Looking Toward 2010: The Content Aggregator World Spins Even Faster!

Jan Sykes, President, Information Management Services, Inc., imagines the tools available to information professionals in six years time.

The crystal ball starts to clear as the year 2010 comes into view. As we contemplate the future of our industry, we see that information professionals are still there, but their role has radically changed. The aggregators, too, are thriving, but their services are both more complex and even simpler. Prices have hardly gone up, but customers are demanding a lot more for their money. The barriers between free and paid-for information and internal and external information, have effectively vanished. Artificial intelligence has not yet exceeded human intelligence. The pace of innovation is even faster. What a wonderful world this is! It clearly needs further examination.

People

Business is about people—how they work and what information they need in order to work smarter and be more productive. For content aggregators to continue the successes they have enjoyed, there is an unending challenge to understand more about the user community (including information professionals, novice users of information products, and recipients of research performed by intermediaries). They must also keep in step with changing business processes and the promises of emerging technology.

It will be spirited conversations with information practitioners, subject matter experts, vendor reps, authors and consultants that will drive product and service developments. There will be no reprieve in the need for constant learning in order to navigate through the content world and intelligently select the best resources and tools at a fair price. At the turn of the century, we talked a lot about adapting to change. In 2010, change will be so ubiquitous and so constant that we will simply accept it as a way of life and as a way of working.

Six years from now, I predict that information professionals ensconced in a corporate library and operating as intermediaries or gatekeepers will not exist. A small, centralized team may handle *ad hoc* requests for information, although the more likely scenario is that these requests will be outsourced.

In general, information professionals will be embedded within an organization's major business units, mapping workflow processes and defining decision points at which external information adds value or is needed for decision support. They will be managing internal content repositories and feeding collaboration tools for communities of practice or knowledge management databases.

Information professionals will also be continually evaluating content resources, matching them to business uses, negotiating fees and distribution rights, and working with IT and the aggregators to inject new content resources and discovery tools into workflow applications.

Over the next few years, information professionals will work closely with content providers to strategically place e-learning modules into desktop applications. In business environments where the focus is on productivity, quality and profitability, e-learning offers the benefits of immediacy and relevancy; aggregators have an obligation to their clients to deliver customizable e-learning applications with their products.

By 2010, the learning object may be a link to a live chat area or a virtual instructor for immediate help. Aggregators will be expected to provide products and support tools that are colorful, engaging, interactive and highly visual to fit the working style of GenXers, and the Millenials—most of whom have been using computers since they were children-- now entering the workforce. The aggregators will also have to allow clients to direct content in varied formats to RSS readers, mobile or wireless devices and computers we might be wearing by 2010.

Because people typically seek information first from their colleagues, there are untapped opportunities for aggregators to develop tools for identifying and codifying expertise—both internal and external—to support innovation. There are also opportunities to leverage emerging social network analysis tools within the framework of information resources to mine an organization's intellectual assets. I anticipate that the aggregators we know today will launch new products to track how ideas and information flow within an organization, and even across organizational boundaries, before 2010.

Content

Will the thirst for more content ever be satisfied? At one end of the spectrum, there are information seekers who claim to be deluged with too much content—to the point of wanting retrieval only from major publications. At the other end of the spectrum, there are those who desire still more content from obscure sources to complement what is readily available today. The bottom line is that more content will be available in 2010 and the content that is available will be more current. There will be no reason for embargoes of any kind if publishers are seriously attuned to the business needs of their clients.

As multinationals look to build their business in developing markets, the threat is not from other multi-national companies who are moving into this territory. The bigger threat is from small local companies that can produce and market products at a much cheaper rate. Although some content aggregators are carving out a niche in providing information from developing areas, there is still a widely held perception of a dearth of secondary content about companies in developing markets. In six years time, the aggregators will undoubtedly provide stronger information trails related to how these businesses operate, how they are registered in their home countries, their investment sources, and local manufacturing approvals required. They will accomplish this either by creating roadmaps to Internet resources published in these markets or by digitizing, indexing and providing access to new content resources.

