offered through a common license, subscription package and support function.

The first point is hugely important and is directly driven by the importance of Web standards. Any successful Web aggregation service has to be true to the proven success of established Web services (**Yahoo**, **Google** etc). This provides a very important benefit for Web users. Think back to CD-ROM and the first generation of "pre-web" online services. The first and second generation of electronic publishing (on-line from a proprietary dialup system and CD-ROM) gave the information providers latitude to invent their own interface for each electronic publication. This predictably led to a confusion of incompat-

ible systems, interfaces, commands and functions. Whatever the merits or demerits of any particular solution or interface, the end-user was in the awkward position of having to learn and "forget" scores of different and incompatible approaches. The enormous pressure of Web standards has helped to stop rival software producers from re-inventing the wheel with each new publication. The Web has encouraged aggregators and it has encouraged them to keep it simple and to limit innovation.

There is however an important role for aggregators who can find ways of adding value without adding needless complication. For example: the Web services which five years ago began to aggregate financial information services soon realised that there was great added value for users in providing automatic and customisable graphing features. At **xrefer** we saw that a consistent model for aggregating reference works would allow us to create an additional type of metadata, xreferences, which provides linkage between reference works created by different compilers. A user who consults **xrefer's** services (e.g., **xreferplus**) is able to browse between the different titles using additional meta-references which are generated by **xrefer's** software and database system. For the user this means that browsing a collection of reference works can be an experience in which intimate connections between entries in different books are revealed.

The very scale and global nature of the Web has encouraged aggregators to focus their efforts within specific domains and to aggregate content sources which might otherwise be thought of as competitive. Thus one sees on the Web a tendency for content of a similar type to cluster within a service rather than being distributed between services. One might call this the "Chinatown" phenomenon. Just as Chinese restaurants frequently cluster within the same street, so content aggregators tend to offer groupings of content, which are both competing and complementary (one aggregator concentrates on newspapers, another on scientific journals, and a third on reference material). I suspect that the reason for this is just the same as the reason that Chinese restaurants tend to cluster together. It is easier for the market to find a good meal if there are several restaurants in close proximity. It is easier for the user to find appropriate content if the competing resources can be expected to be in the same general "location." This is of course a point in which aggregators tend to think very dif-

continued on page 28

ferently from publishers. An aggregator is bound to look for comprehensive services. A publisher in building a list is apt to be careful not to compete directly with his own most successful offerings: a publisher who has established a spectacularly successful journal in a specific field will be careful not to compete directly with that publication, and a publisher of encyclopedias will be content to have one successful single volume encyclopedia, though he may consider developing variants for specific niches.

Finally, one should not underestimate the importance of the aggregators' role in establishing a broad and inclusive commercial structure within which libraries and their users can be assured of a consistent and regular service. It is easier to write one cheque and read one contract than to read 12 contracts and sign 12 cheques. If libraries or consortia of libraries had to negotiate their own terms for copyright clearance, database provision, and 24X7 delivery, users would have much longer to wait before they could have access to a broad range of electronic resources. From the librarian's point of view it is a major attraction that a good aggregator can be trusted to deliver not one newspaper or reference resource to the library's patrons but 5, 10 or 30. The administrative overhead in acquiring the resource(s) or renewing them is much, much simpler than would be involved in dealing with 5, 10 or 20 different licenses and subscriptions.

In reviewing the benefits of aggregation one is inclined to say that for librarians and end-users the benefits are essentially benefits of standardisation. The aggregators are really fulfilling an obvious role that has been created by the prevalence of the Web, and it is the open Web which leads us to expect resources to be highly "findable" and "commonly searchable." The benefits of standardisation are considerable, but there are some problems with the drive to ever more inclusive services. One problem is that for all the wonders of the Web, for all the excellence of **Google**, for all the comprehensiveness of **Yahoo** and the **Open Directory**, a communication system dependent on the complete open-ness of the Web inevitably contains masses of spam, deliberate error, egregious self-advertisement, unreliable opinion and sheer crackpottery. From the user's point of view the information available on the open Web is often too much, too weird, too unreliable or what can be worse of uncertain reliability. From the content point of view, it can appear that the standards to which the Web aspires are the standards of the lowest common denominator.

A publisher who considers the advantages

of aggregation is not likely to focus on the simple advantages of standardisation. Standardisation will be seen as a possibly dangerous issue – especially if it leads to complete open-ness and abdication of a commercial model. What are the advantages of aggregation from the publisher's stand-point? The answer to this question seems to me to depend essentially on the issue whether or not the publisher is doing the aggregation himself.

