GiantSteps
Media Technology Strategies

250 West 57th Street, Suite 2017
New York NY 10107
212 956 1045
fax: 212 581 1352
www.giantstepsmts.com

Google's Settlement with the Publishing Industry

Opportunities and Strategies for Publishers

By Bill Rosenblatt

January 28, 2009

Table of Contents

Introduction

The book publishing industry sued Google, Inc., in 2005 for what it alleged was illegal use of publishers' and authors' content in its Google Book Search service. The parties settled in October 2008.

The agreement that the two sides negotiated in the settlement went far beyond the typical terms in an intellectual property (IP) case – wherein the defendant pays the IP owner and receives a license to use the plaintiff's IP. Instead, the settlement lays out terms for collaboration between publishers and Google of unprecedented scale. It will make a massive amount of book content searchable and discoverable online, and it will enable sales of that content in a way that satisfies service providers like Google as well as publishers.

In addition to an initial set of business models involving Google and publishers, the settlement sets the stage for a future set of new digital business models. It should also engender a new set of service providers that can take advantage of the infrastructure that the settlement requires – including an independent Book Rights Registry that will track content, rights, and royalty payments for digital publishing.

Although the settlement should motivate publishers and service providers to participate in new business models, both types of businesses will need to build digital content infrastructure in order to be ready. Publishers will need to adopt XML-based content repositories, including XML servers as well as tools and workflow processes to enable structured content creation and management. And now more than ever, publishers will need to create complete and consistent metadata for their content and to manage information about rights. Services that repurpose and aggregate content across multiple publishers will bring about the need for standards in these areas as well. XML is the technology that ties all of these requirements together.

The first part of this white paper describes the Settlement Agreement in the litigation, including the Book Rights Registry, the initial set of business models that Google and publishers will implement, and the set of business models that the Settlement Agreement contemplates in the future.

The second part discusses the future opportunities for publishers, particularly those that depend on publishers' ability to build XML-based content architectures and make content available in structured formats with standardized metadata. It then discusses the capabilities that will be necessary for publishers to adopt in order to take advantage of these opportunities, including systems, tools, processes, and standards adoption where appropriate. Of course, a growing number of publishers are already starting to adopt these capabilities.

The Google Settlement

The book publishing industry, represented by the Authors Guild and Association of American Publishers, sued Google in 2005 over Google's Book Search program, in which Google engaged with university libraries to scan millions of books so that their content could be made discoverable online.

Through Google Book Search, Google made it possible for book content to appear in search results. This had not been previously possible to a large degree. The lack of users' ability to find professionally published content through Internet search engines has been called the "discoverability paradox."[1]

Nevertheless, the publishing industry objected to Google's and the libraries' actions, alleging that Google didn't have permission to authorize the scanning and that Google was unfairly making money by "free riding" on publishers' content. Google contended that it was within its fair use rights[2] to make the copies and therefore was not infringing any copyrights.

Google and the publishing industry announced a proposed settlement to the litigation in October 2008[3]. U.S. District Judge John Sprizzo gave preliminary approval to the Settlement Agreement the following month. Final approval is expected in June 2009 or later.

The proposed settlement is a bold statement of intent to cooperate on a range of digital business models. It sets the stage for a new set of online publishing opportunities and, potentially, a realignment of the value chain for online publishing.

There are three particularly important aspects of the settlement:

- The establishment of an independent Book Rights Registry (BRR).
- The initial set of business models specified in the Settlement Agreement.
- The range of future business models contemplated.

We'll describe each of these here.

Book Rights Registry

The Book Rights Registry is specified in Section VI of the Settlement Agreement. It will be a nonprofit organization that maintains a database of information about published works. Although Google will fund the BRR, with more than $30 Million of the $125 Million total

[1] B. Rosenblatt, Rights management and the revolution in e-publishing. INDICARE Monitor, November 25, 2005. Available at http://www.indicare.org/tiki-read_article.php?articleId=152.
[2] 17 USC § 107. Fair use can only be judged by a court, which weighs four factors: the type of work, the amount of the work copied, the nature and purpose of the copies, and the effect of the copies on the market for the work.
[3] Google maintains a set of pages describing the settlement at http://books.google.com/googlebooks/agreement/. The actual Settlement Agreement is available at http://books.google.com/booksrightsholders/agreement-contents.html and elsewhere.

payments specified in the settlement agreement, the BRR will be a third-party independent entity with which anyone can do business.