We can also anticipate that improved machine translation programs will be bundled with the content offering or with desktop workflow applications for accurate translation of complete documents, vs. translation of words and phrases. This will enable quicker response to changing global business conditions.

In addition to new publications and reports, there is a thirst for better access to multi-media resources—executive speeches, press briefings, streaming media, audio clips, videos and image databases. We can safely predict that by 2010 multimedia resources will be indexed and searchable just as easily as newspaper articles and media transcripts are today.

A question that frequently plagues the most savvy information user is "What is still missing?" even after seemingly exhaustive research. The traditional role of the aggregator has been to make available information that is gathered/prepared by other entities. There is an opportunity for the aggregator to unlock the gates to information that is not yet available in an organized, digital format. By 2010, I believe we will see aggregators becoming database publishers—providing access to new content sources such as health statistics or business statistics compiled by countries around the world. Links to statistical information on products licensed in each country or manufacturing plant approvals, for instance, will greatly enrich news stories.

There are opportunities for aggregators to continue to think broadly about "content"—for example, transferring technology used for coding and organizing published information into products for managing all forms of electronic records in compliance with emerging legislation (for example, Sarbanes-Oxley legislation in the U.S. and privacy legislation in the European Union).

Quality

Aggregators can usually skirt complaints about inaccurate data, particularly in directory databases, by pointing back to the publisher or information provider. Given the amount of information covered and the rate at which business conditions change, it is acknowledged that 100% perfection in data quality will never be possible. However, as we look to 2010, users of any commercial information service will be less tolerant of inaccurate content. As users, we will expect to see online forms for reporting bad data. We will expect aggregators to match news stories against directory records and identify those that need to be updated. One scenario is that aggregators will remove individual pieces of inaccurate data until corrections are made. Another scenario is that publishers will take the initiative to aggressively identify errors and correct them immediately. By 2010, the editorial process of catching and correcting errors on quarterly updating cycles, should be a historical artifact.

How information quality impacts business strategy, growth and profitability has not been widely studied to this point. But, increasing costs related to creating, maintaining and deploying internal information repositories, in combination with fees to license relevant external information are likely to bring more intense scrutiny of the

quality of content resources being licensed as well as the value added to business processes.

Search

In 2010, *search* will really work! Clients can have confidence that search accuracy will increase as quickly as the ability to scan increasingly large amounts of data. Highly precise results from typing words and concepts in the search box, due to deep indexing and sophisticated algorithms, may make natural language capabilities unnecessary. Null sets will become obsolete—being replaced by approximate ("in the neighborhood of") answers and recommendations for how to proceed.

The aggregators will provide adaptive spell check features and a *"more like this"* feature that is finely tuned to both metadata fields of the ideal record and to the user profile. Absolute relevancy will result from 100% accurate coding systems for company names, industry descriptors and personal names as well as more extensive subject taxonomies, transparent to the user but integral to the search technology, and definitely on the wish list of searchers. This wish list item covers power searchers and casual searchers alike. In particular, it is important to those business executives represented in recent studies by FIND/SVP who are not confident in the reliability of their search results and those searchers represented in a 2003 IDC study that indicated they found what they were searching for 50% of the time or less.

With content repositories growing in size and their numbers (and names) multiplying at a frenzied pace, information seekers—both information professionals and other knowledge workers—still typically find it very challenging to select the search service, web search engine or internal database that is most likely to contain the desired answer. The process today often ends in frustration at this point. The user may stop pursuing that particular path of inquiry or be forced to re-create information that already exists but cannot be found. This frustration will be just a memory by 2010.

For the past couple of years, content management software companies, library automation software vendors and content aggregators have been devoting research and programming talent to incorporating federated search capabilities into their products. With more content conforming to international information markup and data exchange standards, developers are building simple interfaces that allow a search query to be executed across disparate sources of information rather than repeating the search process in each individual repository.

