

F. SISTEMAS Y TECNOLOGÍAS DE INFORMACION

Informe de situación

Tecnologías y sistemas de información: informe de situación 2013

Juan-Antonio Pastor-Sánchez y Tomás Saorín-Pérez

Pastor-Sánchez, Juan-Antonio; Saorín-Pérez, Tomás (2014). "Tecnologías y sistemas de información: informe de situación 2013". *Anuario ThinkEPI*, v. 8, pp. 291-300.



Resumen: Revisión de las principales novedades y tendencias durante el año 2013 en el ámbito de las tecnologías y sistemas de información. Se muestran los cambios e iniciativas más relevantes por parte de las grandes corporaciones tecnológicas y su incidencia en el planteamiento de aplicaciones y servicios en la nube. Se analiza la tendencia de uso de plataformas de experiencia de usuario en el sector de la gestión de contenidos. Se incluye un resumen de la actividad normativa de ISO y de las principales recomendaciones del W3C. Finalmente se aborda la incorporación de técnicas de *information-discovery* y *entity search* en los procesos de búsqueda y recuperación de información.

Palabras clave: Servicios en la nube, Gestión de contenidos web, Recomendaciones y estándares, Plataformas de experiencia de usuario, *Information discovery*, *Entity search*.

Title: Information technologies and systems

Abstract: In a review of the main developments and trends during 2013 in the field of information technology and systems, relevant changes and initiatives of main technology corporations (*Apple, Google, Microsoft*) are shown and their impact on applications and cloud computing services is discussed. The deployment of user experience platforms in web content management is also examined. The ISO standardization activity and main W3C recommendations are summarized. Finally, the review analyzes the incorporation of information discovery and entity search techniques into search and information retrieval processes.

Keywords: Cloud Computing Services, Web content management, Recommendations and standards, Users experience platforms, Information discovery, Entity search.

1. Más integración... menos opciones

Durante el último año se han planteado una serie de nuevos escenarios en el ámbito de las tecnologías de la información. Se trata de cambios que por un lado refuerzan o plantean nuevos modelos de negocio de grandes corporaciones, o aportan nuevas tecnologías y enfoques al desarrollo de productos y servicios de información. Los modelos de negocio basados en un único producto parecen destinados a desaparecer y buena muestra de ello es el cierre del buscador *Altavista*. En este sentido, resulta obvio afirmar que *Google* conforma el principal núcleo de desarrollo en el sector de internet, no solamente a nivel de servi-

cios, sino también de tecnologías que utilizan a diario miles de millones de usuarios y que tal vez constituyan la punta de lanza de innovaciones que se presenten a lo largo de los próximos años.

Como muestra cabe hablar de uno de los indicadores más destacados: el descenso durante el último año en la venta de PC (*Gartner, 2014*). A nivel de usuario personal o en el entorno doméstico, se prefieren dispositivos que permitan una mayor movilidad e interconexión entre sí. Las *tablets* y *smartphones* constituyen en estos momentos el paradigma, no sólo en cuanto a consumo de dispositivos, sino también en cuanto a los hábitos de acceso, uso y consulta de servicios y productos de todo tipo (**Terrenghi; García-Barrio; Oshlyansky, 2013**). Móviles y tabletas

son ubicuas, los ordenadores parecen condenados a ser arrinconados en el ámbito familiar, y comienzan a atisbarse nuevos cambios en otro icono de la sociedad de masas como el televisor, donde ha comenzado a desembarcar *Android*.

“Un dispositivo de calidad que disponga de aplicaciones que saquen partido de las funciones del hardware y del sistema operativo ya no es suficiente garantía de éxito”

Durante este último año se ha constatado que un dispositivo de calidad que disponga de aplicaciones que saquen partido de las funciones del hardware y del sistema operativo no es suficiente garantía de éxito. También se precisa un ecosistema de aplicaciones de terceros, *APIs* de desarrollo, plataformas de distribución y servicios en la nube que permitan a otras empresas y desarrolladores participar en el crecimiento de un modelo de negocio basado en el consumo de información y servicios fuertemente ligados a internet (**Kavis**, 2014, p. 23). Esto está generando la concentración de la oferta en unos pocos intermediarios, con ecosistemas que sueñan con crear experiencias valladas a las que se accede mediante el pago de aplicaciones, contenidos y servicios. ¿Podría ser el principio del fin de la descentralización masiva de la oferta que facilitaba internet?

Por otro lado, la separación existente entre el uso a nivel personal y a nivel laboral de servicios de internet resulta cada vez menos evidente. Es frecuente que los usuarios consulten tanto su correo personal como el de su trabajo utilizando los mismos dispositivos, y a veces desde una única aplicación como *Gmail*. El uso compartido de documentos de trabajo a través de cuentas personales de servicios de almacenamiento en la nube, como *Google Drive*, *Dropbox*, *iCloud*, *One Drive* o *Ubuntu One* resulta cada vez más común. Se trata de un mundo en el que los servicios, contenidos, datos, documentos, información de carácter personal, herramientas de comunicación y dispositivos físicos están cada vez más entrelazados. La tecnología está explorando las fronteras, extremadamente delicadas, entre la vida personal y la laboral (**Marquina**, 2013).

