

KEY ISSUES

Aggregators – a taxonomy

Chris Beckett

The purpose of this piece is to provide readers with a simple descriptive summary of the various types of services and businesses that are, not always helpfully, described simply as aggregators.

*Chris Beckett, Chris Beckett Associates Ltd, Box Cottage, Tyte End, Great Rollright OX7 5RU
Tel: +44 (0) 1608 730700
E-mail: chris@chrisbeckett.com*

The Original “Packaging” Aggregators

The term aggregator was originally used to describe the secondary publishing function of licensing full-text content from multiple primary print publishers, converting the print into files that could be aggregated on to and distributed by CD-ROM as a package and read on screen using a proprietary interface. Examples of such early aggregators were Adonis, Ebsco Publishing, Gale and Proquest. The original CD-ROM aggregators took advantage of the quantum leap in storage capacity that CD-ROM technology provided to aggregate large bodies of scholarly material, often within large but distinct subject clusters, and deliver them as ASCII files, or original image files in the case of ProQuest, to libraries. These libraries then had a single interface, either standalone or networked, to a large body of content from multiple publishers, and most importantly at significantly discounted prices. The business model pursued by the “packaging” aggregators was one of licensing the content from the primary publisher, carrying out the necessary conversion from paper to electronic formats at no cost to the primary publisher, and distributing, marketing and selling the resulting aggregation of content to libraries. This was carried out through the aggregator’s own sales and marketing operation. All the technology and sales and marketing costs were borne by the aggregator and the majority of the revenue from sales went to them. The “packaging” aggregator defined the pricing and all other sales and marketing issues relating to the aggregation. The primary publishers received a royalty share of the income generated by these sales. This represented a new and additional royalty stream for the primary publisher that, in the early stages of this aggregation model and in the absence of the publisher having their own electronic product, had few significant negative side effects. The original aggregators,

by creating bundled electronic content delivered on CD-ROM, with a single interface, effectively created a whole new market for primary publishers. This arrangement was lucrative for both parties as long as the aggregator's full text database was the only available electronic version of a journal.

However, the arrival of the WWW and the possibility of publishers establishing their own independent web presence by which subscriptions could be delivered, and in some cases sold, meant that there were now at least two channels by which the same content could be received. As the original aggregators incorporated newer formats, specifically PDF file delivery rather than scanned images or ASCII files, the content distinctions between the aggregators' offerings and the publishers' offerings became blurred. This lack of format differentiation was one factor that encouraged libraries to purchase the aggregator's full-text product rather than buying individual subscriptions to journals. This, combined with the much better value that the aggregator product represented when measured by price per article available, inevitably led to some substitution of the aggregator's package for primary subscriptions. One response available to the primary publisher to this possible cannibalisation of their subscription base is to embargo content, effectively making access to the most current output only available via a full subscription. The embargo is a useful tool for the publisher to protect subscriptions, and in the case of Ebsco at least is a valuable tool that enables them to sell aggregated full-text content while still protecting their subscription agency business.

Notwithstanding embargoes at least one publisher has recently withdrawn their entire content from these aggregators. Generally and especially in times of constrained library budgets these aggregators' packages represent extremely good value for libraries. They would appear, according to some primary publishers, to work less well when the publisher has an established electronic product.

A counter argument is that as the major packaging aggregators seek to differentiate their products on the basis of exclusive access to prestigious content, there may be significant commercial benefits for those larger primary

publishers who negotiate vigorously and especially when they are prepared to grant an exclusive arrangement to an aggregator.

A further factor to be considered, especially by the many small and medium-sized publishers, is the sales and marketing reach these aggregators provide. Many of the smaller publishers have invested technologically in their own electronic presence on the Internet but have yet to commit commensurate resources to the promotion of this content. In these circumstances, the original aggregators, because of their well established and effective sales and marketing operations, may still be able to reach markets and achieve sales that the publishers themselves cannot reach or attain. Providing visibility and royalty income to publishers from markets in which the publisher had little or no presence was at the centre of the original value proposition that the packaging aggregators presented. It remains a compelling reason for many small and medium-sized publishers to maintain their agreements with these aggregators.

How long this situation will persist and how these tensions will be resolved in the longer term remain to be seen

There are a number of further differentiating factors that libraries consider when choosing between an aggregator's package and a primary subscription. One of these is content stability. The "packaging" aggregators make no claims as to the long-term content stability of their package. There is no guarantee that a title included in one edition of the package will be available in the next. This is unsurprising given the licensing model underpinning the creation of the aggregation. Therefore while these packages are perceived to be good value for meeting the needs of an undergraduate (or below) readership where precision and absolute relevance are of less importance than retrieving something quickly, they are of much less value to a research institution that needs to guarantee access to a continuum of current research literature.