The BRR will contain metadata about books in copyright, including information about their rightsholders. The primary purpose of the BRR is to administer the initial set of business models specified in the settlement agreement (see below). But clearly the BRR could be used to track rights and payments for a larger set of business models; it would have to in order to support the business models contemplated. The BRR will most likely be free to develop other new business models as it sees opportunities.

The BRR will function like an online rights collecting society. Traditionally, rights collecting societies operated in the offline world, administering various types of rights licensing schemes such as licensing music for broadcast (as BMI does) and compensating periodical publishers for corporate photocopying of their articles (as Copyright Clearance Center or CCC does).

More recently, collecting societies have started to take advantage of the Internet as an efficient means of administering content rights licenses. CCC has been one of the leaders in this regard: its Rightslink division enables online licensing of certain types of rights to publishers' content, while CCC's Ozmo licenses rights to user-generated content such as blogs and podcasts. For-profit companies such as iCopyright also operate online rights licensing businesses.

Figure 1 describes how the BRR is initially intended to work with Google Book Search.

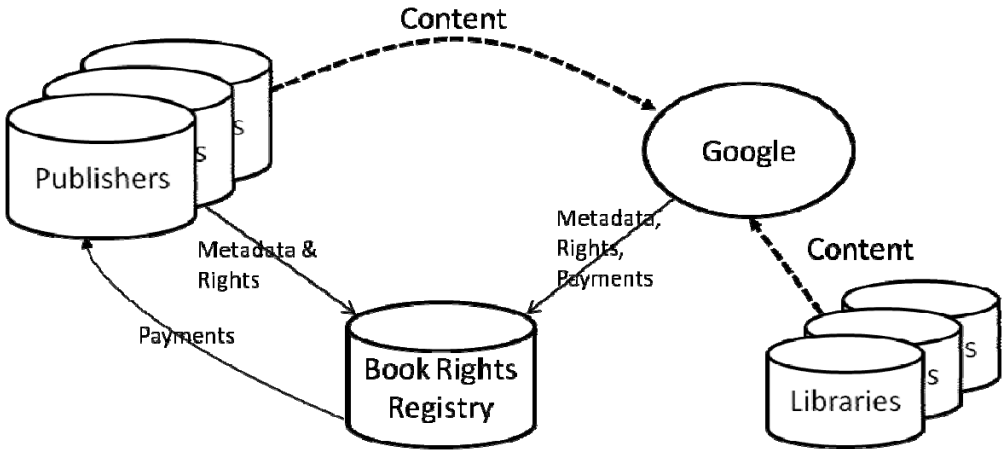

Figure 1: The Book Rights Registry as described in Section VI of the Settlement Agreement.

Google Book Search acquires content from books that participating libraries scan, that publishers physically send, or that publishers upload as PDF files. Google will send the BRR information about content it is making available for sale, including basic metadata (see p. 12) and rights information (see p.13). It will also send the BRR information about content it sells to users as part of the Settlement Agreement (see below). The BRR effectively will also get metadata and rights information from publishers, though in practice, this information will come from third-party metadata providers such as R.R. Bowker.

The BRR has great potential beyond the initial terms of the settlement to work with other search engines and third party service providers. In effect, the BRR could become a new type of collecting society that could support innovative content business models. It could serve as a platform for an unprecedented variety of online publishing services by ensuring that both content and rights are legally available.

