

# Las bases de datos jurídicas y el uso del lenguaje XML en España

**M.<sup>a</sup> Luisa Alvite Díez**

Universidad de León (España)

## Resumen

Se presenta el actual estado de utilización de XML y tecnologías asociadas en las bases de datos jurídicas españolas. Los resultados han sido sistematizados en tres bloques: bases de datos comerciales, portales institucionales y proyectos de investigación. Se percibe una carencia de apuestas institucionales conjuntas, si bien se observan iniciativas aisladas desde diversos organismos públicos. Por otro lado, se subraya la existencia de consolidados y dinámicos grupos de investigación que trabajan en XML legislativo. Se concluye constatando la penetración indiscutible de las tecnologías XML, si bien el nivel de explotación de las mismas es ciertamente dispar. Se sugiere la necesidad de compartir experiencias y aunar esfuerzos.

**Palabras clave:** Bases de datos jurídicas. España. XML legislativo.

## Abstract

The current state of the use of XML technology related to the Spanish legal databases is presented. The results have been gathered in three blocks: commercial databases, institutional portals and research projects. The modest level of institutional implementations is noticeable, though isolated initiatives from several public organisations are perceived. On the other hand, the activity of established and strong research groups that are using legislative XML is highlighted. Undoubtedly, XML has been introduced in the legal databases; nevertheless, the development of its possibilities is irregular. We suggest the need to share experiences and efforts.

**Keywords:** Legal databases. Spain. Legislative XML.

## 1. Introducción

El origen de las primeras bases de datos jurídicas en el ámbito internacional se remonta a la década de los sesenta del pasado siglo, con el nacimiento de Lexis en Estados Unidos, QUIC/LAW en Canadá, CREDOC en Bélgica, etcétera. Desde entonces, la aplicación de tecnologías de la información en el ámbito del derecho ha supuesto transformaciones fundamentales en lo que respecta al soporte y modo

de acceso a la documentación jurídica, proporcionando, asimismo, una oportuna solución a demandas decimonónicas: tratamiento adecuado del volumen creciente de información jurídica, necesidad de una actualización permanente y resolución de las conexiones e interrelaciones legislativas, jurisprudenciales y doctrinales intrínsecas a este tipo de documentos.

Los primeros pasos de las bases de datos jurídicas en España se vienen situando en 1983. A partir de este año se evidencia una continua progresión tanto en la producción como en la distribución de bases de datos encabezada por cuatro instituciones fundamentales: las Cortes Generales, el *Boletín Oficial del Estado* y los Tribunales Supremo y Constitucional. El desarrollo autonómico y comunitario, así como la decisión por parte de empresas privadas de sumarse a la elaboración de productos electrónicos fueron configurando el sector de la edición jurídica en soporte electrónico en nuestro país. La consolidación del mercado de las bases de datos jurídicas, fundamentada en el disco óptico, se asienta entre los años 1990 y 1997. El CD-ROM fue percibido por las editoriales jurídicas como una continuación de su modelo de negocio basado, en gran medida, en las suscripciones a repertorios impresos (Alvite, 2004).

El éxito de Internet y, en concreto, de la World Wide Web, ha desencadenado un cambio tecnológico trascendental en la industria de la información; la Web, convertida en interfaz única, ha alterado radicalmente el mercado de los contenidos en cuanto a la producción, almacenamiento, difusión y recuperación de la información y, todo ello, a una velocidad vertiginosa. En la segunda mitad de la década de los noventa, la Web pasa a convertirse en plataforma óptima para la distribución de contenidos jurídicos. Así, desde 1996 los diversos editores jurídicos comienzan paulatinamente a disponer de sedes web y se inicia la inevitable migración de las bases de datos al nuevo entorno.

Este nuevo contexto ha posibilitado la diversificación del mercado de la información jurídica. De un lado, los tradicionales productores y distribuidores de bases de datos afianzados con anterioridad en el mercado; de otro, las instituciones públicas que han decidido aprovechar la oportunidad para editar en la Web sus bases de datos como servicio a la sociedad en su conjunto; por último, nuevas iniciativas nacidas en el entorno web, normalmente integradas en portales jurídicos, dirigidos prioritariamente al entorno profesional más que al gran público (VLex, Tirant on line, Iustel, Derecho.com, etcétera).

En cuanto al consumo de información jurídica, resultan elocuentes las cifras señaladas por Benjamins et ál. (2005), tomando datos del trabajo de Abel, que revelan un crecimiento exponencial de la documentación jurídica corroborado por el número de profesionales del derecho, que tanto en Estados Unidos como en la Unión Europea se ha triplicado e incluso cuádruplicado desde los años cincuenta del pasado siglo. La distribución de contenidos jurídicos en nuestro país cuenta con

una participación mayoritaria de la iniciativa privada, aspecto este inusual con respecto a la industria de las bases de datos en España. Esta peculiaridad guarda una vinculación directa con la importancia histórica de las empresas privadas en la publicación de repertorios jurídicos, pero ahora han de competir, manteniendo sus objetivos comerciales, con fuentes de información de nuevo cuño, que ofrecen en la Red, en muchos casos, información gratuita. Un nuevo contexto en el que se precisa demostrar un alto valor añadido que singularice contenidos de pago, en apariencia semejantes a los disponibles sin coste alguno.

Como hemos señalado en otra ocasión (Alvite, 2004), la documentación jurídica vertebraba indiscutiblemente la vida cotidiana; sin embargo, en gran parte, las bases de datos evidencian un tratamiento y un diseño orientados a juristas más que genéricamente a ciudadanos, y es que la búsqueda de información para los profesionales del derecho constituye una parte integral de su trabajo, en el que el proceso de recuperación y el de decisión se hallan profundamente imbricados. Ahora bien, la disponibilidad y la fácil accesibilidad a la información jurídica favorecen un incuestionable derecho de la ciudadanía. A los usuarios, digamos tradicionales, se han de sumar los ciudadanos que demandan servicios de e-gobierno y que exigen la interrelación e interoperabilidad con los sistemas de información legal.

Según Castells (2000), hemos pasado de una sociedad industrial a una posindustrial cuya nueva base se sustenta en el manejo y la transmisión de información de toda clase, esto es, la sociedad de la información, en la que la acumulación de productos se ha sustituido por la acumulación de conocimiento. Así, se entiende que el nuevo reto de la Unión Europea fijado en la Cumbre de Lisboa de 2000 sea convertirse en 2010 en la sociedad del conocimiento más competitiva del mundo. Pues bien, en esta nueva sociedad, la información legal emerge como un pilar fundamental, ocupando un lugar prioritario, dentro de los desarrollos de e-gobierno.

En este marco contextual es donde hemos de felicitarnos por la puesta en marcha de un proyecto ambicioso y necesario como el LEFIS —Legal Framework for the Information Society (<http://www.lefis.org>)—, proyecto coordinado por el doctor Fernando Galindo Ayuda, de la Universidad de Zaragoza, y en el que participan más de un centenar de instituciones. Pretende el desarrollo e implementación de una infraestructura para la enseñanza e investigación en el dominio jurídico dentro de sociedad de la información y el conocimiento. Dentro del LEFIS debemos destacar, sin duda, la iniciativa EGOBS (Electronic Government Observatory).

## **2. Adaptabilidad de las tecnologías XML a la documentación jurídica**

Entendemos por documento jurídico, partiendo de la definición establecida por Reviriego y Maciá (1998, pp. 19-20), la expresión de la información jurídica con un contenido parlamentario, legislativo, jurisprudencial o de doctrina jurídica. El documento jurídico, exceptuando los de doctrina, tiene unas características propias

que lo distinguen de otros tipos de documentos y de información: proceden de fuentes muy concretas —atendiendo al órgano de origen del documento—, son documentos tipificados, tienen una estructura definida de acuerdo con su contenido y están sometidos a un control que afecta tanto a su elaboración como a su aprobación y difusión.

La estructura de los documentos jurídicos se ha aprovechado en el diseño de los tradicionales sistemas de recuperación de información para tratar de mejorar su efectividad mediante la creación de diversos campos de búsqueda que respondan a las peculiaridades de este tipo de documentación. Schweighofer (1999) cita como ejemplo más extremo de esta aproximación la base de datos CELEX de la Comunidad Europea, diseñada con 10 índices principales y 80 campos.

Las marcas se han utilizado de un modo simple como ayuda en el procesado informático. En el sentido del ejemplo anterior, permiten la separación de campos de una base de datos; con el tiempo dieron paso a sistemas más complejos como los procesadores de texto y, con pretensiones más ambiciosas, aparecieron los lenguajes de marcas que posibilitan el uso del marcado de los documentos con fines documentales (Alvite, 2003).