Buena muestra de ello son los movimientos continuados de compra/venta de empresas, aplicaciones y tecnologías por parte de corporaciones como *Google*, *Microsoft*, *Apple*, *Facebook* o *Lenovo* entre otras. La adquisición de *Nokia* por parte de *Microsoft*, o la de *Motorola* en primera instancia por *Google* y más adelante por *Lenovo* plantean nuevos interrogantes que sugieren la

gestación de un nuevo escenario. La integración de plataformas, hardware, servicios y productos conforma una necesidad cada vez más acuciante que está replanteando los cimientos de productos de largo recorrido como *Windows*, planteando un modelo de presencia multidispositivo parecido al de *Android* o *iOS*, e incluso ofreciendo una gama de dispositivos propia como *Surface*, con unos planteamientos al estilo de la familia *Nexus* de *Google* que ha visto consolidar sus ventas durante 2013.

Los usuarios están comenzando a habituarse, e incluso a demandar, una experiencia integral en el uso de los diferentes componentes de su experiencia tecnológica cotidiana. El concepto de plataforma de distribución que integre sistemas, servicios, productos y contenidos es una realidad que cobra más fuerza día a día. Este panorama de interconexión e interrelación de elementos no solamente parte desde las grandes empresas del sector, sino que comienza a calar en la conciencia colectiva de consumidores, profesionales, pequeñas y medianas empresas y administraciones públicas. A lo largo del presente informe se abordan ciertos aspectos relevantes que se han producido a lo largo del último año y que precisamente refuerzan la idea de que los sistemas y las tecnologías de la información tienden a encontrar puntos de confluencia que refuerzan la idea de una oferta integral de productos y servicios multiplataforma y multidispositivo, desarrollados a partir de estas nuevas estrategias y modelos de negocio de las empresas líderes en el sector.

“La experiencia de usuario en los sistemas informáticos corporativos empieza a quedar atrás con respecto a la versatilidad que se ofrece en el mercado de consumo personal”

2. Las nubes no dejan ver el bosque

Durante el año 2013 han tomado forma y se han consolidado ciertos cambios introducidos en este tipo de servicios durante 2012. *Apple* con *iTunes/iCloud* define un fuerte entrelazamiento con sus propios productos a través de las funcionalidades de copias de seguridad de archivos y datos y la interconexión de dispositivos. Aunque en un principio parece estar enfocado a la gestión de documentos y ficheros personales, el anuncio de una versión de *iWork* para *iCloud* sugiere que *Apple* ofrecerá a sus usuarios una gama de servicios en la nube similar a la de *Google*. Una posible reflexión es que la experiencia de usuario en los sistemas informáticos corporativos, los que las

organizaciones ponen a disposición de sus trabajadores para realizar sus tareas, empieza a quedar atrás con respecto a la versatilidad que se ofrece en el mercado de consumo personal. Trabajar puede convertirse en un viaje diario al pasado en cuanto a la usabilidad, funcionalidad e interactividad de las aplicaciones y sistemas.

Por su parte, la nube de Google conforma un ecosistema bien definido. Las continuas evoluciones de Gmail, junto con la consolidación como espacio de almacenamiento y entorno de trabajo de Google Drive son el contrapunto ideal a otras aplicaciones de desarrollo más pausado como Calendar o Maps. Google Drive se ha convertido en una herramienta ampliamente utilizada y su integración con el resto de servicios de Google hace que sea una opción muy interesante para usuarios personales o a nivel empresarial. También cabe destacar el cierre de Google Reader debido al "escaso uso" que, para Google, tenía este lector de RSS (Green, 2013). Por otro lado también es significativa la fusión de Google Currents y Google Play Magazines en un único producto: Google Play Newsstand. Tal vez desde Mountain View se pretendan potenciar nuevas formas de agregar y consumir contenidos, por lo que puede que estén poniendo orden dentro de sus aplicaciones en este campo.

Tanto Apple como Google conciben un uso de sus servicios ligado a determinados dispositivos. En el caso de Apple resulta evidente que el principal segmento de interés es el de sus propios usuarios. En el caso de Google no queda del todo claro, puesto que comienzan a postularse alternativas en el entorno de Android a la hegemonía de Google Play, como plataforma de distribución de aplicaciones y contenido. Muy posiblemente se plantee un nuevo e interesante panorama en cuanto a la distribución de aplicaciones, contenidos y servicios dentro del mundo Android, lo cual sin duda afectará al resto de internet. La combinación Google/Android es como un enorme transatlántico alrededor del cual navegan multitud de pequeñas empresas y proveedores de servicios: pueden verse beneficiadas por la corriente de innovación que se genere, pero un cambio de dirección en la evolución de estas tecnologías también puede arrollarlas.

Algunas empresas del sector de servicios en la nube o de internet en general, como Dropbox, Facebook o Amazon, pueden sentirse invulnerables a estos cambios, pero no serían de extrañar nuevas compraventas y cambios en la estructura empresarial que afecten a las plataformas de



Figura 1. Estructura de la plataforma eXo 3.5.

<http://docs.exoplatform.com/public/index.jsp?topic=%2FPLF35%2FOverview.html>

comercio electrónico, distribución de contenidos, e incluso soportes SaaS y PaaS.

Por supuesto, todo lo anterior sin perder de vista la necesidad de privacidad, seguridad, interoperabilidad, portabilidad y uso de estándares que es preciso considerar en todo momento al mencionar los servicios en la nube. Es posible que un usuario personal no contemple todos los factores anteriormente enumerados para utilizar un servicio que inicialmente es gratuito. Pero para una empresa que decide contratar son elementos críticos en su política de sistemas y tecnologías de la información. En este sentido resulta clave el informe de julio de 2013 del NIST y el Departamento de Comercio de Estados Unidos (NIST, 2013) sobre la hoja de ruta a seguir en el desarrollo y aplicación de estándares en el ámbito del *cloud computing*. Es muy interesante el análisis de lagunas y carencias existentes en este campo y el análisis de casos de uso y las propuestas de estandarización.