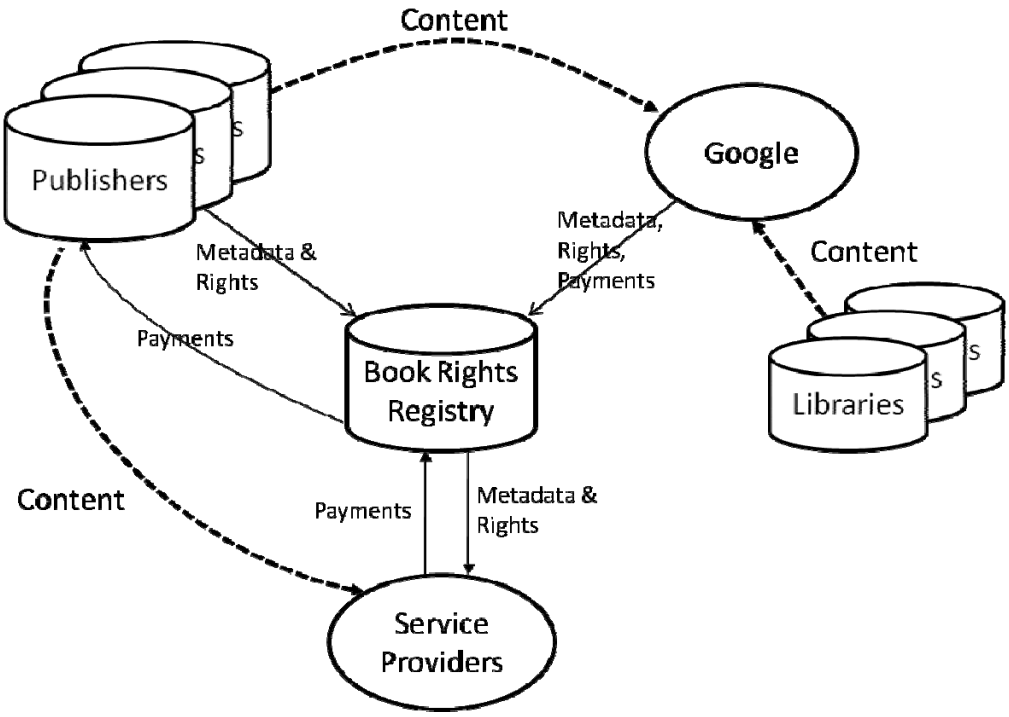

Figure 2: The Book Rights Registry can be extended to work with third party service providers.

Figure 2 shows how the initial BRR arrangement can be extended to work with third-party service providers. It is conceptually simple: Service providers – which could include other search engines that want to get into the business of content sales – can search the BRR for available content, package it according to their customers' needs, and use the BRR to distribute payments.

Crucially, the BRR will maintain a system of unique identifiers for content that will enable service providers to locate the actual content and obtain it from publishers. Although the Settlement Agreement leaves unspecified which identifier scheme the BRR will use, it has a choice of standard identifiers that are already used in segments of the publishing industry. As we will see, the BRR's value in supporting third-party service providers will depend on its ability to store sufficient metadata to enable searching, browsing, and so on. Service providers' use of the BRR's identifier standard will enable access to content.

Initial Business Models

Section IV of the Settlement Agreement specifies the initial set of business models that Google and publishers will implement. There are business models for institutions (schools, government agencies, and corporations) as well as consumers. Institutional pricing is not

explicitly defined in the Settlement Agreement; instead there is a requirement to price comparably to similar services that are currently available.

The consumer business model terms extend Google's current Google Book Search model, which offers publishers the following:

- Display of book content "snippets" in search results.
- Limited free previews configurable by publishers.
- Share of revenue from contextual ads (ads displayed on search result and preview pages that contain book content).
- Publisher-supplied links to web pages where users can purchase books or e-books, e.g., Amazon.com, Fictionwise.com.

The main new development in the Settlement Agreement, compared to the current Google Book Search model, is that consumers will be able to purchase rights to access digital books directly through Google. Google will provide an online reader application for this purpose. Users' access rights will include limited copy/paste (up to four pages with a single copy/paste command) and print (up to 20 pages per print command; print output will contain visible watermarks).

