

SIDRA: XML en la gestión y explotación de la documentación jurídica

Resumen

Se explica la utilización del metalenguaje de marcado XML (Extensible Markup Language) en el Sistema de Información Documental en Red de Asturias (SIDRA), haciendo especial referencia a la documentación legal del sistema, las sentencias y el Boletín Oficial del Principado de Asturias (BOPA) como fuente de información jurídica. Se explica el proyecto SIDRA, y la metodología empleada en la creación de una base de datos de sentencias y el boletín electrónico (E-BOPA) dentro de dicho proyecto.

Palabras clave

XML, Sistema de Información Documental en Red de Asturias, SIDRA, documentación jurídica, bases de datos, sistemas de información

Abstract

This paper explains the use of the XML markup language in the Asturias Online Document Information System (SIDRA), paying special attention to the system's legal documentation, the judgements, and the Asturias Official Gazette (BOPA) as source of legal information. The SIDRA project is also explained as well as the methodology used to create a judgements database, and the online official gazette (E-BOPA)

Keywords

XML, Asturias Online Document Information System, SIDRA, legal documentation, databases, information systems

Podríamos definir el lenguaje XML (eXtensible Markup Language) como un metalenguaje de marcado que permite describir la estructura de tipos específicos de documentos electrónicos. Estas estructuras se definen por medio de DTD (Definición de Tipo de Documento) o XML Schemas (Esquemas XML), definiendo un conjunto de etiquetas y atributos.

La edición de documentos XML persigue los siguientes objetivos:

- Distinguir el contenido y la estructura de los documentos, de su presentación en papel o en pantalla.
- Hacer explícita su estructura y sus contenidos informativos.
- Crear documentos que puedan intercambiarse y procesarse con facilidad en sistemas informáticos heterogéneos.

Para alcanzar estos objetivos XML propone un formato en el que se intercalan marcas en el texto de los documentos con el objeto de distinguir las partes o elementos estructurales del mismo.

La posibilidad de poder intercambiar información entre sistemas informáticos heterogéneos, así como los niveles de seguridad que ofrece este lenguaje, tienen un enorme potencial en el ámbito de la documentación legal. En este sentido, durante los últimos años, han surgido en este ámbito iniciativas que persiguen el intercambio automático de información mediante la utilización de XML. Para ello es necesario crear estructuras de datos que permitan intercambiar información, pero alcanzar la estandarización es tarea difícil debido a la diversidad multicultural, sobre todo en Europa.

En este sentido, las dos organizaciones pioneras en la estandarización de la información legal, son LegalXML en Estados Unidos (sección de OASIS, Organisation for the Advancement of Structured Information Standards) (1) y LEXML (2) en Europa. Las dos organizaciones persiguen la creación de un grupo de DTD para determinados tipos de documentos jurídicos.

LegalXML ha centrado sus esfuerzos en la intermediación entre operadores jurídicos norteamericanos con el fin de elaborar una DTD para todos los documentos de una clase determinada.

Por su parte la Red LEXML se creó como foro para sustentar el interés creciente en el intercambio de datos legales. En este foro abierto se intercambian ideas y experiencias asociadas a XML para el ámbito legal. Se trata de una red descentralizada, pero que tiende a una aproximación paulatina a través de listas de distribución, asambleas y por el trabajo conjunto a través de proyectos internacionales. Este foro es el punto de coordinación y de trabajo para el desarrollo de estructuras estandarizadas, vocabularios y herramientas de intercambio de datos. LEXML persigue un modelo de datos legal y global y el desarrollo de una fuente abierta de un programa oficial legal, que hable y entienda XML.

LEXML no acude a una estructura de datos por tipo de documento, como hace LegalXML, debido a la diversidad cultural, su objetivo es crear un buen número de estructuras de datos para las diferentes comunidades (Vañó y Broseta, 2005).