Users will be delighted if this capability can be extended to include more free web sites, blogs, web-zines in the search, while simultaneously using quality algorithms to evaluate the site. It should be feasible to keep items from what are deemed authoritative sites in search results and eliminate results from sites that do not meet certain validation criteria.

The visionary goal is to execute a query across a broad universe of information, either across sources designated by the user or those automatically selected by the federated search engine based on intelligence derived from the context of the query or identity of the user. Obtaining the best content from any individual repository usually results from using native indexing for that specific source; obtaining the best content from multiple repositories will require mapping of index terms from individual sources to a master authority file. Who is better prepared than the content aggregators to meet this challenge?

Retrieving a blended response set comprised of internal documents, web postings and external publications, ranked by date or relevance, will go a long way to helping users cope with confusion about where to search. For the client, boundaries between fee-free and internal-external need to become more porous, and at some point, invisible. Relying on authentication protocols, results will be easily restricted to resources the client has permission to view. In an ideal world, the client will also be able to click on an icon or links in records in the retrieval set to immediately view an image of the original document. These should be standard capabilities offered by aggregators by 2010.

In fact, it is not unthinkable that by 2010, aggregators will modify their business models to host and manage custom collections of proprietary third-party databases, free resources or subscription-based content—including content from competitor systems, for their clients. Rather than housing everything behind organizational firewalls, clients may find it attractive to outsource management of their unstructured content to the aggregators because of their experience and expertise in content normalization, consistent application of robust classification and indexing schemes and interface design. Traditional aggregators may be relegated or elevated (depending on one's point of view and status in the food chain) to this new paradigm if content providers and database publishers elect to bypass their current distribution channels and align themselves directly with Google or Yahoo! or search engines just now appearing on the horizon.

Discovery

Although there is plenty to strive for in terms of perfecting search capabilities, *discovery* will be the new frontier for the aggregators. Because of their years of experience with managing huge volumes of text, the aggregators are poised to help businesses unlock intelligence by detecting trends or relationships in unstructured documents. Better yet, there will be accompanying entity extraction, visualization and graphical mapping tools to evoke new insights into perceptions, behaviors, and emerging issues based on content analysis. Text mining tools that have debuted recently for reputation management open our imagination to new ways of interpreting patterns revealed in text vs. searching for specific answers in the text. By 2010, text analysis tools will be indispensable for surfacing knowledge contained in internal repositories. The business value will be based on effectiveness of text analytics for early warnings in general, competitor activity and surveillance of any number of topics. There is also the potential to alert the client to questions that it had not even realized were waiting to be asked.

Workflow

Finding accurate answers quickly and easily is more important than ever. It is increasingly evident that organizations will only realize the full value from their investments in information resources if users obtain trustworthy answers when the need for information arises, that is, from within whatever application they happen to be using when that need arises. By 2010, aggregators will surely be offering more options for the seamless application of external information into the workflow of knowledge workers—even beyond the integration of tool bars and search panes and capabilities for clipping and pasting links for articles, reports, and company profiles from aggregator systems into new documents, presentations, and databases.

More linkages between the information in one database or website to related information in another location, likely based on digital object identifier technology or persistent URL's, should be commonplace by 2010. If so, this will make external content come alive in the workspace of the client and add value to business decisions. Examples range from:

- A link from an abstract to the full image journal article on the Internet
- A link from a news announcement of a statement by a CEO to the full analyst update on that company and the entire archived broadcast of the CEO's statement
- A link from a new product announcement to the full text patent and label information
- Links from articles on management topics to related e-books

Ironically, after clamoring for more full-text, information professionals now spend a lot of time analyzing and summarizing key points of interest to provide highlights to persons for whom they are performing the research. I believe the aggregators will develop the functionality to bring the process full circle within the next six years—using technology to sort through, abstract and summarize single or multiple full-text documents pulled together for background information.