Prices for digital books can be set either by rightsholders or (with some restrictions) by Google. The Settlement Agreement contains provisions for discounting and sales through channels such as institutional consortia and affiliate programs.

Google will also be able to offer free previews of limited portions of book content, according to different models that depend on publishers' choices and the type of book involved. For example:

- Poetry and short story collections: no previews allowed.
- Reference volumes: fixed limit of 10% of pages in the book, plus front and back matter (table of contents, index, etc.).
- Fiction: final 15 pages of the book cannot be previewed.
- Otherwise, previews are generally limited to 20% of the pages in the book (plus front and back matter) and ranges of no more than five contiguous pages at a time.

In addition, Google and publishers will share revenue from ads shown on pages with previews or purchased book content, with 70% of ad revenue going to publishers and Google keeping 30%.

Finally, Google will provide free access to book content on a limited basis at not-for-profit higher education institutions and public libraries.

Future Business Models

Section 4.7 of the Settlement Agreement contemplates several future business models. These include:

- Print on Demand.
- Custom Publishing, including but not limited to per-page pricing.
- PDF downloads.
- Consumer subscriptions.
- Summaries, abstracts, and compilations.

Google may not offer any new business models without approval from rightsholders and the Book Rights Registry.

The most important difference between some of these future business models and the initial business models described above is that they are no longer based on pages as the primary unit of content.

The future business models contemplated in the Settlement Agreement will require Google to get content from publishers in digital formats other than PDF page images.

In the following section, we will discuss the implications of these new opportunities on the infrastructure that publishers, the Book Rights Registry, and Google itself will need to provide. We will focus on changes that many publishers will need to make in their content infrastructures and editorial processes in order to take advantage of these opportunities.

Opportunities for Publishers

The future business models contemplated in Section 4.7 of the Settlement Agreement differ qualitatively from the way that Google currently works with publishers – mainly in that they include several opportunities that require the availability of content in structural rather than page-oriented formats.

For example, while Print on Demand is a straightforward business model involving sending the same page images to short-run printing devices (potentially including retail kiosks and home printers) as publishers send to commercial book manufacturers, other future opportunities do not deal exclusively in page images.

One of the most interesting questions for Google in its adoption of these new business models is whether it will agree to acquire content from publishers in structured digital formats rather than in page images.

Even if Google decides not to implement any of the new business models, the Book Rights Registry is set up so that other service providers can. We will describe these new business models and their implications here.

Types of Future Opportunities

Custom Publishing

In the book industry meaning of the term, custom publishing means combining smaller units of content (e.g., chapters or "modules") into publications designed to be used by students in a class or by individuals[4]. The different modules of content could be formatted differently, even when they are supplied by a single publisher, meaning that simply stitching them together into a volume or digital file would not result in a quality product. Furthermore, good-quality custom publications should include indexes, footnotes or endnotes, and other ancillary content that is formatted consistently.

A few businesses already offer custom publications that are formatted consistently, such as VitalSource (Ingram Digital) and Safari Online (a joint venture of the publishers Pearson and O'Reilly & Associates); and some educational publishers operate their own custom publishing programs.

The next logical step beyond current custom publishing models is the ability to produce custom publications that include content from multiple publishers. Currently, McGraw-Hill Education's Primis Online offers the ability to combine modules from McGraw-Hill content with a limited selection of modules from Harvard Business School Publishing and other third parties. Similarly, the Google Settlement could radically expand this market.

Consumer Subscription Services

Subscription-based access to large repositories of book content is currently limited to models for research libraries such as ebrary; it is largely unknown in the consumer world. The subscription model is enjoying moderate commercial success in the music industry, where services like Rhapsody (RealNetworks) and Napster offer access to libraries of

[4] Custom publishing has a different meaning in periodical publishing, where it refers to private-label publications produced by service bureaus for clients such as airlines, investment firms, retail stores, clothing brands, etc.

millions of tracks of music for a monthly price, typically $10-13 per month. SVOD (Subscription Video On Demand) services from cable television networks like HBO are also similar.