Estas iniciativas están consiguiendo importantes avances en la aplicación del lenguaje de marcado XML a la documentación jurídica.

Desde el Gobierno del Principado de Asturias se ha hecho un esfuerzo para la mejora del tratamiento específico este tipo de documentación. En este sentido se han llevado a cabo dos desarrollos: la base de datos de sentencias para el Servicio Jurídico del Principado de Asturias y el Sistema para la Gestión y Edición Electrónica del Boletín Oficial del Principado de Asturias (E-BOPA). Se trata de dos iniciativas en las que se hace utilización de XML con documentación jurídica en el ámbito del gobierno electrónico, y que se enmarcan en el Sistema de información Documental en Red de Asturias (SIDRA), del que hablaremos a continuación.

1. El Sistema de Información Documental en Red de Asturias (SIDRA)

En el año 2004, se crea el SIDRA para dar solución al inadecuado aprovechamiento de la información por parte del personal de la Administración del Principado de Asturias (gestores, directivos, administrativos...) y a la situación de los centros de documentación especializados que no permitía ofrecer un servicio de calidad a los ciudadanos. Existía un claro aislamiento y falta de coordinación entre los centros de documentación y bibliotecas especializadas, y entre las unidades administrativas, sin la necesaria adecuación a las nuevas tecnologías para el cumplimiento de los objetivos marcados por el Gobierno asturiano, en concordancia con los mandatos españoles y europeos respecto al desarrollo de la administración regional en el marco de la sociedad de la información.

El personal de la administración lleva a cabo las más variadas funciones administrativas (jurídicas, técnicas, asistenciales, educativas, sanitarias, de asesoramiento, etc.), pero se identificaron dos necesidades comunes a todos ellos:

- obtener en cada caso la información y el conocimiento precisos para el adecuado desempeño de sus funciones y
- tratar o procesar los documentos que soportan los trabajos realizados o las decisiones adoptadas, a los efectos de la posible reutilización de la información en ellos recogida.

Para dar solución a la situación existente y atender estas necesidades, se decide crear en el marco de sus políticas de información, un sistema de información documental en red, siguiendo las líneas de actuación establecidas en el «Plan de Modernización y Mejora de la Calidad en la Administración

del Principado de Asturias». Es entonces cuando la Dirección General de Modernización y Sistemas de Información (actualmente Dirección General de Modernización, Telecomunicaciones y Sociedad de la Información), a través del Servicio de Archivos Administrativos y Documentación (actualmente Servicio de Publicaciones, Archivos Administrativos y Documentación) pone en marcha el Sistema de Información Documental en Red de Asturias (SIDRA).

En estos momentos el SIDRA es un sistema integral de información documental en red ya consolidado. Está soportado por una única base de datos y en una única aplicación de gestión documental, que permite el tratamiento, la recuperación, la distribución y, en su caso, el préstamo de todo tipo de documentación o información especializada, pública o restringida, necesaria para el ejercicio de las funciones de la Administración del Principado de Asturias. SIDRA es un sistema dinámico que facilita el aprovechamiento de los recursos, el intercambio de información y la cooperación, con dos objetivos fundamentales:

- Dar servicio a la sociedad asturiana, facilitando el acceso web a la información disponible en los centros de documentación y bibliotecas especializadas de la Administración directa e institucional del Principado.
- Dar soporte a la organización del Principado de Asturias para la gestión y la toma de decisiones basadas en el conocimiento.