A finales de los años sesenta del siglo anterior, tres investigadores contratados por IBM, Charles Goldfarb, Ed Mosher y Ray Lorie, recibieron el encargo de diseñar un sistema de edición, almacenamiento, búsqueda y gestión de documentos legales construyendo un sistema de formateo estructural al que, en un principio, se denominó GML. En 1986 se convertiría en un estándar, SGML —Standard Generalized Markup Language—, metalenguaje de etiquetado de texto convertido en norma ISO 8879.

Marta Blaquier (1996) subraya cómo los proyectos relevantes de marcado aplicados a la documentación legal no se presentaron hasta la década de los noventa, de la mano de editores privados como Kluwer o Thomson Legal Publishing y con importantes iniciativas acometidas por organismos públicos: la Government Printing Office, de Estados Unidos; la Her Majesty's Stationery Office, editora oficial del Reino Unido, y la Office for Official Publications (Publications Office) de las Comunidades Europeas. Igualmente reseñables resultan proyectos como el CETL (Center for Electronic Text in the Law), de la Universidad de Cincinnati, así como los proyectos europeos de Finlandia y Suecia; en este último país resulta obligado hacer una referencia al *Corpus Legis Project*, nacido en la Universidad de Estocolmo con el fin de elaborar recursos de textos electrónicos legales para la realización de estudios jurídico-lingüísticos.

Teniendo en cuenta las peculiaridades que caracterizan a los documentos jurídicos y a las bases de datos jurídicas, resulta incuestionable reconocer, como queda establecido en los trabajos de Nogales et ál. (2000) o de Nogales, Martín y

Arellano (2003), las enormes ventajas que aporta la aplicación de la tecnología web a la documentación jurídica, es decir, la utilización de lenguajes de marcas como medio de difusión de este tipo de información en detrimento de las tradicionales bases de datos.

Ha sido HTML, ISO/IEC 15445:2000, una aplicación del lenguaje SGML desarrollada inicialmente por Tim Berners-Lee que indica como se han de codificar los documentos para su distribución en la Web, el lenguaje con mayor presencia en la Red, y ello a pesar de sus notorias limitaciones. En nuestro país han sido varias las recopilaciones de documentos legislativos y jurisprudenciales pioneras en el uso del lenguaje de marcado con su correspondiente tratamiento hipertextual desde hace algunos años: Norm@civil —proyecto del área de Derecho Civil de la Universidad de Gerona que incluye legislación y jurisprudencia—, Noticias Jurídicas —portal de la editorial Bosch que contiene normativa concordada, artículos doctrinales o guía judicial—, el portal IUStel, algunos boletines oficiales autonómicos, etcétera.

En 1996 comenzó su andadura XML (eXtensible Markup Language), respaldado por el Consorcio Web (W3C) con la intención de diseñar un lenguaje de marcas capaz de integrar la simplicidad de HTML con la potencia de SGML. Las etiquetas XML marcan la estructura y la semántica del documento. El W3C publicó la versión 1.0 como recomendación en 1998 y la segunda edición revisada en 2000, y actualmente contamos con la cuarta edición, de 29 de septiembre de 2006 (<http://www.w3.org/TR/2006/REC-xml-20060816/>).

Como apostillan Codina y Rovira (2006), XML es un estándar, una recomendación en palabras del W3C, que, junto con su norma asociada, XML Schema, permite definir tipos de documentos y los conjuntos de etiquetas necesarias para codificar tales tipos de documentos. Una vez los documentos están marcados o codificados con un conjunto de etiquetas XML es posible procesarlos y explotarlos de forma automática con diversos propósitos, que pueden ir desde la minería de datos hasta la recuperación de información (RI).

Con XML se pueden diseñar lenguajes de marcado muy estructurados y muy explícitos. Entre sus ventajas, recogidas en numerosos estudios (Bosak y Bray, 1999; Rosa y Senso, 1999; Eíto Brun, 2005; etcétera), destacamos, en la vertiente dirigida a integrar la gestión e intercambio de datos que aquí nos interesa, las siguientes: *a*) es un estándar internacional; *b*) es una tecnología abierta, no propietaria, independiente de plataforma y sistema operativo; *c*) se centra en la separación de contenido frente a la presentación; *d*) tiene gran potencia para construir vocabularios aplicables a cualquier tipo de documento; *e*) se orienta al tratamiento, transmisión o intercambio de todo tipo de documentos o datos; *f*) permite reutilizar textos y datos ya existentes en unos documentos para la confección de otros nuevos; *g*) aporta potentes mecanismos para la búsqueda y recuperación de la información;

h) permite cualquier dispositivo de consulta: navegador, PDA, teléfonos móviles, etcétera; i) es flexible para permitir la integración con otros sistemas y con objetos no XML; j) es capaz de dotar de seguridad y fiabilidad al documento.

XML permite generalizar la estructura de un documento asignando al mismo tipo de documentos la misma DTD (Document Type Definition). Por su parte, los esquemas XML ofrecen un mecanismo más potente y expresivo y permiten definir con mayor riqueza y complejidad estructuras internas, así como hacer uso de otros estándares XML acompañantes, como Xpath, XSL, XML Namespaces, Xlink, etcétera. Martínez González et ál. (2001) evidencian el interés intrínseco de estas tecnologías para ofrecer una solución a las complejas relaciones entre documentos jurídicos señalando cómo los estándares asociados a XML, Xlink, Xpointer, Xpath proporcionan las herramientas necesarias para direccionar los fragmentos de los documentos y modelar enlaces tan complejos como se desee. La asociación de Xlink y Xpointer permite asociar semántica a los enlaces, crear enlaces multidireccionales, asociar múltiples recursos en un mismo enlace y crear bases de enlaces independientes de los documentos.

Pues bien, la comunidad relacionada con la información legal intuyó tempranamente la potencialidad de XML para el tratamiento de la información jurídica. Son abundantes los ejemplos de aplicación de esta tecnología. Conviene recordar proyectos emprendidos en Canadá o Estados Unidos, como el Electronic Court Filing Project. Este grupo de trabajo se unió con otro de la Universidad de Utah con el fin de debatir y desarrollar DTD de XML para la edición e intercambio de documentos jurídicos, y con Gabe Wachob, del portal jurídico FindLaw, un grupo considerado el promotor de la organización internacional LegalXML.

En noviembre de 1998 se constituyó LegalXML como una organización sin ánimo de lucro en la que participan tanto entidades públicas como privadas; desde 2002 se encuentra unida a OASIS. La sección OASIS LegalXML (<http://www.legalxml.org>) tiene el objetivo de desarrollar estándares técnicos abiertos, no propietarios, para su aplicación en la documentación e información jurídica mediante el uso de XML y tecnologías relacionadas. Se divide en diversos grupos de trabajo atendiendo a los tipos específicos de documentos jurídicos. Colabora, además, con grupos relacionados: ebXML (Electronic Business XML), e-Government, UBL (Universal Business Language) o DSS-X (Digital Signature Services eXtended).

En Estados Unidos, desde 2001, la House of Representatives comenzó a trabajar con XML y en estos momentos cuenta con una sede web (<http://xml.house.gov/>) cuyo objetivo es proporcionar información relacionada con los desarrollos de XML legislativo en los que participan además instituciones como la Library of Congress o la US Government Printing Office, entre otras.

Desde 2001, la organización LeXML se ratificó como foro encargado en Europa del desarrollo de estándares XML aplicados al ámbito jurídico. La sede española, LeXML-ES (<http://www.uv.es/lexml>), es moderada por María José Vañó Vañó y cuenta con el apoyo del grupo de investigación RED Derecho TICs de la Universidad de Valencia y del proyecto de investigación UV-AE-20060720.

La aplicación de las tecnologías XML a los documentos jurídicos en los distintos países europeos ha originado iniciativas y proyectos de forma continuada. La Unión Europea ha encabezado la apuesta por un marco de interoperabilidad para atender los servicios derivados del e-gobierno desarrollando numerosos programas, como el reciente Semantic Interoperability Strategy / XML Clearinghouse.

Queremos subrayar en este punto una serie de proyectos enmarcados dentro de las políticas generales de información que diversos países europeos han puesto en marcha en los últimos años. En el Reino Unido se lanza en 2000 un plan de modernización de las administraciones públicas que integra sus sistemas de información con el programa e-GIF (e-Government Interoperability Framework), tomando como base tecnológica XML. En la actualidad, el portal UK GovTalk (<http://www.govtalk.gov.uk>) pretende desarrollar políticas y estándares para la e-administración y ponerlas a disposición de los sectores público y privado.