3. Publicación de contenidos: servicios frente a servidores

A mediados de septiembre de 2013 tuvo lugar en Londres el *Gartner Portals, Content & Collaboration Summit 2013*. En el encuentro se apuntó que la confluencia de las tecnologías móviles, redes sociales, contenidos, productos de información y servicios en la nube está cambiando el compromiso que deben adquirir los productores de contenidos y servicios, y por lo tanto las organizaciones precisan de estrategias globales que permitan aprovechar este escenario. En las distintas sesiones se propusieron técnicas de participación y colaboración para lograr mejores experiencias de usuario, multiplataforma y multiservicio. Desde el punto de vista operativo de las organizaciones, la productividad de los empleados podría incrementarse, pero únicamente mediante aplicaciones intuitivas, amigables, accesibles y atrac-

tivas. Este compromiso para crear experiencias de usuario debe contemplar a los individuos tanto dentro como fuera de la organización. Por este motivo la disponibilidad de grandes cantidades de contenidos no estructurados representa más una oportunidad que un problema. El uso de técnicas para incorporar estos contenidos, en su mayoría disponibles de forma abierta en internet, puede constituir una nueva fuente de inteligencia corporativa que podría utilizarse para el desarrollo de productos basados en ciclos y sectores de innovación identificados a partir del análisis de dichos contenidos. En este sentido se abordaron dos ideas interesantes, producto de la experiencia de la interacción entre usuarios/clientes y la propia organización: utilizar técnicas de *gamification* como un medio para potenciar la colaboración y la participación entre trabajadores, y el diseño de plataformas corporativas de aplicaciones como un mecanismo para la distribución y gestión de productos de software, ya sean gratuitos o de pago (Gartner, 2013).

Estas propuestas se concretan en las plataformas de experiencia de usuario (*UXP*, *user experience platform*) que integran diferentes tecnologías que permiten alcanzar un alto grado de interacción entre usuarios, aplicaciones, procesos, contenidos y por supuesto con otros usuarios. Una *UXP* ofrece una experiencia integral tanto a usuarios como a los gestores del espacio digital en internet de una organización (Phifer, 2011).

Los sistemas de gestión de contenidos tradicionales se orientan a la difusión de contenidos o servicios a través de una web pública corporativa mediante entornos que ofrecen multitud de funciones a los diseñadores y administradores del sistema, especialmente a través de la ampliación mediante *plugins* o extensiones. Sin embargo este

enfoque no siempre termina de encajar: la selección de extensiones compatibles es un proceso costoso. La adaptación a otros medios diferentes de una web corporativa (versiones móviles, plataformas de aplicaciones) resulta compleja y suele precisar de procesos de selección de ciertos contenidos y muchos desarrollos propios. La visión tradicional de gestión de la comunicación entre usuarios también suele realizarse de forma separada, no solo entre la organización y sus usuarios/clientes, sino también entre los propios usuarios que son potenciales miembros de una comunidad basada en la marca o en los productos adquiridos.

Durante 2013 se ha consolidado la oferta de este tipo de plataformas por parte de empresas especializadas en el sector como *Acquia*, *eZ Systems*, *Exo*, *Oxycyon* o *Liferay* entre otras. Más que un software específico con todas las funcionalidades de serie, lo que proporcionan estas empresas son soluciones ampliables y escalables, adaptadas a las necesidades de una organización en un momento dado o a lo largo del tiempo. Los análisis de *Forrester* señalan *Sitecore* como una excepción a esta tendencia, al pretender desarrollar una gama completa de todas las necesidades de gestión de la experiencia web en un sólo producto (Powers; Aponovich, 2013). En definitiva, se necesita manejar e integrar funcionalmente un conjunto de herramientas:

- gestión de portales y contenidos web;
- sistemas de búsqueda avanzados;
- *mashups* que conecten y reutilicen procesos, contenidos y servicios;
- aplicaciones web enriquecidas con un alto grado de usabilidad;
- sistemas de estadísticas de uso de contenidos y productos en tiempo real;
- herramientas avanzadas de comunicación y participación con los clientes y los proveedores;

- sistemas responsivos que permiten adaptar de un modo muy sencillo los contenidos, productos y servicios al ecosistema de los dispositivos móviles.

Hay un concepto a destacar y es que tras el hecho de “publicar contenidos” queda escondida la verdadera necesidad emergente y crítica: los editores necesi-

Tabla 1. Funciones y tecnologías adyacentes (Fuente: Powers; Aponovich, 2013)

Proceso	Funciones core de la gestión de contenidos web (WCM)	Tecnologías adyacentes
Gestionar	Repositorio de contenidos Autoría de contenidos Clasificación Gestión de taxonomías Permisos Flujo de trabajo Gestión de presentación (<i>layout</i>) Gestión de medios (<i>rich media</i>) Formularios Gestión de la segmentación	Autoría de contenidos Traducción Gestión de las relaciones con clientes Gestión de catálogos de productos Gestión de activos digitales (DAM) Formularios Gestión de la segmentación Gestión de campañas Gestión de procesos de negocio (BPM) Autoría desde el escritorio
Implicar	Autoría de la presentación Redes de contenidos (<i>content targeting</i>) Realización de campañas de correo Encuestas y retroalimentación	Autoría de la presentación Redes de contenidos (<i>content targeting</i>) Transacciones comerciales Redes sociales Desarrollo de aplicaciones móviles Servidores de anuncios
Medir	Test A/B Informes Optimización para buscadores	Test multivariante Análítica web Análítica social Optimización para buscadores

tan manejar lo que sucede después de la publicación del contenido, evaluar alternativas de organización del contenido (Test A-B), extraer información práctica sobre el comportamiento real de los usuarios, es decir, gestionar la experiencia de uso y no sólo la publicación (Kohavi et al., 2009). En palabras de Deane Baker (Tolvanen, 2013), “The big race is not in the side of the publishing button”.

