

CAPÍTULO III.4

DEFINICIÓN DE ESTRUCTURAS Y VOCABULARIOS EN XML

III.4.1. LIMITACIONES EN EL MODELO DE LA DTD DE XML

Uno de los mayores logros del metalenguaje de marcado SGML fue, sin duda, la dotación de una naturaleza dual al modelo de DTD, como ya se indicó con anterioridad. Chahuneau explica perfectamente esta idea al señalar que la DTD es tanto una gramática generativa de lenguajes de marcado, principalmente descriptivos, que pueden emplearse para etiquetar las instancias de los documentos electrónicos así como un esquema con el cual poder caracterizar una clase de documento a través de la definición de una estructura y una serie de restricciones en su utilización¹⁷⁸. En cuanto a la primera capacidad de las DTDs SGML, se analizó con detenimiento en el capítulo anterior que la generación de nuevos vocabularios SGML derivaba de un proceso de análisis documental donde ante todo primaba una modelización orientada al contenido de los documentos en detrimento de los aspectos sintácticos del mismo y de una modelización de los datos¹⁷⁹. En cuanto a la segunda capacidad de las DTDs SGML, y a pesar de esta orientación hacia el establecimiento de restricciones estructurales o, en el mundo de las bases de datos, “esquemas” este tipo de documentos siempre han sido difícilmente procesables por los

¹⁷⁸ François Chahuneau. “SGML & schemas: from SGML DTDs to XML-DATA” [documento HTML]. En: *SGML/XML Europe 98 Conference* (1998. Paris). GCA, 1998. Disponible en <http://www.infoloom.com/gcaconfs/WEB/paris98/chahuneau.HTM> (consultado el 5 de octubre de 2000).

¹⁷⁹ El concepto de *data modelling* procede del campo de actuación de las bases de datos relacionales donde los datos que ha de contener la base de datos son modelizados obteniendo de este modo las entidades (siendo su correspondiente representación física en este modelo las tablas) y los atributos (siendo su correspondiente representación física en este modelo las columnas con los tipos de datos contenidos). Resulta fácil entender, como así lo defiende C. Corkern, la similitud existente en cuanto a su funcionalidad entre el análisis documental y la modelización de los datos pues en ambos casos de lo que se trata es de establecer los mecanismos adecuados para una perfecta definición del tipo, estructura y relaciones que se dan entre los elementos informativos contenidos en un documento o en un repositorio de datos. Para una más amplia información sobre esta idea, véase Carla Corkern. “XML and Information Modeling” [documento HTML]. En: *XML 98 Conference* (1998. Chicago). GCA, 1998. Disponible en <http://www.infoloom.com/gcaconfs/WEB/chicago98/corkern.HTM> (consultado el 21 de enero de 2001).

sistemas de bases de datos relacionales y orientadas a objetos, debido a aspectos tan controvertidos como, entre otros muchos, las “excepciones” SGML¹⁸⁰.

XML ha pretendido corregir estos aspectos ciertamente controvertidos tratando de ser, como ya se ha expuesto con anterioridad en este capítulo, un modelo de metalenguaje de marcado tanto para documentos como para datos que han de ser servidos y procesados en un espacio de trabajo web potente, pero a su vez flexible. Esta flexibilidad viene dada en gran medida por la opcionalidad que se le brinda al usuario de adscribir el documento XML que está elaborando a un determinado tipo documental; esto es, conforme a una serie de restricciones en los nombres de elementos y atributos (un vocabulario) y en la secuencia de aparición y de contenido de dichos elementos y atributos (documento válido con una DTD más simplificada y, por tanto, más sencilla de procesar por los *parsers*) o, por el contrario, elaborar documentos XML sintácticamente correctos pero sin una declaración formal del tipo de documento (documento bien formado).

Esta posibilidad de crear documentos XML sin DTD, en principio loable en aras de la simplificación en el uso de este metalenguaje en la Web y de dar cabida a todo tipo de orientaciones en su aplicación, ha sido duramente criticada por algunos especialistas procedentes del campo de aplicación del metalenguaje SGML cuando dicho empleo de XML se circunscribe a entornos corporativos de trabajo. Uno de los casos más significativos lo constituye C. Ensign quien establece la necesidad del empleo de las DTDs en los documentos XML aduciendo tres importantes razones para ello¹⁸¹:

1. Previene posibles luchas por el establecimiento de las etiquetas (*tags war*) por parte del personal al establecerse un modelo normalizado de tipos documentales dentro de la organización.
2. La creación de una especificación formal de tipo de documento, aunque larga y compleja, proporciona una mayor eficacia a las organizaciones dado que se aseguran

¹⁸⁰ *Ibid.*, <http://www.infoloom.com/gcaconfs/WEB/paris98/chahuneau.HTM>

¹⁸¹ Chet Ensign. “Structure rules! Why DTDs matter after all”. *Markup Languages: Theory and Practice*, v. 1, n° 1, winter 1999, p. 104.

aspectos tan importantes como la integridad y veracidad de los documentos y datos tratados.