En Francia la apuesta por la modernización de la Administración pública ha dado frutos como Legifrance (<http://www.legifrance.gouv.fr>), un portal que pone a disposición de los ciudadanos toda la legislación y la jurisprudencia emanadas de los diversos órganos competentes y cuya arquitectura está organizada en torno a bases de información jurídicas estructuradas en XML, o ANTALIA, repertorio de recursos semánticos para la administración en línea.

En Dinamarca, Lex Dania es el sistema nacional para la creación y el intercambio de documentación legislativa. Se trata de un sistema sustentado en el modelo OIO-XML (<http://isb.oio.dk/info>) que dio sus primeros pasos en el Ministerio de Ciencia Tecnología e Innovación y fue adoptado por el Ministerio de Justicia y el Parlamento danés.

En Italia, instituciones como el Istituto di Teoria e Tecniche dell'Informazione Giuridica cuentan con una larga trayectoria en el ámbito que nos ocupa. Proyectos en marcha, como LOIS, sobre ontologías para compartir información legal, o el desarrollo del editor XMLeges pueden servir de ejemplo. Además, el trabajo del Centro Nacional de Tecnología de la Información en la Administración Pública (CNIPA) ha contribuido a la puesta en marcha del portal NormeinRete (<http://www.normeinrete.it>) o del proyecto Ex 107, dirigido a lograr el acceso al corpus legislativo italiano. Los estándares XML son respetados igualmente en ABC for the Law, una herramienta de ayuda diseñada para que el ciudadano no inmerso en el mundo jurídico pueda realizar búsquedas legales.

Desde la Universidad de Ámsterdam se creó en 2002 el portal Metalex (<http://www.metalex.eu>), en la actualidad un servidor de documentos y herramientas en formato abierto que sirve de marco genérico y extensible para la codificación de la estructura y contenidos de documentos legales en XML y RDF. Metalex es igualmente el origen del proyecto europeo ESTRELLA: European project for Standardized Transparent Representations in order to Extend Legal Accessibility (<http://www.estrellaproject.org>).

En este periplo es obligado reconocer un punto de encuentro trascendental para difundir proyectos e iniciativas, los Legislative XML Workshops, realizados en el marco de los encuentros de la Foundation for Legal Knowledge Based Systems (Jurix: <http://www.jurix.nl/>).

### **3. Tratamiento semántico de la documentación jurídica**

En 2001 el W3C lanzó el ambicioso y prometedor proyecto de la Web Semántica. Desde entonces XML se halla ligado indisolublemente al futuro de la Web. En palabras de Codina y Rovira (2006), la infraestructura de esta Web Semántica se sustenta en tres elementos: XML, metadatos y ontologías. Apostillan los mismos autores cómo XML es indudablemente el componente de la Web Semántica que mayor repercusión ostenta hasta el momento.

El segundo gran pilar de la Web Semántica son los metadatos. La marca descriptiva en XML es, en sí misma, una forma de metadatos; sin embargo, la semántica que define es solo parcial, y de ahí que desde el Consorcio Web se desarrollase una de las primeras aplicaciones surgidas del metalenguaje XML, el RDF (Resource Description Framework), considerado un metamodelo de metadatos, que permite codificar distintos esquemas de metadatos y además crear otros vocabularios específicos a través del lenguaje para la descripción de vocabularios RDF o RDFS (RDF Schema).

Recoge Vañó Vañó (2005) ejemplos de los primeros prototipos de diccionarios propuestos; cita así a Muller, quien en 2002 propone la creación de un diccionario legal RDF en Europa, iniciativa en la que se pretendía aplicar el modelo de metadatos Dublin Core. Por su parte, en Estados Unidos, MacClure propone un diccionario similar con todos los términos legales que puedan aparecer en la DTD XML Legal.

En la actualidad, el sistema danés Lex Dania, al que ya hemos aludido, implementa metadatos expresados en RDF. Asimismo, cabe destacar la creación del grupo de trabajo denominado *Common use of XML for the production and distribution of official gazettes*, creado en el contexto del European Forum of Official Gazettes con la intención de que la cooperación entre los editores de boletines oficiales europeos conduzca a la creación de un vocabulario que simplifique el desarrollo y uso de los modelos XML específicos de los documentos legislativos. El

grupo ha propuesto un vocabulario común de metadatos, dado que ninguno de los modelos preexistentes se adaptaba a la descripción de documentos jurídicos. La implementación del modelo, que no excluye la posibilidad de incorporar metadatos específicos, entre otras indudables ventajas supone un apoyo decidido al empleo de estructuras XML normalizadas para la búsqueda e intercambio de metadatos en el entorno de la información legal (Nielsen, 2006).

En cuanto a las ontologías, conviene subrayar que su formalización se asienta en el lenguaje propio de la Web Semántica (OWL: Web Ontology Language). OWL utiliza RDF/XML para representar y codificar las ontologías. En el ámbito del derecho este tipo de conceptualizaciones ocupan ya un espacio muy relevante. Como apuntan Vallbé et ál. (2007), el área jurídica se ha mostrado especialmente interesada en el desarrollo de ontologías que propicien la implementación de aplicaciones “más inteligentes”. Los mismos autores subrayan dos diferencias claves en las ontologías jurídicas frente a otras ontologías: *a)* el lenguaje empleado en los textos legales combina lenguaje técnico y lenguaje natural; *b)* la estrategia de construcción de ontologías ha de tener en cuenta el modelo concreto de derecho que se ha seleccionado. Por eso se distingue entre ontologías superiores (*upper-top*) y de dominio específico (ontologías prácticas).

#### **4. Aplicaciones de XML a la documentación jurídica en España: estado del arte**

La decidida apuesta de la Unión Europea por XML ha servido de acicate imprescindible para las distintas administraciones en la apuesta por la estandarización. El plan de acción i2010 o ePractice.eu, Observatorio Europeo de la Administración Electrónica, son un buen ejemplo de estas iniciativas. Sin duda, la Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los servicios públicos supone el espaldarazo definitivo al reconocimiento del derecho de los ciudadanos a relacionarse electrónicamente con las administraciones públicas, así como la obligación de estas a garantizarlo. Las administraciones tendrán la obligación de hacer efectivos estos derechos a partir de 2009.

La iniciativa en nuestro país recae en el Consejo Superior de Administración Electrónica, dentro del Ministerio de Administraciones Públicas, con planes específicos para el desarrollo de la administración electrónica en la Administración española: Plan Moderniza, Plan Avanza, eModel, etcétera. Una de las acciones más relevantes ha sido la puesta en marcha del proyecto Ventanilla Única de la Acción Pista (1997) SCIRES, en virtud del cual se han generado formatos de intercambio XML. Conviene hacer una mínima referencia, por su relación con el ámbito jurídico, a XBRL (eXtensible Business Reporting Language), norma que establece el uso de XML para el intercambio de información financiera. En XBRL España participan, entre otras, entidades tan relevantes como red.es o el Banco de España.

Como hemos apuntado en las líneas precedentes, la comunidad jurídica internacional atisbó muy pronto la trascendencia de XML para el tratamiento de la documentación legal, acuñándose el término *XML legislativo* para referirse estrictamente al tratamiento de documentación jurídica, legislativa o jurisprudencial mediante XML.

En este trabajo nos proponemos dibujar el actual estado del arte de las bases de datos jurídicas en nuestro país y, en concreto, la utilización del metalenguaje XML para codificar los documentos contenidos en estos sistemas en orden a un procesamiento y una explotación eficaz de los mismos. Sin embargo, hablaremos de XML legislativo en sentido lato, sumando a los desarrollos estrictos que constituyen el eje de interés ciertas iniciativas, experiencias o proyectos relacionados que complementan el retrato de las bases de datos legales.

Para este objetivo hemos partido del seguimiento realizado por Vicente y Martínez (2007), único análisis sistemático sobre la utilización de XML legislativo en España. Tomando este punto de inicio, hemos tratado de ampliar el estudio lanzado una solicitud de información sobre los desarrollos tecnológicos y herramientas terminológicas implementados que se han enviado a empresas distribuidoras de bases de datos jurídicas y a portales institucionales con bases de datos legales de ámbito estatal y autonómico. Se ha complementado igualmente con una revisión bibliográfica de la literatura científica y un seguimiento de sedes web de empresas e institucionales relacionadas bien con el ámbito jurídico, bien con el investigador.

La primera dificultad evidenciada para llevar a cabo el análisis procede del propio dominio interdisciplinar desde el que se aborda nuestro objeto de trabajo, un contexto en el que confluyen, al menos, tres áreas de conocimiento: derecho, informática y documentación. A esto hemos de unir, con la excepción lógica de los desarrollos efectuados desde el ámbito investigador, la escasez de publicaciones sobre las aplicaciones de empresas e instituciones en el terreno que nos ocupa.