El informe Forrester de 2013 (Powers; Aponovich, 2013) también insiste en que las funciones core de la gestión de contenidos no pueden concebirse separadas de otras funciones adyacentes cada vez más críticas.

La mejora de la experiencia de usuario con independencia del sistema operativo o dispositivo utilizado está en el centro de la simplificación de la interfaz de usuario. Mediante una interacción muy orientada a la tarea hemos descubierto que poníamos demasiadas opciones en cada página web y en este sentido las interfaces cada vez son más claras, amigables y adaptadas a las tecnologías táctiles. Podría decirse que comienzan a ser más limpias y adaptadas a la interacción táctil (Rogowski, 2013). Un detalle curioso es la extensión del concepto de “web visual” no solamente hacia plataformas que etiquetan y comparten imágenes, sino también a otras centradas en mostrar información sobre productos y servicios. No se trata de una tecnología, sino de una tendencia en el diseño que hace uso de grandes imágenes, cabeceras que ocupan todo el ancho de la página, metáforas del mundo de las revistas gráficas digitales, fuerte entrelazamiento entre títulos, imágenes y contenidos (Dans, 2013).

Desde el punto de vista de la gestión de contenidos web tradicional -la que nos cuenta que cualquiera tiene a su alcance los medios para montar su presencia en internet- hemos hablado en otras ocasiones de la extensión de diferentes plataformas de gestión de contenidos como WordPress, Joomla o Drupal. Desde luego sigue existiendo esa larga cola de sitios web (casi un 75% de los que usan un CMS están gestionados con alguno de los tres citados) pero cada vez se precisa una mayor profesionalización y medios para gestionar una presencia web multiplataforma con posibilidades sofisticadas de control de la interacción y capacidad de reacción para conectar entre partículas de contenido y conversaciones dispersas entre múltiples contenedores y servicios

The screenshot shows the searchFAST interface. At the top, there is a search bar with the keyword 'borges' entered. Below the search bar, there are two main sections: 'FAST TERMS' and 'TERM DETAILS'. The 'FAST TERMS' section displays a table of search results for 'borges', including terms like 'Borges, Jorge Luis, 1899-1986', 'Borge, Victor', 'Borg, Björn, 1956-', 'Kékszakállú herceg vára (Bartók, Béla)', 'Borges, Jacobo, 1931-', 'Aleph (Borges, Jorge Luis)', 'Borge, Tomás, 1930-2012', 'Borg-Warner Corporation', 'Ficciones (Borges, Jorge Luis)', and 'Borg, Carl Oscar, 1879-1947'. The 'TERM DETAILS' section provides information for the selected term 'Borges, Jorge Luis, 1899-1986', including 'USED FOR:' (listing various works and authors), 'SEE ALSO:', 'USAGE:', 'RECORD ID:', 'SOURCES AND OTHER LINKS:', and 'LINKS TO FULL RECORD:'.

Figura 2. Ejemplo de búsqueda en searchFAST

Fuente: <http://fast.oclc.org/searchfast>

en la nube. ¿Cuántos actores apuestan por el uso directo de servicios online frente al alojamiento en sus propios servidores? Los datos de Datanyze nos indican que Blogger y WordPress.com suponen el 10%, frente al 57,7 de WordPress.org, el gran dominador en todos los rankings¹.

La tendencia que parece imparable es la de un mercado de componentes para sitios web. Los sitios de medios con altas necesidades de gestión de la participación integran servicios de terceros como Discuss, Google Analytics, Mailchimp, Intensedebate, Doopress, Uservoice, o una plataforma especializada en cada función específica -votaciones, decisión, preguntas y respuestas, concursos, recomendaciones, etc.- mediante una integración transparente que desplaza la inversión en tecnologías en acuerdos con proveedores².

“Las interfaces cada vez son más claras, amigables y adaptadas a las tecnologías táctiles”

El mercado de proveedores de servicios web es determinante para desplegar toda la potencia de los contenidos que se quieren publicar (Walters et al., 2013). La combinación de un set de piezas de muy alta calidad, y bajo coste de mantenimiento al ser externas, permite reducir los costes tecnológicos de funcionamiento y que los productores se centren en la información y no en las plataformas.

4. Nuevas recomendaciones y estándares

ISO ha desarrollado una intensa actividad en el ámbito de la normalización.

La norma *ISO 25577:2013* aporta una estructura generalizada para el intercambio de registros bibliográficos y otro tipo de metadatos mediante XML.

La esperada segunda parte de la norma sobre tesauros *ISO 25964-2:2013* aborda la interoperabilidad de los tesauros con otros vocabularios. En el contexto de la web semántica y *linked open data (LOD)* resulta un documento muy valioso, ya que ofrece una serie de recomendaciones para la definición y mantenimiento de relaciones de mapeado entre los tesauros y otros tipos de vocabularios controlados.