3. El establecimiento de DTDs dentro de la organización resulta igualmente vital para definir la metainformación contenida en los documentos electrónicos, expresada en la forma que dicha organización establezca e imbuida en los elementos y/o atributos de dichos documentos marcados.

Si bien es cierto que son pocos los que dudan de los beneficios que proporciona el mecanismo de definición de estructuras y vocabularios en XML, más aún si este análisis se produce desde la óptica de los profesionales de la información y la documentación, constituyendo en gran medida la piedra angular del modelo de gestión de los documentos electrónicos XML en entornos corporativos¹⁸², no es menos cierto que el mecanismo de la DTD resulta insuficiente para lo que XML pretende ser: XML está concebido de tal modo que sea capaz de manipular tanto el texto de los documentos como cualquier otra forma de estructuración de datos¹⁸³.

Son múltiples las limitaciones que el modelo de DTD heredado y reducido de SGML presenta para este nuevo entorno de aplicaciones XML donde documentos y datos se combinan, entre las que se pueden destacar las ofrecidas por P. Murray-Rust¹⁸⁴:

- **Falta de soporte para la inclusión de semántica a las declaraciones de elementos:** A pesar de que se pueden incluir anotaciones textuales o comentarios en las

¹⁸² Este aspecto de vital importancia para todos aquellos que nos dedicamos profesionalmente al tratamiento y gestión de documentos electrónicos en entornos corporativos fue tratado con detalle en la comunicación de Bonifacio Martín Galán, David Rodríguez Mateos. “Estructuración de la información mediante XML: un nuevo reto para la gestión documental”. En: *Jornadas Españolas de Documentación* (7º. 2000. Bilbao). Bilbao: Universidad del País Vasco, 2000, p. 119 y ss.

¹⁸³ Demetrios Ioannides. “XML schema languages: beyond DTD”. *Library Hi-Tech*, v. 18, n° 1, 2000, p. 10.

¹⁸⁴ Peter Murray-Rust. “Do we need DTDs?” [documento HTML]. En: *XML Europe 99 Conference* (1999. Granada). GCA, 1999. Disponible en <http://www.infoloom.com/gcaconfs/WEB/granada99/mur.HTM> (consultado el 22 de enero de 2001).

DTDs, estos comentarios no son procesables por las herramientas informáticas y no aportan, además, mecanismos formales de enlace a recursos ontológicos o programáticos.

- **Gran limitación de los modelos de contenidos:** En las DTDs de XML los modelos de contenido de un elemento sólo pueden hacer mención a la inclusión de texto procesable (PCDATA) y/u otros elementos por lo que muchos desarrollos actuales optan por definir sus modelos de contenido como ANY (pues con los actuales documentos electrónicos un determinado elemento puede y suele contener texto, gráficos, imágenes, datos numéricos, otros ficheros en formatos electrónicos diversos, etc.).
- **Incapacidad para reutilizar un documento completo que declare una DTD:** Resulta del todo imposible reutilizar un documento XML de forma completa dentro de otro si ambos tienen sus propias definiciones de tipos de documentos pues el mecanismo de asociación a través de la propiedad ENTITY sólo permite la inclusión de entidades paramétricas que hayan sido previamente definidas pero nunca de una DTD completa (y, por tanto, del documento)¹⁸⁵.
- **No se proporciona soporte para tipos de datos diversos:** Mediante la DTD de XML sólo se puede establecer dos tipos de datos contenidos en el documento, bien texto procesable o bien datos binarios u otros tipos que no han de ser procesados sintácticamente. Así, la ausencia de un método aceptable para la restricción de los valores que pueden tomar los datos en los elementos o en los atributos de los mismos

