June 2008

I Hear the Train A Comin' -- pub2web and MetaStore

Greg Tananbaum

Scholar Next, gtananbaum@gmail.com

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

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

Recommended Citation

Tananbaum, Greg (2008) "I Hear the Train A Comin' -- pub2web and MetaStore," *Against the Grain*: Vol. 20: Iss. 3, Article 34.
DOI: <https://doi.org/10.7771/2380-176X.2426>

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.

I Hear the Train A Comin' — pub2web and MetaStore

Column Editor: **Greg Tananbaum** (Consulting Services at the Intersection of Technology, Content, and Academia)
<gtananbaum@gmail.com> www.scholarnext.com

This month, I have the pleasure of speaking with **Louise Tutton**, Head of Client Management for **Ingenta**. Louise has ten years' experience in the electronic publishing industry, primarily focusing on client relations but with a résumé incorporating project management, production and editorial experience. She joined **CatchWord** in 1999 and was heavily involved in transitioning publisher customers through the merger with **Ingenta**. Her current remit as Head of Client Management includes overall responsibility for **Ingenta's** business with its publishing partners throughout the world. In addition to membership of the **ALPSP Professional Development Committee**, Louise serves on the **SSP Education Committee**, and is an Editorial Board member for *The Serials Librarian*.

What are pub2web and MetaStore, and how do they interact?

pub2web is — at its simplest — a Website for publications. It's a next-generation publishing platform, a combination of technologies designed to maximize the value and visibility of information.

Metastore is the data repository that supports the platform's innovative data capabilities — for example, semantic mining and manipulation of data to reveal new connections and research paths.

Metastore breaks content down into its most granular parts — freeing it from the restrictions of article, issue, journal, chapter and book structures — which means that publishers can be much more creative in how they license and distribute information. **Metastore** — and thus **pub2web** — is also format-agnostic in that it can support a multitude of data types — books, journals, reports, statistics, raw data, audio, video and more — so a publisher can give users a holistic, seamless view of their information assets, and meet their needs for a comprehensive research resource. By storing data at such a granular level **Metastore** opens up a number of opportunities in terms of content discoverability and the ability to re-package content for online sale. For example a biological title could be data mined for instances of species names which are then stored as distinct data objects within **Metastore**. Species "homepages" can be automatically generated within a **pub2web** site as a result, displaying a range of related information including:

- Metadata and graphics relevant to a particular species.
- Internal **pub2web** links to all other content referencing the species name — easing navigation to related and relevant material.
- Integration of external links to authoritative resources within a given subject area — a huge user benefit, easing navigation

to key subject specific resources within a publisher specific site and across the Web.

- Anything which is stored as a data object within **Metastore** can be repackaged as part of a virtual product for online sale — therefore extremely tailored packages can be built including movie clips, chapters, articles, species information (for example!).

Species names have been used as an illustration here but the possibilities across different subject areas are fascinating.

What are some real-world examples of how these services are changing the delivery of scholarship?

The new **OECD iLibrary** is a great example of a publisher drawing together multiple information types into a single publishing platform, which enriches the user's research experience by enabling them to discover and collate a variety of authoritative, and previously dispersed,

data sources. It also showcases some of **pub2web's** other useful features; for example, the site's multilingual interface facilitates easier content discovery and access for speakers of other languages. And let's not forget the basics: simple, uncluttered design is one of our hallmarks — it's about making the user experience more intuitive, optimizing discovery, and letting the content achieve its potential.

Ingenta recently announced an advertising partnership with Ten Alps. What is this all about?

Ten Alps will be selling advertising on behalf of **Ingenta's** publisher partners. A paradigm shift is underway in scholarly publishing and established business models are being reevaluated. Our clients are rightly concerned about how the value they add to the publishing process will be funded in the future, and many are seeking to explore alternative revenue streams for scholarly content. Presenting discreet advertising around their content helps them to balance subscription erosion. We have taken a coordinating role, representing a consortium of our publishers to enable them to break into established advertising networks and attract more interest than they would individually.