También dentro de la organización del conocimiento, se ha publicado la norma *ISO/IEC 13250-3:2013* que define una sintaxis XML para el intercambio de datos relativos a *topic maps*.

Igualmente resultan de gran interés las cinco partes de la norma *ISO/IEC 20944* sobre la interoperabilidad de registros de metadatos. Entre otros contenidos, la norma se centra en la definición de los atributos básicos que se requieren para la descripción de elementos de metadatos, su interoperabilidad, definición de un modelo de datos conceptual de un registro de metadatos, estructura y consistencia de los registros de metadatos, etc.

“Durante 2013 se ha trabajado en las recomendaciones de procesamiento y sintaxis para el almacenamiento de la firma digital y el cifrado de datos en documentos XML”

En el contexto de la web, *ISO/TR 14873:2013* aborda la definición de estadísticas, términos y criterios de calidad para el archivo de recursos web, que frecuentemente se realizan en el seno de bibliotecas, archivos, museos y otras organizaciones. Se orienta a los profesionales que participan en estas prácticas y aporta una perspectiva amplia que integra aspectos técnicos informáticos, procesos de gestión y biblioteconomía.

Las normas *ISO/IEC 17963:2013* e *ISO/IEC 18450:2013* abordan los servicios web desde dos enfoques. En el primer caso se describe un protocolo de servicios web de gestión (*WS-management*) basado en SOAP, que puede ser de utilidad en el ámbito de la interoperabilidad para el intercambio de recursos de información. En el segundo se normaliza el lenguaje de descripción

de servicios web (WSDL) que permite especificar el modo de acceso a los mismos y sus mecanismos de uso.

Durante 2013 el W3C ha publicado la familia de recomendaciones *RIF (rule interchange format)*, compuesta por seis documentos y que completa, junto con *OWL*, *RDF Schema* y *Sparql*, la capa de tecnologías necesarias para la definición interoperable de predicados de lógica de primer orden y la consiguiente ejecución de procesos inferencia y descubrimiento de información. En los próximos años veremos si esta tecnología tiene un impacto significativo en el funcionamiento de motores de búsqueda o en el diseño de nuevos servicios y productos basados en la explotación de conjuntos de datos abiertos expresados mediante *RDF/OWL* (W3C, 2013b).

Sparql ha sido objeto también de un intenso trabajo por parte del correspondiente grupo de trabajo del W3C (2013c). Los once documentos que conforman la recomendación *Sparql 1.1*, realizan aportaciones interesantes, en especial en cuanto a las posibilidades de esta tecnología para la actualización de conjuntos de datos *RDF* (la recomendación anterior únicamente contemplaba consultas para recuperar datos), la posibilidad de realizar consultas federadas que impliquen el uso de diferentes *Sparql endpoint* en una misma búsqueda y la ampliación de los formatos que se pueden utilizar para representar los resultados.

Por otro lado se ha ampliado la familia de recomendaciones y tecnologías relativas a xml. Durante 2013 se ha trabajado en la elaboración de las recomendaciones que abordan el procesamiento y la sintaxis para el almacenamiento de la firma digital y el cifrado de datos en documentos xml (W3C, 2013d; 2013e). Este nuevo estándar tiene una aplicación inmediata en el ámbito de la administración electrónica y las transacciones seguras en internet.

El marcado semántico de contenidos web también ha conformado un punto de interés del W3C. Las segundas ediciones de las recomendaciones *RDFa Core 1.1* y *xhtml+RDFa 1.1* han venido acompañadas de la nueva recomendación *html+RDFa 1.1* (W3C, 2013a), que supera el cisma existente entre *xhtml* y *html5*. Por lo tanto ya es posible utilizar un mismo modelo de marcado (en este caso *RDFa 1.1*) en contenidos web, con independencia del lenguaje de marcado utilizado. Será interesante observar en el futuro cómo afectan estas recomendaciones al uso de los microdatos, una tecnología por la que apostaron *Google*, *Yahoo* y *Microsoft* debido a la inexistencia de otra alternativa para el marcado semántico.

Todas estas tecnologías han sido objeto de mención y aplicación en la Resolución de 19 de febrero de 2013, de la *Secretaría de Estado de Administraciones Públicas*, sobre la *Norma técnica*

de interoperabilidad de reutilización de recursos de la información (BOE núm. 54, de 4 de marzo de 2013, pp. 17045-17071). En dicha resolución se establece el conjunto de pautas básicas para reutilizar documentos y recursos de información disponibles en el ámbito de la administración pública por parte de cualquier agente interesado. También se define el contexto de la reutilización de la información, especificando los mecanismos de identificación de recursos, protocolos y tecnologías a utilizar, y los esquemas y términos de metadatos para la descripción mínima de los documentos y recursos de información reutilizables que se pondrán a disposición pública. Se trata de una norma muy concreta y centrada en aspectos técnicos que hacen referencia de forma explícita al uso de tecnologías de largo recorrido como *RDF*, *OWL* o *Skos*. Además la resolución es aún más pertinente, teniendo en cuenta la publicación de la *Directiva 2013/37/UE* del Parlamento Europeo y del Consejo, de 26 de junio de 2013 que modifica, profundiza e insiste en muchos de los aspectos de la *Directiva 2003/98/CE* relativa a la reutilización de la información del sector público.

5. Más allá de los repositorios: *information discovery* y *entity search*

Tras el esfuerzo sostenido durante años para levantar y alimentar repositorios digitales, es necesario poner atención en las tecnologías óptimas para asegurar su explotación por parte de las comunidades de usuarios. Podemos identificar una tendencia fuerte hacia los sistemas de *information discovery*.

