La “descubribilidad” de los libros en la era de Internet

Máster en Economía Digital
e Industrias Creativas

2011

Marcos Pérez Sánchez
marcos.perez.sanchez@gmail.com
itruminations.wordpress.com
@grankabeza

Esta publicación está bajo licencia Creative Commons Reconocimiento, No Comercial, Compartir Igual (BY-NC-SA). Usted puede usar, copiar y difundir este documento o parte del mismo siempre y cuando se mencione su origen, no se use de forma comercial y no se modifique su licencia. Más información: http://creativecommons.org/licenses/by-nc-sa/3.0/

Índice

1 Introducción 4
1.1 De la escasez del contenido a la escasez de la atención 4
1.2 El nuevo papel de los editores 6
1.3 La "descubribilidad" 7
2 Los caminos hacia la "descubribilidad" 7
2.1 Buscadores 7
2.1.1 Google Books 8
2.2 Redes sociales 12
2.3 Comunidades de lectores y tiendas online 13
2.3.1 Comunidades de lectores 13
2.3.2 Amazon 16
2.3.3 Google eBookstore 17
2.3.4 Libranda 18
2.4 Interfaces de Programación de Aplicaciones (APIs) 20
2.4.1 Qué es una API 20
2.4.2 APIs en el mundo editorial 21
2.4.3 Google Books 22
2.4.4 O'Reilly Media 24
2.4.5 Goodreads 25
3 Metadatos 25
3.1 Qué son los metadatos 26
3.2 Tipos 27
3.3 Para qué sirven 27
3.3.1 Descubrimiento 27
3.3.2 Organización 28
3.3.3 Identificación 28

3.4 Estándares 28

3.5 ONIX 29
3.5.1 Origen 29
3.5.2 Quién lo controla 29
3.5.3 Quién lo utiliza 29
3.5.4 Por qué se creó 30
3.5.5 Evolución 31
3.5.6 ONIX y XML 31
3.5.7 ONIX en España 32

3.6 Calidad de los metadatos 33
3.6.1 Certificación 34

3.7 La identificación de textos en formato digital 34
3.7.1 ISBN (*International Standard Book Number*) 35
3.7.2 DOI (*Digital Object Identifier*) 36
3.7.3 ISTC (*International Standard Text Code*) 37

3.8 Sistemas de clasificación de materias 37

4 XML (*eXtensible Markup Language*) 38
4.1 Transformaciones de XML 39

4.2 Esquemas XML 39
4.2.1 DTD (*Document Type Definition*) 40
4.2.2 XSDL (*XML Schema Definition Language*) 40

4.3 Por qué usar XML 40

4.4 Qué aporta XML a las editoriales 41
4.4.1 ONIX y XML 41
4.4.2 Contenido y XML 41
4.4.3 Contenido vs. metadatos 43

Referencias adicionales 44

1 Introducción

1.1 De la escasez del contenido a la escasez de la atención

Durante más de 500 años, el negocio editorial se ha basado en la fabricación, distribución, difusión y venta de objetos físicos, los libros, que incorporaban las obras sujetas a propiedad intelectual.

Durante este tiempo, y especialmente desde la industrialización de la edición en el siglo XVIII, las editoriales controlaban los costosos medios de producción de los libros, y con ellos la oferta de títulos disponibles en cada momento.

La progresiva informatización de las editoriales, a partir de los años 70 del siglo XX, y en particular la difusión de las tecnologías digitales de impresión ha ido reduciendo gradualmente las barreras de entrada en la industria editorial. Esta tendencia se agudiza con la popularización del libro electrónico, lo que con toda probabilidad llevará a un crecimiento exponencial de la cantidad de títulos disponibles, con el añadido de que a los de nueva aparición se sumarán los de fondo de catálogo que volverán a la vida con la digitalización.

Las estadísticas dan cuenta de este crecimiento constante y cada vez más acelerado de la cantidad de contenido disponible:

Tabla. Número de ISBNs inscritos en España (2006-2010)

Año	Producción de libros en soporte papel	Producción de libros en otros soportes	Total
2006	82.940	4.500	87.440
2007	88.679	8.091	96.770
2008	95.508	8.715	104.223
2009	96.955	13.250	110.205
2010	95.959	18.500	114.459

Fuente: Panorámica de la edición española de libros[1]

Los números del mercado estadounidense son aún más llamativos, pues tienen en cuenta la explosión en los últimos años de la impresión bajo demanda[2] (incluida dentro de la rúbrica "no tradicional"; ver nota al pie), que ha disparado extraordinariamente el número de títulos publicados cada año.

Tabla. Número de títulos nuevos en Estados Unidos (2002-2010)

Año	Tradicional	No tradicional[3]	Total
2002	215.138	32.639	247.777
2003	240.098	26.224	266.322
2004	275.793	19.730	295.523
2005	251.903	30.597	282.500
2006	274.416	21.936	296.352
2007	284,370	123,276	407,646
2008	289,729	271,851	561,580
2009	302.410	1.033.065	1.335.475
2010 (previsto)	316.480	2.776.260	3.092.740

Fuente: Bowker[4]

[1] "Datos globales de la edición. Cuadros de evolución" (Observatorio del Libro y de la Lectura. Ministerio de Cultura) Consultado el 20 de junio de 2011 (http://www.mcu.es/libro/MC/PEE/estadisticas/globalesEvo.html)

[2] Impresión bajo demanda (*Print On Demand, POD*) es un proceso de impresión en el que las copias no se imprimen mientras no se ha recibido la comanda. Está basada en tecnologías digitales de impresión, que han permitido que este modelo sea económicamente rentable.

[3] "No tradicional consiste en su mayoría en reimpresiones, a menudo de obras en dominio público, y otros títulos en impresión bajo demanda.", Bowker (http://www.bowkerinfo.com/pubtrack/AnnualBookProduction2010/ISBN_Output_2002-2010.pdf)

Pero estas estadísticas no reflejan aún el impacto en el número de títulos de la publicación directamente en formato digital que, especialmente en Estados Unidos (debido al impulso de las editoriales pero, sobre todo, de los programas de autopublicación de importantes actores como Amazon o Barnes & Noble), incrementan aún más el número total de títulos que se publican al año y que invita a pensar que la tendencia se exacerbará en los próximos años.

Aunque lo más probable es que los libros impresos y electrónicos convivan durante mucho tiempo, la situación está cambiando de forma radical y a un ritmo vertiginoso: por una parte, la fabricación del libro en papel, que aún es la forma principal, va a pasar a ser un medio entre otros de distribuir el contenido sobre el que la editorial posee derechos de propiedad intelectual; por otra, la distribución digital de contenidos hace posible una relación directa entre la editorial y el lector, aunque también permite que otros actores (agentes, canales de venta como Amazon, Barnes & Noble, entre otros) compitan directamente con los editores, permitiendo la autopublicación o incluso creando o comprando sellos editoriales.

1.2 El nuevo papel de los editores

Si quieren mantener su posición en este nuevo escenario de sobreabundancia de libros (que, además, compiten de una forma cada vez más directa con otras formas de cultura o entretenimiento por la atención de los posibles lectores), las editoriales, cuyo negocio había acabado basándose en la venta de los libros a otros intermediarios (distribuidores, librerías), y no a los consumidores finales, los lectores, han de reinventarse para convertirse en creadores y gestores de comunidades, encauzando y aprovechando la atención que estos presten a sus libros.

En palabras de Hugh McGuire[5], además de sus labores tradicionales de filtrado y revisión de los textos, ahora las editoriales "deben ayudar a los lectores a pasar más tiempo con los libros", facilitando la conversación que se genera a su alrededor, entre los lectores e incluso con los autores, y recopilando y agrupando toda la información relacionada con los libros que pueda ser de interés para los lectores (noticias, entrevistas con el autor, premios...). Deben buscar la manera de que los lectores "puedan hacer más cosas con los libros", facilitando la difusión e incluso reutilización del contenido (en redes sociales, mediante widgets o interfaces de programación).

---

[4] "Print isn't dead, says Bowker's Annual Book Production Report", Bowker (http://www.bowker.com/index.php/press-releases/633-print-isnt-dead-says-bowkers-annual-book-production-report)

[5] "Trailblazing: Leading The Way to A New Kind of Supply Chain", Hugh McGuire y Mark Lefebvre (vídeo), Booknet Canda Tech Forum 2010 (http://blip.tv/booknet-canada/hugh-mcguire-and-mark-lefebvre-trailblazing-leading-the-way-to-a-new-kind-of-supply-chain-3527158)

Esta transformación de las editoriales de empresas B2B (*Business To Business*) a B2C (*Business To Consumer*) conlleva para ellas la necesidad de adquirir nuevas habilidades que les permitan conocer mejor a sus lectores como, por ejemplo técnicas de análisis de datos para obtener información de los hábitos de los lectores, o "escucha activa" en redes sociales y gestión de las comunidades de lectores.

1.3 La "descubribilidad"

Pero incluso más importantes que los cambios directos que las nuevas tecnologías provocan en la propia industria son las nuevas formas de relación con el contenido que propician en los lectores.

Entre ellas, una paradigmática de la era de Internet: la búsqueda. Los años de uso cotidiano de la Red han ido acostumbrando a los lectores no solo a leer en pantalla (de ordenador, de teléfono móvil, de ereader o tablet) sino a buscar y compartir información y opiniones sobre sus lecturas en la Red, ya sea a través de los propios buscadores, de las redes sociales, o incluso de plataformas de venta online como Amazon.

A esto se une la progresiva disminución del número de librerías, que han sido tradicionalmente el medio más importante a través del cual los posibles lectores entraban en contacto con los libros y tenían noticia de las novedades del mercado. Y, en el caso del contenido nacido directamente en formato digital, ni siquiera cabe la posibilidad de que el posible lector sepa de él fuera de la Red.

Por todo esto, las editoriales deben buscar nuevas maneras de hacer que los lectores conozcan, en primera instancia, la existencia misma de los libros y puedan obtener fácilmente toda la información que necesiten para decidirse a comprarlos. En resumen, utilizando un neologismo ad hoc, las editoriales han de potenciar la "descubribilidad" de los libros.

2 Los caminos hacia la "descubribilidad"

2.1 Buscadores

Un requisito fundamental para cualquier contenido que pretenda hacerse visible en Internet es el de optimizar su posicionamiento en los resultados que los principales buscadores (Google, Yahoo!, Bing) ofrecen a sus usuarios, de forma que aparezca entre los primeros resultados (en particular, en la

primera página; mejor aún, entre los tres primeros[6]) para las búsquedas más frecuentes relacionadas con el propio contenido.

Los buscadores distinguen entre sus resultados orgánicos, por los que los gestores de los sitios web (*webmasters*) no pagan a los buscadores por el posicionamiento, y sus resultados de pago, "enlaces patrocinados" o, directamente, "anuncios".

A los primeros va dirigida la optimización de resultados en buscadores (*Search Engine Optimization*, SEO), que engloba todo un conjunto de técnicas y prácticas con el objetivo común de mejorar la posición del contenido en cuestión en los resultados orgánicos de los grandes buscadores cuando el usuario introduce las palabras clave de interés en la caja de búsqueda de su navegador.

Por su parte, el marketing en buscadores (*Search Engine Marketing*, SEM) tiene como objetivo mejorar el posicionamiento del contenido en los resultados de pago de los buscadores.

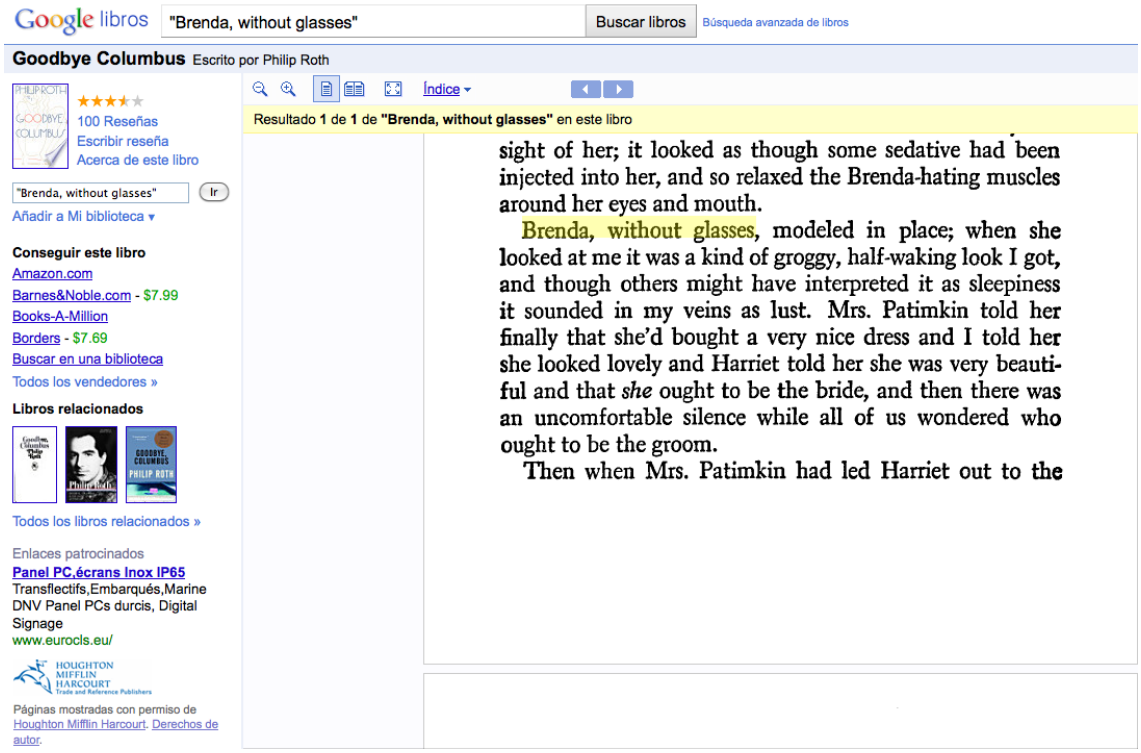
En ambos casos, la optimización del posicionamiento en los buscadores es una tarea continuada, estrechamente relacionada con la analítica web[7] (esto es, la "recopilación, medición, evaluación y explicación racional de los datos obtenidos de Internet, con el propósito de entender y optimizar el uso de la página web de la organización"[8]), que permite tomar decisiones al respecto fundamentadas en datos objetivos que caracterizan el comportamiento de quienes visitan los sitios web.

