The importance of aggregators

Simon Inger
CatchWord Ltd., UK

Abstract. Aggregators take on several forms in the electronic publishing arena, and having a single collective name for all of these forms has caused some significant confusion to publishers and librarians alike. This paper will show the distinctiveness of each type of aggregator today and why each one brings benefits to publishers of all sizes and to libraries and individuals throughout the world. Aggregators are a life-line for the small publisher in the electronic age.

Introduction

It seemed strange to me that, when I was invited to write this paper for the second ICSU conference on electronic publishing, the organisers wanted to categorise it amongst papers on secondary publishing. Much of what I am to say seems so close to the core businesses of today’s journal publishers that one tends to consider it to be primary publishing. But not very long ago the only kind of aggregator in existence was the kind that licensed full text content from primary publishers and sold it as a collection to libraries and researchers - a truly secondary publishing function.

In reassessing the role of aggregator one needs to look once again at the basic concept of aggregations and uncloak some of the varied business models now somewhat misleadingly known as aggregation models. The term “aggregation” has become too widely deployed in the electronic publishing sector.

Is It an Aggregator?

So what is this aggregator aggregating? To the user the term aggregator is taken to mean the aggregation of full text content. But the companies who are collectively termed aggregators today range from those who aggregate full text on a selective basis, organised by subject, to those who simply provide a non-selective hosting service for full text publishers, to those who aggregate abstracts and metadata. More sublimely, there are those who just aggregate links to full text. To merely group all of these functions within the industry into one collective term has been the cause of some considerable confusion.

In the beginning, back in the days when compact disk aggregations of material appeared to be the answer to every collection-developer’s dreams, the new term “aggregator” meant just one thing. That was, a company that licensed content from primary publishers and sold it as a collection to libraries and researchers - a truly secondary publishing function.

A Closer Definition of Aggregators

There are three clear, distinct classes of company that have become “aggregators” in this new world. Firstly there are those companies whose primary focus is to provide a hosting service for publishers – the content host. Secondly there are those who index or categorise disparate content on other content host services – the gateways. And lastly the “traditional” aggregators of licensed full text content – the full-text aggregators.
The Role of Content Hosts

The role of the content host should be primarily a service to publishers. This is the role occupied by CatchWord, Highwire Press, Allen Press, American Institute of Physics, the hosting services of Ingenta and so on.

The fundamental business model of a full text hosting service provider dictates that its primary revenue stream will come from the services it provides to publishers. Primary publishers simply pay content hosts for the services that they need. This means that, in general, these companies cannot afford to be selective about the subject area (or indeed quality) of the content that they host.

The exception to this is Highwire Press. They have made a success of cornering one subject area but surrounded it in mantel of exclusivity that made many publishers believe that it was the only sensible place to host their content. This fact, coupled with the apparent exclusive nature of its clients, limited to society and university presses, has further enabled the host to concentrate on the subject areas that it wished to target and not those which target it.

The lack of selectivity of the majority of content hosts has not impinged on their success. On the contrary, their size alone warrants that gateway services and libraries alike place these organisations high on their target lists of services to which they need to link or index.

The Role of Gateways

This role is the one that almost all of the major subscription agents of the world embarked upon from late 1995 as their model for inclusion into the electronic journal market. Latterly, this is also the territory of the abstracting and indexing companies, such as Cambridge Scientific Abstracts, Silverplatter and ISI. This is also a goal of some of the dot com companies in our market, e.g. TheScientificWorld, although many of them seek to make significant document sales from this linking operation as opposed to subscription sales for the subscription agents.

The gateway is a large collection of links to publishers’ full text content. The gateway does not host the full text. The gateway does own (or at least accumulate) information about the full text, usually an abstract and other key “header” information, such as author, article title and other standard article metadata. It uses this information to provide its users with an adequate set of information to support a browsing function and a searching function, but not inclusive (generally) of searching the full text.