How have Ingenta services embraced Web 2.0 functionality?

Well, for the most part we try to avoid talking about "embracing **Web 2.0**" because it's becoming perceived as a fad, a bandwagon to which all sorts of tired technologies are trying to hitch themselves. That's not to say we don't investigate features which are tagged with the **Web 2.0** label, and we have implemented several that we think actually add value for

publishers, such as integration with social networking sites (we integrated this fairly early on in 2006 and are keeping a close eye on usage trends in this area) and deployment of blogs and wikis to help our publishers grow and engage with their user communities. The same is true for the systems and processes we use to run our business — for example, we use the collaborative tool **Basecamp** to manage our projects and client communications. As a technology company it's apt for there to be this consistent smartness across both our products and our processes. Really, though, our focus is on the semantic Web, which is generally accepted to be the original objective of the Web and of several of the collaborative, analytic technologies that have been popularized by **Web 2.0**. We're planning now for the next revolution in scholarly publishing that will be enabled when machine-readable data is published and shared as part of the research process. We'll be unveiling some of these features at our **Publisher Forum** in Boston on May 28th.

In 2007, Ingenta merged with VISTA, a publishing systems conglomerate. How has this impacted Ingenta's products and strategic vision?

Our competitive position and business footing have both been strengthened by the merger. We've been able to tap our new colleagues' alternative perspective on publishing, for example to develop better support for different types of content (**VISTA** has historically focused primarily on books and trade publishing). Their complementary knowledge, experience and connections enable us to reach new markets, and the integration of our product lines means we occupy a unique position as the only end-to-end provider of publishing software. During integration, the businesses were complementary enough that each group was able to focus on its strategy. For **Ingenta** it was to proceed with the development of **pub2web** and **Metastore** and bring some of the benefits of that technology to **IngentaConnect** also. Looking ahead, with our increased market share we are able to be bolder in our strategic planning, and have invested in key new staff and resources to drive further development and growth.

Where do your library customers and end users sit in all of this?

We wanted the merger to be pretty much transparent to our library customers and end users of **IngentaConnect**, as it's important that the service should not be disrupted or unduly changed. We were able to devote some time last year to reviewing and streamlining our library services in line with feedback from this community. For example, **IngentaConnect** now offers free tools for libraries to apply branding to the site, while our **IngentaConnect Complete** package is now a set of discrete

continued on page 85

I Hear the Train A Comin' from page 84

modules which a library can license independently to meet document delivery or current awareness needs in the most effective way. And of course, libraries and end users are the customers whom our publisher partners want to serve through the publication platforms we build, so all of the services I've just talked about are ultimately designed to meet their needs — whether it's by integrating software and content with the tools used by these groups (such as bibliographic managers or RSS readers), or by adhering to industry standards such as COUNTER and OpenURL.

In a world where technology is easier to manage and increasingly inexpensive, why do publications work with companies like Ingenta?

We're increasingly finding that the evidence does not bear out the assumption that technology is becoming easier to manage. In a world of evolving industry standards, demand for more advanced "bells and whistles", seman-

tic Web developments and visibility amongst the vast array of content on the Web, publishers are under pressure to conform to the latest standards, regularly roll out new features and functionality in line with technical advances as well as ensuring their technology is robust, scalable and future proof. A challenge which can be a distraction from publishers' core area of expertise (publishing) which in turn can impact on ROI as technology choices are critical to the success of publishers' businesses. As a result, we're finding that demand for the support of an established technology partner remains strong.