Subscription models for book content would have to be more complex than they are for music, where the there is a single, simple unit of compensation for rightsholders: the track or song. Legacy search services in professional information markets like Lexis-Nexis and Factiva provide access to news stories and other types of content on various bases, including fees per document, lead paragraph, keywords in context, etc.

In publishing, it may be possible to compensate rightsholders based on the number of pages that a customer views, but this model doesn't apply very well to various types of content, such as short fiction, poetry, and various types of reference content. It would be more appropriate to apply compensation models to logical units of content in those cases.

Summaries, Abstracts, and Compilations

Abstracting and indexing (A&I) services are common in both print and online formats, especially in STM (scientific, technical, and medical) and other academic and professional segments. Customized A&I services can save a lot of effort if they have online access to content in uniform structures and formats as well as easily available information about rights.

Compilations are another form of custom publishing (see above), with even more implication for granularity of content below the chapter level and availability of content from multiple publishers.

The availability of the Book Rights Registry, and Google's ability to make copyrighted content discoverable, will radically enable the creation of scalable services that will implement these new business models.

These new business models will all be more successful if the services that provide them can offer content from multiple publishers.

With this in mind, there are three themes that are common to all future opportunities other than Print on Demand:

1. **Logical Structure**: Content should be available in logically structured and repurposable units rather than (or in addition to) page-oriented formats.

2. **Metadata**: Consistent and, in some cases, standardized metadata is required for describing content (keywords) as well as enumerating its bibliographic characteristics (author, date, publisher, etc.).

3. **Rights**: A uniform way of obtaining information about rights to content as well as rightsholder compensation information is necessary.

These are common themes in every publisher's content infrastructure strategy. And there is one technology that binds all three themes together: XML (eXtensible Markup Language). XML is standard technology for representing content, its logical structure, and metadata – including rights information – in a way that is independent from how it is

rendered (displayed to users) or handed off to downstream services such as search and aggregation.

The most important requirement for publishers in order to prepare for the business models described above is to adopt an XML-based content architecture that combines content with structure and metadata, as well as XQuery capabilities for powering the kinds of downstream services implicit in future business models.

XML-based content architecture also provides flexibility for *scalable product development* – implementing other new business models without having to build needlessly expensive, risky technology and process "silos." A growing number of publishers are finding that the right XML-based content architecture leads to major cost avoidance in new product and service development, the key to growth of any media business.

We will see how each of the three themes relates to the idea of XML-based content architecture and XML-related processes and standards.

Logical Structure

Publishers will need to supply content to Google and/or third-party service providers in logically structured formats.

XML is capable of representing content structure at as deep a level of detail as a publisher wishes. A growing number of publishers, especially smaller ones, are beginning to store content in XML that identifies logical structure down to paragraph level. STM publishers have generally made the greatest strides in this area.

In addition to tagging for logical structure, it is also important to identify each element's place in a sequence, such as paragraph numbers. This is necessary in order to represent business model constraints such as those that Google places on previews (see p. 7) and to ensure that logical elements can be viewed in the right order. For example, a user may retrieve a paragraph (or other logical element) in search results and then want to see the preceding and following elements to get a sense of context. Tagging elements with sequence information solves this problem without having to resort to page images.

Many publishers already have content stored in page-image format -- including native page-layout formats like QuarkXPress and Adobe InDesign as well as PDF – in digital asset management (DAM) systems such as Canto Cumulus, Documentum, OpenText Artesia TEAMS, and North Plains TeleScope. These systems have varying abilities to store content in XML, but in general, they treat XML as byproducts of traditional publishing processes based on proprietary page layout formats. Thus they are less than ideal for use as the basis of XML-based content architectures.

A growing number of publishers are implementing native XML repositories like Mark Logic. These are preferable to traditional DAM systems because they are more efficient at storage, retrieval, and especially querying of XML that will enable rapid access to content by Google or third party services.