En segundo lugar, la tasa de respuesta a nuestra solicitud de información ha sido ciertamente limitada, tanto en el ámbito público como en el privado; al silencio, en numerosas ocasiones, se ha unido, en el caso de varias empresas, la comunicación expresa de la negativa a facilitar información que se considera comercial.

Los resultados, con las limitaciones subrayadas, han sido sistematizados en tres bloques: bases de datos comerciales, portales institucionales (sistemas parlamentarios, diarios oficiales, plataformas de e-gobierno y ámbito judicial) y proyectos de investigación, en este caso con dos subgrupos, XML legislativo, y ontologías legales y aplicaciones semánticas.

#### **4.1. Bases de datos comerciales**

El *Boletín Oficial del Estado* comienza a trabajar con XML en 2002, empleando desarrollos propios con librerías estándar en Linux. Utiliza listas de palabras

clave y está trabajando en el desarrollo de tesauros. Aunque desde el 1 de enero de 2004 las bases de datos gestionadas por el BOE son de acceso libre y gratuito, mantiene las tradicionales suscripciones de pago en los formatos en CD-ROM / DVD. Recientemente ha comunicado que a partir de 2009 solo será consultable en formato electrónico.

Derecho.com, portal jurídico aparecido en 1999, hace uso de XML desde 2003 en virtud de los desarrollos llevados a cabo por la empresa Cometatech. Según datos de la empresa, el sistema corre sobre Linux, emplea bases de datos MySQL y J2EE como lenguaje de programación, contando además con un servidor web Apache y Tomcat; el marco de la aplicación web es Cocoon, basado en el uso de XML XSLT y Java.

La editorial Lex Nova cuenta con una DTD XML propia y utiliza para la publicación plantillas XSLT. Asimismo, hace uso de metadatos. El sistema ha sido desarrollado en colaboración con el grupo de investigación en Recuperación de Información y Bibliotecas Digitales de la Facultad de Informática de la Universidad de Valladolid.

Tirant on Line salió al mercado en junio de 2002 con acceso exclusivamente vía web. La responsable de la puesta en marcha del sistema fue la empresa ISOCO. El desarrollo se basa en Java y hojas de transformación XSL, y la tecnología empleada se basa en la combinación de técnicas de inteligencia artificial para tratamiento de textos, técnicas avanzadas de búsqueda, seguridad en las conexiones, autenticación de usuarios y XML. El formato XML se usa para introducir y clasificar de forma automática los documentos en el sistema, permitiendo una conveniente separación entre la capa de datos, la lógica de la aplicación y el sistema de visualización. Además, emplean vocabularios controlados: listas de palabras clave, tesauros, metadatos y ontologías.

El grupo Wolters Kluwer (La Ley, Ciss) comienza a hacer uso de XML en 1999, empleando ORACLE, MSXML 6.0 y desarrollos propios. Hace uso de lenguaje controlado (palabras clave, tesauros, metadatos) y ontologías.

## 4.2. Portales institucionales

### 4.2.1. Sistemas parlamentarios

El Senado ha implementado con gran acierto el tesauro Eurovoc para efectuar búsquedas “semánticas” en las ocho bases de datos que gestiona. Ha empleado la edición 4 del tesauro Eurovoc, un desarrollo propio del Senado y la *Relación de municipios y códigos por provincias* publicada por el Instituto Nacional de Estadística. Creemos que la utilización de un instrumento de control terminológico como Eurovoc permite incorporar una red semántica inteligente en la que los términos mantienen relaciones asociativas, jerárquicas y de sinonimia y que mejora

sustancialmente el rendimiento en la RI. XML se utiliza, por el momento, en el archivo de imágenes y sonidos.

Las Cortes de Aragón hacen uso de la aplicación Lotus Notes, que genera automáticamente ficheros XML; cuentan, además, con listas de autoridades. El portal fue creado en 2006 y se está trabajando en un nuevo desarrollo en la actualidad.

Por su parte, la Junta General del Principado de Asturias puso en marcha el nuevo portal en 2006, empleando XML con desarrollos propios. El sistema de gestión de contenidos utilizado ha sido desarrollado por Intermark Tecnologías y corre sobre Apache Tomcat y base de datos MySQL. Emplea metadatos, listas de palabras clave y tesauros.

El Parlamento de Baleares emplea desde 2001 XML y desarrollos propios. Utilizan la edición 3 de Eurovoc adaptada por el Senado y traducida al catalán por la biblioteca del propio Parlamento.

El portal del Parlamento de Cantabria data de finales de 2006 y emplea un sistema de gestión de contenidos desarrollado por la empresa EUREKA. El *software* utilizado es Nhibernate, y utiliza XML en algunos puntos del sistema. Incorpora metadatos y listas de palabras clave.

Las Cortes de Castilla y León han comenzado a trabajar con XML en el año 2000. Utilizan XML de forma nativa. El último avance ha sido el tratamiento de fragmentos de vídeo de las sesiones parlamentarias en XML.

El Parlamento del País Vasco hace uso de XML desde 2003 y emplea *software* a medida. Asimismo, utiliza Eurovoc para el control del vocabulario.

#### 4.2.2. *Diarios oficiales*

El Principado de Asturias ha puesto en marcha el Sistema de Información Documental en Red de Asturias (SIDRA) como sistema corporativo de gestión documental y de información en red. Comenzó a utilizarse XML en el año 2000. La aplicación de gestión documental elegida es DOGMA (Document Gallery Management), desarrollada por la empresa Software AG. Se emplea XML de forma nativa. DOGMA está basado en Tamino y el marco de la aplicación es del Principado de Asturias. En concreto, la aplicación que gestiona el *Boletín Oficial del Principado de Asturias* (e-BOPA) es la combinación de PUBLIBOPA + SIGBOPA. Además de la versión web del BOPA, los índices semestrales se publican en DVD. Emplea lenguajes controlados: palabras clave, tesauros y metadatos.

El *Boletín Oficial de la Región de Murcia* se puso en marcha en formato web en 1998. Desde 2006 es un servicio prestado por Aranzadi Westlaw. Emplea XML y tecnología J2EE. Además de la versión web, edita un CD-ROM anual. Hace uso de lenguajes controlados: palabras clave, tesauros y metadatos.

#### 4.2.3. Plataformas de e-gobierno

El portal de la Administración General del Estado 060.es, desarrollado por la empresa XimetriX Network Thoughts y presentado en 2006, cuenta con un sistema de gestión de contenidos, ximDEX, que emplea XML. Permite la completa separación entre contenidos y presentación, la publicación en distintos canales y la explotación y uso automático de la información para tareas diversas: sindicación automática, resúmenes automáticos de documentos, publicación web, etcétera. Incluye una base de datos legislativa de potencialidad limitada.

El Gobierno Vasco ha utilizado BEA WebLogic Server 8.1 como servidor de aplicaciones en un modelo de administración electrónica y Bea WebLogic Integration de BEA Systems. Soporta estándares XML y Java, y emplea Linux y una base de datos ORACLE.

El portal de la Generalitat de Cataluña desde 2005 ha integrado la filosofía Web 2.0. Tecnológicamente, el intercambio y la manipulación de datos se efectúan mediante XML y XSLT, y la recuperación de datos asíncrona, por medio de XMLHttpRequest. Se complementa con JavaScript para su articulación.

Nos constan los esfuerzos de la Junta de Andalucía, que en 2002 inauguró el proyecto w@nda, donde se concentran todos los esfuerzos por acercarse electrónicamente al ciudadano. Por su parte, el Gobierno Valenciano, continuando la labor iniciada en PEMAV (1996-1999) y Moderniza (2000-2003), impulsa dentro de Avantic dos nuevos planes estratégicos para avanzar de manera firme hacia la nueva sociedad tecnológica y del conocimiento, en los aspectos específicos de telecomunicaciones (PEVTA) y en las tecnologías de la información y sociedad del conocimiento (PETIC).

Dentro de la Administración local, ponemos de relieve la actuación del Ayuntamiento de Zaragoza, institución que ha implantado un mecanismo de búsquedas basado en Web Semántica (Fernández, Angós y Salvador, 2007). La ontología distingue agentes, eventos, objetos y procesos, además de las relaciones entre ellos; ha sido desarrollada por ISOCO y se ha realizado en RDFS, utilizando como editor de ontologías Protégé. Cubre cinco subáreas de los trámites municipales: urbanismo, salud pública, asuntos personales, medio ambiente y tributos municipales.