ScholarNext

Consulting Services at the Intersection of
Technology, Content, & Academia

- Web 2.0 Functionality
- Strategic Planning
- Business Development
- Policy Analysis

- Publishers
- Information Providers
- Technology Companies
- Libraries

(510) 295-7504 Greg Tananbaum greg@scholarnext.com

www.scholarnext.com

Technology for publishers is **Ingenta's** core competence, our sole focus, which is why a growing number of publishers (more than 250 now) are seeking **Ingenta's** support for their technical strategy. 🌱

Vendor Library Relations from page 81

Harvard's FAS vote are on everyone's radar. Could be time for materials vendors to have another look at what it is their customers most care about. Getting themselves into the offices of library decisionmakers — vendors have always known the importance of doing that. Without at the least having a few thoughtful things to say about open access and its **Ranganathanian** cousin, fair use, and how in their accustomed in-the-middle position vendors might make a difference, vendors could lose the one kind of open access they've always understood. 🌱

Back Talk from page 86

I think we cannot help but feel some of each emotion. I am leaning toward sending the existing downloaders emails asking them to provide proper attribution and to also strengthen the language in the click-through instructions stating that in the future readers MAY NOT download materials for further distribution. I think this is justified since while our students may have given us permission to put things up on the Web, we didn't ask, and I don't think they had in mind giving permission for 15 or 1,500 libraries and other organizations to make copies of their theses for posting on

their sites. Unfortunately I am also considering assigning someone to go through the 4,000 plus pre-1923 **Google Book Select** entries in which the words Hong Kong appear to find full text materials for our own electronic collection. Can I forbid others to do what I want to do? What do you think we should do? Please drop me a line if you have an opinion <ferguson@hkucc.hku.hk>. 🌱

Endnotes

1. Stryker, Cheri (2001) About the new, OT group, and its FAQ. Retrieved June 2, 2008, from **UseNet Replayer**, <http://www.usenet-replayer.com/faq/alt.binaries.multimedia.xena-herc.html>.

Rumors from page 71

and information on the site, though the material created by contributors and the user community, which each member will control and be credited for, will be published alongside the encyclopedia. **Encyclopaedia Britannica** itself will continue to be edited according to the most rigorous standards and will bear the imprimatur '**Britannica Checked**' to distinguish it from material on the site for which **Britannica** editors are not responsible." See "Encyclopaedia Britannica Goes — Gasp! — Wiki," by **Josh Fischman**, *Chronicle of Higher Education*, June 6, 2008. www.chronicle.com/britannicanet.com/?p=86

Tis the season to be collaborating ... Look at our interview with the astute **Rommel Nunn** about **Crossroads** in this issue, p.56. And, another interesting development. The Association of College and Research Libraries (ACRL), the Association of Research Libraries (ARL) and

SPARC (Scholarly Publishing and Academic Resources Coalition) have released a new series of **bookmarks** in the **Create Change** campaign, which targets scholars in different disciplines with messages about the benefits of wider research sharing. Librarians can use these freely available files to enhance their efforts to engage faculty interest in changing the way scholarly information is shared. The **Create Change Website** emphasizes the rapid and irreversible changes occurring in the ways faculty share and use academic research results. www.createchange.org www.acrl.org www.arl.org www.arl.org/sparc

Did you see the information that we posted on the **ATG News Channel** (5/13/08)? I am posting a **Rumor** most every day. Bad, bad, if you didn't!! Anyway, there was a lawsuit filed against **Georgia State University** by three publishers — **Oxford University Press**, **Cambridge University Press**, and **Sage Publications**. The publishers take issue with how **Georgia State** is handling electronic reserves. The *Chronicle of Higher*

Education interviewed **Lolly Gasaway**, **ATG's** expert on copyright, about this lawsuit which alleges that Georgia State professors infringed publishers' copyrights by "inviting students" to download, view, and print material from thousands of copyrighted works. The outcome of this lawsuit could have implications for how colleges distribute course material online.

We told you last time about **Choice's** move into new digs in late 2008 or early 2009 (**ATG**, V.20#2, p.12). Check out these photos of the construction project and see how **Irv** looks in a hard hat! www.flickr.com/photos/acrl/sets/72157604368374700/

And — last but not least — wanted to let you know that the **New England Journal of Medicine** has selected **Atypon** for its new integrated content delivery platform. There is a certain symmetry to this which is why I picked it as our last **Rumor**. **ATG** has interviews in this issue with both **Tom Richardson** of **NEJM** and **Chris Beckett** of **Atypon**. Like, cool! www.atypon.com/content.nejm.org/ www.massmed.org/ 🌱