2.1.1 Google Books

Además de su buscador generalista, Google ha desarrollado uno especializado en libros, Google Books[9], que permite la búsqueda no sólo a partir de información bibliográfica (título, autor…) sino también del propio texto completo del libro (ver captura de pantalla) y ofrece información sobre lugares donde obtener el libro, tanto tiendas como bibliotecas, así como libros relacionados con él. Los resultados de Google Books aparecen también cuando se realiza una búsqueda en el buscador generalista de Google.

---

[6] "Why Publishers Should Care About SEO", Jamie Low, fundador de SearchEngineMarketing.com, 22 de octubre de 2008 (http://www.youtube.com/watch?v=gAkEilpmdSE)
[7] "Web Analytics Primer for Publishers", Mac Slocum, O'Reilly Media, 2 de septiembre de 2008 (http://radar.oreilly.com/2008/09/web-analytics-primer-for-publi.html)
[8] "Carta de bienvenida", Asociación Española de Analítica Web (http://www.aeaw.es/que-es-la-aeaw/carta-de-bienvenida/)
[9] http://books.google.com/



Fuente: captura de pantalla de búsqueda por texto completo en Google Books, 19 de junio de 2011.

La polémica de Google Books

En octubre de 2004, durante la Feria del Libro de Fráncfort, Google presentó *Google Print*, un acuerdo con varias importantes editoriales (McGraw-Hill, Cambridge University Press, Oxford University Press, Pearson o Penguin, entre otras) por el que el buscador podría mostrar entre los resultados de las consultas, junto a los enlaces a las páginas web relevantes, pequeños extractos (*snippets*) de los libros en los que apareciesen los términos buscados. En diciembre de ese año, presentaba asimismo su proyecto *Google Library* en el que, de acuerdo con varias de las principales bibliotecas estadounidenses (de las universidades de Harvard, Stanford y Michigan y de la New York Public Library) y la biblioteca de la Universidad de Oxford, Google escanearía digitalmente unos 30 millones de volúmenes de sus colecciones, incluyendo tanto las obras en dominio público como ejemplares aún bajo copyright, para incorporarlos a la base de datos de su buscador. Mientras que de los libros en dominio público Google mostraría el texto entero del libro, de libros aún protegidos por la propiedad intelectual únicamente permitiría ver hasta tres pequeños fragmentos por libro,

cada uno de ellos de unas pocas líneas. Ambas iniciativas constituyen lo que actualmente se conoce como *Google Books*[10].

Aunque Google ofrecía a los titulares de los derechos la posibilidad de solicitar que sus obras se excluyesen del programa de escaneo de las bibliotecas (sistema *opt-out*), así como de incorporarse al programa para editores, por el que recibirían parte de los ingresos que Google obtuviese por los anuncios situados en las páginas donde se mostrasen los fragmentos de sus libros, encontró gran resistencia a *Google Library*, pues muchos de los titulares de los derechos de propiedad intelectual sobre los libros, autores o editores, consideraban que tanto el escaneo completo de los textos para incorporarlos a la base de datos del buscador como el hecho de mostrar pequeños fragmentos de los libros alrededor de los términos buscados constituían una vulneración de sus derechos exclusivos, ya que Google debería haber contado con su autorización previa para poder hacerlo legalmente.

La disputa con autores y editores dio lugar a sendas demandas en Estados Unidos. El 20 de septiembre de 2005, el Gremio de Escritores (*Authors Guild*) demandó a Google alegando que *Google Library* vulneraba sus derechos de propiedad intelectual. Un mes más tarde, varias editoriales estadounidenses hicieron lo propio. Ambos casos acabaron integrándose en uno solo.

Sin embargo, lo que prometía ser un importante caso de aplicación del *fair use* (un límite a los derechos de propiedad intelectual contemplado en la legislación estadounidense, cuya aplicación ha de determinar el juez caso por caso, siguiendo los criterios establecidos en la ley) dio un vuelco cuando, tras tres años de litigio, el 28 de octubre de 2008, las partes presentaron ante el juez un acuerdo, revisado en 2009, que iba mucho más allá de lo que en principio se estaba dirimiendo con la demanda. En resumen, el acuerdo permite que Google prosiga con el escaneo de los libros para incorporarlos a su buscador, a cambio de lo cual pagará a los titulares. De aprobarse el acuerdo, Google podría no sólo mostrar fragmentos de las obras (hasta el 20% de obras protegidas), sino también comercializar el acceso a los libros tanto por clientes finales como por instituciones.

El acuerdo trata de distinta manera las obras actualmente a la venta (*commercially available*) y las descatalogadas (*not commercially available*), estableciendo las siguientes reglas por defecto:

- Para libros a la venta: no se incluirán en los servicios de venta ni de suscripción institucional; sólo se mostrará información bibliográfica.

[10] "El futuro del libro 8. El proyecto Google Books", Marcos Pérez Sánchez, 21 de febrero de 2011, (http://itruminations.wordpress.com/2011/02/21/el-futuro-del-libro-8-el-proyecto-google-books/); "El futuro del libro 9. El acuerdo sobre Google Books", Marcos Pérez Sánchez, 28 de febrero de 2011, (http://itruminations.wordpress.com/2011/02/28/el-futuro-del-libro-9-el-acuerdo-sobre-google-books/)

- Para libros descatalogados: se permitirán los servicios de visualización, salvo que el titular se oponga.

- Además, Google pondrá libremente a disposición de sus usuarios las obras en dominio público, incluida la posibilidad de descargarlas.

Un aspecto especialmente controvertido es el relativo a las obras huérfanas, pues sus detractores entienden que el acuerdo supone de facto una licencia exclusiva a Google para explotarlas comercialmente; frente a lo cual hay quienes ven en el acuerdo la posibilidad de que parte de la cultura del siglo XX escape al "agujero negro" en que habría quedado sumida, provocado por la prolongación de los plazos de protección de la propiedad intelectual y los altísimos costes de transacción en que habría de incurrir quien acometiese la tarea de localizar a todos los titulares de las obras que aún se encuentran protegidas pero que dejaron de explotarse comercialmente hace tiempo para obtener su permiso para convertirlas y utilizarlas en formato digital.

Por tratarse de una demanda en forma de *class action*, en la que los demandantes representan a todos los titulares de derechos sobre las obras afectadas que no manifiesten su voluntad de no participar del acuerdo, para que sea este sea válido ha de ser aprobado por el juez. El 22 de marzo de 2011, el juez rechazó el acuerdo y emplazó a las partes a modificarlo para que fuese aceptado, argumentando que su ámbito, en particular en lo referente a los usos futuros por parte de Google de los libros digitalizados, excedía con creces lo que la demanda inicial debía dirimir, a saber, si el escaneo de los libros para su incorporación a la base de datos del buscador y la muestra de pequeños fragmentos en respuesta a consultas de los usuarios suponía una infracción de los derechos de los titulares o bien cabría dentro del límite del *fair use*. A fecha de hoy (junio de 2011), el futuro del acuerdo es incierto.

Situación en Europa

Dado que la legislación europea sobre propiedad intelectual no contempla la particularidad del *"fair use"*, en Europa Google Books sólo se alimenta de las dos fuentes que se mencionan en la imagen: por un lado, los libros aún protegidos por la propiedad intelectual cuyos titulares han alcanzado acuerdos con Google para aparecer en sus resultados de búsqueda; por otro, los libros en dominio público escaneados como resultado de los acuerdos a los que Google ha llegado con varias grandes bibliotecas. Por tanto, no aparecen en los resultados ni las obras huérfanas, ni las descatalogadas, ni las de aquellos editores que han optado por no sumarse al programa para editoriales de Google.

Fuente: "Internet: Un nuevo ecosistema del libro en Internet", presentación de Luis Collado (Google España) en EOI, 27 de enero de 2011.

2.2 Redes sociales

Uno de los fenómenos dominantes en Internet en los últimos años ha sido el espectacular crecimiento de las redes sociales, sitios web que permiten a sus usuarios crear un perfil personal y establecer relación con otros usuarios ("amigos" en Facebook, "seguidores" en Twitter).

Las empresas han visto en las redes sociales una oportunidad para conectar de una forma más directa con sus potenciales clientes, y es rara la organización de cierto tamaño que no tiene presencia en alguna de ellas, en particular las más utilizadas, como son por ejemplo son Facebook y Twitter, entre las generalistas, o LinkedIn, entre las de propósito más específico.

En particular, interesa a quienes pretender vender cualquier producto o servicio el poder de la recomendación, pues parece demostrado que la mayoría de las personas da más valor y confianza a lo que le recomienda un "amigo" que a la información promocional de las marcas.

Otra de las características de las redes sociales que las marcas tratan de aprovechar es la denominada "viralidad", la posibilidad de que la circulación de que un contenido aumente exponencialmente gracias al boca a oreja digital.

Para sacar provecho efectivo de las oportunidades que las redes sociales ofrecen, las técnicas utilizadas en SEO y SEM tuvieron que adaptarse a estas nuevas plataformas, dando lugar a lo que se conoce actualmente como Social Media Optimization (SMO) y Social Media Marketing SMM)

En general, se trata de crear contenido que atraiga la atención e incite a los usuarios a compartirlo con su círculo de contactos, que supuestamente le darán mayor credibilidad al mensaje al provenir de alguien de confianza, y no de la propia de la marca.

Por tanto, no sólo se trata de tener presencia en redes sociales (ya sea la editorial, el autor o incluso el propio libro), sino de diseñar el contenido para favorecer y facilitar que se comparta, ya sea mediante enlaces, widgets, o cualquier otro medio que reduzca al mínimo el esfuerzo que un usuario tenga que hacer para compartirlo.

2.3 Comunidades de lectores y tiendas online

2.3.1 Comunidades de lectores

Como ya se ha dicho, tradicionalmente la industria editorial tenía como clientes directos a los puntos de venta que eran quienes, a su vez, tenía relación con el cliente final, el lector, y por tanto quienes mejor podían conocer sus gustos e intereses.

Esta situación cambia con la transición al mundo del libro digital, pues a los productores del contenido (y a nuevos intermediarios) se les abre ahora la posibilidad de mantener relación directa con los lectores e incrementar así su capacidad de influencia a la hora de moldear sus compras.

Para que esta influencia sea efectiva, sin embargo, no basta con conocer las afinidades de los lectores (a través sus historiales de compra, por ejemplo) y el gráfico social del usuario (sus contactos). Hace falta un "motor de recomendación" que una todas las piezas y proporcione al comprador potencial información útil y valiosa para ayudarle a responder a la gran pregunta: ¿qué libro leo ahora?