#### 4.2.4. *Ámbito judicial*

Conviene subrayar la preocupación temprana de la Unión Europea por potenciar y dinamizar la distribución de la información jurídica en general y de la jurisprudencial en particular, tanto para los profesionales jurídicos como para el resto del público. Sirven de ejemplo la Recomendación 1983/3, relativa a la protección de los usuarios de los servicios de informática jurídica, o la 1995/11, sobre la selección, tratamiento, presentación y archivo de las resoluciones judiciales en los sistemas de documentación jurídica automatizados. En el año 2001 apareció la

Recomendación 2001/3 del Comité de Ministros a los Estados miembros sobre los servicios de tribunales y otras instituciones jurídicas ofrecidos a los ciudadanos por las nuevas tecnologías. En ella se configura e impulsa la distribución de la información jurídica en general y la jurisprudencial en especial. Por fin, la Directiva 2003/98/CE del Parlamento Europeo y del Consejo de 17 de noviembre de 2003, relativa a la reutilización de la información del sector público, ha afianzado los principios establecidos en las recomendaciones precedentes.

En España, el Consejo General del Poder Judicial decidió en 1997 la creación del Centro de Documentación Judicial (CENDOJ), asumiendo, entre otros desafíos, dos fundamentales: superar la publicación de la jurisprudencia impresa como se venía haciendo desde el siglo XIX y tomar las riendas con respecto al acceso a las resoluciones judiciales en manos de editoriales comerciales, sin garantía de acceso a esta información por parte del ciudadano.

En la actualidad, la base de datos de jurisprudencia del Tribunal Supremo se publica íntegramente en la Web de modo gratuito e incluyendo el correspondiente análisis jurídico. Jueces y magistrados cuentan con un acceso restringido al sistema. Las editoriales jurídicas privadas acceden igualmente al sistema abonando un precio por sentencia si desean la comercialización de la misma; además se ha llegado a acuerdos con organismos públicos y colegios de abogados.

El CENDOJ emplea LIVELINK Collection Server —antes BASIS— (<http://www.opentext.com>), gestor de grandes colecciones documentales de alta calidad en la búsqueda y recuperación de información, un sistema completo para el tratamiento de información bibliográfica, textual, estructurada en SGML, XML, HTML, etcétera. Diseñado para almacenar y gestionar muy diversos tipos documentales, el sistema cuenta con un potente motor de recuperación, implementa un módulo de tesoro y ofrece muchas posibilidades de personalización.

### **4.3. Proyectos de investigación**

#### *4.3.1. XML legislativo*

Afortunadamente, en nuestro país contamos con investigaciones relevantes sobre la aplicación de XML legislativo a la documentación jurídica, como se puede constatar en la sinopsis siguiente.

En la Universidad de Valladolid, las investigaciones referidas al campo que nos ocupa se inician en 1996, con dos proyectos de investigación sobre el acceso a servidores de información legislativa: “Estudio, diseño y puesta en marcha de un servidor público de información legislativa sobre red Internet CICYT (TEL96-1296)” y “Estudio y realización de un servidor público de información legislativa de Castilla y León sobre red Internet (VA10196)”. Esta línea de investigación continuó en colaboración con el Institut National de Recherche en Informatique et

en Automatique (INRIA), de Francia, y culminó en 2001 con la defensa de la tesis doctoral de M.<sup>a</sup> Mercedes Martínez González, *Principios para la explotación dinámica de relaciones entre documentos en las bibliotecas digitales: aplicación al entorno jurídico*, dirigida por los doctores Pablo de la Fuente y Jean-Claude Derniame, del INRIA.

Dos grupos de investigación convergen en esta institución, desde el Departamento de Informática, en concreto el grupo de investigación en Recuperación de Información y Bibliotecas Digitales al que nos hemos referido, y desde el Área de Derecho Internacional Privado, Dámaso-Javier Vicente Blanco y José Manuel Sánchez Felipe, de la Facultad de Derecho. Son numerosos los trabajos publicados por este grupo interdisciplinar (Martínez et ál., 2001; Martínez, Fuente y Derniame, 2003a; Martínez, Fuente y Derniame, 2003b; Martínez, Fuente y Vicente, 2004; Martínez et ál., 2007; Martínez y Fuente, 2007, etcétera). Queremos destacar uno de los primeros estudios, el de Martínez et ál. (2001), referido al uso de XML y estándares asociados Xlink, Xpointer y Xpath en la documentación jurídica; en él se propone explotar las relaciones entre textos legislativos de forma que se puedan realizar consultas sobre estas y se presenta la posibilidad de generar automáticamente documentos de la versión definitiva de la disposición legal.

En la actualidad, los dos grupos de investigación señalados llevan a cabo el proyecto de investigación VA010B06, financiado por la Junta de Castilla y León y el Colegio de Procuradores de Valladolid: “Creación de una aplicación de Web Semántica para el tratamiento y manipulación electrónicos de los textos jurídicos de la Unión Europea y el Espacio Económico Europeo (en su versión en español) en materia de conflicto internacional de jurisdicciones”.

El grupo DELi, de la Universidad de Deusto, comenzó su actividad en 1997 por iniciativa de profesores de las facultades de Letras y la Facultad de Ingeniería (ESIDE), con intereses comunes en las áreas de edición digital e ingeniería lingüística. Componen el equipo los lingüistas Joseba Abaitua, Patricia Fernández Carrelo y Koldo Garai Bilbao; por su parte, de ESIDE proceden Unai Aguilera, David Buján, Josuka Díaz Labrador, Inés Jacob y Txus Sánchez. Destacamos el proyecto XML-Bi (Abaitua et al., 2001), en el que, durante los años 2001 y 2002, se desarrolló el LEGE-Bi, un corpus bilingüe XML aplicado al *Boletín Oficial del País Vasco*.

A partir de 2002 el interés se ha concitado alrededor de las tecnologías presemánticas y semánticas, y en 2005 se crea la empresa EmergiaTech, que produce y comercializa productos, soluciones y servicios relacionados con las tecnologías semánticas: herramientas XML para la gestión empresarial, taxonomías, ontologías o buscadores semánticos como KnowBoard. Una recopilación de las publicaciones de este grupo de investigación está accesible en <http://www.deli.deusto.es/AboutUs/Publications>.

En el Departamento de Biblioteconomía y Documentación de la Universidad Carlos III de Madrid se pusieron en marcha en el año 2000 dos proyectos dirigidos por los profesores Mercedes Caridad y Tomás Nogales que atendían a la normativa del Mercosur y a las Disposiciones Generales publicadas en el *Boletín Oficial de la Comunidad de Madrid*, respectivamente.

En febrero de 2002 defiende uno de los miembros del grupo de investigación, Bonifacio Martín Galán, su tesis doctoral, *Tratamiento y difusión en Internet de información jurisprudencial mediante tecnologías XML: aplicación al caso del Tribunal Constitucional*, en la que se aborda un buen número de estándares integra-dos dentro de la familia de tecnologías XML.

En abril de 2003, Carmen Arellano Pardo, otro de los integrantes del grupo, defiende su tesis, *La organización hipertextual del ordenamiento jurídico por medio de tecnologías XML: aplicación a la normativa del IRPF*, en la que se desarrolla una maqueta que demuestra la viabilidad y utilidad de Xlink para establecer y hacer operativas las complejas relaciones que pueden darse entre diversas normas del ordenamiento jurídico.

Además de numerosas publicaciones (Nogales et ál., 2000; Martín, Nogales y Arellano, 2002; Nogales, Martín y Arellano, 2003; Nogales, Arellano y Martín, 2004, etcétera), este grupo de investigación, en colaboración con el Departamento de Ingeniería Telemática y la empresa PSd, efectuó un estudio de asesoría para el Centro de Documentación Judicial del Consejo General del Poder Judicial (CENDOJ) sobre la viabilidad de las tecnologías XML para la creación de un sistema de gestión, publicación y distribución de jurisprudencia.

Desde el Departamento de Lenguajes y Sistemas Informáticos de la Universidad del País Vasco, Sara Sanz, Silvia Sanz, Mikel Villamañe, Julián Gutiérrez Serrano, José Ángel Vadillo Zorita y Tomás A. Pérez son los responsables del proyecto de creación de BD-IDHV, un repositorio de derecho histórico vasco, ejemplo se-ñero en España de aplicación de tecnologías XML a documentación legal históri-ca (Sanz et ál., 2003).