Access

For customers to enjoy more seamless access to information resources, there remains much work to be done in the areas of authentication and encryption. Significant progress has been made over the past few years in authentication based on ranges of IP addresses, user ID and passwords, and LDAP (lightweight directory access protocol) standards. Some information providers claim to offer all of the above. The reality is that those offering built-in authentication protocols, designed to plug into enterprise application software (like Microsoft's Active Directory ® service, tools from PeopleSoft or SAP, and portal software products) with the least amount of effort and disruption on the part of the client IT organization stand to have an advantage in the selection process. It is essential for aggregators to make it easier to get their product in the door and then on

the desktop of designated users without expecting large commitments of time and effort from the buying organization.

Beyond implementation convenience, sophisticated authentication protocols will assume greater importance as the basis for filtering information feeds or search results by client identity. The sign-on information will indicate the client's location, department, communities of practice in which this person participates and will perhaps, even connect to his/her calendar. With this amount of information about the individual, the online service can tailor search results and alerts so that they are personalized and relevant. I think it is very likely that the aggregators will offer a premium service by 2010—proactively calling up the client on a cell phone or PDA to deliver breaking news on high-priority topics, either by displaying text or actually reading a news item.

Expectations for "seamlessness" apply to technology and product upgrades as well. Clients know that technology changes will occur. They actually prefer to work with information providers that continually upgrade their products and those that quickly adapt and implement new technologies. The upgrades should never be a burden for the client organization—rather an "easy-to-integrate" enhancement.

Pricing

Content providers, aggregators and clients all wish they had the answer to what pricing will look like in 2010. Our crystal ball is pretty fuzzy in this area. It is safe to say that the few companies still using connect time pricing will no longer do so. It is also safe to say that the market will not bear much of an increase, if any, in per record rates. The availability of so much free content on the web makes many users question why they should pay for any content. Our responses about costs for creating content, organizing it, and providing search technology, post-processing capabilities and deep archives do not carry much weight in the face of *free*.

I believe that more primary publishers and aggregators alike will experiment with distributing content via Google or Yahoo! or similar search engines, testing the concept of mingling fee-based items with free items in search results. By 2010, aggregators and publishers will have worked around the need for every client to sign a user agreement and they will be able to accept micropayments in very small sums. The GenXers and the Millenials are comfortable with debit/credit cards and with using them for small transactions.

For enterprise pricing agreements, aggregators should expect to provide their clients with even more detailed usage data as a basis for pricing. Unless the economy is particularly robust for the next six years, the review and adjustment of subscription levels may happen more frequently than on an annual basis. For heavily used content resources, companies will be looking for more attractive pricing (cheaper) and redistribution rights (more flexible) and thus, will continue striking deals directly with primary publishers. To balance this loss, aggregators will adjust their pricing models to include fees for consulting, integration, customized products, reporting and perhaps hosting of content.

Digital object identifiers incorporated into the content will enable tracking of original documents and will allow information providers to be rewarded fairly for use of their content. Because of this tracking capability, clients will be delighted that licensing agreements permit reasonable levels of reposting and redistribution

In Conclusion

Looking back six years to 1998, if sophisticated information professionals had been asked what they expected from online content aggregators by the year 2004, how might they have responded? There would have been requests for larger and more global content sets, more browser-based access to premium content, access to more full-text documents, removal of duplicate items, natural language processing, and better systems features for reformatting and presenting output.

The companies profiled on the pages of this book (some of which did not even exist six years ago) have delivered content, search and retrieval tools and expanded services in response to customer needs and market opportunities. In some cases, they have surpassed what could have been imagined in the early, and even recent, development of online services.

With our predictions for six years into the future, there is a keen awareness, even hopefulness, that transformational technologies, as yet undefined, will again totally alter expectations and capabilities.