¹⁸⁵ Resulta ciertamente revelador a este respecto el capítulo 4 “Extending a Document Type Definition with Local Modifications” del manual de Lee Anne Phillips, citado con anterioridad en este capítulo, pues en él se muestra a través de numerosos ejemplos prácticos la gran dificultad que existe para hacer “extensible” una DTD; esto es, añadir nuevos elementos y/o atributos definidos en otras DTDs a través de referencias a entidades paramétricas externas. Piénsese, además, que el mecanismo de los XML Namespaces, ya comentado con anterioridad, surge con posterioridad a la aparición de la Recomendación XML de 1998, tratándose en gran medida de un “parche” provisional para solventar esta gran limitación de la DTD. Para una mayor información véase, L. A. Phillips. *Op. cit.*, pp. 93-110.

es, sin lugar a dudas, uno de los mayores y más importantes inconvenientes del modelo de la DTD de XML¹⁸⁶. Hoy en día resulta fundamental en muchas aplicaciones diferenciar los datos textuales de los numéricos (de todo tipo, como fechas, números enteros, números decimales, etc.) o de cualesquiera otra forma de datos, y esto es algo a lo que el modelo de DTD de XML no ha podido dar respuesta¹⁸⁷.

- **Escasa flexibilidad:** Para algunas aplicaciones que requieren conformidad en las estructuras la rigidez de la DTD puede resultar útil. Pero existen otras muchas aplicaciones donde es necesario que los modelos de contenido sean abiertos y flexibles.
- **Escaso control del valor de los atributos:** Únicamente existe un control del valor de los atributos cuando estos son fijos o existe una enumeración de los mismos.
- **El propósito y la sintaxis de la DTD resultan arcaicos:** La DTD XML al ser heredada de SGML se orienta principalmente al tratamiento de documentos electrónicos donde mayoritariamente predominen los elementos textuales. Además, la sintaxis empleada para la redacción de las DTDs (la gramática EBNF) es distinta en muchos aspectos a la utilizada para la redacción de los documentos XML.

¹⁸⁶ Opinión manifestada unánimemente por todos los especialistas e investigadores en la materia y destacada sobremanera en el artículo de Mark H. Needleman. "XML Schema Language: The New Way of Coding XML". *Serials Review*, v. 27, n° 2, August 2001, pp. 49-52.

¹⁸⁷ Es importante señalar que han existido diversas propuestas para la definición e incorporación de nuevos tipos de datos dentro del modelo de DTD. Tal vez, la propuesta más conocida fue la denominada DT4DTD (*Datatypes for DTDs*) realizada por L. Buck, Ch. F. Goldfarb y P. Prescod al W3C para su debate en enero de 2000. Esta propuesta, fuertemente influenciada por el esquema XML-Data de la compañía Microsoft, define un modelo para la conversión de las declaraciones de marcado de las DTDs a un modelo de esquema XML, permitiendo a los autores especificar ciertos tipos de datos para los valores de los atributos y los contenidos de los datos a través de referencias en la DTD a los tipos de datos definidos en el modelo de esquema antes citado. Para una mayor información, véase World Wide Web Consortium. *Datatypes for DTDs (DT4DTD) 1.0. W3C Note* [documento HTML]. Lee Buck, Charles F. Goldfarb, Paul Prescod (eds.). W3C, January 13, 2000. Disponible en <http://www.w3.org/TR/2000/NOTE-dt4dtd-20000113> (consultado el 25 de enero de 2001).

III.4.2. NUEVOS MODELOS DE ESQUEMAS PARA DOCUMENTOS XML

Debido a estas grandes limitaciones, y fruto del pujante peso que los desarrolladores informáticos de bases de datos y las compañías con intereses en el intercambio de datos entre aplicaciones a través de la Web, empezaron a surgir nuevas iniciativas al amparo del W3C para la sustitución del modelo de DTD en los documentos XML por mecanismos más potentes y precisos, todas ellas englobadas dentro del nombre genérico de *XML Schemas* (esquemas XML).

El término “esquema” hace referencia a las restricciones impuestas en la estructura de datos dentro de las tablas en el modelo de bases de datos relacional¹⁸⁸. En estos esquemas, los modelos de contenido son descritos en términos de “restricciones”, las cuales definen lo que puede aparecer en un determinado contexto. De forma general, se pueden establecer dos tipos de restricciones a través de los esquemas: las restricciones en el modelo de contenido, encaminadas a describir el orden y la secuencia de los elementos, y las restricciones en los tipos de datos, encaminadas a la descripción de las unidades válidas de datos. Así, si un documento XML puede ser concebido como un “contenedor” de datos, dicho documento participará de características similares a las tablas en las bases de datos relacionales y, de forma más directa, a las clases en las bases de datos orientadas a objetos.