En la Universidad de Murcia, Javier de Andrés Rivero y Antonio F. Gómez Skarmeta, del Departamento de Ingeniería de la Información y las Comunicacio-nes, en el marco del grupo liderado por Rafael Hernández Marín, han desarrolla-do el sistema CRNOLEX para la gestión el ciclo completo de vida de las dis-posición normativas, desde su etapa como borrador hasta su puesta a disposición vía web (Hernández, López e Iniesta, 2005; Andrés y Gómez, 2007). Para ello se ha elaborado una DTD XML capaz de representar todos los tipos de documentos normativos existentes en nuestro país. El sistema emplea una base de datos XML nativa y se ha articulado en varios módulos (legislación consolidada, módulo de búsqueda, módulo de edición, módulo web y módulo de base de datos). El proyecto

trabaja intensamente en la introducción de metadatos estructurados en XML para resolver las interconexiones normativas y mejorar la recuperación semántica.

#### 4.3.2. Ontologías jurídicas y aplicaciones semánticas

Recogemos en este punto, atendiendo a las cuestiones metodológicas previamente especificadas, aquellos desarrollos que guardan relación con tecnologías XML o los desarrollos semánticos dirigidos a la información jurídica pero que no se pueden considerar estrictamente investigaciones en XML legislativo.

En nuestra opinión, esta línea de trabajo resulta heredera de los estudios sobre aplicaciones derivadas de la inteligencia artificial (IA) destinadas al ámbito jurídico, que comienzan en los años ochenta del pasado siglo. El entorno jurídico ha sido objeto de interés de la inteligencia artificial, con un consolidado dominio de investigación dedicado a la IA y el derecho y relevantes resultados derivados de aplicaciones de la ingeniería del conocimiento. Curran y Higgins (2000) señalan que las técnicas de inteligencia artificial suponen el reto más ambicioso emprendido hasta el momento para mejorar el proceso de búsqueda legal, y creen que este tipo de aplicaciones recondujeron sus objetivos desde la ambición inicial de actuar como expertos legales hacia la idea más modesta de constituir sistemas de apoyo a la toma de decisiones legales.

En España, Galindo y Lasala (1995) proponen la utilización de tecnología de IA para el desarrollo de lo que denominan *Sistemas Inteligentes de Recuperación de Documentación Jurídica* (SIREDOJ), en cuyo planteamiento incluyen comprensión del lenguaje natural e integración de distintos sistemas de bases de datos con sistemas expertos. En este entorno presentan el prototipo ARPO-2, diseñado para utilizar argumentos legales relativos a incumplimiento de contratos de obra.

Hemos de mencionar el papel fundamental de instituciones como la International Association for Artificial Intelligence and Law (IAAIL) y la Foundation for Legal Knowledge Based Systems (Jurix), así como las sucesivas conferencias auspiciadas por dichas entidades, la International Conference on Artificial Intelligence and Law y la Jurix: International Conference on Legal Knowledge-Based Systems, sin olvidar las revistas más específicas en este ámbito, *Artificial Intelligence and Law* y *Journal of Information Law and Technology* (JILT).

Desde finales de los noventa, y particularmente dentro del Programa Marco de la Unión Europea, se ha venido desarrollando un creciente interés por parte de la comunidad dedicada a la ingeniería del conocimiento hacia el desarrollo e implementación de ontologías jurídicas que permitan una recuperación semántica de la información legal. En concreto, bajo los auspicios del VI Programa Marco se han puesto en marcha importantes proyectos de investigación dedicados a la Web Semántica. La tendencia inicial se dirigió hacia la construcción de ontologías de abstracción máxima (Vallbé et ál., 2007; Casanovas et ál., 2005); sin embargo, en los

últimos años el rumbo parece desviarse hacia la formalización de ontologías en ámbitos específicos del conocimiento jurídico.

De modo sucinto, recogemos en las líneas siguientes algunos de los proyectos y grupos de investigación más relevantes.

ISOCO Lab (<http://www.isoco.com>) nació en 1999 en el Instituto de Investigación en Inteligencia Artificial (IIIA-CSIC) del Consejo Superior de Investigaciones Científicas, en el campus de la Universidad Autónoma de Barcelona. Es una empresa y centro de I + D de gran prestigio internacional en el área de la sociedad de la información, y en particular en las tecnologías de Web Semántica, servicios web, grid semántica y adquisición de conocimiento. Ha participado en numerosos proyectos de investigación y en iniciativas comerciales.

En el proyecto Reimdoc: Recuperación e Integración Multilingüe de Documentos Jurídicos (proyecto PROFIT-340100) 2004-2005 participaron Atos Origin, la Universidad Politécnica de Madrid, Derecho.com y la Universidad de Zaragoza. Cofinanciado por el Ministerio de Industria, Turismo y Comercio, dentro del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica, tenía por objeto estudiar las características de la tramitación de los procedimientos jurídicos tomando como referencia la seguida por la escritura pública de compraventa de bienes inmuebles recogida en soporte digital en los procedimientos que se realizan en tres ámbitos: los registros de la propiedad, la administración tributaria de las comunidades autónomas (impuesto de transmisiones patrimoniales y actos jurídicos documentados) y la administración de justicia.

El Institut de Dret de la Universidad Autónoma de Barcelona se ha consolidado como un centro interdisciplinar con líneas de investigación centradas en el derecho y la inteligencia artificial, la Web Semántica o las ontologías jurídicas, entre otras.

Citamos algunos proyectos, como NetCase (2003), financiado por el Ministerio de Ciencia y Tecnología, en el que en colaboración con ISOCO se desarrolla un sistema de *software* inteligente para la gestión de servicios jurídicos en redes transnacionales.

Dentro del marco de diversos proyectos PROFIT estatales y del proyecto europeo SEKT —Semantically Enabled Knowledge Technologies— destaca la implementación de Juriservice (Vallbé et ál., 2007), una plataforma FAQ inteligente de soporte para la Escuela Judicial Española (2006-2007), ideada como herramienta de asistencia a nuevos jueces. Se ha desarrollado con la colaboración de ISOCO un *software* capaz de resolver dudas relacionadas con la práctica judicial proporcionando respuestas justificadas y uniformes. Para ello se ha interrelacionado la OPJK —Ontology of Professional Judicial Knowledge (Ontología del Conocimiento Judicial)— con un sistema FAQ inteligente (Juriservice) que contiene unas mil preguntas frecuentes y las correspondientes respuestas legales.

Existe un proyecto I + D de un metabuscador semántico legal, *clustering* de resultados y autoclasificación de fuentes jurídicas financiado dentro del programa PROFIT para el bienio 2007-2008. Se trata de un proyecto consorciado de Wolters Kluwer España (coordinador), Institut de Dret i Tecnologia (IDT-UAB) e Intelligent Software Components (ISOCO).

DALOS (Drafting Legislation with Ontology-based Support) es un proyecto de la Comisión Europea (ICT for Citizens and Businesses eGovernment) coordinado por el Institute of Legal Information Theory and Techniques del Italian National Research Council y en el que participa el Institut de Dret i Tecnologia. El objetivo es crear herramientas lingüísticas y de gestión del conocimiento que puedan ser utilizadas en los procesos de redacción de documentos legales, contribuyendo a la armonización y coherencia de los textos legislativos. Además, las ontologías facilitarán a posteriori la recuperación de la información legal por parte de los ciudadanos.

En cuanto al grupo de ingeniería ontológica Ontology Engineering Group (OEG) del Laboratorio de Inteligencia Artificial de la Facultad de Informática de la Universidad Politécnica de Madrid, entre los numerosos proyectos sobre Web Semántica llevados a cabo por el equipo de investigación dirigido por Asunción Gómez Pérez, cuenta con una línea de trabajo dedicada al e-gobierno. Ha trabajado en la adaptación de la taxonomía de Breuker para construir una ontología adaptada al ámbito legal español. Para la construcción se emplea METHONTOLOGY como metodología y el *software* WebODE.

Trabajan en la actualidad en la ontología EGODO (Egov Documentation Ontology), ideada para actuar interrelacionada con aplicaciones semánticas que apoyen el e-gobierno: EgoIR para la recuperación de información legal y Egoster para ayudar a compartir e intercambiar documentos. En una comunicación reciente sobre este proyecto, Ortiz-Rodríguez (2007) subraya la complejidad del conocimiento legal frente a otros dominios como pueden ser el de la física o las matemáticas y propone la evaluación comparativa del modelo con iniciativas como LegalXML, así como la necesidad de trabajar en las capacidades de razonamiento de estas ontologías.

El equipo de Web Semántica de la Fundación CTIC, en colaboración con profesorado de la Universidad de Oviedo, ha desarrollado el buscador Esperteyu (Fundación CTIC, 2007), un prototipo que se enmarca dentro de las políticas generales de sociedad de información del Principado de Asturias para acercar la Administración pública a los ciudadanos. En particular, este buscador intenta mejorar y facilitar el acceso de la ciudadanía al *Boletín Oficial del Principado de Asturias* (BOPA). El proyecto aplica distintas tecnologías semánticas, en particular ontologías y tesauros. La construcción del buscador se ha realizado buscando la máxima interoperabilidad con sistemas existentes, utilizando los estándares del W3C: XML, XHTML, CSS, OWL-DL o SKOS-CORE.