If the aggregator is a service which is supplied by an independent business (**EBSCO**, **Ovid**, **Ingenta** or **xrefer** for example) the publisher is likely to focus rather closely on the issue of the advances, fees, royalty or profit-share that will be derived from the deal. Since the deal will certainly be a "non-exclusive" arrangement, the publisher will be predisposed to do the deal if he is sure that it will not damage existing profitable lines of business and if he has no ambition to establish his own "exclusive" aggregation business. So why and to what degree will a publisher be tempted to follow the aggregation model? There are at least two compelling and attractive reasons for developing an aggregation service:

- There are considerable economies of scale in developing an aggregated service. When we started **xrefer**, we estimated that it would cost us at least \$75,000 to develop a reference service using one substantial reference work. As it happens, we intended to benefit from scale from the beginning. So we launched our first free service with a total of 20 texts at a combined development cost of roughly \$500,000. But as we added more resources the cost per new resource has fallen towards \$4,000 per title. Our costs are unlikely to fall much beneath this (each title requires detailed analysis and individual treatment before it is added to the system). Nor should one underestimate the requirement for ongoing investment in a system that can handle hundreds of complex texts in a consistent manner. Once an aggregator has developed an efficient production and development system it should be much cheaper to add content at the margin than to build the system again, or start afresh.
- In the long run an aggregators' investment in developing a content management and database system will be dwarfed by the investment in marketing supporting and delivering the service. The economies of scale are much greater on the marketing and delivery side of the business. A reasonable Web

marketing and delivery service is unlikely to cost less than \$1M dollars per annum (and this is a "minimum," since the marketing and promotion cost could easily reach \$1M a year, just maintaining a bare bones delivery and support service with guaranteed 24X7 service and a modicum of hardware redundancy could also cost \$300,000 per annum).

These advantages are essentially benefits of scale. But as our estimates of the costs of achieving scale indicate, the investment can be quite considerable. Furthermore the risks are somewhat open-ended. A publisher contemplating his own aggregation strategy will be undertaking a new kind of business and the investment costs are likely to rise. A non-exclusive licence to a third party can simply be terminated if it is not profitable. But a publisher who embarks on his own aggregation strategy is likely to be wary of developing a full service which may cost \$1 million in the short term, but could become a much larger and more unmanageable investment if it is moderately successful. Furthermore, most publishers will recognise that they are not in a sufficiently dominant position that they can create a comprehensive aggregation service from their own material. Most of the major national and international newspapers have outsourced the delivery of their backfiles to independent aggregators. In the field of scientific journals, only **Elsevier** and some of the specialist societies have clearly committed themselves to a proprietary aggregation platform (there are other proprietary platforms – but are we sure that they will still be there in 5 years?). In the reference field a few publishers have also embarked on a proprietary aggregation strategy: **Gale** and **Oxford University Press** are two examples.

It is an interesting corollary of the cost and uncertainty involved in developing a proprietary aggregation strategy that a publisher who does this is unlikely to be in a strong position either to take resources from another publisher on a non-exclusive basis or to offer his own publications on a non-exclusive basis to another aggregator. The proprietary publisher/aggregator will be reluctant to subsidise a third party publisher, so the royalties offered will probably not appear generous to the independent publisher. He will also be reluctant to grant non-exclusive licenses to aggregators who may be potentially competitive. If publisher/aggregators such as **Elsevier** in the journals field or **Oxford** in the reference field do decide to start licensing in resources from other publishers, one might expect two things to happen:

- They will need to pay significant royalties to copyright holders sufficient to attract independent publishers.
- They will become more open to the idea of licensing their own aggregated resources to other aggregators.

continued on page 30

If/when this happens, it can be taken as a sign that their own aggregation strategies are working! Publisher-based aggregation strategies face a challenge. They cannot afford to become isolated islands. They will only be successful in the long run if they are seen to interoperate efficiently with other Web services: which must include other and rival aggregation strategies. It will be interesting to see how this situation evolves.

Limitations of Aggregation

Aggregation helps the librarian and end-user directly through accelerating a process of standardisation in the provision of electronic resources; and indirectly through economies of scale that encourage the publisher and the aggregator to add more content to the service. But there are some potential dangers. By and large librarians are not overly concerned by the thought that aggregators are to some extent invading the traditional function of the librarian (perhaps they are simply pleased that some aspects of their traditional function can be outsourced). But librarians are very wary of becoming hostage to a quasi-monopoly supplier who may be tempted to increase subscription prices once a significant market has been established. The scientific serials market is a particularly painful precedent or warning. University librarians are particularly sensitive to this possible development.