Another distinguishing feature of many gateway services is that these gateway companies know the access rights to the content that they index. They achieve this either by being a subscription agent, or by acquiring the subscription information from the publishers whose content they index and link to, or having obtained this information by a combination of inputs from publisher and library, e.g. OCLC FirstSearch ECO. This provides the end user with a certainty of being able to access content presented to him after having done a search.

The Role of Full-Text Aggregators

Abstracting and indexing companies, however, tend not to acquire this subscription information, but instead rely on their clients, usually libraries, to indicate instead the titles (rather than title combined with year) that their users are allowed to see. This is the route taken by ISI, for example. While this method does not provide the gateway with irrefutable information on access rights, it is extremely likely that the library will set up appropriate access for its users, even if it means that their users might be challenged for payment before being able to view the resource.

Having said all that, many libraries are themselves becoming gateways. The simple fact that no commercial gateway normally facilitates access to all of the content that any one library requires means that the world’s larger libraries, particularly those in the USA, are creating library web pages that link to all their subscribed electronic content, or in some cases, are able to enhance their web-base catalogues to link directly to the full text resources. For example the University of Massachusetts Medical School has set up web pages detailing all of the titles that students on and off campus can access via the proxy server of their networked site. Indeed, the creation of such library web pages, or library portals as they are increasingly known, has become part of the business model of companies like Ingenta.

Furthermore, libraries are supported in their quest to be the most appropriate gateways for their students and researchers by a very important initiative called SFX.

In a traditional gateway environment the gateway provides the user with his first port of call only; since once a user has made a search, located an article or title and navigated his way to the full text, there is a significant likelihood that his next click of the mouse will take him to a text referenced by the article, probably on a publisher’s web site or within another aggregation. The gateway has lost control of its user and indeed, worse still, the user may now be approaching content through the web pages of a rival service.

SFX, however, allows a library, and hence a library portal, to override the linking suggested within a document and replace it with something more appropriate for its user. In other words, the user may be re-directed not to content on a publisher web site, but perhaps to a manifestation of that content through the library’s chosen gateway or in a locally held full-text aggregation. Whatever the mechanics, the crucial element is that library gateways are in the unique position to dictate that they retain the user as he browses and navigates from one document to another. This currently puts libraries in a uniquely strong position in the provision of information portals to its clients.

The Table 1 describes the key differences between the major classes of e-journal gateway providers.

The Role of Full-Text Aggregators

The full-text aggregator, as the name suggests, is a company that creates databases of full text articles, defined by subject area and sold as a single product, rather than as individual subscriptions to components of the database.
Whilst the companies in this sector were the first to be called aggregators, increasingly their business model is causing some difficulty for primary publishers. On the one hand these companies provide a valuable service to the internet-inexperienced publishers. They can take input from as little as scanned images of the printed text and make an online product quickly and easily. But on the other hand, their very presence has the potential to limit the growth of the primary publishers as they begin to experiment with flexible pricing to “new” markets for their material. A new market for a small publisher would quite likely include the very institutions to which the full-text aggregator had been selling the publisher’s content. In that sense, many new opportunities are stifled since the aggregator has “got there first”.

There are many companies that provide full-text aggregations. Perhaps the best known of these are Adonis, Ebsco Publishing and UMI (Bell and Howell). The business model is fairly straightforward. The full-text aggregator licenses content from its owner and pays a royalty for the sale of the content to libraries and information centres. This has the benefit of allowing the generation of some income for journal publishers from those libraries that would not normally pay a full subscription price.

However, herein lies a problem. The very concept of a subscription price is being challenged as primary content is placed on the internet. In the print medium a journal subscription has a list price, occasionally with some small variations based on institutional type or regional considerations. But in the internet subscription age, an ever-increasing number of publishers price their content on a case by case basis. In other words they are creating a market for their subscriptions within the very institutions that would have purchased their content via a full-text aggregator. It would be quite conceivable for a publisher to charge just 5rate to a small community college and still make money from it. As a consequence, many of the larger publishers are reviewing their agreements with the full-text aggregators, often limiting the terms of the licence to include only older materials, leaving them free to negotiate a discounted price directly with the client for current information.