Otra tendencia profunda del mundo digital impulsa también a las editoriales, y no solo a ellas, hacia la creación y gestión de comunidades de lectores (y de escritores, en muchos casos): la sobreabundancia de contenido provocado por el descenso de las barreras de entrada en el mundo de

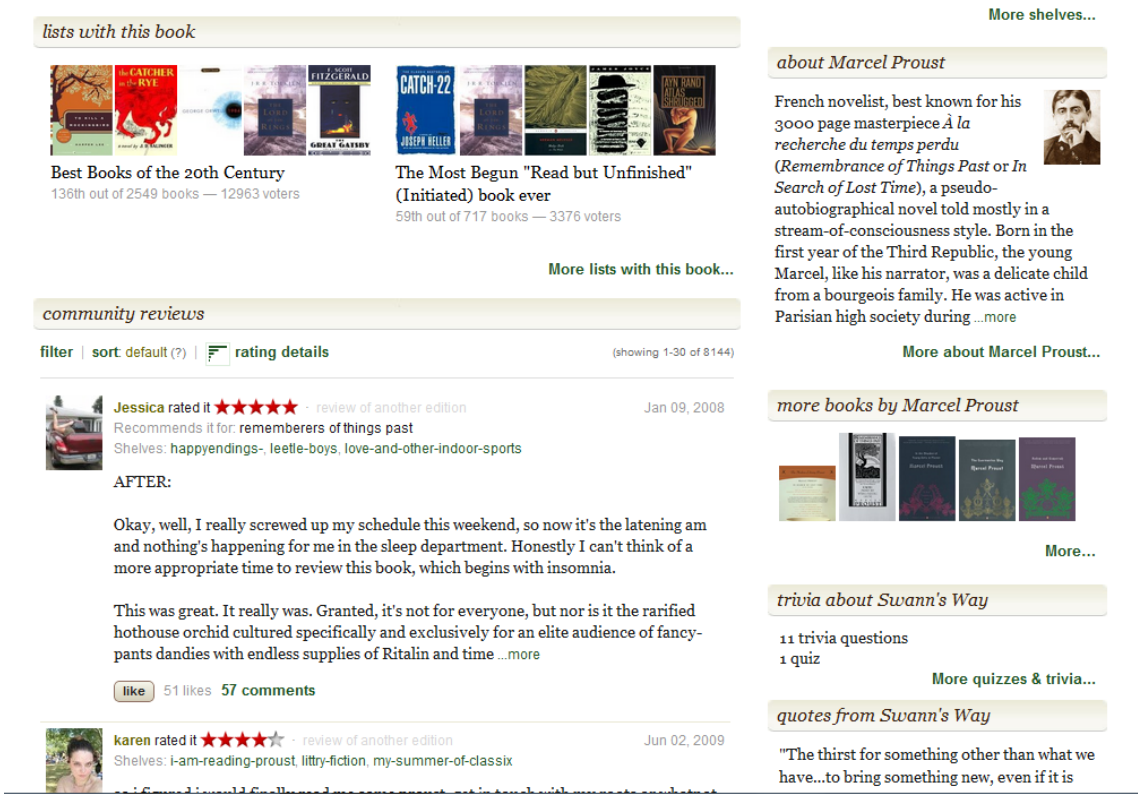
la edición (en particular, de la digital). Este exceso de oferta hace que el valor para el cliente se traslade del propio contenido es sí a los servicios relacionados con él. Servicios que, según Kevin Kelly[11], fundador de la revista *Wired,* tendrán algunas de estas características: inmediatez, personalización, interpretación, autenticidad, accesibilidad, personificación, mecenazgo, “encontrabilidad”. En resumen, harán más valioso un contenido que, por sí solo, puede que no sea suficiente para que los lectores paguen por él. Y para ofrecer esos servicios al lector es imprescindible haber establecido con él una relación directa.

Así, en los últimos años han ido apareciendo multitud de ejemplos de comunidades de lectores de distinto tipo.

Algunas son “independientes”, promovidas por entidades tradicionalmente ajenas al mundo editorial, como Goodreads, Shelfari, BookGlutton, LibraryThing o Copia, o, en España, Lecturalia.

[11] "Better than Free", Kevin Kelly, enero de 2008 (http://www.kk.org/thetechnium/archives/2008/01/better_than_fre.php) (traducción española, “Mejor que gratis”: http://www.uncafelitoalasonce.com/mejor-que-gratis/)

Fuente: Goodreads.com[12]



Fuente: Goodreads.com

Otras han sido promovidas por editoriales tradicionales, como MeGustaLeer y MeGustaEscribir (dirigida a escritores), tras las que se encuentra Random House Mondadori, en España; o Book Country, que pretende aglutinar a lectores exclusivamente de ficción y que, si bien es obra de Penguin USA, asegura que promoverá en igualdad de condiciones libros de otros sellos editoriales[13]. O aNobii, una red social de lectores nacida en Hong Kong y adquirida[14] por un grupo de inversores británicos en el que figuran HarperCollins, Penguin y el grupo Random House junto con el grupo HMV, propietario no solo de la mayor cadena de tiendas de entretenimiento del Reino Unido sino también de la mayor cadena de librerías, Waterstone's.

Otro ejemplo en ciernes, pues se espera su estreno para el verano de 2011, es Bookish, fruto de la colaboración entre Penguin, Hachette y Simon & Schuster, tres de las mayores editoriales

[12] "A Connected Bookshelf" (presentación), Hadrien Gardeur (Feedbooks), Books in Browsers 2010 (http://www.slideshare.net/HadrienGardeur/a-connected-bookshelf)
[13] "About Us. What's the Penguin About?", Book Country (http://bookcountry.com/AboutUs.aspx#sub2)
[14] "Good News", The aNobii Blog, 2 de marzo de 2011 (http://blog.anobii.com/post/3597628742/good-news)

estadounidenses, y que significativamente tendrá su sede en Silicon Valley[15], centro neurálgico de las grandes empresas de Internet, como Apple o Google.

Otras incluso están promovidas por toda una industria editorial, como es el caso de Canadian Bookshelf[16], tras la que se encuentran la *Association of Canadian Publishers* (Asociación de Editores Canadienses), el *Canadian Publishers Council* (Consejo de Editores Canadienses) y BookNet Canada, organización para el fomento de la innovación y la aplicación de la tecnología en la industria editorial canadiense, y que cuenta también con apoyo institucional.

Muchas de estas comunidades incorporan funcionalidades de tienda virtual y otras probablemente lo acaben haciendo, o busquen otras maneras de "monetizar" la relación entre los lectores con el contenido, o con los propios autores. Este es el caso de la editorial Redlemona.de, de Richard Nash, "una red social de autores que comparten sus manuscritos con el fin de mejorarlos o enriquecerlos de manera colectiva" [17], la primera de las "50.000 editorial independientes"[18] que Nash pretende propulsar mediante su plataforma Cursor: "Mitad sitio web, mitad proceso de negocio, combina la discriminación editorial tradicional con la sabiduría de la multitud, aprovechando la pasión compartida de escritores y lectores. Hace posible así una amplia variedad de nuevas fuentes de ingresos para el autor y la editorial y nuevas experiencias culturales para escritores y lectores".

2.3.2 Amazon

Amazon es la tienda online por antonomasia. Aunque empezó dedicándose a la venta a través de Internet de libros impresos, en los más de quince años desde su fundación no ha dejado de crecer, ampliando su catálogo de productos y servicios mucho más allá de los libros, y al mismo tiempo, en el ámbito editorial, tratando de ocupar cada vez más espacio.

Así, en 2007 presentó su lector de libros electrónicos, el Kindle, que fue el primer gran catalizador del gran crecimiento que se ha producido en el mercado de los libros electrónicos (primero en Estados Unidos, y ahora cada vez más en el resto del mundo). Después, además de sus programas de

[15] "Bookish CEO: 'Thrilled To Have A Dual Revenue Stream'", PaidContent, 7 de mayo de 2011 (http://paidcontent.org/article/419-bookish-ceo-thrilled-to-have-a-dual-revenue-stream/)
[16] "Canadian Bookshelf, eh?", BookNet Blog, 9 de junio de 2011 (http://www.booknetcanada.ca/index.php?option=com_wordpress&p=2290&Itemid=319)
[17] "Redlemonade. Nueva editorial basada en la web social", Comunicación Cultural, 10 de mayo de 2011 (http://www.comunicacion-cultural.com/2011/05/10/redlemonade-nueva-editorial-basada-en-la-web-social)
[18] "A soft drink's soft launch, Red Lemonade goes live…", Richard Nash, 11 de mayo de 2011, (http://rnash.com/article/a-soft-drinks-soft-launch-red-lemonade-goes-live/)

autopublicación para Kindle, ha ido creando una serie de sellos editoriales[19] especializados, por ejemplo, en novela romántica o en traducción al inglés de obras publicadas originalmente en otros idiomas.

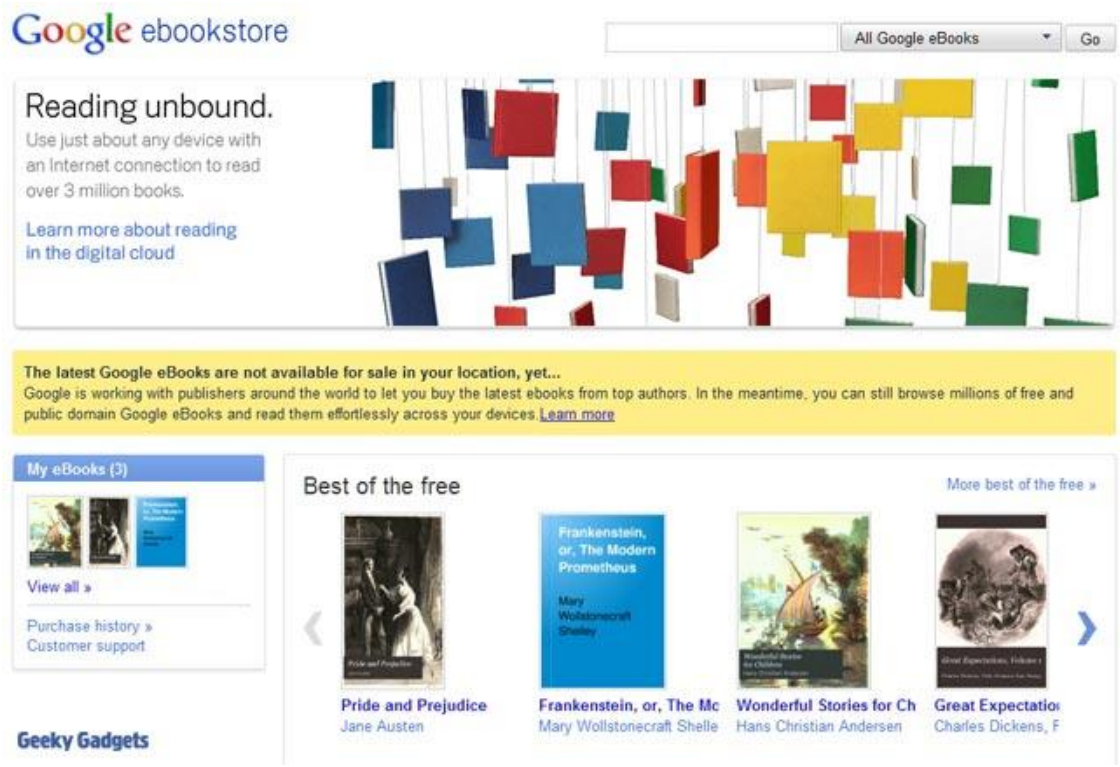
Parte del éxito de Amazon, junto a la amplitud de su oferta y lo ajustado de sus precios, radica en el conjunto de mecanismos que ayudan a sus usuarios a orientarse en sus compras.

Estos mecanismos incluyen una parte relacionada con el propio historial de navegación y compra del usuario en Amazon; otra parte social, como son las críticas y recomendaciones que otros usuarios hacen los libros; y otra relacionada con el contenido en sí, como es la posibilidad de hojear parte del contenido del libro antes de comprarlo, así como de realizar búsquedas sobre el texto completo del libro[20] o, en el caso de libros electrónicos, de descargarse las primeras páginas.

2.3.3 Google eBookstore

En diciembre de 2010, Google abrió al público estadounidense su tienda de libros electrónicos[21], que aún (junio de 2011) no está disponible en España. Su modelo[22] está basado en la lectura en la nube, de forma que los libros se leen desde el navegador web o desde aplicaciones específicas, para lo cual los dispositivos han de estar conectados a Internet.

[19] "How many imprints does Amazon run?", Jenn Webb, O'Reilly Radar, 5 de mayo de 2011 (http://radar.oreilly.com/2011/05/amazon-imprints-publishing.html)
[20] "Amazon Search Inside"(http://www.amazon.com/Search-Inside-Book-Books/b?ie=UTF8&node=10197021)
[21] Google eBookstore (http://books.google.com/ebooks)
[22] "Introducing Google eBooks", Google (http://www.youtube.com/watch?v=ZKEaypYJbb4)



Fuente: Google eBookstore

Antes de su lanzamiento, Google había llegado a un acuerdo con la American Booksellers Association, la asociación de librerías independientes estadounidenses, por que el unas 200 librerías utilizarían la plataforma Google eBooks para vender libros electrónicos[23], y en junio de 2011 entró en funcionamiento el programa de afiliados de Google para ebooks[24], de forma que no sólo las librerías asociadas sino cualquier otro sitio web (blogs, portales temáticos…) podrá obtener una comisión en la venta de libros electrónicos a través de Google eBooks.