From the end-users and the librarian's point of view one of the most important safeguards here is the insistence that, wherever possible, aggregators and other information suppliers should be encouraged and required to design services which are not "closed content boxes." If the advantages of aggregation are to be achieved there may inevitably need to be a degree of concentration in each particular information field (ejournals, eBooks, newspaper backfiles, reference services etc.) but the competitive situation will remain healthy if the suppliers are required to develop and sustain models of co-operation and interoperability. The **CrossRefer** initiative in the ejournals field is an important example of what can be done. This has been a sensible and efficient (albeit partial) solution to an obvious need: that users should be able to seamlessly navigate between electronic citations in different journals.

As an aggregator of reference material one is especially alive to the importance of avoiding "closed content boxes." There are reasons for thinking that reference works are especially *unsuited* to being delivered to users in closed aggregations. The reasons for this flow from two of the defining features of "reference" titles:

- Reference books tend to be episodically consulted rather than read from cover to cover (incidentally this is the

reason that reference works are not well-suited to eBook formats of the PDF or Adobe Acrobat type).

- Reference works tend to be consulted whilst users are engaged with other (often "non-reference") library resources.

If we put these two points together we might say: reference works are consulted episodically and these episodes are punctuated by moments when other works are the focus of the user's attention.

Reference books tend to be *consulted* or *browsed* or *referred* to when a user is *engaged* with *other material* (primary resources or secondary resources, ie periodicals or textbooks etc). Reference resources are needed to help us understand and navigate the electronic library and they are only truly useful and valuable if they can be immediately available as and when they are relevant. This puts a great deal of responsibility on the librarian and the aggregator to ensure that reference resources are well-integrated with other resources in the electronic library. Having spent a lot of time at **xrefer** figuring out a consistent way of developing a relevance linking framework for reference material, we now sense that the big challenge we face as a reference aggregator is this: it is not enough for the interlinked reference resource to be available in the electronic library. It needs to be immediately available to the reader in the most intelligent, relevant and discriminate form. It needs to be available through the appropriate Web linkage services. When a student reader is learning about the **Punic Wars**, they should be able instantly to "click

through" to reference items about **Hannibal**, **Carthage**, **Scipio** and to historical maps of the battle of **Cannae**. When a Ph.D. is looking at an abstract of an article titled "*Effects of intrapartum zidovudine therapy on fetal heart rate parameters in women with human immunodeficiency virus infection*" she should be able to select from a dozen of the most relevant entries with a bearing on this publication: a definition of HIV, and of fetal heart rate, a short overview of the toxicity of zidovudine, etc. Making this happen may not be easy, but it should be do-able.

In conclusion: aggregation helps. But it is a first step. Having aggregated resources into a convenient Web service it is important that these Web services are mobilised to become as useful as possible for their users and consumers. Libraries and librarians will play a vital part in this. If this is right, we conclude that elibraries should be libraries without walls in a double sense:

- The elibrary loses its walls because users can consult the library from any place/time, provided they have a Web connection and authorisation to use the resource.
- The elibrary will also inevitably dissolve some of the internal walls which keep content in closed content boxes. The constituent services of an electronic library necessitate inter-connection and interoperation.

The library without walls should become a library that encourages the maximum deployment of the resources it *provides*, but no longer *contains* — since the walls have come down.

Online Auction Services Orders: An Order Librarian's Reflections on OLAS

by **Nick J. Vincelli** (Order Librarian, Davis Library, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514-8890; Phone: 919-962-1120; Fax: 919-962-4450) <vincelli@email.unc.edu> www.lib.unc.edu/

Beginning in October 2000, the **UNC-Chapel Hill Academic Affairs Library** began acquiring material sold through **eBay.com** auctions. Although the Library had acquired material through auctions in the past (through auction houses such as **Sotheby's** and **Christie's**), obtaining material from **eBay.com** established a precedent for the Library, engendering a new workflow that was eventually named **Online Auction Services (OLAS)**. As the Academic Affairs Library's Order Librarian, I registered

with **eBay.com**, often placed the bids (under guidance from the selectors) and coordinated the creation of the firm order record, payment to the seller and receipt of the selected item. **OLAS** has proven a challenging, rewarding, fascinating and sometimes frustrating experience for me and other staff members.

This article will provide a narrative of our experience with **OLAS**, as well as a description of the guidelines we developed to ensure biblio-

continued on page 32