

Intelligent Content and Technology Integration

By Mike Tansey, President and CEO, Thomson Scientific

To create a unified digital library environment, information managers can no longer select database products based purely on content. Instead, they must seek out implementations from leaders who can also offer new technologies for organization, searching and links navigation. Information providers are developing fully integrated solutions, including links management systems and non-traditional search technologies. Meeting the challenge of content management, therefore, means selecting the right content, and ensuring that the tools and technologies that accompany it build on the research environment already in place.

Linking Gateways

In the Web world, the first piece of the digital library technology puzzle is the links infrastructure. Information managers have a daunting task: to ensure that links management within specific vendor platforms offers the best value-added benefits, and that those same vendor platforms work seamlessly with any portal-level, context-sensitive linking system in use by the library.

A well-conceived vendor platform is one that allows a researcher to follow an idea wherever it may lead, allowing the underlying linking system to integrate, extend and organize the research environment. A successful linking infrastructure acts “behind the scenes” to ensure that the natural relationships between content sources are highlighted for the user. The ISI Web of Knowledge platform is an example of how a linking infrastructure can provide those connections.

Interproduct links: Connect a record in one content source to the same record in another. By seeing how one article can be found in numerous resources, researchers are able to explore a set of related databases in a targeted way, and to quickly and easily gather the unique information provided in each. A researcher has a variety of ways to explore a topic within an individual database, but with interproduct links the possibilities increase dramatically. The ISI

Links infrastructure within ISIWeb of Knowledge permits this type of exploration by automatically showing special link buttons whenever a paper appears in two or more platform resources. ISI Links manages the connections between content—within the context of the institution’s subscriptions—so that a researcher doesn’t need to.

Shared Citation Links: As serendipity is as much a part of the research process as effort, vendors must find new ways to help researchers along the discovery path. For us, this means using the ISI Links management system to “share” citation information across platform databases. Special buttons have been added to the full record of hosted content sources to allow novice users to “stumble” upon the benefits of citation indexing information. Direct links to full bibliographies, lists of citing articles and even a “find more like this” feature (called Related Records, formerly only available in Web of Science) are now available within hosted databases such as BIOSIS Previews and INSPEC.¹

Full Text Links: For vendor platforms based on bibliographic databases, management of full text links is critical. The role of bibliographic databases is to provide an efficient way to filter an ocean of information down to a pool of relevant articles, papers and patents needed at a given moment. The next step is to locate the full text of those items—and in a well-designed platform, doing so is a matter of a few mouse clicks.

Here again, the ISI Links management system comes into play within the ISI Web of Knowledge platform, offering full text links via direct publisher feeds and a unique pre-verified algorithmic linking called “RoboLinks.” Link resolution is always assured through this stable yet extensible system that has been specifically designed to ensure reliable links to the appropriate copy of an institution’s full text.

Context Sensitive Links: A final consideration for the information professional wrestling with the evaluation of a vendor



Mike Tansey

Mike Tansey is the CEO of Thomson Scientific. Previously, Mike served as the President and CEO of ISI. He has been involved with the evolution of electronic publishing for almost 20 years. Prior to becoming President of ISI,

Mike was responsible for all product management and was instrumental in the development and launch of the *ISI Web of Science*. Before joining ISI, Mike was responsible for all technology operations at BRS Information Technologies and prior to that he was responsible for all technical publishing activities at Aspen Systems Corporation—a leading supplier of information management solutions to the Federal Government and Legal Markets.

platform is that of links compatibility with the greater library mission. More institutions are realizing the importance of a context-sensitive link package, or “links server,” to a digital library. A links server offers a way to provide a “menu” of ideas to help researchers decide the best next step in the research process. For example, it can identify which databases index a particular journal, direct a user to all the places where the full text of an article can be found, or to work directly with a document delivery system. The sophistication of links servers range from basic (focusing on relationships between standard electronic resources) to comprehensive (focusing on complete serials management).

To fully support a digital library, a vendor platform must be able to seamlessly integrate with an institution’s context-sensitive linking package. To this end, the ISIWeb of Knowledge platform has been enhanced to offer the integration of OpenURL-based links servers. Web of Science is currently OpenURL-enabled, and all other content sources within the platform will soon follow suit.

Beyond the Traditional Search

The second piece of the digital library technology puzzle is the search infrastructure. Whereas links offer the opportunity for content relationships to be highlighted, search options offer the researcher a way to use those relationships in a personal, targeted context for precise information retrieval.

A well-developed vendor platform allows different types and levels of searching to meet the needs of different types of research methods. In today’s digital research environments, traditional (Boolean) searching is complemented by new relevance-based

natural language searching, cross-search technologies and even new portal-level cross-collection discovery tools.