En esta analogía de los esquemas XML con las bases de datos, un lenguaje de esquema en XML cumple la misma función que el lenguaje de definición de datos (*Data Definition Language*) para las bases de datos relacionales: un esquema XML específico resulta similar a la definición de una tabla específica en una base de datos relacional y, por tanto, la instancia del documento XML corresponderá a un registro dentro de dicha tabla. De igual modo, el lenguaje de esquema en XML es idéntico en su función a un lenguaje de programación (por ejemplo, Java o C++) utilizado para definir una clase en las bases de datos orientadas a

¹⁸⁸ Norman Walsh. *Schemas for XML* [documento HTML]. XML.com, July 1, 1999. Disponible en <http://www.xml.com/pub/1999/07/schemas/index.html> (consultado el 14 de noviembre de 2000).

objetos: un esquema XML específico es como una clase con sus propiedades y, por tanto, la instancia del documento XML tiene su correspondencia con la instancia del objeto de dicha clase¹⁸⁹.

Un esquema XML es en cuanto a su relación con los documentos XML un mecanismo de propósito idéntico a la DTD, esto es, un modelo útil tanto para la descripción de la estructura de la información contenida en un documento XML como para la definición de nuevos vocabularios XML, permitiendo, por consiguiente, la validación de la estructura y del contenido de dicho documento¹⁹⁰. Pero los esquemas XML dan un paso más allá en la evolución del metalenguaje XML permitiendo superar las limitaciones de las DTDs antes expuestas y ofreciendo un mecanismo mucho más potente y expresivo para que las aplicaciones web puedan intercambiar datos de forma más consistente sin tener que renunciar a mecanismos *ad hoc* de validación de los mismos.

A grandes rasgos, se pueden definir los esquemas XML en cuanto a su funcionalidad como¹⁹¹:

- Un modelo de datos dado que a través de los esquemas XML se puede especificar el modo en el que los datos XML van a ser organizados así como los tipos de datos contenidos para los mismos.

¹⁸⁹ Yasser Shohoud. "XML's Grand Schema" [documento ASP]. *XML Magazine*, v. 1, n° 3, summer 2000. Disponible en <http://www.xmlmag.com/upload/free/features/xml/2000/03sum00/ys0300/ys0300.asp> (consultado el 5 de julio de 2000).

¹⁹⁰ Simon St. Laurent. *Describing Your Data: DTDs and XML Schemas* [documento HTML]. XML.com, December 1, 1999. Disponible en <http://www.xml.com/pub/1999/12/dtd/index.html> (consultado el 9 de diciembre de 1999).

¹⁹¹ Roger L. Costelo. *XML Schemas. XML Technologies Course* [documento PPT]. XFront.com, rev. May 6, 2001. Disponible en <http://www.xfront.com/xml-schema.html> (consultado el 25 de mayo de 2001). Existe, como se puede comprobar, una disparidad entre tipo documento declarado para este recurso (una presentación en PowerPoint) y el tipo de documento al cual hace mención el URL (un documento HTML). Este hecho es debido a que en realidad dicho documento PPT se encuentra incluido dentro de un vasto conjunto de ficheros electrónicos de todo tipo que conforman el excelente manual "XML Schema Tutorial" que el citado autor suministra de forma gratuita (y en formato comprimido para agiliza su descarga) a través de la red.

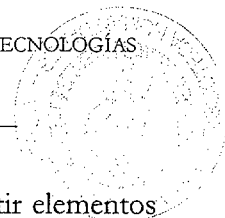
- Un contrato para elaborar tipos documentales dentro de las organizaciones al aceptar los miembros de éstas una serie de estructuras normalizadas a través de los esquemas para los documentos XML que se elaboran en el seno de las mismas.
- Una fuente de gran riqueza para los metadatos dado que el documento que define al esquema XML contiene una gran cantidad de información (datos) sobre los datos que se han de incluir en los documentos XML.

Con todo ello, las principales aportaciones que los modelos de esquemas XML realizan frente al mecanismo anterior de las DTDs las podemos situar en las siguientes¹⁹²:

- **Utilización de una sintaxis XML:** Los esquemas XML están redactados en la sintaxis propia de XML (se emplean elementos, atributos y valores para los mismos con el fin de expresar la semántica del esquema) lo que facilita enormemente la elaboración de los mismos por parte de los usuarios al no tener ahora que aprender dos sintaxis distintas para producir definiciones de tipos documentales e instancias de documentos XML ajustadas a dichos tipos. De igual forma, al ser los esquemas XML sintácticamente idénticos a documentos XML cualquiera de las múltiples herramientas informáticas XML disponibles hoy en día en el mercado pueden ser utilizadas de forma inmediata con ellos (*parsers* XML convencionales, el DOM, SAX, XSLT o cualquiera de los navegadores preparados para trabajar con documentos XML).