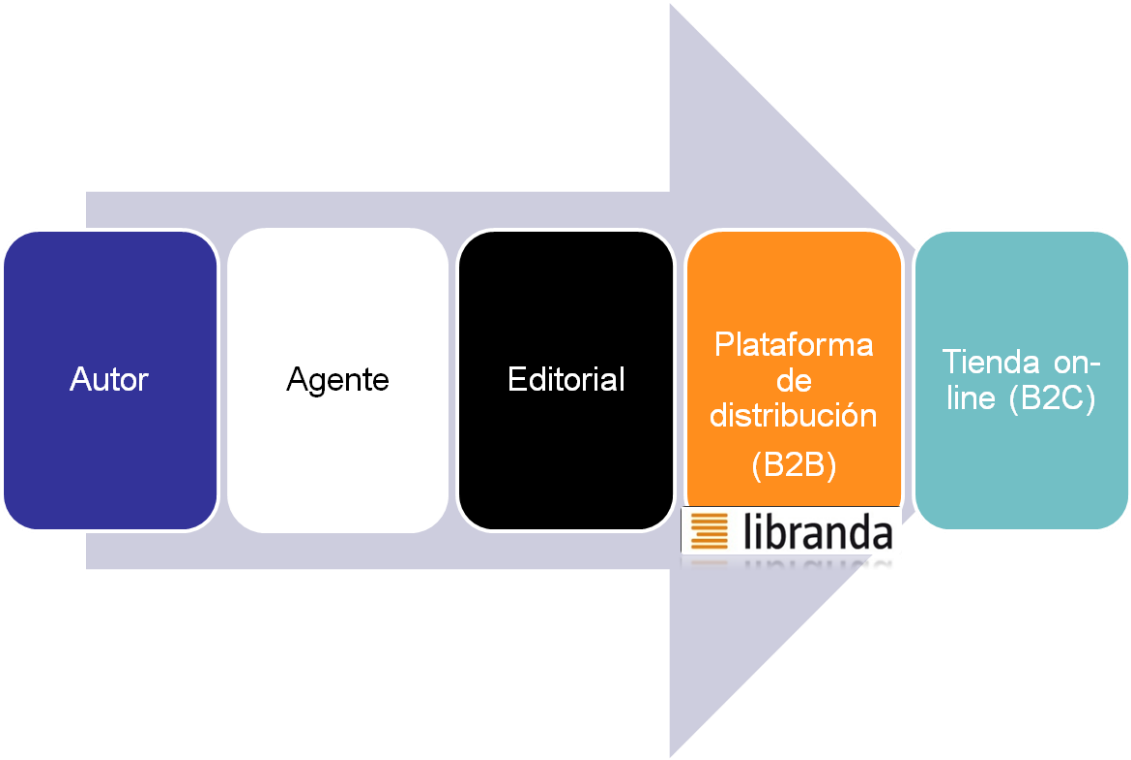
2.3.4 Libranda

En julio de 2010 entró en funcionamiento Libranda[25], una plataforma de distribución de libros electrónicos y otros servicios digitales, constituida inicialmente por 7 grupos editoriales españoles (Grupo Planeta, Grupo Random House Mondadori, Grupo Santillana, Roca Editorial, Grup62, Grupo

[23] "Independent bookstores selling Google Ebooks", American Booksellers Association, 6 de diciembre de 2010 (http://www.indiebound.org/google-ebooks)
[24] "Google eBooks Affiliate program - Open for business", Pratip Banerji, Product Manager, Google Books team, 16 de junio de 2011 (http://booksearch.blogspot.com/2011/06/google-ebooks-affiliate-program-open.html)
[25] "El 15 de julio, explosión digital (controlada)", El País, 9 de junio de 2010 (http://www.elpais.com/articulo/cultura/julio/explosion/digital/controlada/elpepucul/20100609elpepucul_7/Tes)

SM y Grupo Wolters Kluwer) y cuyos clientes son las propias editoriales y las tiendas online, pues Libranda, pese a una cierta confusión inicial entre el público, no se dirige al consumidor final.

La intención de Libranda es "ayudar a editoriales y tiendas online a "gestionar el entorno digital ofreciéndoles un conjunto de servicios globales" y así "contribuir a la difusión del libro digital [en español y en catalán] a nivel internacional, haciendo llegar los contenidos al mayor número de puntos de venta y procurando que sean legibles en la mayoría de dispositivos de lectura existentes en el mercado"[26].



Fuente: Libranda.

Actualmente (junio de 2011), Libranda trabaja con más de 100 editoriales[27] y más de 60 tiendas online[28], la mayoría de ellas dirigidas a clientes europeos, si bien también las hay de varios países latinoamericanos (Argentina, Colombia y Chile); Barnes & Noble, la gran cadena de librerías estadounidense, se sirve de Libranda para su oferta de libros electrónicos en español[29].

[26] "¿Quiénes somos y qué ofrecemos?", Libranda (http://www.libranda.com/default.aspx?id_seccion=803)
[27] "Editoriales", Libranda (http://www.libranda.com/default.aspx?id_seccion=819)
[28] "Tiendas", Libranda, (http://www.libranda.com/default.aspx?id_seccion=822)
[29] "Nook Books en español", Barnes & Noble (http://www.barnesandnoble.com/ebooks/category.asp?PID=28507)

Libranda ofrece a las editoriales, entre otros servicios[30], la gestión del "catálogo digital y sus metadatos para posicionarlo en las distintas tiendas on-line de forma rápida y segura", así como "la creación de herramientas de marketing para la promoción de contenidos (ej. Widgets) o la creación de páginas web blancas para que las librerías que no tengan tienda *on-line* puedan disponer de ella"[31].

2.4 Interfaces de Programación de Aplicaciones (APIs)

El siguiente paso para extender la distribución (controlada) del contenido de las editoriales en Internet más allá de las redes sociales y las tiendas y comunidades de lectores, ya sea de información sobre los libros o del propio contenido, pasa por la creación de APIs para los libros. (En los últimos meses, se han empezado a celebrar, de momento en Nueva York[32] y en Toronto[33], encuentros entre gente del mundo del libro tradicional y programadores, cuyos resultados probablemente se empiecen a ver en los próximos meses.)

2.4.1 Qué es una API

Una API (*Application Programming Interface*; Interfaz de Programación de Aplicaciones)[34] es un conjunto de reglas y especificaciones que una aplicación ofrece para permitir el acceso mediante lenguajes de programación a los datos que maneja, de forma que desarrolladores, tanto de la organización que ha creado la interfaz como ajenos, pueden hacer uso de los datos y servicios que proporciona.

Esta es una práctica muy habitual en Internet, que ha permitido el desarrollo de ecosistemas que comparten datos y contenido creado en un lugar, pero distribuido e incluso actualizado de forma dinámica a lo largo y ancho de la Red.

Uno de los casos paradigmáticos es Twiter, cuya API[35] ha sido fundamental para que se haya convertido en una de las herramientas de comunicación más importantes de la Red, pues ha

[30] "¿Qué ofrece Libranda a las editoriales?", Libranda (http://www.libranda.com/default.aspx?id_seccion=805)
[31] "Dossier de prensa", Libranda (www.libranda.com/docs/dossier.pdf)
[32] "Code Meet Print NY: Cold Beer, Power Point and the Future of Reading", Publishers Weekly, 12 de mayo de 2011 (http://www.publishersweekly.com/pw/by-topic/digital/conferences/article/47190-code-meet-print-ny-cold-beer-power-point-and-the-future-of-reading.html)
[33] "When Code Met Print", blog de BookNet Canada, 16 de junio de 2011 (http://www.booknetcanada.ca/index.php?option=com_wordpress&p=2308&Itemid=319)
[34] "What is open? A simple description of APIs", BBYOpen, Youtube (http://www.youtube.com/watch?v=7r7QpIDEl_o)
[35] "API Documentation", Twitter (http://dev.twitter.com/doc)

permitido que sobre ella se hayan creado miles de aplicaciones (clientes para diferentes plataformas, servicios de analítica web…) que han hecho crecer en gran medida la utilidad para los usuarios de la información que fluye a través de Twitter. Los ejemplos son innumerables: Facebook, Youtube y muchos otros servicios de Google, Amazon[36], Flickr, Slideshare, LinkedIn… Todos ofrecen APIs sobre las que construir nuevas aplicaciones que aporten valor a sus usuarios.

2.4.2 APIs en el mundo editorial

Mac Slocum, de la editorial O'Reilly Media, identifica[37] 3 formas en que las APIs pueden ser útiles para las editoriales:

1. Como forma de enfrentarse al problema del descubrimiento

Cualquier tipo de contenido digital se enfrenta al mismo problema del descubrimiento: por muy bueno que sea el contenido, de poco vale si la gente no puede encontrarlo.

Como ya hemos dicho, esta es una cuestión que irá ganando en importancia a medida que se consolide la creciente aceleración en el ritmo de creación de nuevo contenido, que habrá de competir con todo el que ya existía.

Una API, si bien no resolverá el problema por sí sola, sí puede contribuir a hacerlo al permitir la diseminación del contenido de una forma estructurada y abierta: mantener el contenido dentro de los límites de un sitio web, un ebook, o cualquier otro tipo de contenedor digital implica que serán menos los posibles consumidores que llegarán a tener conocimiento de su existencia. Una API puede convertirse en un medio efectivo de llevar el contenido allí donde están sus potenciales lectores, en lugar de forzarlos a visitar un sitio web específico, ya sea el de la editorial, el del autor, o incluso el de la propia tienda online.

2. Como medio de experimentación

La industria editorial se encuentra en pleno proceso de reinvención. Dentro de un tiempo, la existencia de una API para cualquier tipo de contenido editorial se dará por supuesta. Entretanto, la relativa oscuridad de las APIs ofrece mayor flexibilidad para experimentar con nuevos modelos: API solo de metadatos, API de texto completo, o incluso con ganchos promocionales.

---

[36] "Product Adverstising API", Amazon (http://docs.amazonwebservices.com/AWSECommerceService/2010-11-01/GSG/)

[37] "3 ways APIs can benefit publishers", Mac Slocum, 21 de marzo de 2011 (http://radar.oreilly.com/2011/03/3-ways-api-help-publishers.html)

Se puede incluso utilizar la API como base para organizar competiciones entre desarrolladores externos, a la manera de ciertas Administraciones imbuidas del espíritu del Open Data, para tratar de sacar partido de la creatividad ajena a la organización editorial, que probablemente aporte formas de ver y pensar el contenido diferentes de las que, por la propia inercia de toda organización, se imponen dentro.

3. Como manera de propiciar intencionadamente consecuencias inesperadas

En la línea del párrafo anterior, Slocum cita ejemplos de APIs que propiciaron la generación a su alrededor de ecosistemas con consecuencias muy probablemente inesperadas incluso para sus propios creadores.

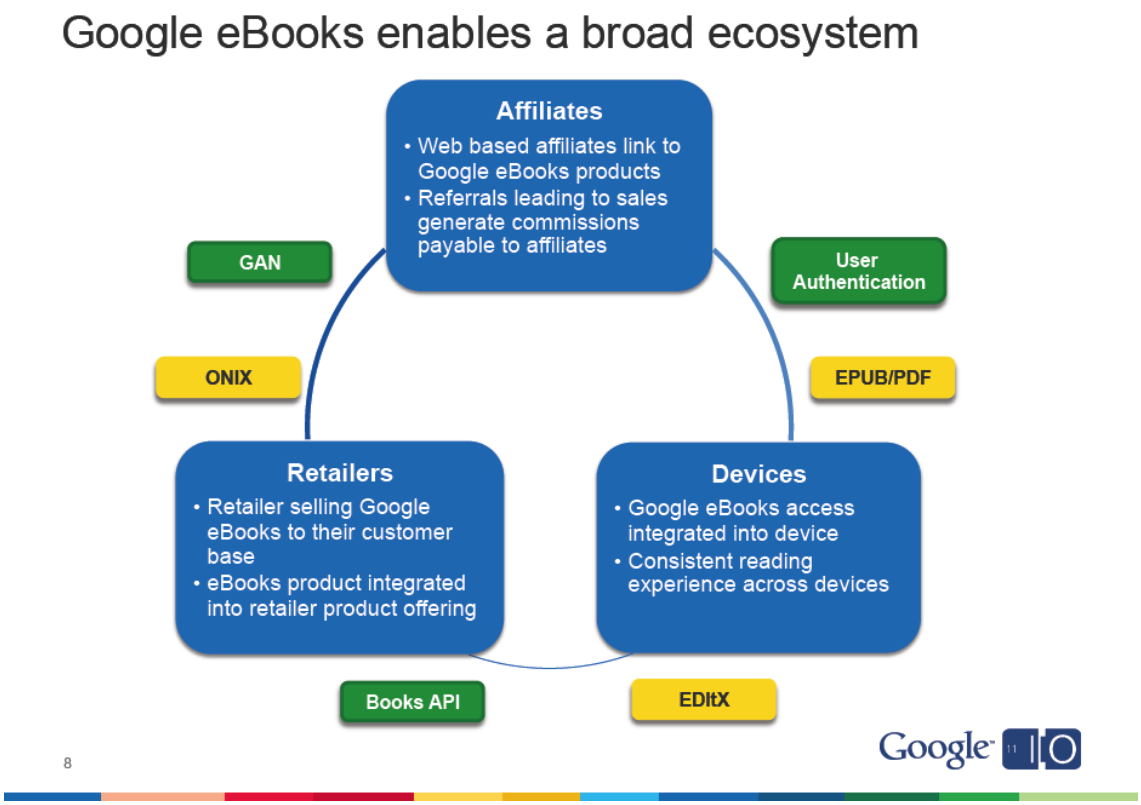
Uno de los casos con mayor repercusión es el de Google Maps. En un primer momento, cuando Google presentó su herramienta de mapas, ni siquiera ofrecía una API. Pero era tal su utilidad que, pese a las dificultades, hubo quien creó servicios basados en sus mapas. Google reaccionó abriendo su API y esto propició que floreciese un vibrante ecosistema

Dice Slocum: "Las APIs favorecen consecuencias inesperadas como éstas. Son combustible para la creatividad." Eso sí, según Slocum, "no se trata de copiar lo que hacen las grandes compañías de Internet, sino de inyectar nuevas ideas en una organización sin necesidad de reorganizar la empresa o de lanzar un negocio completamente nuevo".