In addition, many publishers now have successful relationships with consortia of libraries. In many cases these deals with consortia include discounts for low-use, peripheral-interest materials, or materials destined for small institutions. Academic Press was the clear leader in this way of thinking.

As we speak, there are a number of initiatives under way to bring together groups of smaller publishers who can collectively negotiate with library consortia. This in turn will lead these publishers to the same differential pricing strategy as their larger forerunners and thus to review their relationships with full-text aggregators.

### Table 1.

<table>
<thead>
<tr>
<th>Type of gateway</th>
<th>Search Headers</th>
<th>Search Full Text</th>
<th>Know Permissions</th>
<th>Maintain Control of User</th>
<th>Coverage</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subscription Agent, some Dot Coms</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Multi-disciplinary, partial and non-selective</td>
<td>SwetsBlackwell, Rowecom IQ, OCLC Firstsearch ECO</td>
</tr>
<tr>
<td>A&amp;I Company, other Dot Coms</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Single discipline, selective</td>
<td>ISI WOS, CSA, TheScientificWorld</td>
</tr>
<tr>
<td>Library</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Everything selected (by definition)</td>
<td>All libraries</td>
</tr>
</tbody>
</table>

Who Benefits From the Presence of Aggregators?

**Libraries**

Purchasing of mass collections or organising access to a mass collection that has been purchased as a separate act, both allow libraries quickly to address the information needs of their patrons.

**Small publishers**

Small publishers gain very much through careful dealmaking with aggregators. In particular, by using an appropriate content host, the small publisher achieves the same
“shop-window” status of its larger counterpart. In addition it can use gateways to further improve its visibility. The very existence of value-for-money content hosts ensures the continuation of the tradition of the small publishing house in being able to secure a niche market for its niche products.

**Large Publishers**

Larger publishers save money by outsourcing many of their non-core competencies, such as printing and typesetting, leaving their valuable management time to concentrate on key differentiators from publisher to publisher, namely editorial and organisational differences. Electronic journal hosting can and should be one of those functions, just like printing, typesetting and distribution. In addition its presence in the right gateways can be used to carefully enhance its brand as long as care is taken not to allow the gateway to subsume the publisher brand.

**Scholarship**

Scholarship as a whole gains from a combination of the above benefits for libraries and publishers alike. Aggregators facilitate the diversity of publication from large numbers of publishers rather than promoting the continued conglomeration of publishing houses.

**Lessons for Small to Medium Commercial Publishers**

The primary lesson for all publishers is to get its content online quickly, but at a price that it can afford. Publishers should use gateways to maximise content visibility, and take control of the ownership of their own content by making sure to keep tight control over licensing arrangements. In addition they should probably join a consortium of publishers with the goal of increasing their market by differential pricing to smaller institutions. Likewise, it is important to make sure the content is never sold too cheaply, especially as wealthy multi-nationals stretch the meaning of the “site-licence”.

**Lessons for Large Commercial Publishers**

Most large commercial publishers are already online, but should keep a close eye on cost to make sure the most cost-effective method of having content online is being deployed. Many larger publishers already take great care of their licensing terms and are at least experimenting with differential pricing.

**Lessons for Not-for-Profit Publishers**

While the commercial imperative of differential pricing and sales to consortia is less apparent to the not-for-profit publisher, maximising readership remains an intrinsic must. This can be best achieved through cost-effective content hosting, the careful management of its own (often powerful) brand and maximising visibility through gateways appropriate to its readership.

**Conclusions**

This paper has been one largely intended to bring improved clarity to the varying roles of a group of companies collectively termed aggregators. The three classes of aggregator all bring about different benefits to publishers and libraries alike, and their business models are quite distinct.

The future for libraries appears to be very bright indeed, and with careful planning, can be equally bright for publishing as a whole.

---

Simon Inger, CatchWord Limited PO Box 261 OX14 3BP Abington, UK, e-mail: simon.inger@catchword.com

(Received 14 February 2001)