¹⁹² El siguiente listado es fruto de las aportaciones realizadas al respecto de esta cuestión por diversos especialistas en la materia y recogidas en varios de los artículos y manuales ya referenciados con anterioridad. Entre estos documentos caben destacar el de N. Walsh. "Schemas for XML". *Op. cit.*, <http://www.xml.com/pub/1999/07/schemas/index.html>, el de D. Ionannides. "XML schema languages: beyond DTD". *Op. cit.*, p. 12, y el de K. Williams. "XML Databases". *Op. cit.*, pp. 148-149. De igual modo, y desde un punto de vista más práctico, resulta especialmente interesante la consulta del artículo de R. Jelliffe en el cual se establecen una serie de cuadros comparativos entre los elementos soportados por el modelo de esquema del W3C (a la postre, el más importante) frente a lo soportado por la instancia del documentos XML, por las declaraciones de marcado XML (XML DTD) y por las declaraciones de marcado SGML (SGML DTD), véase Rick Jelliffe. *The W3C Schema Specification in Context* [documento HTML]. XML.com, January 10, 2001. Disponible en <http://www.xml.com/lpt/a/2001/01/10/schemasincontext.html> (consultado el 16 de enero de 2001).

- **Gran riqueza en los tipos de datos:** Se trata de uno de los aspectos más importantes en estos nuevos modelos dado que se definen tipos de datos tan importantes como los booleanos, numéricos (enteros, decimales, reales, etc.), de fechas, de horarios, intervalos de tiempo, URIs, etc.
- **Tipos de datos definidos por el usuario:** Los esquemas XML permiten, además, aumentar la riqueza de los tipos de datos a emplear en los documentos XML posibilitando al usuario la definición tanto del nombre de nuevos tipos de datos así como el contenido y estructura de los mismos (por ejemplo, un tipo de dato denominado “CodigoPostal” con un número exacto de dígitos comprendidos dentro de un determinado rango).
- **Tratamiento y agrupación de atributos:** La definición de atributos resulta aquí más elocuente y sencilla. Además, es posible que ciertos atributos que han de ser aplicados a todos o una gran parte de los elementos puedan ser agrupados haciendo explícita dicha relación entre elementos y atributos (esta funcionalidad era obtenida en las DTDs a través de las entidades paramétricas, simplificando la tarea de creación de las mismas, pero la información no es trasladada más allá del procesador).
- **Herencia de contenidos entre elementos:** Otra característica de gran importancia en los esquemas XML es la capacidad que tienen para definir un elemento a través de la reutilización de la información ya existente en otro elemento del mismo esquema o de otro distinto. Este aspecto resulta de especial interés en el acercamiento hacia los modelos de datos relacionados.
- **Modelos de contenidos abiertos:** Además del modelo de contenido “cerrado” que se establecía en las DTD, los esquemas suelen permitir otras dos posibilidades: “abierto” (*open*) y “susceptible de ser refinado” (*refinable*). En un modelo de contenido abierto todos los elementos definidos han de estar presentes pero resulta admisible que aparezcan otros que no lo han sido previamente. Un modelo de contenido refinado o



pulido está a medio camino entre los dos anteriores, esto es, pueden existir elementos adicionales pero siempre y cuando el esquema defina qué es lo que son dichos elementos (por ejemplo, un esquema que extienda a otro, el cual puede refinar el modelo de contenido de algún tipo de elemento para añadir nuevos elementos).

- **Soporte para los XML namespaces:** La DTD daba un soporte rudimentario, y no exento de cierto truco, para los espacios locales XML lo que conlleva, las más de las veces, serios problemas de validación de los documentos XML que incluyesen este mecanismo. Con los esquemas XML se proporcionan mecanismos sencillos de soporte para la inclusión de múltiples nombres locales que, combinados de modo racional, permiten la correcta validación de documentos XML complejos.
- **Descripción semántica de los elementos:** Los esquemas XML permiten el uso de elementos de anotación para describir la función o propósito de un elemento. Como dichos elementos de descripción de contenidos de elementos están escritos en XML resulta muy sencillo para cualquier aplicación recoger dichos elementos y crear, por ejemplo, manuales de uso y aplicación de los elementos anotados.
- **En definitiva, los esquemas XML son verdaderamente extensibles:** Con el modelo de DTD resultaba paradójico que siendo extensible el propio lenguaje XML no lo fuera en sí mismo el mecanismo para la descripción de estructuras y nuevos lenguajes. Con las capacidades que tienen los esquemas XML para definir tipos de datos *ad hoc*, definir complejas estructuras, referenciar múltiples esquemas desde una única instancia de documento, reutilizar partes de esquemas en otros esquemas, así como para otro largo etcétera de posibilidades, es ahora cuando se puede hablar de un XML verdaderamente extensible.