Quien construya productos alrededor de la API probablemente verá el contenido de maneras que a la editorial nunca se le hubiesen ocurrido, pues no se verá influenciado por su tradición y objetivos. Esto es precisamente lo que necesitan las editoriales en momentos de reinvención como este.

2.4.3 Google Books

Como en muchos otros ámbitos, también en el ámbito editorial Google es un referente en innovación con su intención de crear mediante sus APIs y sus programas de afiliación todo un ecosistema alrededor de los libros electrónicos en Internet con su propia tienda, la Google eBookstore, en el centro.



Fuente: Integrating to eBooks: APIs to Sell and Read eBooks for Affiliates, Retailers and Device Makers, presentación de Pratip Banerji, Product Manager, y Kevin Landry, Tech Lead, en Google IO, mayo de 2011.

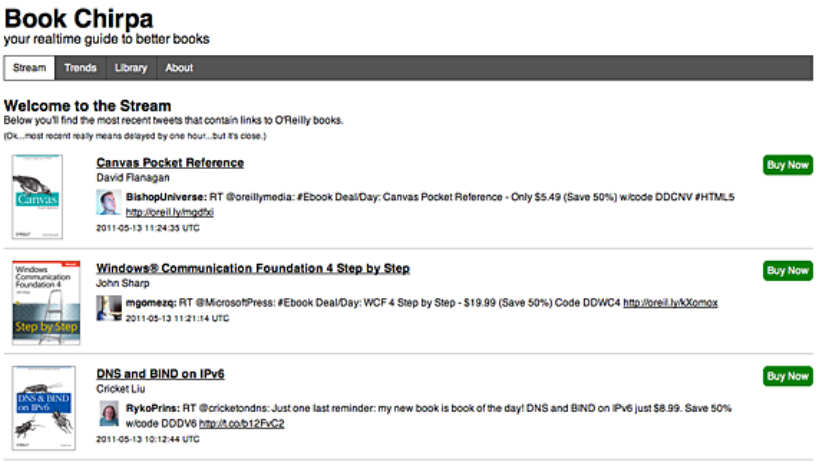
Las APIs de Google Books permiten integrar en aplicaciones o sitios web de terceros muchas de las funciones disponibles en el sitio de Google Books, como:

- Buscar y navegar a través de la lista de libros que se obtiene como respuesta a una determinada búsqueda.
- Ver información sobre un libro, incluyendo metadatos, disponibilidad y precio.
- Incrustar contenido de Google Books directamente en una página web, al estilo de Google Maps o Youtube, lo que permite desarrollar una experiencia de usuario propia alrededor de las vistas previas de los libros.

2.4.4 O'Reilly Media

La editorial O'Reilly Media es otro de los referentes en este espacio. En 2009, lanzó una API[38] de metadatos de todos sus libros en formato RDF[39]. Más recientemente, en marzo de 2011, publicó una nueva API, basada en la tecnología de la empresa Fluidinfo[40], cuya característica más destacada es que es “escribible”[41], esto es, permite que sus usuarios incorporen nueva información (por ejemplo, etiquetas con las que clasificar los libros) a la proporcionada por la editorial.

En la línea de lo que sugería Mac Slocum en el artículo citado más arriba, O'Reilly organizó un concurso de desarrolladores[42] basado en su nueva API, cuyo primer premio fue para *Book Chirpa*, "creado para visualizar qué están diciendo los usuarios de Twitter sobre los libros de O’Reilly"[43] y que permite ver el flujo de menciones a los libros de O’Reilly, cuáles son populares (*trending*), o una biblioteca virtual de todos los libros de O’Reilly mencionados en Twitter, e incluye un botón de compra que remite a la tienda online de la editorial.



Fuente: Book Chirpa

---

38 “O'Reilly Product Metadata Interface” (http://labs.oreilly.com/opmi.html)

39 Resource Description Framework (RDF) es un lenguaje para representar información sobre recursos en la Red utilizando XML. Es uno de los elementos básicos de la Web Semántica que propugna Tim Berners-Lee, inventor de la World Wide Web.

40 “The structure of O’Reilly book and author data in Fluidinfo”, blog de Fluidinfo, 21 de marzo de 2011 (http://blogs.fluidinfo.com/fluidinfo/2011/03/21/the-structure-of-oreilly-book-and-author-data-in-fluidinfo/)

41 "A writable API for O'Reilly", Terry Jones, 21 de marzo de 2011 (http://radar.oreilly.com/2011/03/oreilly-writeable-api-fluidinfo.html)

42 "Winners of the writable API competition", Terry Jones, 13 de mayo de 2011 (http://radar.oreilly.com/2011/05/oreilly-writable-api-contest-winners.html)

43 “About Book Chirpa” (http://www.bookchirpa.com/about)

2.4.5 Goodreads

Goodreads, que hace uso extensivo de la API de Google eBooks[44] (de hecho, sirve de ejemplo de integración con ella en la presentación antes mencionada[45]), ofrece a su vez una interfaz de programación[46] a través de la cual quienes la utilicen (entre otros, el propio Google[47]), además de la propia información bibliográfica, pueden obtener:

- Sindicación de las críticas de libros hechas por sus usuarios ("10 millones de críticas de 700.000 títulos").

- Acceso a toda la información de las cuentas de los usuarios de Goodreads (*Goodreads Connect*), previa autorización por parte del propio usuario. Esta información incluye sus libros, puntuaciones, críticas, así como sus contactos ("el gráfico social de lectura"), que puede utilizarse, como sugieren desde Goodreads, "para personalizar un sitio de comercio electrónico, potenciar recomendaciones, mostrar un widget[48] con los libros favoritos del usuario, desarrollar una aplicación cliente para móvil o para ordenador, y más".

3 Metadatos

Todas las posibilidades de fomentar el conocimiento y la "descubribilidad" del contenido de las editoriales tienen una base común, la información que estas gestionan y difunden, ya sea el propio contenido de los libros o cualquier tipo de información relacionada con ellos, desde material promocional a información bibliográfica o relacionada con su disponibilidad o precio.

En el entorno que hemos venido describiendo, de creciente importancia del libro electrónico y, sobre todo, de cambios en los hábitos de búsqueda de libros, tanto físicos como electrónicos, que cada vez se produce en mayor medida a través de Internet, la importancia de los metadatos para dar a conocer los libros, para facilitar su descubrimiento por los potenciales lectores, no hace más

---

44 "Who's using it: GoodReads", Google Books API (http://code.google.com/intl/es-ES/apis/books/casestudies/goodreads.html)

45 "Integrating to eBooks: APIs to Sell and Read eBooks for Affiliates, Retailers and Device Makers", Pratip Banerji, Product Manager; Kevin Landry, Tech Lead, Google IO, 10-11 de mayo de 2011 (http://www.google.com/events/io/2011/sessions/integrating-to-ebooks-apis-to-sell-and-read-ebooks-for-affiliates-retailers-and-device-makers.html)

46 Goodreads API (http://www.goodreads.com/api)

47 "Sony Reader and Blio now showing Goodreads reviews", Goodreads blog, 27 de octubre de 2010 (http://www.goodreads.com/blog/show/239.Sony_Reader_and_Blio_now_showing_Goodreads_reviews)

48 Pequeña aplicación que permite mostrar cierta información o proporcionar alguna funcionalidad en un sitio web proveniente de otro.

que crecer, por lo que su gestión y control se ha convertido en un asunto estratégico para las editoriales. Y esto no solo para la preservación de los modelos de negocio tradicionales sino, con mayor peso si cabe, para la generación de nuevas oportunidades de negocio que las tecnologías de la información hacen posibles.

Por si fuera poco lo anterior, las editoriales obtienen ventajas adicionales, en forma de mayor eficiencia, e incluso reducción de costes de operación, al automatizar y simplificar sus procesos internos de creación y utilización de metadatos. En la industria editorial aún es habitual ver cómo empleados de distintos departamentos tienen que volver a introducir la misma información en distintos sistemas para distintas finalidades. Si se implantasen las herramientas y procedimientos adecuados, la editorial podría tener una base de datos unificada con los metadatos recopilados por los distintos departamentos, de la que se alimentarían los sistemas que generan los metadatos que se transmiten a los socios comerciales en los formatos estandarizados con diversas finalidades. Suponiendo que existiese un sistema así, la responsabilidad de comprobar y validar los datos se podría distribuir entre departamentos y, al mismo tiempo, cualquier actualización de la información, como el título o el precio, se propagaría automáticamente allá donde fuese necesario.

Los metadatos son así tanto una herramienta de marketing como una forma de añadir funcionalidad a las publicaciones electrónicas, pues permiten a las editoriales "abrir" su contenido para el comercio electrónico y la indexación, búsqueda y enlace que facilitan su descubrimiento, al tiempo que mantienen el control de su negocio.

3.1 Qué son los metadatos

Los metadatos son literalmente datos sobre datos, datos que describen o proporcionan contexto sobre otros datos. En la práctica, el término ha venido a significar información estructurada que describe, explica, localiza o facilita la obtención, el uso o la gestión de un recurso de información.

Aunque han existido prácticamente desde que el ser humano empezó a organizar la información que iba recopilando, hasta hace apenas unas décadas eran prácticamente coto vedado de bibliotecarios y conservadores de museos, encargados de catalogar, clasificar e indexar los elementos que componían sus colecciones. Con la aparición de Internet y la Web, y la creciente digitalización de diferentes objetos culturales o de información, no sólo por profesionales sino también por el grueso de la población internauta, los metadatos han dejado de ser algo desconocido y minoritario para pasar a formar parte, si bien con frecuencia de una forma transparente, de las herramientas habituales de todos los que crean o utilizan información.

3.2 Tipos

Existen múltiples formatos de metadatos para describir distintos tipos de objetos textuales y no textuales, como libros, documentos electrónicos, objetos de arte, materiales educativos o conjuntos de datos científicos. Se pueden distinguir tres tipos principales:

- Descriptivos: describen un recurso con la finalidad de facilitar su descubrimiento e identificación, mediante elementos como: resumen, autor, palabras clave…

- Estructurales: indican cómo se forman objetos compuestos. Por ejemplo, cómo se ordenan las páginas para formar capítulos, y los capítulos para dar lugar a libros.

- Administrativos: proporcionan información que permite gestionar un recurso, tal como cuándo y cómo se creó, su tipo de fichero y otra información técnica, o quién puede acceder y qué puede hacer con él. Los metadatos administrativos pueden ser:
  - Metadatos para la gestión de derechos: por ejemplo, de propiedad intelectual.
  - Metadatos para la conservación: que contienen información necesaria para archivar y conservar el recurso.

Los metadatos pueden incorporarse al propio objeto digital o guardarse por separado. Almacenarlos junto al objeto que describen asegura que no se perderán, elimina el problema de vincularlos con él y facilita que el objeto y sus metadatos se actualicen al mismo tiempo. Sin embargo, resulta técnicamente imposible adjuntar los metadatos a ciertos tipos de objetos. Además, almacenarlos por separado puede simplificar su gestión, por lo que es habitual que se haga así y se vinculen de alguna manera los metadatos con los recursos a los que hacen referencia.

3.3 Para qué sirven

Una razón importante para la creación de metadatos descriptivos es que facilitan el descubrimiento de información relevante, al tiempo que ayudan a organizar los recursos electrónicos y facilitan su identificación del contenido.

3.3.1 Descubrimiento

Los metadatos facilitan el descubrimiento de varias maneras:

- Permitiendo encontrar los recursos según diversos criterios relevantes

- Identificando los recursos
- Agrupando recursos similares
- Distinguiendo entre recursos diferentes
- Proporcionando información sobre dónde encontrar los recursos

3.3.2 Organización

La cantidad de contenido existente crece cada vez más rápido, lo que hace que la capacidad de filtrarlo y organizarlo sea cada vez más importante.

3.3.3 Identificación

La identificación unívoca de los recursos es imprescindible para su gestión efectiva. Este asunto implica complicaciones adicionales en el caso de recursos electrónicos.

3.4 Estándares

Los metadatos no son nuevos en la industria de la edición. Lo que sí es relativamente nuevo, y crucial para la interoperabilidad en la época digital, es su estandarización. Esto es, el proceso por el que se alcanza el consenso alrededor de buenas prácticas en cuanto al formato y a la utilización de metadatos para aplicaciones específicas, de forma que las máquinas pueden interpretar e intercambiar esta información de una forma eficiente.

Lo que ha movido a la industria editorial a la adopción de estándares es la posibilidad de rentabilizar la venta agregada de libros (tradicionalmente, mediante distribuidores; desde hace unos años, también a través de Internet), que hace necesario que los editores proporcionen a distribuidores y otras plataformas información promocional sobre los libros de una forma estructurada y normalizada.