Que los repositorios -tanto científicos como patrimoniales- cumplan sus objetivos depende de que se llegue a sus contenidos en contextos reales de uso. Que millones de objetos digitales se incorporen a los procesos de producción científica, estudio cultural y comunicación pública. *Europeana* tiene estrategias para potenciar este compromiso, basadas en exposiciones en línea y la oferta de *APIs* para que sus contenidos alimenten otros proyectos culturales. A través de la potencia del *Europeana data model*, se pretende



Figura 3. Big data y big content <http://info.aiim.org/digital-landfill/newaiimol2012/03/15/big-data-and-big-content-just-hype-or-a-real-opportunity>

que el esfuerzo de agregación de contenidos se manifieste en mejoras tangibles para descubrir recursos de información. El posicionamiento en buscadores es sólo una parte, por debajo late una apuesta por el aprovechamiento de metadatos para interrelacionar recursos, tratando de sacar valor del escenario de LOD y la web semántica. *Europeana* es un ecosistema, y éstos sólo sobreviven si son capaces de aportar más valor del que capturan.

“Se hace necesaria una mayor calidad del control de autoridades y del etiquetado en los repositorios”

Cada vez se le da más importancia al principio de que la calidad de un sistema está en sus metadatos. Por este motivo en el universo de los repositorios se hace necesaria una mayor calidad del control de autoridades y del etiquetado. Estos son la materia que permitirá elaborar sistemas con una experiencia de usuario avanzada, o con posibilidades de *discovery* potenciadas. Un ejemplo de esto es la atención que le dedica la *AIIM* en su *whitepaper: How to develop taxonomies to support navigation, information discovery and findability* (Weise, 2013).

El motor de búsqueda semántica de *Europeana* aún permanece en sus laboratorios, pero a partir de la mejora en el manejo de metadatos, se percibe una promesa de mayor capacidad de descubrimiento de objetos digitales. En consecuencia

la extensión de sistemas para la indexación como *FAST (Faceted Application of Subject Terminology)* respaldados por la *Library of Congress* y *OCLC*, puede mejorar la buscabilidad de la información catalogada a gran escala multiplicando las relaciones semánticas entre recursos. ¿Nos estamos acercando a lo que *OCLC* llamaba “el santo grial de la búsqueda” (**Prescott**, 2011)?

En el ámbito de las bibliotecas científicas ya podemos considerar maduras a las herramientas de última generación para la búsqueda integrada en fuentes digitales heterogéneas. Las herramientas de descubrimiento web para bibliotecas digitales (*WSD, Web Scale Discovery Services*), como *Primo Central* o *Summon*, unifican la experiencia de búsqueda en catálogos, proveedores, agregadores y repositorios (**Ávila-García**, 2013; **Breeding**, 2013). Se basan en el filtrado por facetas y la integración previa de índices mediante recolección (a diferencia del modelo clásico de metabuscadores); la interfaz de usuario apuesta por la caja única de búsqueda y el ranking de resultados. Por supuesto, en combinación con servicios de resolución de enlaces para el acceso al texto completo. Los trabajos de la *Open Discovery Initiative (ODI)* de *NISO* tratan de consolidar estándares en este campo para la interoperabilidad de formatos, esquemas y mecanismos de transformación (*ODI*, 2013).

“La fidelidad es tenue y un desajuste en la experiencia de usuario o de canal puede lanzar a los usuarios a los brazos de la competencia”

Aunque estas tecnologías para la búsqueda se promocionan bajo el lema de sencillez (“funciona igual que *Google*”), lo cierto es que requieren un esfuerzo complementario en formación de usuarios, para aprovechar la riqueza que pueden aportar (**Richardson**, 2013). Puede resultar de interés el concepto de “interfaces generosas”, en relación con la necesidad de crear experiencias de búsqueda y descubrimiento que aprovechen, sin desbordar, la ductilidad y potencialidad de la información en bases de datos (**Whitelaw**, 2013). En términos más generales, los expertos consideran que la búsqueda corporativa (*enterprise search*) sigue siendo un territorio francamente mal resuelto y un quebradero de cabeza para los diseñadores de sistemas de información. Por otro lado hemos de tener en cuenta que las necesidades de búsquedas internas de una organización (*business search*) son muy diferentes de las del cliente del sitio web (*consumer search*) y que **Rosenfeld** (2011) denomina *findability gap*. La

búsqueda interna de productos, documentos y, en general, cualquier información de valor sobre una organización, ha de desarrollar aquellos aspectos que no pueden ser tenidos en cuenta por un buscador masivo externo, especialmente para la información de transacciones almacenada en bases de datos no expuestas a los índices (la *deep web*, la *internet invisible*).

En este sentido herramientas como *Elasticsearch* muestran una adaptación adecuada. Su arquitectura se basa en *Lucene*, pero mejora el rendimiento, en cuanto a la escalabilidad e integración de múltiples índices. *Elasticsearch* proporciona una serie de *APIs* bien definidos para el acceso a los índices y los resultados de las búsquedas en tiempo real. Desde el punto de vista de la escalabilidad, también resulta novedoso e interesante que este sistema incorpore un componente denominado *Percolator* que ofrece un procedimiento inverso a la tradicional indexación de documentos, a través del almacenamiento de consultas y la identificación de las más adecuadas a cada documento. Como muestra de la potencia de este nuevo tipo de enfoque, la fundación *Wikimedia* tiene previsto incorporar este sistema en todos sus proyectos entre los meses de marzo y abril de 2014.