Pero hemos venido hablando aquí de esquemas XML, o lo que es lo mismo, de diversos modelos y propuestas que han ido surgiendo desde que diversas instituciones y empresas

detectaron las limitaciones y carencias del modelo formado por la DTD, todas ellas bajo la estricta observación y consideración del XML Schema Working Group¹⁹³ del W3C.

La primera propuesta de esquema, conocida por el nombre de **XML-Data**, fue elaborada por un grupo de compañías del sector informático, estando a la cabeza de dicho grupo la todopoderosa compañía norteamericana Microsoft. Esta propuesta fue presentada al W3C y lanzada como “nota” para su estudio y debate en enero de 1998¹⁹⁴. Una versión reducida y más manejable de este modelo, conocida por las siglas de XDR (*XML-Data Reduced*), es la que ha venido incorporando la compañía Microsoft en su parser XML a partir de la versión 2 (MSXML 2) y, por tanto, soportada por el navegador IExplorer desde la versión 5. Este esquema ofrece unos potentes mecanismos para la descripción de bloques de estructura así como tipos de datos que no son posibles describir a través de las DTDs. Fuertemente inspirado en el modelo de definición del lenguaje SQL, su principal aportación radica en haber servido de gran inspirador a otros muchos modelos de esquemas que fueron apareciendo con posterioridad. Microsoft junto a otras empresas y organizaciones ha venido trabajando en la evolución y consolidación como estándar para la industria de este modelo dando lugar al marco de trabajo **BizTalk**¹⁹⁵, que tanta polémica ha levantado en

¹⁹³ La información relativa a la actividad llevada a cabo por este grupo de trabajo, al frente del cual se encuentran Dave Hollander y C.M. Sperberg-McQueen, se encuentra disponible en la dirección <http://www.w3.org/XML/Activity#schema-wg>. Igualmente importante y representativa (dada su exhaustividad y su permanente actualización) resulta la información suministrada por Robin Cover en el apartado dedicado a los esquemas XML dentro de su ya más que citado espacio web, véase <http://www.oasis-open.org/cover/schemas.html>

¹⁹⁴ World Wide Web Consortium. *XML-Data: W3C Note* [documento HTML]. Andrew Layman... [et al.] (eds.). W3C, January 5, 1998. Disponible en <http://www.w3.org/TR/1998/NOTE-XML-data/> (consultado el 20 de junio de 2000). De igual modo, se encuentra disponible en la red un manual en línea sobre este importante modelo de esquema XML en la dirección <http://msdn.microsoft.com/xml/XMLGuide/schema-overview.asp>

¹⁹⁵ Véase el sitio web en <http://www.biztalk.org/>

algunos sectores de la comunidad científica por la influyente presión hacia la adopción de modelos “normalizados” procedentes de las grandes compañías de software¹⁹⁶.

Otro modelo de gran interés fue el **Document Content Description** (DCD), propuesta lanzada al W3C en agosto de 1998¹⁹⁷ conjuntamente por las compañías IBM y Microsoft, siendo básicamente, como en el caso anterior, una versión simplificada del XML-Data pero en la cual ya se tenía en mente el modelo RDF para la descripción de metadatos (de hecho, esta propuesta fue redactada como un vocabulario RDF). Uno de los grandes avances de esta propuesta fue la capacidad para la reutilización de elementos entre diversos esquemas.

El **Schema for Object-Oriented XML** (SOX), presentado al W3C en septiembre de 1998 y llegando a ser debatida hasta una versión avanzada debido a la importancia que adquirió rápidamente entre los desarrolladores de bases de datos XML.¹⁹⁸ Esta propuesta fue desarrollada por la compañía Veo Systems (actualmente adquirida por la compañía CommerceOne), con una fuerte orientación a objetos, como su propio nombre indica.