Publishers are also beginning to convert their backlist content from page layout formats to XML, often on an opportunistic basis. There are several options for XML conversion, including large-scale conversion tools (such as CambridgeDocs and Exegenix),

converters used within authoring tools (such as Inera eXstyles for Microsoft Word), and third-party conversion/rekeying services such as Innodata Isogen and Publishing Dimensions.

Some publishers are opting to maintain their traditional DAMs alongside XML servers. This makes sense for certain types of publications that are heavily dependent on graphical elements, such as elementary school textbooks and art books; it is cumbersome to create such layouts through XML pagination systems and best to stick with traditional page layout tools for those, yet XML conversion is still necessary for online feeds.

However, an increasing number of publishers in other segments – such as STM and higher education – are moving to an "XML first" workflow with XML servers replacing DAM systems entirely. In many cases, those publishers have to stick with legacy content authoring tools such as Microsoft Word because authors are comfortable with them. Some publishers are beginning to adopt Word 2007, whose native .docx format is XML-based, for new authors. XML can be paginated for print (or PDF) output through tools such as Antenna House Formatter and RenderX XEP Engine, which use the XSL-FO (eXtensible Stylesheet Language – Formatting Objects) standard.

In addition to tools and processes, conversion to repurposable XML entails various editorial considerations. For example, custom publishing of textbooks can run into trouble over definitions of terminology. Chapters, sections, or modules must be edited so that they are as "standalone" as possible, so that when users combine them into custom publications, readers (e.g., students) don't find undefined terms.

Metadata

Publishers will need to make appropriate metadata available to the Book Rights Registry in order to enable future business models. It will be important to adopt metadata standards, particularly to enable business models that aggregate content from multiple publishers.

Metadata – information about information – is a very broad field. Publishers have had varying degrees of success in integrating the creation and management of complete, consistent metadata with their production processes. XML-based content architectures afford publishers unique opportunities to bind metadata to content and incorporate it into editorial processes, thereby enabling increased quantity and quality (consistency) of metadata.

The BRR will undoubtedly require feeds of standardized metadata from publishers in order to support its services. There are many metadata standards in the publishing industry, with varying degrees of applicability in market segments and varying degrees of success.

Two types of metadata are especially important for enabling the business models discussed here: bibliographic and descriptive.

The book publishing industry has adopted two standards that contain bibliographic metadata: ONIX and Dublin Core. Dublin Core[5] was invented in 1995 at a workshop hosted by OCLC (Online Computer Library Center) in Dublin, Ohio, a library research

[5] http://dublincore.org/.

organization. It consists of 15 elements that cover basic bibliographic metadata, plus fields for Identifier, Subject, Rights, and Type. Dublin Core has been incorporated into many more complex metadata standards including SCORM (Sharable Content Object Reference Model)[6] for educational content, PRISM (Publishing Requirements for Industry Standard Metadata)[7] for periodical content, and Adobe's XMP (eXtensible Metadata Platform)[8] for inter-application metadata portability.

ONIX[9] is maintained by the standards organization EDItEUR in conjunction with BISG (Book Industry Study Group) in the US, BIC (Book Industry Communication) in the UK, and other organizations in other countries. The original ONIX standard was designed to capture product metadata for books and ancillary products, as a means of automating book supply chain management from publishers through distributors to retailers. The latest release, 2.1, dates to 2006. ONIX contains bibliographic metadata fields as well as those for physical characteristics of books, such as weight, size, number to a carton, and so on.

Although ONIX and Dublin Core developed separately and for different purposes, it is possible to map core fields in one to the other[10].

There are many metadata standards that apply to more specialized types of publishing, such as SCORM, PRISM, and CrossRef[11] for scientific journal content. It would not be necessary for the BRR to adopt these. Instead, specialized metadata registries could build on the BRR's metadata scheme and use the BRR's unique identifiers[12] to link to content that the BRR tracks. For example, CrossRef already operates such a registry for journal article reference linking, although it does not handle financial aspects of journal article retrieval.