Pero esta información, para que sea verdaderamente útil, debe permitir su tratamiento automatizado. Por ello, esos estándares han de expresarse en un lenguaje que los ordenadores puedan manejar. Es aquí donde entra en escena XML (*eXtensible Markup Language;* Lenguaje de Marcado Extensible), también establecido como estándar internacional, que facilita el tratamiento y la comunicación de información estructurada.

3.5 ONIX

ONIX (*Online Information eXchange*; Intercambio de Información Online) es un formato de metadatos desarrollado por editoriales y agentes de venta de libros para describir sus productos mediante un registro digital con alto contenido de información que permite dos cosas: ofrecer la información necesaria para proporcionar al consumidor la experiencia de hojear libros online y compartir de una forma rápida y precisa información sobre los productos entre los participantes en la cadena de suministro editorial. ONIX comprende un conjunto estructurado de elementos, una serie de vocabularios controlados (las listas de códigos) y un esquema que establece cómo ha de expresarse en XML para su comunicación automatizada entre ordenadores. Consta de más de 200 elementos, reunidos en grupos de información relacionada, como formato del producto, título, autoría, etc.

3.5.1 Origen

El origen de ONIX se remonta a un estudio del *British National Bibliography Research Fund* titulado "Elements for Book Product Information" ("Elementos de Información del Libro como Producto"), cuyo primer resultado fue EPICS (*EDItEUR Product Information Standards*; Estándares de Información sobre el Producto de EDItEUR), producido por BIC (*Book Industry Communication*), asociación del sector editorial en el Reino Unido, y su contraparte estadounidense, BISG (*Book Industry Standards Group*). ONIX, un subconjunto de EPICS, surgió a su vez de una reunión en 1999 entre las editoriales y las librerías online organizada por la AAP (*American Association of Publishers*; Asociación Americana de Editores). La AAP publicó la primera versión de ONIX en enero de 2000. La cooperación internacional tardó poco en llegar, y BIC organizó un seminario en noviembre de 2000 para presentar la primera revisión importante del estándar, ONIX International.

3.5.2 Quién lo controla

El *ONIX International Steering Committee* (Comité Ejecutivo Internacional de ONIX) de EDItEUR (organización internacional, con sede en Europa, que coordina el desarrollo de estándares para el comercio de libros, libros electrónicos y publicaciones por entregas), en el que tienen representación grupos de usuarios de más de 15 países, incluida España, por medio de la Federación de Gremios de Editores de España (FGEE).

3.5.3 Quién lo utiliza

En sus inicios, ONIX fue un estándar anglo-estadounidense, apoyado desde el principio por pesos pesados de la industria como Cambridge University Press, que ya en 2001 ofrecía la información de

todos sus productos en este formato, o Amazon.com, uno de sus mayores impulsores. Sin embargo, en menos de un año el interés por implementar ONIX era ya notorio en otros países, como Francia o Alemania. A finales de 2002, se había convertido en el estándar de facto en la industria editorial anglosajona y era ya el estándar internacional de referencia para información bibliográfica.

3.5.4 Por qué se creó

Para conseguir que toda una industria se ponga de acuerdo en un formato de metadatos en tan poco tiempo, poderosas fuerzas tuvieron que alinearse, y más aún si se tiene en cuenta la historia de falta de cooperación a este respecto de la industria editorial. Antes de que apareciera ONIX, era habitual que editoriales, distribuidores y puntos de venta utilizaran distintos formatos para representar la información sobre los productos. Dos hechos distintos pero relacionados entre sí provocaron esta importante transformación en las costumbres de la industria respecto a los metadatos: por un lado, el libro electrónico; por otro, la venta online de productos impresos.

Aunque la industria del libro electrónico ha tenido clara desde hace mucho tiempo la importancia de los metadatos para la distribución de sus productos, fue curiosamente en el mundo de la edición impresa, tradicionalmente remiso a escuchar los llamamientos a utilizar estándares abiertos, donde la necesidad de un estándar común de metadatos se hizo más evidente. Y esto se debe en gran medida a Amazon y demás librerías online, que supieron hacer ver a las editoriales la importancia de metadatos de calidad para las ventas de sus productos. Amazon demostró cómo el secreto de las ventas online no pasa sólo por abrumar al cliente con la cantidad ingente de contenido disponible, sino por permitirle buscar y hojear libros de una forma lo más parecida posible a la que está acostumbrado a hacer en una librería física.

Así, ONIX apareció como respuesta directa al gran crecimiento de las ventas de libros por Internet y a la constatación de que estas eran mucho mayores cuanto más completa era la información sobre los mismos que se mostraba online, incluyendo imágenes, críticas u otro tipo de referencias publicitarias. Hasta entonces, los formatos de metadatos o bien no eran lo suficientemente ricos en las áreas adecuadas (como EDI, *Electronic Data Interchange*), o su implementación resultaba demasiado costosa para la industria editorial (como en el caso de MARC, *MAchine-Readable Cataloging*).

ONIX proporciona, además de la información blibiográfica tradicional, cosas como críticas, biografías y fotografías de los autores, imágenes de portada, información promocional, sobre derechos de propiedad, sobre precio y disponibilidad para diferentes mercados. También incluye la

descripción física del libro impreso (dimensiones, peso), especialmente útil para la gestión de inventario en almacenes.

3.5.5 Evolución

En un principio, ONIX evolucionó muy rápido (la versión 1.0, como hemos dicho, data de enero de 2000; la 1.1, de apenas seis meses después; la 1.2 apareció en noviembre de ese mismo año). A finales de 2003 se publicó la versión 2.1, que suponía un cambio de enfoque, pues, para evitar que continuase el aumento del número del elementos que el estándar contempla (que habían pasado de 148 en la versión inicial a 235 en la 2.1), se optó por hacer que la semántica de los elementos fuese menos específica, haciendo uso en su lugar de vocabularios controlados (las listas de códigos), que permitirían llevar a cabo de una forma más flexible los cambios necesarios para adaptarse a las cambiantes necesidades de los usuarios del estándar.

Hubo dos factores que contribuyeron a la rápida expansión y revisión de ONIX: por un lado, la participación de bibliotecarios de instituciones importantes, como la British Library o la Biblioteca de la Universidad de Florida, por otro, la expansión de las tiendas online, como Amazon, que hacía que se difuminasen las fronteras entre la venta tradicional de libros y la de otro tipo de medios de entretenimiento. Así, la versión 1.2 tenía la intención de ser útil para describir no sólo libros físicos sino también libros electrónicos; la 1.3, ampliaba su ámbito a formatos de vídeo; la revisión 2.0 fue necesaria para que ONIX permitiese describir publicaciones por entregas y DVDs, que podían contener varios tipos de medios. De hecho, fueron precisamente las dificultades al tratar de describir las publicaciones por entregas las que hicieron necesaria la separación de ONIX en dos formatos: *ONIX for Books*, para los libros, y *ONIX for Serials*, para las publicaciones por entregas.

Un tiempo después, en 2008, un nuevo estándar se incorporó a la familia: *ONIX for Licensing Terms* (ONIX para condiciones de licencia), pensado para facilitar la gestión de las licencias sobre recursos electrónicos por parte de bibliotecas académicas e institucionales.

La última revisión de *ONIX for Books*, que dio lugar a la actual versión 3.0, vino motivada principalmente por la necesidad de mejorar la descripción y el tratamiento de los libros electrónicos, así como para introducir nuevos elementos que amplían la variedad de material de marketing que las editoriales pueden incluir.

3.5.6 ONIX y XML

Para facilitar el tratamiento automatizado de la información contenida en registros ONIX, EDItEUR publica directrices sobre cómo expresar ONIX en XML, un lenguaje pensado precisamente para el

intercambio de información estructurada entre ordenadores. (En el capítulo siguiente se explica con más detalle en qué consiste XML y qué beneficios puede proporcionar su utilización a las editoriales.)

3.5.7 ONIX en España

La Federación de Gremios de Editores de España (FGEE) ha coordinado el desarrollo de una implementación de ONIX adaptada al mercado español (ONIX-ES), que tiene en cuenta determinadas particularidades como, por ejemplo, la codificación de niveles educativos en libros de texto o a la obligatoriedad y forma de codificación de determinados campos.

DILVE

DILVE (Distribuidor de Información del Libro español en VEnta) es una plataforma web desarrollada por la FGEE en colaboración con la Fundación Germán Sánchez Ruipérez para todos los profesionales de la cadena del libro, que permite la gestión y distribución de metadatos bibliográficos y comerciales (cubiertas, resúmenes, fragmentos…) del libro en venta.

Se basa en estándares internacionales (ONIX y clasificación de materias BIC) y permite la integración de datos de cada obra junto a contenidos ricos orientados a la venta (imágenes, resúmenes y otros textos, grabaciones de audio o vídeo...), pero no impone como requisito que la información se introduzca directamente en ONIX, ya que permite que las editoriales introduzcan la información de sus títulos en diversos formatos (ONIX, CSV, formulario web).

Además, DILVE ofrece a sus usuarios servicios de conversión y distribución de metadatos, de forma que la información puede extraerse en diferentes formatos y por distintos medios. Las editoriales tienen también la posibilidad de enviar información a varias plataformas (como la Agencia española del ISBN o Libranda) en los formatos y por los medios que estas requieren.

Los datos que maneja DILVE son propiedad y responsabilidad de las editoriales, que han encargarse de su controlar su calidad, consistencia y actualización. DILVE sólo contiene una réplica de los datos que cada editorial tiene en sus sistemas y puede transformar formatos pero no contenidos, si bien los analiza para comprobar su consistencia interna.

Fuente: DILVE

El valor de DILVE ha quedado de manifiesto por el hecho de que, en los últimos meses, organizaciones como Bowker[49] o Amazon[50], han buscado acuerdos con la FGEE para obtener información sobre los libros editados en España.

CEGAL en red

CEGAL (Confederación Española de Gremios y Asociaciones de Libreros) gestiona "CEGAL en red"[51], la otra gran base de datos de libros editados en España. Pero, aunque CEGAL en red se alimenta de información proveniente de DILVE (junto con otras fuentes, como la Agencia española del ISBN y distribuidores y editores), no utiliza ONIX, sino que está basada en el estándar de ficha del libro de SINLI[52] (Sistema de Información Normalizada para el LIbro), a partir del que se genera la "ficha básica", que se enriquece con información adicional, como la portada, el índice de contenidos de la obra y la reseña publicada por el editor, dando lugar a la "ficha ampliada".

3.6 Calidad de los metadatos

Para un uso efectivo de los metadatos es fundamental la implantación de sistemas y procedimientos integrados dentro del proceso general de publicación, para así hacer que la creación y actualización

[49] "La FGEE firma un acuerdo con Bowker para obtener datos de DILVE", DILVE, abril de 2011 (http://www.dilve.es/dilve/dilveweb/dilve_noticias_dest.jsp?anio=2011#texto_881)
[50] "Amazon integrará en su base de datos los libros de DILVE", Actualidad Editorial, 16 de febrero de 2011 (http://www.actualidadeditorial.com/amazon-integrara-en-su-base-de-datos-los-libros-de-dilve/)
[51] "Qué es CEGAL en red", CEGAL (http://www.cegal.es/lib/informes/introInformes.php?codTipoInformeAsoPrincipal=20)
[52] "Qué es SINLI" (http://www.fande.es/sinli/sinli.asp)

continua de los metadatos sea una parte normal del flujo de trabajo de producción. Por el contrario, la creación de metadatos de forma automática, o por personas que no tengan experiencia en catalogación o indexación puede afectar negativamente a su calidad.

BISG ha publicado sendas guías para emisores y receptores de metadatos[53] en las que, dentro de los más de 200 campos de que consta un registro ONIX, señala 31 como los fundamentales (*core metadata)*, sobre los que se debe extremar el cuidado, y entre los que están, por ejemplo: ISBN, título, autoría, editorial/sello/marca, precio, fecha de publicación, código de materia o derechos territoriales.

A estos se suman los metadatos enriquecidos o mejorados (*enhanced metadata*), que añaden información adicional a los registros, en particular material promocional o de otro tipo que ayude a su comercialización.

3.6.1 Certificación

Para mejorar el nivel de calidad de los metadatos en sus respectivos ámbitos de influencia, tanto BISG[54] en Estados Unidos, como BookNet[55] en Canadá y BIC[56] en Reino Unido han puesto en marcha programas de certificación de los metadatos que sus asociados envían a la cadena de suministro. Se trata de programas voluntarios, organizados por niveles, que evalúan objetivamente la calidad y completitud de la información sobre los productos.

3.7 La identificación de textos en formato digital

Aunque existen muchas semejanzas, desde el punto de vista de la cadena de suministro, entre las necesidades de identificación de los libros físicos y la del contenido digital, los nuevos modelos de negocio y los nuevos canales suponen un reto para las prácticas establecidas.

El *International Standard Book Number* (ISBN) se considera apropiado para el comercio, el descubrimiento y la comunicación de información comercial sobre productos digitales en la cadena

[53] "Product Metadata Best Practices", BISG, (http://www.bisg.org/what-we-do-21-8-product-metadata-best-practices.php)
[54] "Product Data Certification Program", BISG (http://www.bisg.org/what-we-do-3-4-product-data-certification-program.php)
[55] "Bibliographic Certification", Booknet Canada (http://www.booknetcanada.ca/index.php?option=com_content&view=article&id=487&Itemid=312)
[56] "BIC Accreditation Schemes", BIC (http://www.bic.org.uk/9/Product-Data-Accreditation/)

de suministro. Se debe asignar un ISBN distinto a cada manisfetación digital que se comercialice por separado.