El **Document Description Markup Language** (DDML) desarrollado por los miembros integrantes de la activa y poderosa lista de distribución XML-DEV fue presentado al consorcio en enero de 1999¹⁹⁹. A este modelo se le conoció en un principio bajo el nombre de *Xschema* pero tuvo que ser renombrado como DDML para evitar

¹⁹⁶ Sin entrar en dicho debate pues excede a las pretensiones y alcance de esta tesis doctoral, sí recomendamos, no obstante, la lectura de John Taschek. “The basic failure of XML is its premise”. *PC Week*, v. 17, n° 17, April 24, 2000, p. 61.

¹⁹⁷ World Wide Web Consortium. *Document Content Description for XML. Submission to the World Wide Web Consortium* [documento HTML]. Tim Bray, Charles Frankston, Ashok Malhotra (eds.). W3C, July 31, 1998. Disponible en <http://www.w3.org/TR/NOTE-dcd/> (consultado el 28 de noviembre de 2000).

¹⁹⁸ World Wide Web Consortium. *Schema for Object-Oriented XML 2.0. W3C Note* [documento HTML]. Andrew Davidson... [et al.]. W3C, July 30, 1999. Disponible en <http://www.w3.org/TR/NOTE-SOX/> (consultado el 28 de noviembre de 2000). La versión preliminar de esta especificación fue remitida al W3C bajo el epígrafe de “submitted” de fecha de 15 de septiembre de 1998, encontrándose disponible en la dirección <http://www.w3.org/TR/1998/NOTE-SOX-19980930/>

¹⁹⁹ World Wide Web Consortium. *Document Definition Markup Language (DDML). Specification, Versión 1.0. W3C Note* [documento HTML]. Ronald Bourret... [et al.] (eds.). W3C, January 19, 1999. Disponible en <http://www.w3.org/TR/NOTE-ddml/> (consultado el 28 de noviembre de 2000).

confusiones con otras propuestas de esquema que se estaban desarrollando en esos momentos. Este desarrollo tiene una orientación más práctica y sencilla, encaminada a proporcionar un mecanismo simple para la transformación de las antiguas DTDs al nuevo paradigma de trabajo, además de hacer posible que las aplicaciones XML interpreten y analicen rápidamente estos esquemas. Pero este modelo fue paralizado rápidamente dado que coincidió con el lanzamiento de la propuesta oficial de esquema por parte del W3C.

Otros de los últimos modelos de esquema implicados en esta incesante carrera, aunque no sometidos a la consideración del W3C, han sido el **Schematron**, desarrollado por Rick Jellife de la Academia Sinica Computing Centre en Taiwán y el **Document Structure Description** (DSD), actuación conjunta de los Laboratorios AT&T y la Universidad de Aarhus de Dinamarca; en ambos casos se aborda el problema desde una perspectiva más compleja, haciendo uso de técnicas y herramientas procedentes de las hojas de estilo²⁰⁰.

Con todo esto, la labor de consenso que ha venido llevando a cabo el W3C Schema WG desde la publicación en mayo de 1999 del primero borrador de la propuesta oficial de un esquema para XML²⁰¹ no ha resultado nada sencilla dado los múltiples intereses particulares

²⁰⁰ Existen, además de éstas, otras tantas propuestas de interés, como RELAX (*Regular Language Description for XML*) o TREX (*Tree Regular Expressions for XML*), pero que no serán tratadas aquí por no alargar en exceso este punto del capítulo, emplazando al lector interesado en estos últimos desarrollos fuera del dominio normativo del W3C a la consulta del artículo de S. St. Laurent. "Describing your data..." *Op. cit.*, <http://www.xml.com/pub/1999/12/dtd/index.html> y al recurso de R. Cover sobre esquemas XML en <http://www.oasis-open.org/cover/schemas.html>

²⁰¹ La primera propuesta del grupo de trabajo del W3C fue lanzada como "Working Draft" en mayo de 1999, dividida en dos partes (estructuras y tipos de datos):

World Wide Web Consortium. *XML Schema Part 1: Structures. W3C Working Draft* [documento HTML]. David Beech... [et al.] (eds.). W3C, May 6, 1999. Disponible en <http://www.w3.org/1999/05/06-xmlschema-1/> (consultado el 28 de noviembre de 2000).

World Wide Web Consortium. *XML Schema Part 2: Datatypes. W3C Working Draft* [documento HTML]. Paul V. Biron, Ashok Malhotra [et al.] (eds.). W3C, May 6, 1999. Disponible en <http://www.w3.org/1999/05/06-xmlschema-2/> (consultado el 28 de noviembre de 2000).