Rights

The Book Rights Registry will need to store information about the rights that publishers make available to Google and service providers to enable new business models.

Content rights can be stored alongside content as XML taqs and/or in separate databases. Many XML-based metadata standards, including PRISM and XMP, have fields for storing rights information for each content item. There are also XML-based rights metadata standards, including rights expression languages (RELs) for specifying rights in machine-readable form to digital rights management (DRM) technologies on users' PCs or

[6] http://www.adlnet.gov/scorm/.
[7] http://www.prismstandard.org/.
[8] http://www.adobe.com/products/xmp/.
[9] http://www.editeur.org/onix.html.
[10] See for example http://www.jisc.ac.uk/whatwedo/programmes/pals2/synthesis/themes/mapping.aspx. The author has also created this type of mapping .
[11] http://www.crossref.org/.
[12] The Settlement Agreement does not specify what identifier scheme the BRR will use. The book industry's longtime standard is ISBN (International Standard Book Number); ISBN will work for the initial business models but not for most of the future ones. For those, a more flexible identifier scheme that is also backward-compatible with ISBNs should be used, such as DOI (Digital Object Identifier).

consumer electronics devices[13], as well as standards for conveying rights licensing information to downstream business processes[14].

Some of the rights that the BRR will need to manage are incorporated or implied in the initial business models defined in the Settlement Agreement; see p. 7. These include rights to purchased content (e.g., copy/paste up to 4 pages, print up to 20 pages at a time) and the various rights to free previews that depend on the type of content (fiction, reference, poetry, etc.).

The BRR may not need to store precise, granular information about rights such as copy/paste, print, etc., when implementing the initial business models; it will suffice to refer to the type of content – as determined by BISAC (Book Industry Standards and Communications) classification codes[15]. Accordingly, publishers will not need to store detailed rights information when sending content to Google for the initial business models.

But future business models will require that the BRR track rights more explicitly as publishers may be able to decide whether to supply content to these services or not. Google Book Search and the initial business models laid out in the Settlement Agreement are generally not "opt in" for publishers, but future business models may be; and there will inevitably be disputes over whether certain business models are covered under copyright law versus those that require explicit permissions from rightsholders. For example, publishers may decide whether to allow service providers to create compilations, which are considered derivative works in copyright law.

Rights information is a slippery slope: it can very quickly become a complex mess of details about rights, and questions may even arise about whether service providers would be required to use technology such as DRM to enforce rights in users' devices or software. Google's ability to restrict users to a certain number of pages when they copy/paste or print is essentially a DRM function[16].

Some details about rights can be handled contractually rather than via technology. The BRR will probably be best off taking an approach to storing and handling rights information that is specific to the business models it supports while also being easily extensible to new ones. It will be necessary, for example, to track rights to use individual content items in subscription services. Publishers might like the idea of offering such rights for their entire catalogs but may well run into obstacles related to specific authors or legacy contracts[17].

Yet the success of complex rights information standards over the past few years has not been great, especially compared to the success of metadata standards like Dublin Core. Therefore it should be better to design the BRR's rights information capabilities to handle business models explicitly contemplated in the Settlement Agreement than to design them to support a wider range of hypothetical future business models. Publishers will eventually need to manage some types of rights metadata in their XML-based content architectures,

---

[13] Examples include XrML (eXtensible Rights Markup Language), used in certain Microsoft DRM systems, and ODRL (Open Digital Rights Language), a variant of which is used in the Open Mobile Alliance DRMs for wireless content delivery.

[14] Examples include ACAP (Automated Content Access Protocol) for news publishers and PLUS (Picture Licensing Universal System), used by digital stock image agencies.

[15] See Attachment F to the Settlement Agreement, http://www.googlebooksettlement.com/intl/en/Attachment-F-Preview-Uses.pdf.

[16] The same is true for Amazon.com's AmazonOnlineReader for Search Inside the Book and other functions.