Natural Language Searching: With the development of search engines specifically designed to meet the needs of Web-based information, there has been a shift away from the traditional Boolean search paradigm towards a probabilistic model. When retrieving information, a traditional search system manipulates the exact algebraic relationship between the terms entered by the user. In contrast, probabilistic (or “natural language”) search systems focus on the concept behind the terms, by weighting each term and then applying relevance to select documents. Natural language searching complements traditional searching. A platform that provides both greatly enhances the research experience.

“With the advent of federated searching, portal-level search technology options are about to change dramatically.”

Within ISI Web of Knowledge, the MuscatDiscovery probabilistic search engine supports two tools: Current Contents eSearch and ISI CrossSearch. In a single search, Current Contents eSearch allows users to retrieve journal articles through a traditional engine and evaluated Web sites (and individual Web documents) through a probabilistic engine. The researcher enters terms into the Current Contents Connect search interface, which queries them against a set of journals. Current Contents eSearch then transforms the Boolean search into a probabilistic one by adding weighting and relevance criteria. The resultant query is matched against Web sites and Web documents in the Current Web Contents database; relevant hits are returned. Because this second search is completed “behind-the-scenes,” the user can uncover valuable Web documents and Web site reviews as a natural extension of a typical journal search.

Cross-Searching: Cross-searching of multiple resources comes into play when there is a need to complement individual database searching (whether traditional or natural language) with a next-level discovery tool.

ISI CrossSearch provides a way of discovering relevant documents—journals,

proceedings papers and patents—found in the databases produced by Thomson as well as those hosted within the platform through partnerships with other information producers. A researcher has a choice between conducting a traditional cross-search or a natural language cross-search. For the latter, the easy-to-use “concept” box welcomes users to enter a phrase, sentence, or entire paragraph. This allows the user to approach the research process in a different way, starting with a general idea or concept rather than a specific set of words. The concept CrossSearch is run against the databases chosen by the user, and returns a deduplicated results list sorted by relevance. From there, a researcher decides which individual resource to drill down into by selecting whichever individual database best suits his/her needs.

Federated Searching: Enabling true cross-collection discovery, however, demands even more than a cross-search mechanism. It requires a meta-search mechanism at the portal level, a system referred to as “broadcast,” “multi-protocol,” “meta-” or “federated” searching.

Federated searching provides a single search interface for all of an organization’s electronic resources. Unlike in a cross-search system or a single protocol-based system (such as Z39.50), each database remains in its native format and is not expected to be enabled with a certain query language. Instead, a federated search system houses a set of translators to complete each search—one translator for each database. The system takes the user’s search terms, translates the search string into the proper syntax for each electronic resource the user has selected, and then sends each query out separately to the appropriate content source. The federated search system has no search engine of its own—it relies upon the capabilities of the search engines found within the individual databases themselves to retrieve results.

Designed to complement rather than replace the searching within individual databases, this discovery system offers powerful benefits for a digital library environment. It allows a content manager to facilitate easy access to an organization’s electronic resources—acting as a bridge to lead researchers from the library or organization portal homepage quickly and easily into the electronic resources they need most for their day-to-day information gathering activities. It provides a new tool for both novice and experienced information users in a way that allows a library professional to direct them to the proper resources in an efficient and

focused manner. A federated search system also aids e-resource managers by increasing usage of their underutilized resources in order to increase return-on-investment for those content expenditures.

We have chosen to incorporate federated searching in two distinct ways. First, a proprietary federated search infrastructure is a fundamental part of ISI Web of Knowledge. Using the ISI CrossSearch feature as a foundation, a researcher can opt to have a search query automatically translated into the syntax necessary for two external content sources: PubMed and AGRICOLA. Other free resources in various disciplines will be added in the future, as well as optional subscription-based resources. Second, we have entered into a partnership with WebFeat, Inc., a leader in federated search systems, to offer solutions directly on the library or organization portal.

Future Directions

With the adoption of the OpenURL standard, the information industry has the foundation it needs to improve and extend linking infrastructures in new directions. Information vendors are OpenURL-enabling their products so that a library’s context-sensitive links server can be easily integrated with their product offerings.

With the advent of federated searching, portal-level search technology options are about to change dramatically. NISO has already formed a “MetaSearch” standards initiative, and this new type of resource discovery will certainly become an important part of any digital library environment.

The bottom line is that content managers are no longer thinking purely about database content, and information technology specialists are no longer thinking simply in terms of systems. Instead, they are working together to look at the bigger digital library picture, and are taking a comprehensive approach toward the development of electronic resource environments. The only way to ensure intelligent integration within the research organization is to choose content from information companies that offer value-added linking and searching with the larger digital library environment in mind. □

Thomson ISI products and features mentioned herein are trademarks, service marks and registered trademarks used under license. Thomson ISI has no proprietary interest in the marks or names of others.

1. BIOSIS Previews is from the publisher of Biological Abstracts. INSPEC is produced by the Institution of Electrical Engineers.