En la actualidad el SIDRA está integrado por los siguientes organismos:

- Red de Centros de Documentación. Esta red está compuesta actualmente por 21 unidades de información entre las que se encuentran, entre otras, las de los siguientes organismos: Ente Público de Servicios Tributarios del Principado de Asturias, Consejo Económico y Social, Oficina de Investigación Biosanitaria, SADEI (Sociedad Económica de Estudios Económicos e Industriales), etc.
- Servicio de Publicaciones, Archivos Administrativos y Documentación del Gobierno del Principado de Asturias. Este servicio es el coordinador del sistema y se ha encargado del desarrollo de los trabajos objeto de este artículo, la base de datos de sentencias del Servicio Jurídico del Principado de Asturias y el Sistema para la Gestión y Edición Electrónica del Boletín Oficial del Principado de Asturias (E-BOPA).
- Red Asturiana de Información Juvenil (RAIJ), compuesta por 39 Oficinas de Información Juvenil (OIJ).
- Gabinetes de las Consejerías del Gobierno del Principado.
- Servicio Jurídico del Principado de Asturias.

1.1 Software

El SIDRA se ha desarrollado con tecnología estándar universal XML bajo la plataforma de Tamino y Dogma. Este software permite definir de forma abierta y flexible distintos modelos de datos y asociar a los registros todo tipo de documentos originales (archivos textuales, video, imagen, sonido...) tratados (captura, digitalización, indexación, versionados...) para garantizar su localización inmediata y la recuperación de su contenido. Los registros se almacenan en formato XML.

1.1.1 Tamino

Se trata de una base de datos centrada en documentos (*data-centric database*) que respeta la estructura de los mismos, permite hacer consultas sobre dicha estructura y hace posible la recuperación del documento tal como fue insertado originalmente. Algunas de sus características más importantes son:

- Soporta información en diferentes formatos, debido a que les permite pasar su información a un formato XML rápidamente, utilizando la aplicación de su preferencia, y de esta manera almacenarla en una de estas bases de datos y evitarse el proceso de conversión de un formato a otro.
- Provee tanto almacenamiento XML como mecanismos de almacenamiento relacional SQL. Esta característica permite a los usuarios consultar datos heterogéneos mediante XQL y recibir conjuntos de resultados en formato XML.
- Facilita el desarrollo de aplicaciones basadas en XML, como por ejemplo servicios web.
- Almacena casi cualquier tipo de documento incluyendo información formateada XML, páginas HTML, cartas, audio, video, imágenes y datos de bases de datos SQL o de objetos. Los documentos se agrupan en función de la información que contienen, en colecciones que a su vez pueden contener otras colecciones.
- Soporta varios lenguajes de consulta.
- Indexación XML. Se ha de permitir la creación de índices que aceleren las consultas realizadas frecuentemente
- Crea de identificadores únicos. A cada documento XML se le asocia un identificador único por el que será reconocido dentro del repositorio.

1.1.2 Dogma

Se trata de la aplicación de gestión documental del SIDRA, soportada por Tamino, sus características principales son las siguientes.

- I*NET: servicios web de acceso multicanal.
- Módulos parametrizables y escalables.
- Repositorio único de información.
- Integración con Portal www.asturias.es, BD de Terceros, LDAP, CRM.
- Tratamiento documental: catalogación, indización...
- Captura automática y almacenamiento de documentos de texto, imagen, sonido y video en diferentes formatos: .doc, .pdf, .jpg, .wma, .wmv .mp3,...
- Digitalización asistida por OCR.
- Campos controlados y ayudas configurables.
- Gestión de tesauros, clasificaciones y listas de materias.
- Soporte de flujo de trabajo (Workflow).
- Versionado de documentos.
- Difusión selectiva de la Información (DSI).
- Circulación de documentos y gestión del préstamo.
- Definición de perfiles de usuarios y niveles de seguridad.
- Generación de informes y estadísticas a través de la herramienta XML-MART.

El concepto de “documento”, para DOGMA, es totalmente abierto agrupando desde descripciones (información de los atributos o metadatos) a múltiples objetos (imagen, vídeo, sonido, etc.) para un mismo documento. El resultado es que el documento es visto como una agrupación simple o compleja de información. DOGMA facilita de forma integrada el acceso a cualquier tipo de información e incorpora toda una serie de funciones que permiten el mantenimiento de la información así como su análisis y tratamiento.