<http://www.elasticsearch.org>

El descubrimiento de información se apoya en otra área de la representación de información: el reconocimiento de entidades (*entity recognition*). Aquí está el futuro inmediato del posicionamiento en buscadores (**Bruemmer**, 2013), una cuestión de significados y no simplemente de *keywords*. La extensión de tecnologías de la web semántica permitirá diferenciar entre un documento que trata sobre *Wikipedia*, de aquel que usa la coletilla “...como dice *Wikipedia*”. En el primero se reconoce tras el término a la entidad, el concepto relevante, individualizado en una ontología y con una URI de referencia. Esta extracción de conceptos desde el lenguaje natural ya puede encontrarse en la oferta de aplicaciones de gestión documental³. La posibilidad del procesamiento semántico a gran escala de contenido ha generado la etiqueta *big content*, como contrapeso a *big data* (**Backaitis**, 2013). Por lo tanto conviene prestar atención a cómo los sistemas de gestión de contenidos incorporan la gestión de la semántica (**Behrendt; Damjanovic**, 2013). La publicación semántica tiene implicaciones más allá del campo del SEO y del uso de *Knowledge graph* por parte de *Google*. La idea que comprime el reto de la publicación semántica, y que conduce a *information discovery*, es construir “relaciones que expanden significados” (**Milhollin**, 2012). Además, detrás de plataformas de recomendación como *Outbrain* o *Autonomy*, que encontramos incorporadas en medios como *El diario*



<http://www.google.com/insidesearch/features>

o *El país*, se encuentran desarrollos punteros en *e-discovery* y oportunidades de negocio digital. <http://www.google.com/insidesearch/features/search/knowledge.html>

6. Conclusiones

Quizá sería exagerado decir que estamos en la era post-web, pero sí que resulta cada vez más importante desarrollar una presencia digital que integre todos los canales y flujos de información para posicionarse en la atención al usuario y monetizar los proyectos. Quedarse fuera de un dispositivo puede ser la oportunidad para la competencia de quedarse con los usuarios de una compañía rival. La fidelidad es tenue y un desajuste en la experiencia de usuario o de canal puede lanzar a los usuarios a los brazos de la competencia.

En un futuro podría resultar interesante explorar una mayor interrelación entre paquetes cerrados de contenido (como el *ebook* de alto coste unitario y escasa interacción con el lector) y la publicación digital enriquecida, fluida y publicada en un medio abierto y navegable. La integración de la gestión de contenidos y la distribución de ficheros puede maximizar los procesos de publicación digital.

Por este motivo tal vez sea adecuado comenzar a invertir en mejorar la búsqueda, adoptando e implantando nuevos sistemas de recuperación de información o contratando proveedores de búsqueda para sitios corporativos. El objetivo es potenciar la experiencia de búsqueda y consulta, tanto de contenidos de diferentes fuentes como de objetos digitales cerrados, a través de un único proceso de búsqueda y aportando una experiencia óptima de usuario.

Hay que considerar que es el contexto de uso el que determina la propuesta de experiencia de usuario, mediante una capa diferenciada, que opera sobre una base común de contenidos digitales, que gracias a una adecuada codificación y manejo de datos, puede presentarse de forma adaptada a múltiples dispositivos e incluso plata-

formas de distribución. No es sólo que un contenido se visualice de forma adecuada en cualquier pantalla, sino que se adapte a un contexto (sitio web, tienda de aplicaciones, etc.) con datos e informaciones que derivan del acto de consulta del contenido. Las experiencias digitales cada vez son más contextuales y esto supone que se multiplica la cantidad y calidad de los datos que intervienen en un, aparentemente sencillo, proceso de consulta de una página web.

“Hay que potenciar la experiencia de búsqueda y consulta, tanto de contenidos de diferentes fuentes como de objetos digitales cerrados, a través de un único proceso de búsqueda”

7. Notas

1. CMS market share in the Alexa top 1M, *Datanyze*. <http://www.datanyze.com/market-share/CMS>
2. A este respecto es destacable el servicio que ofrece *Medium*, una reinención de los comentarios sobre contenidos y las referencias cruzadas. <https://medium.com>
3. Como muestra esta nota comparativa del *Athento's text analyzer* frente al *Stanbol semantic motor*. <http://smartdocumentmanagement.athento.com/2013/03/athentos-text-analyzer-versus-stanbol.html>

8. Bibliografía

- Ávila-García, Lorena** (2013). *Herramientas de descubrimiento en bibliotecas universitarias*. Trabajo fin de Master, Universidad Carlos III, Octubre. <http://hdl.handle.net/10835/2569>
- Backaitis, Virginia** (2013). “Big content’s big deal? What Acquia, Alfresco, Alta Plana, Bloomfire, EMC IIG, Gartner Say”. *CMS Wire*, mayo. <http://www.cmswire.com/cms/customer-experience/big-content-big-deal-what-acquia-alfresco-alta-plana-bloomfire-emc-iig-gartner-say-021039.php>

Behrendt, Wernher; Damjanovic, Violeta (2013). *Developing semantic CMS applications. The IKS Handbook*. <http://www.iks-project.eu/sites/default/files/IKS%20Handbook%202013.pdf>

Breeding, Marshall (2013). "Library technology: the next generation". *Computers in libraries*, v. 33, n. 8, October.