El *Digital Object Identifier* (DOI), en modo alguno sustituto del ISBN, sí es un valioso identificador adicional que permite el descubrimiento de contenido en la Red.

Por su parte, el *International Standard Text Code* (ISTC) puede suponer para las editoriales una manera de agrupar distintas manifestaciones de una misma obra.

Una vez que el contenido del libro se encuentra en formato digital, se abre un abanico casi infinito de nuevos modelos de negocio para su consumo tanto online como offline, asociados a una gran variedad de derechos de uso otorgados o denegados al usuario. Estas múltiples combinaciones implican retos para la identificación tanto en la cadena de suministro como en el mantenimiento de los propios sistemas de las editoriales.

Muchas editoriales utilizan ya sofisticados sistemas para la gestión del contenido y del flujo de trabajo asociado (DAMs, *Digital Asset Management Systems*; Sistemas de Gestión de Activos Digitales) o para su agregación y distribución (DAMs, *Digital Asset Distribution Systems*; Sistemas de Distribución de Activos Digitales).

Estos sistemas requieren en marco normalizado de identificación de los activos digitales. En la relación entre empresas, las editoriales, librerías y bibliotecas, necesitan saber qué versiones están disponibles y cuáles están promocionando, solicitando y comprando. Los consumidores necesitan saber qué versiones digitales de los títulos están disponibles, y si éstas son compatibles con sus necesidades o con sus programas y dispositivos, así como entender qué derechos de uso implican.

3.7.1 ISBN (*International Standard Book Number*)

Desde la introducción a gran escala de la informática en la industria editorial, en los años 70, la mayoría de las necesidades de identificación las ha cubierto el ISBN. Su éxito, junto a la proliferación del contenido en forma electrónica hizo que en 2005 se pasase de su formato original de 10 caracteres al actual de ISBN-13.

Se aprovechó el proceso de revisión del estándar para estudiar y redefinir el ámbito de aplicación del ISBN. Por primera vez, el nuevo estándar contemplaba explícitamente las publicaciones electrónicas, incluidos capítulos, artículos y otras partes de publicaciones cuando se comercializasen por separado.

El ISBN es un identificador para la cadena de suministro. La razón de su existencia es identificar cada manifestación de una publicación susceptible de ser comercializada por separado. Sus finalidades principales son: facilitar el comercio de libros, hacer posible y facilitar el descubrimiento de libros en bases de datos bibliográficas y catálogos de bibliotecas, posibilitar la obtención de estadísticas de venta para toda la industria y facilitar la gestión de derechos de propiedad intelectual. Aunque algunas editoriales lo utilicen así, no se diseñó para identificar contratos o *royalties*. De hecho, en los sistemas de muchas editoriales, el ISBN se ha convertido de facto en el atributo clave del producto, que une ventas, distribución, logística, gestión del flujo de trabajo, gestión de contratos y derechos y pago de *royalties*.

La pregunta es si el ISBN es el mejor identificador posible para las publicaciones digitales, a pesar de una importante limitación: no sirve para vincular las distintas manifestaciones del mismo contenido, algo cuya importancia es cada vez mayor a medida que proliferan múltiples versiones, impresas o digitales, del mismo título. A cubrir esta deficiencia viene el ISTC, aparecido en 2009.

No obstante lo anterior, el ISBN satisface de forma eficiente casi todas las necesidades de identificación de la cadena de suministro. Las editoriales, como generadores del producto en la cadena de suministro, son las responsables de decidir a cuáles de sus productos se les asignarán ISBNs. Las recomendaciones de la Agencia Internacional del ISBN son que se hagan:

- Cuando el producto va a ser comercializado directamente
- Cuando sea necesario "descubrir" el producto
- Cuando sea necesario obtener informes de las transacciones alrededor del producto

3.7.2 DOI (*Digital Object Identifier*)

La función del DOI es la de proporcionar una "dirección" controlada por la editorial, dinámica y actualizable, para cualquier recurso, ya sea físico o digital. Nunca se pensó como identificador de contenido de cara a la cadena de suministro.

Puede servir de ayuda para el descubrimiento de prácticamente cualquier cosa que las editoriales quieran comercializar, tanto libros enteros (impresos o digitales), como capítulos u otros fragmentos. En muchos casos, el DOI funcionará en combinación con el ISBN.

3.7.3 ISTC (*International Standard Text Code*)

El ISTC es un identificador de "obras de texto". No identifica productos físicos o digitales, tales como libros o ebooks, y no interviene en la comercialización del contenido, ya sea físico o digital. Está pensado para facilitar el intercambio de información entre autores, agentes, editores, puntos de venta, bibliotecarios y otras partes interesadas, permitiendo agrupar las diferentes manifestaciones del mismo título.

Está diseñado para ser flexible y granular: se puede usar para identificar, por ejemplo, El Quijote, como obra original (ofreciendo un mecanismo para agrupar todas las ediciones impresas y digitales, así como manifestaciones en otro tipo de medios) y también para agrupar las diferentes ediciones de un mismo título o todos los capítulos de una obra. Esto implica que puede ser de mucha utilidad en la gestión de contratos y *royalties,* así como en los procesos de búsqueda en las tiendas online, entre otras muchas aplicaciones.

3.8 Sistemas de clasificación de materias

Otro de los tipos de metadatos más importantes para facilitar el descubrimiento de los libros por los lectores es el que indica de qué materia o materias tratan. En los libros físicos, esto tiene repercusión normalmente sobre dónde se coloca el libro físico en la librería (narrativa, poesía, no ficción…). Para el libro electrónico, su clasificación implica no sólo que figure en unas u otras "estanterías virtuales", sino que influye directamente en que el libro aparezca en los resultados de búsqueda.

La industria editorial, consciente de la importancia de una correcta y normalizada clasificación de materias, ha consensuado una serie de vocabularios controlados que incluyen una lista de términos para la descripción del contenido de los libros, lo que permite la transmisión de información sobre las materias entre socios comerciales, incluso a nivel internacional, utilizando una sintaxis consistente; disponer de un lenguaje común para los informes de ventas; servir como puntos de acceso en las búsquedas en Internet y servir de criterio para la ordenación de los libros impresos en las librerías físicas.

Se trata de sistemas codificados (lo que facilita su tratamiento automatizado), jerárquicos (con diversos niveles de detalle), unívocos (que evitan ambigüedades en la clasificación), acotados (con un número finito de materias a los distintos niveles) y controlados por la propia industria, que los

actualiza con regularidad, incluyendo nuevas materias que se consideren pertinentes y eliminando, o al menos marcando como no recomendables, las que han dejado de ser relevantes.

A escala internacional, en el ámbito comercial, son dos los sistemas de clasificación de materias más importantes:

- BISAC (*Book Industry Standards and Communications*; Estándares y Comunicaciones para la Industria Editorial), mantenido por BISG (*Book Industry Study Group*), la asocación de la industria editorial estadounidense, y utilizando sobre todo en Estados Unidos y Canadá.

- BIC (*BIC's Subject Categories and Qualifiers Scheme*; Sistema de Calificadores y Categorías de BIC), que gestiona BIC (*Book Industry Communication*), la asocación de la industria editorial británica y se utiliza principalmente en Reino Unido, Australia, Nueva Zelanda y, parcialmente, en otros países, como Italia.

En España, un comité formado por representantes de todo el sector comercial (editoriales, distribuidoras, librerías y agencia del ISBN), coordinados por la Federación de Gremios de Editores de España a través de DILVE, ha optado[57] por el sistema de clasificación de materias BIC para su utilización en el mercado español del libro y ha procedido a su traducción y adaptación[58]. No obstante, existen en la actualidad reglas para el mapeado entre BIC y BISAC y no cabe descartar que acaben confluyendo en una sola clasificación en un futuro no muy lejano.

4 XML (*eXtensible Markup Language*)

XML (*eXtensible Markup Language;* Lenguaje de Marcado Extensible) es un lenguaje de marcado (basado en etiquetas) que permite describir la estructura de la información, comunicarla y almacenarla. A diferencia de HTML, el lenguaje más utilizado en las páginas web, XML no tiene etiquetas propias, sino que permite a sus usuarios crear las etiquetas que necesiten, siempre que cumplan con la sintaxis de XML. Es decir, XML es un conjunto de reglas que permiten a sus usuarios definir lenguajes de marcado adaptados a sus necesidades.

[57] La primera versión en español del sistema de clasificación de materias BIC", DILVE, 10 de mayo de 2011 (http://www.dilve.es/dilve/dilveweb/dilve_noticias_dest.jsp?anio=2011#texto_901)
[58]"BIC", DILVE, (http://www.dilve.es/dilve/dilveweb/dilve_BIC.jsp)

4.1 Transformaciones de XML

Como ya se ha dicho, mientras que HMTL se utiliza para dar formato a la información que se quiere mostrar, XML describe (y constituye) la propia información, pero no proporciona por sí mismo información sobre el formato con que esta ha de mostrarse. Para especificar cómo se debe mostrar un documento XML se puede utilizar XSL (*eXtensible Stylesheet Language;* Lenguaje de Estilos Extensible), que se compone en realidad de tres lenguajes:

- XSLT, para transformar documentos XML;
- XPath, para identificar las distintas partes de un documento XML; y
- XSL-FO, para dar formato de impresión a un documento XML.

XSLT (*eXtensible Stylesheet Language Transformations*) permite manipular la información contenida en un documento XML para transformarla en cualquier otro formato, como HTML (es muy habitual en la Red que se utilice XML para manejar y organizar la información, y se aplique XSL para convertir el XML en HTML), o en un documento XML con diferente estructura del original, algo que puede ser muy útil para las editoriales, como veremos.

XSL-FO (*eXtensible Style Language Formatting Objects*) se diseñó específicamente para dar formato de impresión, por ejemplo creando un PDF a partir de XML. Es básicamente un lenguaje de composición, que permite definir el diseño de la página, incluidos los márgenes y el espacio entre líneas; crear cabeceras, pies de página y notas en los márgenes; portadas y tablas de contenido.

4.2 Esquemas XML

Para definir un lenguaje propio en XML es necesario identificar cuáles son sus elementos y atributos, indicando cuáles son obligatorios y cuáles no. Esta información sobre la estructura de un documento XML es lo que se denomina un esquema. Si el documento no cumple las reglas definidas en el esquema, no se considera válido para el lenguaje definido por dicho esquema.

Hay dos formas principales para definir esquemas en XML: DTD (*Document Type Definition*; Definición de Tipo de Documento) y XSDL (*XML Schema Definition Language*; Lenguaje de Definición de Esquemas XML). Ambas especifican las etiquetas que pueden utilizarse para crear documentos XML de un determinado tipo, así como qué contenido y atributos puede incluirse en esas etiquetas.

4.2.1 DTD (*Document Type Definition*)

Una DTD es un conjunto de reglas que definen así un determinado lenguaje de marcado basado en XML. Identifica qué elementos pueden aparecer en un documento, qué elementos "hijos" pueden depender de ellos, qué atributos puede tener cada elemento, si estos atributos son obligatorios u opcionales y cuáles son sus valores aceptables. De esta forma, la DTD define la estructura de un documento para que pueda considerarse válido en un determinado lenguaje. Aunque la DTD es un documento de texto, no es XML.

4.2.2 XSDL (*XML Schema Definition Language*)

En 2001, el *World Wide Web Consortium* (W3C) creó un nuevo lenguaje de esquemas, XML Schema[59], para hacer frente a las limitaciones de las DTDs. Aunque ahora, en su versión 1.1., su nombre es XSDL (*XML Schema Definition Language*; Lenguaje de Definición de Esquemas XML), se sigue utilizando mucho su denominación original.

XSDL, que sí es XML, es también mucho más potente que las DTDs, pues proporciona mucho más control sobre los contenidos de un documento XML, y es muy probable que se acabe imponiendo como el lenguaje de esquemas más aceptado. Por ejemplo, en una DTD, todos los elementos se definen mediante un tipo de elemento, mientras que en XSDL los elementos pueden ser de tipo simple, si solo contiene texto, o complejo, si contiene elementos hijos y/o atributos.

4.3 Por qué usar XML

XML permite separar completamente el contenido y su descripción del formato con el que se presente, pues al contrario que HTML no está diseñado pensando en la presentación de la información y sus etiquetas no tienen por qué contener información sobre cómo mostrar el contenido en un soporte determinado.

Como ya hemos dicho, XML permite diseñar lenguajes de marcado adaptados a los requisitos de sus usuarios, que luego se utilizarán para almacenar y comunicar la información de forma estructurada, facilitando su tratamiento automatizado. Estos lenguajes están compuestos de etiquetas y reglas que permiten describir los datos.

[59] Este nombre puede provocar cierta confusión, puesto que tanto las DTDs como como los documentos en XML Schema (que a veces también se denominan XSD [*XML Schema Definition*; Definición de Esquema XML]), son esquemas XML.