2. Base de datos de *Sentencias* del Servicio Jurídico del Principado de Asturias

La función de esta base de datos es la de permitir la recogida, tratamiento, recuperación y difusión de la Jurisprudencia en la que son parte el Gobierno y la Administración del Principado de Asturias. No se han creado estructuras de datos tipo DTD o XML Schema para las *Sentencias* si no que se ha definido un modelo de datos sobre la infraestructura tecnológica preexistente en el SIDRA (Tamino y Dogma) para cargar los registros a los que se adjuntarán las sentencias a texto completo.

Los objetivos con los que se creó esta base de datos son los siguientes:

- Dotar al Servicio Jurídico de un sistema de información, mediante su incorporación al SIDRA.
- Mejorar el nivel de eficiencia, efectividad y satisfacción en el acceso a la información jurídica.
- Facilitar el acceso inmediato a los activos documentales relevantes.
- Generar nuevos servicios de valor añadido.

Partiendo de estos objetivos se realizó un análisis de las necesidades del Servicio Jurídico mediante una Auditoría de Información que se estructuró en 7 fases:

- Análisis del fondo, requisitos especiales del servicio y necesidades.
- Identificación de las entidades documentales básicas.
- Definición de los atributos de las entidades documentales.
- Definición del nivel de seguridad de los documentos.
- Configuración de búsquedas.
- Análisis de los instrumentos de descripción (tesauros, clasificaciones...) utilizados en el Servicio.
- Definición de nuevos instrumentos de descripción (tesauros, clasificaciones listas de materias...) en aquellos casos en los que el servicio no utilice ninguno.

2.1 Análisis del fondo, requisitos especiales del Servicio y necesidades

En colaboración con el Servicio Jurídico se detectaron unas necesidades específicas de tratamiento de documentos, flujos de trabajo e información estadística. Estas necesidades se concretaron en la decisión de crear una base de datos de *Sentencias* que permitiese recoger, tratar, recuperar y difundir la Jurisprudencia en la que son parte el Gobierno y la Administración del Principado de Asturias.

2.2 Identificación de las entidades documentales básicas. Análisis, tratamiento y propuesta de gestión normalizada

Se identificó un grupo homogéneo de documentación, las *Sentencias*. El Servicio Jurídico proporcionó varios modelos de dichas *Sentencias*, con una serie de elementos que se repiten en todas y cada una de ellas y que proporcionaron un marco sobre el que trabajar.

Sobre estos datos se definió la siguiente propuesta:

- La creación de una base de datos específica para el Servicio Jurídico, denominada *Sentencias*, compuesta por una entidad principal para esta clase de documentos.
- La creación en dicha base de datos de una serie de entidades auxiliares que permitieran por un lado la normalización de los datos y por otra agilizar el proceso de carga de una serie de campos incluidos en la entidad principal.

Teniendo en cuenta que los nombres de los diferentes campos reflejan su contenido desde el punto de vista documental, no jurídico y que pueden/deben ser modificados por el Servicio Jurídico, el esquema general de dichas entidades fue el siguiente:

- Entidad principal: *Sentencias*.
- Entidades auxiliares:
 1. *Fallos de sentencias*.
 2. *Juzgado / Tribunal*.
 3. *Recursos contra sentencia*.
 4. *Sentencias asuntos penales*.
 5. *Sentencias dictadas contra recursos contencioso-administrativos*.

2.3 Definición de los atributos de las entidades documentales

Para cada una de las nuevas entidades se definieron una serie de atributos (metadatos) específicos que se detallan a continuación:

- Entidad principal (*Sentencias*): Como entidad principal era la más compleja y la que contenía un mayor número de datos, además estaba relacionada con todas y cada una de las entidades auxiliares. El modelo de datos definido para esta entidad se fundamentó en la información proporcionada por el Servicio Jurídico del Principado de Asturias.
- Entidades auxiliares: Se trata de entidades mucho más sencillas y los atributos definidos para cada una de ellas fueron los siguientes:
 - Fallos de sentencias
 - Número de control, asignado por el sistema.
 - Fallo: tipo de fallo de la sentencia.
 - Descripción: comentario en texto libre sobre el fallo.
 - Juzgado / Tribunal
 - Número de control, asignado por el sistema.
 - Nombre del Juzgado / Tribuna.