Bruemmer, Paul (2013). "Future SEO: understanding entity search". *Search engine land*, 7 octubre. <http://searchengineland.com/future-seo-understanding-entity-search-172997>

Dans, Enrique (2013). "El año de la web visual". *El blog de Enrique Dans*, 19 diciembre. <http://www.enriquedans.com/2013/12/el-ano-de-la-web-visual.html>

Gartner (2013). *Gartner portals, content & collaboration summit 2013: trip report*. <http://www.gartner.com/imagesrv/summits/docs/emeal/portals/PCC-EMEA-2013-Trip-report.pdf>

Gartner (2014). *Gartner says worldwide PC shipments declined 6.9 percent in fourth quarter of 2013*. <http://www.gartner.com/newsroom/id2647517>

Green, Alan (2013). "Powering down Google Reader". *Google reader blog*, 13 marzo. <http://googlereader.blogspot.co.nz/2013/03/powering-down-google-reader.html>

Kavis, Michael J. (2014). *Architecting the cloud: design decisions for cloud computing service models (SaaS, PaaS, and IaaS)*. New Jersey, Wiley. ISBN: 978 1 118 61761 8

Kohavi, Ron; Longbotham, Roger; Sommerfield, Dan; Henne, Randal M. (2009). "Controlled experiments on the web: survey and practical guide". *Data mining and knowledge discovery*, v. 18, n. 1, pp. 140–181. <http://ai.stanford.edu/~ronnyk/2009controlledExperimentsOnTheWebSurvey.pdf>

Marquina, Julián (2013). "La vida en la nube: big data y cloud computing". *Julian Marquina*, 22 abril. <http://www.julianmarquina.es/la-vida-en-la-nube-big-data-y-cloud-computing>

Milhollin, Kristen (2012). "Dynamic semantic publishing for beginners. Part 3". *Semanticweb.com*, 2 julio. http://semanticweb.com/dynamic-semantic-publishing-for-beginners-part-3_b30352

NIST (2013). *NIST cloud computing standards roadmap*. NIST Special Publication 500-291, version 2. U. S. Department of Commerce & National Institute of Standards and Technology. http://www.nist.gov/itl/cloud/upload/NIST_SP-500-291_Version-2_2013_June18_FINAL.pdf

ODI (2013). *ODI survey report reflections and perspectives on discovery services*, January. http://www.niso.org/apps/group_public/document.php?document_id=9977

Phifer, Gene (2011). *Hype cycle for web and user interaction technologies*. Stamford, CT: Gartner.

Powers, Stephen; Aponovich, David (2013). *The Forrester Wave: web content management for digital sus-*

tomer experience, Forrester Research, Q2, april 2013.

Prescott, Leah (2011). *Single search: the quest for the Holy Grail*. OCLC, July. <http://www.oclc.org/research/publications/library/2011/2011-17r.html?urlm=161047>

Richardson, Hillary A. H. (2013). "Revelations from the literature: how web-scale discovery has already changed us". *Infotoday*, May. <http://www.infotoday.com/cilmag/may13/Richardson-How-Web-Scale-Discovery-Has-Already-Changed-Us.shtml>

Rogowski, Ron (2013). "Digital customer experience trends to watch". *Forrester research*, January. http://customercommunications.com/wp-content/uploads/2013/10/Digital_Customer_Experien.pdf

Rosenfeld, Louis (2011). *Search analytics for your site*. Rosenfeld Media. ISBN. 1 933820 20 9

Terrenghi, Lucía; García-Barrio, Laura; Oshlyansky, Lidia (2013). "Tablets use in emerging markets: an exploration". En: *MobileHCI'13 Proceedings of the 15th international conference on human-computer interaction with mobile devices and services*, pp. 594-599. <http://static.googleusercontent.com/media/research.google.com/es//pubs/archive/41609.pdf>

Tolvanen, Perttu (2013). "Future, present and history of CMSs as told by CMS expert Deane Barker". *North patrol*, 11 noviembre. <http://northpatrol.com/2013/11/11/future-present-and-history-of-cmss-as-told-by-cms-expert-deane-barker>

W3C (2013a). *HTML+RDFa 1.1. Support for RDFa in HTML4 and HTML5*. W3C Recommendation 22 August 2013. <http://www.w3.org/TR/2013/REC-html-rdfa-20130822>

W3C (2013b). *RIF overview (Second Edition)*. W3C Working Group Note 5 February. <http://www.w3.org/TR/2013/NOTE-rif-overview-20130205>

W3C (2013c). *SPARQL 1.1 Overview*. W3C Recommendation 21 March. <http://www.w3.org/TR/2013/REC-sparql11-overview-20130321>

W3C (2013d). *XML Encryption syntax and processing. Version 1.1*. W3C Recommendation 11 April. <http://www.w3.org/TR/2013/REC-xmlenc-core1-20130411>

W3C (2013e). *XML Signature syntax and processing. Version 1.1*. W3C Recommendation 11 April. <http://www.w3.org/TR/2013/REC-xmlsig-core1-20130411>

Walters, Tim; McKnight, Cathy; Kay, Marianne; Elliot, Tiffany; Liewehr, Scott (2013). *Guide to service providers for web content and customer experience management*. Digital Clarity Group, 2013 http://www.citytechinc.com/content/dam/citytechinc/pdf/DCG_CITYTECH130812.pdf

Weise, Carl (2013). *How to develop taxonomies to support navigation, information discovery and findability*, AIIM. <http://www.trindocs.com/Portals/3/Taxonomy-Briefing-2012-v01.pdf>

Whitelaw, Mitchell (2011). *Generous interfaces*. National Digital Forum. <http://es.slideshare.net/mtchll/generous-interfaces>