[17] The major music companies have tried to do this with respect to music subscription services like Rhapsody and Napster. But certain well-known hold-outs – such as the Beatles for EMI and Led Zeppelin for Warner Music Group – have prevented them from doing so.

but this should be done on a relatively modest and gradual scale. XML provides the extensibility necessary to expand rights metadata to meet future needs.

Many publishers will need to build content infrastructure capabilities in order to support the future business models contemplated in the Settlement Agreement. Building such capabilities involves more than just acquiring technology and tools; it also requires evolving a publisher's processes, skill sets, and organizations.

As we have seen, the future business models contemplated in the Google Settlement lead to requirements for such capabilities, specifically an extensible XML-based content architecture that incorporates content, structure, metadata, and rights; includes strategic XML content servers; and provides the flexibility for scalable product development. GiantSteps Media Technology Strategies' Cross-Media Transformation methodology, for example, provides a framework for capturing capability requirements and translating them into actionable implementation plans.

Conclusions

Ever since the beginning of the networked digital age in the mid-1990s, publishers have been anticipating a world in which content can be sliced, diced, repurposed, and recombined for incremental reach and market share[18]. Publishers have had to move deliberately in adopting new digital business models for a number of reasons, including limitations on:

- Content created and managed in logically structured, repurposable form.
- Sufficient and consistent metadata to assist in searching, browsing, combining, and aggregating content.
- Contract and permission issues with authors and third party sources of content (stock image agencies, other publishers, etc.) that limit repurposability of content beyond its original form.
- Standardized and simple mechanisms for licensing content rights to providers of value-added content services.
- Standardized metadata for content discovery.
- Use of XML as a fundamental technology for managing content, metadata, and rights.

Google's settlement with the publishing industry does not eliminate all of these problems overnight, but it should help quite a bit with the last two by motivating developments on an unprecedented scale. This in turn should motivate the entire digital publishing ecosystem to move forward with new business models at a greater pace. Publishers will finally have the opportunity to make their content discoverable and conveniently available, thereby offering viable alternatives to the oceans of free and often inferior content available online.

Publishers, as owners of the content that makes all of these services and business models possible, must take the initiative to build capabilities that address the above limitations. Publishers have been working to adopt structure content management, metadata creation, and rights management issues. But now is the time to accelerate the pace of development.

About Mark Logic Corporation

Mark Logic Corporation is the provider of the industry's leading XML server. The company's flagship product, MarkLogic Server, allows customers to store, manage, search, and dynamically deliver content. The company has two patents on its innovative technology and is privately held. Sequoia Capital is the company's lead investor. Mark Logic has garnered several industry awards and recently been named the fourth fastest growing Silicon Valley Software and Information Technology (IT) company in Deloitte's Technology 2008 Fast 50 Program. To read the Mark Logic CEO Blog, visit http://marklogic.blogspot.com. To learn more about Mark

[18] For example, the AAP convened its Enabling Technologies Committee to look into this in 1994; some of the standards discussed in this white paper arose out of that activity.

Logic, or to download a free community or trial edition of MarkLogic Server, go to www.marklogic.com.

About the Author

Bill Rosenblatt, president of GiantSteps Media Technology Strategies, is a recognized authority on digital media technologies, including digital rights management, content management, cross-media strategy, and content production systems, as well as on issues related to intellectual property in the online world. Before founding GiantSteps in 2000, Bill was a business development executive at a leading technology vendor, an IT executive at major publishing companies, and chief technology officer of an e-learning startup. Bill is the author of several books, including *Digital Rights Management: Business and Technology* (John Wiley & Sons, 2001), and he is editor of the blog Copyright and Technology (http://copyrightandtechnology.com).

About GiantSteps Media Technology Strategies

GiantSteps Media Technology Strategies is a management consultancy focused on the content industries that help its clients achieve growth through market intelligence and expertise in business strategy and technology architecture. GiantSteps' clients have included branded content providers, digital media technology vendors ranging from early-stage startups to Global 500 firms, and technology public policy entities in the United States and Europe.