## 5. Conclusiones

Creemos que la tecnología ha desdibujado los límites entre las bases de datos entendidas como corpus jurídicos de documentos estructurados, analizados y dotados de índices adecuados para su acceso y consulta para interrelacionarse con aplicaciones y herramientas dirigidas a la gestión y a la toma de decisiones. Podemos hablar en estos primeros años del siglo XXI de una superación de la idea de repertorio de legislación o jurisprudencia heredado del XIX, trasladado con éxito al formato electrónico y al que hoy se le exigen cambios cualitativos. Se precisan soluciones integrales para la gestión del conocimiento jurídico.

El impacto de la Web Semántica en los ámbitos jurídico y de e-gobierno se prevé muy significativo; sin embargo, en este momento tenemos que hablar de una fase meramente experimental. Un caso distinto es el de XML: su penetración es indiscutible, lo que ha de traducirse, como señalan Codina y Rovira (2006), en el triunfo de un modelo de datos no propietario, independiente de plataforma y que garantiza la compatibilidad con futuros sistemas.

El ámbito de la documentación jurídica constituye un dominio extraordinario para la implementación de tecnologías XML y aplicaciones semánticas. Convenimos en la consideración categórica de Martínez, Fuente y Dermame (2003a) de que la adopción de XML para representar documentos jurídicos es tan general que se puede afirmar, sin temor a equivocarse, que es el lenguaje del e-gobierno. Los documentos legales, por su carácter fuertemente estructurado y por las complejas conexiones e interrelaciones que requieren expresar, se adaptan de modo inmejorable a esta tecnología.

Asimismo, coincidimos con la apreciación de Vañó (2005): en España, las pequeñas comunidades encargadas de desarrollar distintas áreas y documentos han dado paso a iniciativas institucionales públicas y semipúblicas. Estamos empezando a traspasar la línea que nos confirma que XML ha pasado de tecnología emergente a omnipresente.

En la evolución de las bases de datos jurídicas españolas en este nuevo siglo encontramos dos puntos de inflexión trascendentales: el primero se corresponde con la iniciativa decidida de las instituciones públicas de poner el corpus legal a disposición de los ciudadanos (destacamos en este punto la actividad del CENDOJ y del BOE a nivel estatal); en el segundo creemos que se han de situar las emergentes plataformas de e-gobierno, a las que se exige estandarización y una adecuada integración con la información legal. Es indudable que el acceso a la información jurídica, por su trascendencia directa sobre la ciudadanía, ha de constituir una preocupación prioritaria para los distintos organismos públicos y se erige como un corpus estratégico dentro de los desarrollos del e-gobierno.

En el terreno de las bases de datos comerciales, la tendencia parece dirigirse al empleo de tecnologías semánticas, toda vez que la propia industria del *software* apuesta por la gestión del conocimiento y la ingeniería documental. Las búsquedas semánticas permitirán la gestión y la recuperación de conocimiento y la posibilidad de incorporar medios no solo textuales. Como señalan Benjamins et ál. (2005), la recuperación de información legal va más allá de la recuperación de documentos; esto es, la respuesta a una pregunta dada suele requerir algún tipo de deducción o inferencia.

Los sistemas comerciales han de demostrar más que nunca sus valores añadidos y parece que es el momento de acercar las bases de datos hacia sistemas expertos en la resolución de tareas legales.

Los portales legislativos institucionales han pasado progresivamente de meros repositorios de documentos normativos a sistemas de RI, con mejores prestaciones, gracias a un mayor cuidado en el tratamiento documental y jurídico, y se encuentran en una etapa de transición en la que creemos se verán beneficiados por las tecnologías semánticas que se imponen en la administración electrónica. Como ya se observa en algunas de las implementaciones, las tecnologías XML supondrán un salto cualitativo en la gestión y en la eficacia en la recuperación de información. Refrendamos en este punto las consideraciones de Biasiotti y Nannucci (2006), para quienes las instituciones han llevado a cabo en estos últimos años un esfuerzo ingente por digitalizar contenidos, un esfuerzo que ahora debe complementarse creando desde la Administración servicios y herramientas que transformen la información en conocimiento, y en el corazón de estos servicios se han de situar indudablemente las bases de datos jurídicas.

En nuestro país, salvo alguna excepción, los resultados muestran que las distintas Administraciones públicas no se han involucrado, al contrario de lo sucedido en países como Dinamarca o Italia, en una política institucional orientada al tratamiento y gestión de la información mediante estándares XML, lo que sin duda retrasará la efectividad de las herramientas de e-gobierno que demanda la sociedad de la información.

Por último, existen grupos de investigación enormemente dinámicos en nuestro país que vienen trabajando tanto en el ámbito nacional como en el internacional desde finales de la década de los noventa, en algunos casos en proyectos compartidos con empresas pioneras en las aplicaciones de XML a la información jurídica o en proyectos para distintas instituciones públicas o semipúblicas que han servido de pilares fundamentales de los desarrollos con los que empezamos a contar en la actualidad. Resulta palmario que ha llegado el momento de compartir experiencias y aunar esfuerzos.

## Referencias

- Abaitua, J., et ál. (2001). XML-Bi: procedimientos para la gestión de flujo documental multi-lingüe sobre TEI/XML. // Procesamiento del lenguaje natural. XVII Congreso de la SEPLN. 27 (2001) 291-292. <http://www.sepln.org/revistaSEPLN/revista/27/27-proyecto2.pdf> (2008-01-12).
- Álvaro Bermejo, C.; Sanz de Ormazábal, I.; Cueto Aparicio, M. (1998). El Centro de Documentación Judicial (CENDOJ) del Consejo General del Poder Judicial. // VI Jornadas Españolas de Documentación. Valencia: FESABID, 1998. 73-84.
- Alvite Díez, M.ª L. (2003). Tendencias en la investigación sobre recuperación de información jurídica. // Revista Española de Documentación Científica. 26:2 (2003) 191-212.
- Alvite Díez, M.ª L. (2004). Evolución de las bases de datos jurídicas en España. // Anales de Documentación. 7 (2004) 7-27.
- Andrés Rivero, J. de; Gómez Skarmeta, A. F. (2007). CRONOLEX: a system for a dynamic representation of laws. // Biagioli, C.; Francesconi, E.; Sartor, G. (eds.). Proceedings of the V Legislative XML Workshop. Florencia: European Press Academic Publishing, 2007. 99-106.
- Arellano Pardo, M.ª C.; Nogales Flores, J. T.; Martín Galán, B. (2003). La organización hipertextual del ordenamiento jurídico. Posibilidades de XML y estándares relacionados. // Revista General de Información y Documentación. 13:2 (2003) 181-191.
- Arellano Pardo, M.ª C. (2005). Aportaciones de la técnica legislativa y XML a la informática jurídica documental. Madrid: Universidad Carlos III, Boletín Oficial del Estado, 2005.
- Blaquier Ascaño, M.ª (1996). Aplicaciones del SGML a la información jurídica. // Informática y Derecho: Revista Iberoamericana de Derecho Informático. 12-15 (1996) 1481-1504.
- Benjamins, V. R., et ál. (eds.) (2005). Law and the Semantic Web: legal ontologies, methodologies, legal information retrieval, and applications. Berlín: Springer, 2005.
- Benjamins, V. R., et ál. (2005). Law and the Semantic Web: an introduction. // Benjamins, V. R., et ál. (eds.). Law and the Semantic Web: legal ontologies, methodologies, legal information retrieval, and applications. Berlín: Springer, 2005. 1-14.
- Biasiotti, M. A.; Nannucci, R. (2006). Converting online public legal information into knowledge: "ABC del Diritto", an Italian e-government citizen-oriented service. // Proceedings of the 2006 International Conference on Digital Government Research. ACM International Conference Proceeding Series. 151 (2006) 62-66.
- Bosak, J.; Bray, T. (1999). XML and the second-generation Web. // Scientific American. (May 1999). <http://www.sciam.com/article.cfm?articleID=0008C786-91DB-1CD6-B4A8809EC588EEDF> (2008-01-14).
- Casanovas, P., et ál. (2005). Supporting newly-appointed judges: a legal knowledge management case study. // Journal of Knowledge Management. 9:5 (2005) 7-27.
- Castells, M. (2000). La era de la información, vol. 1: La sociedad red: economía, sociedad y cultura. Madrid: Alianza, 2000, 2.ª ed.
- Codina, L.; Rovira, C. (2006). La Web Semántica. // Tramullas Saz, J. (coord.). Tendencias en documentación digital. Gijón: Trea. 9-54.