The Content Hosting Aggregators

This was the role filled by CatchWord (now part of Ingenta), Ingenta, HighWire Press, Allen Press, American Institute of Physics and others. The underlying business model here is significantly

different from the original aggregators. These hosting aggregators are essentially service providers that promote services rather than journals. They are publisher-facing, deriving the majority of their revenue from providing publishers with data conversion, hosting, distribution, access control and linking services. Essentially they are service organisations that get paid by publishers for building and maintaining their electronic presence. They do not, unlike the original aggregators, provide a significant additional supplemental revenue stream through royalty payments. They do provide the ultimate point of access for subscription-based access; however, the revenue associated with these subscriptions is not handled by the content hosting aggregators, but by the publishers themselves directly or via subscription agents. The content hosting aggregators do, in some instances, provide supplemental revenues from other sources such as document delivery. However, the majority of the content they deliver is on a the basis of a subscription to the full text, where the publisher defines both the price and the product, unlike the "packaging" aggregators. The presentation of content to the end user by the content hosting aggregators varies from aggregator to aggregator, with the individual journal brands being more or less prominent. They all tend to offer cross-journal searching, browsing by subject, and in some cases, such as HighWire, a specific subject focus. The larger content hosting aggregators such as Ingenta are less subject-focussed since their collection development policy is essentially determined by which publishers they sign up for their service. This can lead to a very mixed subject collection. This in itself is not a disadvantage, especially when publishers are trying to extend their reach into the academic library market, since the availability of a very large body of content, such as that available through Ingenta, through a single interface with one registration and set-up procedure is a considerable advantage. In many instances, especially for the larger libraries in the USA, the content hosting aggregators are targets for the library's own in-house gateway products and therefore unwanted or irrelevant content can be screened out within the library gateway. There are a variety of technologies available to do this, the most prominent at the moment being the use

of Open URL resolvers such as SFX. Making the content on the content hosting aggregator site available through a large variety of secondary and other services such as abstracting and indexing services (such as PubMed), general-purpose gateways (such as SwetsNetnavigator), and document delivery services is an increasingly important role that content hosting aggregators undertake.

Subscription Agent Aggregators and Gateways

Almost all the subscription agents have adopted the gateway model. These gateways generally provide access to a larger body of content than the content hosting aggregators (although not perhaps in the case of Ingenta since it is both a content hosting aggregator and a gateway). The subscription agency gateway is a large collection of links to primary publishers' content hosted on the publisher's own site or on a content hosting aggregator's site. Generally the subscription agent gateway does not host full text. It does, however, host metadata about the full text, usually the bibliographic header and the abstract. This information provides for local searching and browsing (although not of the full text) with links behind the scenes that retrieve the full text from the publisher's or content hosting aggregator's site. Apart from their comprehensiveness, the particular advantage that the subscription agencies' gateways have is that they "know" about the rights that their library customers have to access full text, because the subscription agent is in possession of the subscription information. This advantage is qualified in so far as it is only true for those subscriptions that an individual subscription agency handles. This has limited somewhat their appeal to any library that uses a combination of subscription agents and direct subscriptions to the publisher to obtain their print and electronic journals. In such circumstances an individual subscription agency and therefore that subscription agent's gateway only "knows" about subscriptions bought through that agent, and therefore presents a less than perfect solution to a library with such a split acquisition pattern. However, recent developments from Ebsco demonstrate one solution to this problem by allowing individual libraries to register with

Ebsco subscriptions that are not handled by them. This increases the appeal of the Ebsco gateway product significantly by providing a single point of access to all a library's electronic journal subscriptions, irrespective of whether or not the subscription is handled by Ebsco. This comes at a small price premium, thereby still fulfilling the original business purpose of the subscription agent gateway, which is to provide an incentive for the consolidation of all subscriptions through a single subscription agent. It is anticipated that other agents will follow this lead.

The business model of these gateways varies. Until recently most subscription agent gateways were made available to the subscription agent's customers free of charge as one of the benefits of doing business with that agent. This may be changing as the costs of deploying ever more sophisticated functionality, and the value that the gateway delivers to the library, both increase.

Abstracting and Indexing Gateways

The abstracting and indexing (A&I) services are perhaps not traditionally thought of as gateways. However, as they have sought to consolidate and increase their appeal to the individual end user by building links from the metadata that they hold to the full text, some of the functions they perform are sufficiently gateway-like to justify their inclusion in this brief review. Abstracting and indexing services are distinguished from general-purpose subscription agent gateways

through many features. Two important ones are their subject focus and the controlled language searching that they present to the end user. A potential weakness is that they do not "know" what rights that an individual library has to access the full text. Therefore without additional work by the A&I service, while they can lead the reader from the highly relevant search result set (which their subject focus and sophisticated search interface can generate) to the full text, they cannot guarantee that the individual reader will be able to access the full text. This is simply because the A&I service is not in possession of the necessary subscription information. The A&I services have overcome this in ways similar to the content hosting aggregators, by requesting from libraries lists of journals to which they have valid subscriptions, and often using this information in conjunction with information supplied by primary publishers to facilitate seamless access.

Conclusions

Despite its title, the purpose of this short piece is to discourage the general use of the term aggregator and encourage greater specificity when seeking to identify and describe the many different organisations and business models encompassed by the using the word aggregator. The taxonomy offered above is no doubt one of many that could be devised. Readers with specific thoughts on this are encouraged to contact the author.