XML puede utilizarse para comunicar información entre sistemas y organizaciones dispares. Los documentos en XML son simplemente ficheros de texto con una estructura definida, lo que hace que sea fácil entenderlos (tanto por máquinas como, en cierta medida, por humanos que los lean), analizarlos y transformarlos.

Por último, XML es un estándar internacional, creado y mantenido por el W3C, una organización dedicada al desarrollo y promoción de estándares Web (como HTML, CSS o RDF). Esto hace que muchas organizaciones de todos los tamaños y de muchos países ya lo utilicen como medio para compartir información, y que exista un gran número de aplicaciones, algunas de software libre, capaces de manejar documentos en XML.

En los últimos años, se han desarrollado infinidad de lenguajes basados en XML, como RSS, Atom o XHTML. Asimismo, la mayoría de aplicaciones ofimáticas, como Microsoft Office (Office Open XML) u OpenOffice.org (OpenDocument), utiliza por defecto formatos basados en XML.

4.4 Qué aporta XML a las editoriales

4.4.1 ONIX y XML

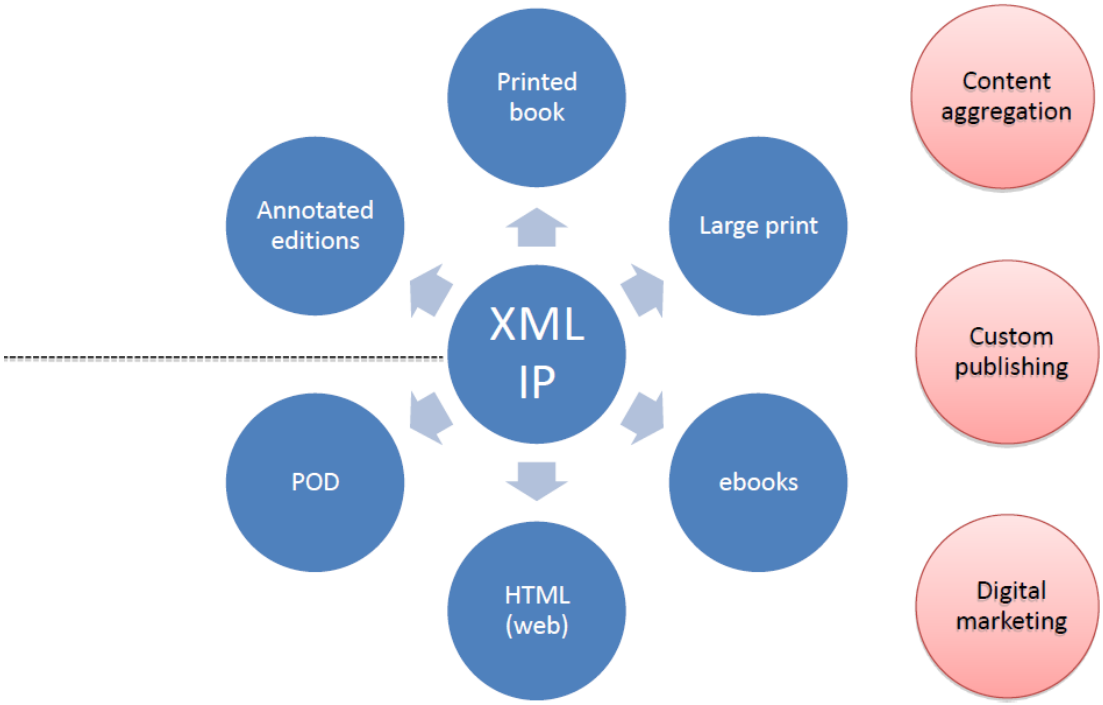
Un mensaje ONIX (ver sección 3.5) consiste en un conjunto de elementos definidos por etiquetas escritas en XML que siguen las especificaciones contenidas en una plantilla específica (DTD o Schema), que detalla, entre otras cosas, el orden y la relación entre los elementos.

4.4.2 Contenido y XML

Pero esta capacidad de XML para estructurar la información que ayuda a describir, que facilita su tratamiento automatizado, puede aplicarse no sólo a los metadatos, como ONIX, sino también al propio contenido. Esto permite, por ejemplo, partir de un documento en XML y convertirlo sin demasiada dificultad a diversos formatos, como PDF para impresión o para visualización en pantalla, HTML o ePub (ver

Transformaciones de XML). Esta separación entre formato y contenido posibilita a su vez una mayor independencia de los generadores de contenido respecto de la variedad de canales y formatos de distribución del mismo, que además están en continua evolución.

Figura: El contenido en XML como centro del proceso de producción de la industria editorial



Fuente: Brian O'Leary[60]

De hecho, con el crecimiento del libro electrónico y, en general, de la publicación digital, la gestión del propio contenido irá progresivamente desplazando a la producción de libros impresos del núcleo del negocio editorial. En el proceso para ganar independencia respecto de la producción de libros físicos, XML puede ser una herramienta muy útil, pues precisamente una de sus principales

[60] "Start with XML: A Practical Guide to When, Why and How", Brian O'Leary, BookNet Canada Tech Forum 2009 (http://www.slideshare.net/booknetcanada/bnctechforumbrianolearyxml)

funciones es la de separar el contenido y su forma de presentación, ya sea impresa o digital, permitiendo[61]:

- La integración de contenido: permite combinar y normalizar contenido estructurado y sin estructurar en diferentes formatos;
- La reutilización de contenido: utilizar el contenido en distintos productos sin tener que recrearlo o reformatearlo;
- La producción de múltiples formatos: satisfacer la demanda por parte de los clientes de formatos específicos de entrega del contenido sin necesidad de crear diferentes flujos de publicación;
- Una búsqueda más granular: más allá de la búsqueda por palabras clave, al permitir agregar etiquetas de distinto tipo al contenido.
- La indicación del contenido: capacidad de integrar, normalizar y reutilizar contenido de múltiples fuentes.

4.4.3 Contenido vs. metadatos

Pero acaso esta distinción entre datos o contenido y metadatos sea innecesaria y artificial, vestigio de una época en que las limitaciones de los sistemas de tratamiento de la información hacían imprescindible una separación que hoy podría ser abolida pues, en el fondo, como dice David Weinberger, todo son metadatos o, mejor dicho[62]: "*los metadatos son lo que ya conoces y los datos lo que estás buscando*".

[61] "A Blueprint for Book Publishing Transformation. 7 Essential Processes to Re-Invent Publishing" (pág. 109), BISG, 2010 (http://www.bisg.org/contentweb/recommended-reading/gilbane_blueprint/)
[62] "Data and metadata: together again", David Weinberger, 6 de septiembre de 2009 (http://www.hyperorg.com/blogger/2009/09/06/data-and-metadata-together-again/)

Referencias adicionales

Introducción

- "Publishing 3.0" (vídeo), Richard Nash, BookNet Canada Technology Forum de 2010 (http://blip.tv/booknet-canada/richard-nash-publishing-3-0-3472801)

- "DBW Insights: Richard Nash" (vídeo), 15 de junio de 2001, http://www.digitalbookworld.com/2011/dbw-insights-richard-nash/?et_mid=508163&rid=80960253

- "Context First: A Unified Theory of Publishing" (vídeo), Brian O'Leary (http://vimeo.com/20179653)

- "Piracy is Progressive Taxation, and Other Thoughts on the Evolution of Online Distribution", Tim O'Reilly, 11 de diciembre de 2002 (http://openp2p.com/lpt/a/3015)

- "Whoever Builds A Good Tool To Help Us Find New Books To Read Will Get Rich", Kevin Smokler, PaidContent (http://paidcontent.org/article/419-whoever-builds-a-good-tool-to-help-us-find-new-books-to-read-will-get-r/)

- "The Demand for Convenience — Attention Scarcity Reverses the Equations of Publishing", Scholarly Kitchen, 7 de junio de 2011 (http://scholarlykitchen.sspnet.org/2011/06/07/the-demand-for-convenience-attention-scarcity-reverses-the-equations-of-publishing/)

- "What's for Sale? Moving from Selling Content to Monetizing Attention", The Scholarly Kitchen, junio de 2011 (http://scholarlykitchen.sspnet.org/2011/06/02/whats-for-sale-moving-from-selling-content-to-monetizing-attention/)

- "The line between book and Internet will disappear", Hugh McGuire, O'Reilly Radar, septiembre de 2010 (http://radar.oreilly.com/2010/09/beyond-ebooks-publisher-as-api.html)

- "Understanding the Ebook Consumer: The Results of the BISG Consumer Survey", TOC 2010 (http://www.toccon.com/toc2010/public/schedule/detail/10724)

- "An Open, Webby, Book-Publishing Platform", Hugh McGuire, 19 de mayo de 2011 (http://radar.oreilly.com/2010/05/-wordpress-as-book-publishing.html)

- "The Changing World of Publishing and the Role of XML", Mike Shatzkin, 2009 (http://www.slideshare.net/StartWithXMLLondon/mike-shatzkin-the-changing-world-of-publishing-and-the-role-of-xml)

- "A View of the Future: Multi-Format Publishing at O'Reilly Media", Andrew Savikas, 2009 (http://www.slideshare.net/StartWithXMLLondon/a-view-of-the-future-multiformat-publishing-at-oreilly-media-by-andrew-savikas/download)

- "As E-book Sales Explode, Consumption Patterns Change", GigaOm, marzo de 2011 (http://www.businessweek.com/technology/content/mar2011/tc20110323_556521.htm)

Vías para la "descubribilidad"

Redes sociales

- "The 5 NEW Rules Of Social Media Optimization (SMO)", Rohit Bhargava, 2010 (http://www.rohitbhargava.com/2010/08/the-5-new-rules-of-social-media-optimization-smo.html)

Interfaces de Programación de Aplicaciones (APIs)

- "New Books API for developers" Google Books Search blog, 10 de mayo de 2011 (http://booksearch.blogspot.com/2011/05/new-books-api-for-developers.html)

- "The line between book and Internet will disappear", Hugh McGuire, O'Reilly Radar, septiembre de 2010 (http://radar.oreilly.com/2010/09/beyond-ebooks-publisher-as-api.html)

- "Publishing In The Digital Devices Boom: The Importance of APIs", Publishing Perspectives, 16 de abril de 2011 (http://publishingperspectives.com/2011/04/publishing-importance-of-api/)

Metadatos

- "Understanding Metadata", NISO, 2004, (www.niso.org/publications/press/UnderstandingMetadata.pdf)

- "Introduction to Metadata" segunda edición, The Getty Research Institute, 2008 (http://www.getty.edu/research/publications/electronic_publications/intrometadata/index.html)

- "Streamlining Book Metadata Workflow", NISO, 2009 (http://www.niso.org/publications/white_papers/StreamlineBookMetadataWorkflowWhitePaper.pdf)

- "How to Sell More Books with Metadata", Publishing Perspectives, enero de 2011 (http://publishingperspectives.com/2011/01/sell-more-books-with-metadata/)

ONIX

- "An Overview of the ONIX Metadata Scheme" (pdf), David Hamrick, 2004 (http://courses.unt.edu/SLIS5223Resources/Spring2005/WebCT/Assignments/ResearchPapers/HamrickResearchPaperFinal.pdf)

- "ONIX", EDItEUR (http://www.editeur.org/8/ONIX/)

- "Increase Sales and Lower Costs With Better Metadata: A Step-By-Step ONIX Guide for Publishers", BookNet Canada (pdf) (http://www.booknetcanada.ca/images/media/StC/bnc_onix_marketing_v1.pdf); blog (http://booknetcanada.ca/index.php?option=com_wordpress&p=1714&Itemid=312)

La identificación de textos en formato digital

- "ISBN. It's about metadata!", Pensódromo 21, 28 de enero de 2011 (http://www.pensodromo.com/21/2011/01/isbn-its-about-metadata/)

- "ISBN", EDItEUR (http://www.editeur.org/35/ISBN/)

Sistemas de clasificación de materias

- "BIC Standard Subject Categories", BIC (http://www.bic.org.uk/7/BIC-Standard-Subject-Categories/)

- "BISAC Subject Headings. 2010 Edition", BISG (http://www.bisg.org/publications/product.php?p=14)

XML (*eXtensible Markup Language*)

- "A Blueprint for Book Publishing Transformation. 7 Essential Processes to Re-Invent Publishing", BISG, 2010 (http://www.bisg.org/contentweb/recommended-reading/gilbane_blueprint/)

Qué aporta XML a las editoriales

- "The Changing World of Publishing and the Role of XML", Mike Shatzkin, 2009 (http://www.slideshare.net/StartWithXMLLondon/mike-shatzkin-the-changing-world-of-publishing-and-the-role-of-xml)

- "Toward Agility & Efficiency: Best Practices for 'Future-Proofing' New Content", Bill Kasdorf, 2011 (http://www.bisg.org/mip/program-description/schedule-of-events/)

Contenido y XML

- "Metadata vs Data: a wholly artificial distinction", Terry Jones, Blog de Fluidinfo, 5 de septiembre de 2009 (http://blogs.fluidinfo.com/fluidinfo/2009/09/05/metadata-vs-data-a-wholly-artificial-distinction/)