- Código: código correspondiente a cada Juzgado o Tribunal.
- Recursos contra sentencia:
 - Número de control, asignado por el sistema.
 - Tipo de recurso.
 - Descripción: comentario en texto libre sobre el fallo.
- Sentencias asuntos penales:
 - Número de control, asignado por el sistema.
 - Tipo de sentencia.
 - Descripción: comentario en texto libre sobre el fallo.
- Sentencias dictadas contra recursos contencioso-administrativos:
 - Número de control, asignado por el sistema.
 - Tipo de sentencia.
 - Descripción: comentario en texto libre sobre el fallo.

2.4 Definición del nivel de seguridad de los documentos

SIDRA permite establecer los siguientes tipos de seguridad en documentos:

- **Información pública:** información que podrá ser accedida por todo tipo de usuario a través de Internet.
- **Información interna:** información que solo es accesible a través de Intranet.
- **Información restringida:** información que solo puede ser accedida por los usuarios que pertenezcan al Servicio Jurídico.
- **Información restringida/ compartida:** podrá estar disponible para otros colectivos o centros si así lo indican los responsables del Servicio Jurídico.

En el caso de las *Sentencias* el nivel de seguridad que se definió fue el restringido, con la posibilidad de compartir en algún momento con otros colectivos si así lo indican los responsables del Servicio.

2.5 Configuración de búsquedas

De nuevo, para cada una de las entidades, principal y auxiliares, se definieron una serie de campos de búsqueda en función de las necesidades de que se detectaron en el Servicio Jurídico y de un análisis documental de las mismas. Sobre cada uno de los atributos (metadatos) definidos para todas las entidades se pueden realizar tres tipos de búsquedas: *textual*, *alfabética* y *jerárquica*. Se configuró la búsqueda *textual* como criterio de búsqueda por defecto. En todas aquellas entidades

que tengan definido más de un campo de búsqueda se pueden utilizar operadores lógicos (y, o, no) para combinar campos.

2.6 Análisis de los instrumentos de descripción utilizados en el Servicio

El Servicio Jurídico del Principado de Asturias no utilizaba ningún instrumento de descripción (tesauro, lista de materias) normalizado.

2.7 Definición de nuevos instrumentos de descripción (tesauros, listas de materias...) en aquellos casos en los que el Servicio no utilice ninguno

En la nueva base de datos de *Sentencias* se ha incluido como instrumento de descripción normalizado el Tesauro Eurovoc, con el desarrollo propio del Senado (4). Esta decisión se tomó una vez analizadas otras posibles opciones (5).

3. Sistema para la Gestión y Edición electrónica del Boletín Oficial del Principado de Asturias (E-BOPA)

El proyecto E-BOPA, es un proyecto actualmente en desarrollo, que acomete la modernización de todos los procesos implicados en la publicación del Boletín Oficial del Principado de Asturias, desde el momento en que se produce la solicitud de la publicación de un anuncio en el boletín, hasta su publicación en el mismo. También gestiona los productos de salida: el propio BOPA, su publicación en el portal corporativo, en el catálogo de servicios del portal, en la base de datos legislativa INVESDOC y en otros posibles servicios futuros al ciudadano, como el acceso a un repositorio de instancias tipo para gestiones con la Administración del Principado o la posibilidad de recibir un servicio de Difusión Selectiva de la Información (DSI) del BOPA.

Con este proyecto se pretende atender a las siguientes necesidades:

- Gestión de los anuncios a publicar.
- Edición de anuncios.
- Montaje de boletines.
- Mantenimiento de un histórico de anuncios y/o Boletines.
- Elaboración de informes.
- Exportación a distintos formatos.