- Curran, K.; Higgins, L. (2000). A legal retrieval information system. // *Journal of Information, Law and Technology (JILT)*. 3 (2000). [http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/2000\\_3/curran](http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/2000_3/curran) (2008-01-15).
- Éito Brun, R. (2005). XML y la gestión de contenidos. // *Hipertext.net*. 3 (2005). <http://www.hipertext.net/web/pag256.htm> (2008-01-12).
- Fernández Ruiz, M. J.; Angós Ullate, J. M.; Salvador Oliván, J. A. (2007). Metodología de trabajo en un sitio web de la Administración pública: el Ayuntamiento de Zaragoza. // VIII Congreso ISKO-España. León: Universidad, 2007. 511-522.
- Fundación CTIC (2007). Búsqueda semántica en bases documentales gubernamentales. <http://www.uniovi.es/teso/pdfs/technical-report.pdf> (2008-01-25).
- Galindo Ayuda, F.; Lasala Calleja, P. (1995). Metodología para el desarrollo de sistemas jurídicos de inteligencia artificial: el prototipo ARPO-2 como ejemplo. // *Scire: Representación y Organización del Conocimiento*. 1:2 (1995) 73-103.
- Gómez Skarmeta, A. F. (2000). Una arquitectura de metadatos para la gestión de información en el web. // *Novática*. 146 (2000) 9-11.
- Hernández Marín, R.; López Hernández, J.; Iniesta Delgado J. J. (2005). CRONOLEX: a computing representation of law dynamics. // XXII World Congress of Philosophy of Law and Social Philosophy. Granada, 2005. <http://www.lefis.org/meetings/workshops/2005/granada/docs/cronolex.pdf> (2008-02-04).
- Marius Monton, J., et ál. (2007). Accelerating semantic search with application of specific platforms. // *Proceedings of the Workshop on Semantic Web Technology for Law at International Conference on AI and Law ICAIL 2007*. Stanford, 2007. 47-50. <http://www.cs.vu.nl/~mcaklein/SW4Law/proceedings.pdf> (2008-01-20).
- Martín Galán, B.; Nogales Flores, J. T.; Arellano Pardo, M.<sup>a</sup> C. (2002). Modelos formales para la definición estructural y semántica en documentos XML: comparación de posibilidades en un corpus textual de documentación jurisprudencial. // III Jornadas de Bibliotecas Digitales (JBIDI' 2002). Madrid: Universidad Politécnica, 2002. 97-106.
- Martínez González, M.<sup>a</sup> M., et ál. (2001). Explotación dinámica de relaciones en las bibliotecas digitales: aplicación a una biblioteca jurídica. // II Jornadas españolas de bibliotecas digitales (JBIDI' 2001). Almagro: Universidad de Castilla – La Mancha, 2001. 169-180.
- Martínez González, M.<sup>a</sup> M.; Fuente, P. de la; Derniame, J. C. (2003a). Una propuesta integrada de extracción de información para gobierno electrónico: estructura, referencias y evolución de los documentos jurídicos. // IV Jornadas de Bibliotecas Digitales (JBIDI' 2003). Alicante, 2003. 201-210.
- Martínez González, M.<sup>a</sup> M.; Fuente, P. de la; Derniame, J. C. (2003b). Relationship-based dynamic versioning of evolving legal documents. // *Lecture Notes on Artificial Intelligence (LNAI)*. 2543 (2003) 290-306.
- Martínez González, M.<sup>a</sup> M.; Fuente, P. de la; Vicente, D.-J. (2004). Dealing with the automatic extraction of references from legislative digital libraries. // *Veille Strategique, Scientifique et Technologique 2004 (VSST 2004)*. 281-288.
- Martínez González, M.<sup>a</sup> M.; Fuente, P. de la (2007). Introducing structure management in automatic reference resolution: an XML-based approach. // *Information Processing and Management*. 43:6 (2007) 1808-1832.

- Martínez González, M.ª M., et ál. (2007). Electronic manipulation of european texts about conflicts of jurisdiction: a Semantic Web tool. // Seminar on Legal Informatics and e-Governance as Tools for the Knowledge Society. Reikiavik, 13-14 July 2007.
- Muller, M. (2002). Legal RDF Dictionary. // XML Europe 2002. [http://www.lexml.de/legal\\_rdf\\_dictionary\\_barcelona.htm](http://www.lexml.de/legal_rdf_dictionary_barcelona.htm) (2008/01/20).
- Nielsen, S. B. (2006). Use of XML for the production and distribution of the official gazettes. // 3.<sup>rd</sup> Meeting European Forum of Official Gazettes. Luxemburgo: Office for Official Publications of the European Communities, 2006. 33-76. [http://circa.europa.eu/irc/opoce/ojf/info/data/prod/html/forum%202006%20\\_%20wd.pdf](http://circa.europa.eu/irc/opoce/ojf/info/data/prod/html/forum%202006%20_%20wd.pdf) (2008/01/20).
- Nogales Flores, J. T., et ál. (2000). Un repertorio legislativo hipertextual mediante marcado de texto: las disposiciones generales del Boletín Oficial de la Comunidad de Madrid. // I Jornadas Españolas de Bibliotecas Digitales. Valladolid: Universidad, 2000. 89-104.
- Nogales Flores, J. T.; Martín Galán, B.; Arellano Pardo, M.ª C. (2003). Informática, derecho y documentación. Experiencias y posibilidades de aplicación de los lenguajes de marcado de texto (SGML, HTML y XML) a los documentos jurídicos. // Actas del XVII Encuentro sobre Informática y Derecho. Madrid: Instituto de Informática Jurídica – Universidad Pontificia de Comillas, 2003. 355-374.
- Nogales Flores, J. T.; Arellano Pardo, M.ª C.; Martín Galán, B. (2004). Propuesta de aplicación de los criterios de técnica legislativa a un sistema de información de legislación usando tecnologías XML. // Actas del XVIII Encuentro sobre Informática y Derecho. Madrid: Instituto de Informática Jurídica – Universidad Pontificia de Comillas, 2004. 273-288.
- Office for Official Publications of the European Communities (Publications Office). FORMEX – Formalized Exchange of Electronic Publications. <http://formex.publications.eu.int/> (2008-01-12).
- Ortiz-Rodríguez, F. (2007). EGODO and applications: sharing, retrieving and exchanging legal documentation across e-government. // Proceedings of the Workshop on Semantic Web Technology for Law at International Conference on AI and Law ICAIL 2007. Stanford, 2007. 21-26. <http://www.cs.vu.nl/~mcklein/SW4Law/proceedings.pdf> (2008-01-16).
- Páez Mañá, J. (1994). Bases de datos jurídicos: características, contenido, desarrollo y marco legal. Madrid: CINDOC, 1994.
- Reviriego, J.; Maciá, M. (1998). Introducción general a la documentación jurídica. // Maciá, M. (ed.). Manual de documentación jurídica. Madrid: Síntesis, 1998. 19-34.
- Rosa Piñero, A. de la; Senso, J. A. (1999). XML como medio de normalización y desarrollo documental. // Revista Española de Documentación Científica. 22:4 (1999) 488-504.
- Sanz, S., et ál. (2003). BD-IDHV: una biblioteca digital de derecho histórico vasco. // IV Jornadas de Bibliotecas Digitales (JBDI'2003). Alicante, 2003. 125-134.
- Schweighofer, E. (1999). The revolution in legal information retrieval or the Empire strikes back. // Journal of Information, Law and Technology (JILT). 1 (1999). [http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/1999\\_1/schweighofer](http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/1999_1/schweighofer) (2008-01-23).

- Vallbé, J. J., et ál. (2007). Iuriservice: ontologies per a la representació del coneixement jurídic. // I Congrés Català de Filosofia. Barcelona: Institut d'Estudis Catalans, 2007. <http://idt.uab.es/docs/2007/Iuriservice-FILCAT.pdf> (2008-01-23).
- Vañó Vañó, M. J. (2005). XML, una herramienta al servicio del buen gobierno corporativo. // Revista Aranzadi de Derecho y Nuevas Tecnologías. (2005) 101-115.
- Sjöberg, C. M. (1998). Critical factors in legal document management: a study of standardised markup languages. Estocolmo: Jure AB, 1998.
- Vicente Blanco, D.-J.; Martínez González, M.<sup>a</sup> M. (2007). Spain on going legislative XML projects. // Biagioli, C.; Francesconi, E.; Sartor, G. (eds.). Proceedings of the V Legislative XML Workshop. Florencia: European Press Academic Publishing, 2007. 23-38.

Recibido: 2008-02-30. Aceptado: 2008-04-30