- Recepción de anuncios multicanal.

Para conseguir estos objetivos se han creado dos aplicaciones basadas en la misma infraestructura tecnológica utilizada en el SIDRA (Tamino y Dogma):

- PUBLIBOPA. Sistema de inserción digital de anuncios.
- SIGBOPA. Sistema para la Gestión y Publicación del Boletín Oficial del Principado de Asturias.



Figura 1. E-BOPA.

Estas aplicaciones además de estar integradas entre sí, deberán integrarse con otras aplicaciones preexistentes:

- Portal corporativo www.asturias.es. Esta integración permite la publicación web del Boletín Oficial en el portal corporativo, para hacer posible esta publicación se ha creado un XML Schema.
- Catálogos de Servicios y Trámites del Principado de Asturias. Actualización de la ficha del Catálogo de Servicios en el portal corporativo con el modelo de solicitud.
- Base de datos legislativa INVESDOC. La finalidad de la integración con esta base de datos es el registro y almacenamiento a texto completo en formato PDF de los anuncios publicados en el Boletín Oficial.

4. Conclusiones

La aplicación del lenguaje XML durante los últimos en los procesos de gestión de contenidos ha significado un importante avance, tanto para el almacenamiento de los mismos y sus metadatos, como para la integración y el intercambio de contenidos procedentes de aplicaciones o de organizaciones externas.

Las iniciativas desarrolladas desde la Administración del Principado de Asturias que hemos expuesto en este artículo, pretenden conseguir precisamente esto, una mejora de los procesos de gestión de los contenidos que emanan de nuestra organización, de forma que ayuden al personal de la administración en la toma de decisiones y en las desarrollo de sus tareas diarias. Todo ello orientado a ofrecer mejores servicios a los ciudadanos.

En esta línea, desde nuestra Administración seguimos adaptándonos a estas nuevas tecnologías, que favorecen una adecuada gestión del conocimiento en la organización y una gestión de calidad.

5. Bibliografía

- Alvite, M^a Luisa (2004). Evolución de las bases de datos jurídicas en España. // Anales de Documentación. 7 (2004) 7-27. <http://www.um.es/fccd/anales/ad07/ad0701.pdf> (2008-02-25).
- Arellano, María del Carmen (2005). Aportaciones de la técnica legislativa y XML a la informática jurídica documental. Madrid: Universidad Carlos III de Madrid, Boletín Oficial del Estado, 2005.
- Bosch, Mela (2001). Documentos y lenguaje de marcado: conceptos, problemas y tendencias. // El profesional de la información. 10:11 (Noviembre 2001) 4-9.
- Eíto, Ricardo (2005). XML y la gestión de contenidos = Content Management and XML: the needed interaction. Hipertext.net.3 (Mayo 2005). <http://www.hipertext.net> (2008-02-21).
- Morrison, Michael. (2000). XML al descubierto. Madrid: Prentice Hall, 2000.
- Vañó, María José; Broseta, Manuel (2005). XML, una herramienta al servicio del buen gobierno corporativo. // Revista Aranzadi de Derecho y Nuevas Tecnologías. 8 (2005) 101-116.

Notas

(1) Legalxml: <http://www.legalxml.org/>

(2) LEXML-ES: <http://www.uv.es/lexml/>

(3) Página del Catálogo de acceso en línea del Sistema de Información Documental en Red de Asturias: <http://sidra.princast.es>.

(4) Acceso al Tesauro Eurovoc: <http://www.senado.es/basesdedatos/index.html>.

(5) Otros tesauros de interés:

- www.cindoc.csic.es (Tesauro de derecho).
- www.westlaw.es (Tesauro de Aranzadi). Disponible a través de Intranet.
- www.bosch-online.net (Tesauro específico de Jurisprudencia).