En febrero de 2000 aparecería el primer borrador de trabajo de lo que hoy en día constituye la parte preliminar y descriptiva (no normativa) de esta recomendación (de ahí que lleve la numeración cero):

de las diferentes compañías implicadas en este asunto, así como las diversas formas, orientaciones y modelos existen en la actualidad para el control y procesamiento de la información electrónica. La propuesta del Consorcio, conocida por el nombre de **XML Schema Definition Language** (XSD), o simplemente por el genérico de *XML Schema*, es desde fechas recientes (mayo de 2001) una realidad normativa al ser reconocida como Recomendación del W3C, hecho éste con el cual esperamos y confiamos muchos de los investigadores interesados en este campo que se imponga un poco de orden y cordura en este agitado entorno de intereses múltiples.

III.4.3. EL MODELO DE ESQUEMA XML DEL W3C

Como ya se comentó con anterioridad en el apartado dedicado a los estándares acompañantes, la especificación del modelo XSD se estructura formalmente en tres partes.

La primera de ellas, la Parte 0, es un texto no normativo pues se trata tan sólo de un extenso documento (cinco secciones más un apartado de apéndices) en el cual se facilita la comprensión y forma de uso a través de numerosos ejemplos de todo lo expuesto en las otras partes (estructuras y tipos de datos) de la Recomendación.

La segunda parte (numerada como Parte 1) está dedicada a la descripción del lenguaje de definición de esquemas así como los mecanismos desarrollados por esta Recomendación para la definición de estructuras en los documentos XML estructurando para ello su contenido en cinco capítulos más una serie de apéndices. Esta segunda parte de la Recomendación hace especial hincapié en el empleo y facilidades aportadas por los XML Namespaces en toda esta labor y en los procesos para la validación de estos esquemas.

Pero es sin duda la tercer parte de la Recomendación (numerada como Parte 2) la más interesante y la que aporta las principales novedades de este modelo del W3C para la definición de vocabularios y estructuras en documentos XML al estar dedicada por completo a la definición de los tipos de datos a emplear en los esquemas XML así como en otras especificaciones XML que puedan hacer uso de ello, como son, y de forma más significativa, los casos de XSL o RDF.

Gracias al afán con el que el W3C ha tratado de dar cobertura técnica a todos los tipos de datos que suelen emplearse en las aplicaciones, lenguajes y sistemas habituales de bases de datos, hace que el conjunto resultante, como es señalado por K. Williams, sea la principal ventaja del *XML Schema* frente al modelo anterior de la DTD²⁰². En la propia especificación se cita de forma expresa que este modelo propuesto debe ser capaz de:

²⁰² Kevin Williams. *Soapbox: Why XML Schema beats DTDs hands-down for data: A look at some data features of XML Schema* [documento HTML]. IBM DeveloperWorks, June 2001. Disponible en <http://www-106.ibm.com/developerworks/xml/library/x-sbsch.html> (consultado el 12 de julio de 2001).

- Dar soporte a tipos de datos primitivos, incluyendo los tipos byte, fecha, entero, secuencia, los propios de los lenguajes SQL y de Java, entre otros muchos.
- Definir un tipo de sistema que sea adecuado para la importación y la exportación de datos procedentes de los diversos modelos de bases de datos existentes hoy en día (relacional, objetos, OLAP, etc.).
- Diferenciar los mecanismos asociados a la representación léxica de los datos frente a aquellos otros que controlan el conjunto de información subyacente.
- Permitir a los usuarios la creación de nuevos tipos de datos, como pueden ser aquellos que se derivan de tipos de datos existentes a los cuales se les pueden aplicar ciertas restricciones en sus propiedades (por ejemplo, alcance, precisión, longitud, formato, etc.).

Antes de entrar en detalles con la sintaxis de construcción de esquemas XML del W3C parece oportuno reflejar mediante un ejemplo práctico las diferencias existentes entre el modelo de DTD de XML y este nuevo modelo de esquema para que, de este modo, se comprenda de forma rápida y visual lo que posteriormente se realizará a través de comentarios escritos. Utilizaremos para ello un ejemplo de una entrada de datos para un catálogo básico de libros marcados en XML donde, por un lado, se genera una DTD sencilla para definir la estructura y el contenido de los elementos y, por otro lado, se genera un esquema a través del modelo de esquema XML del W3C. Este último modelo resulta más largo y complejo pero añade el beneficio de una mayor potencialidad y exactitud a la hora de definir contenidos de datos concretos (como es el caso, por ejemplo, del elemento “ISBN”, del elemento “Fecha” o del elemento “Precio”) de la siguiente instancia de documentos XML:

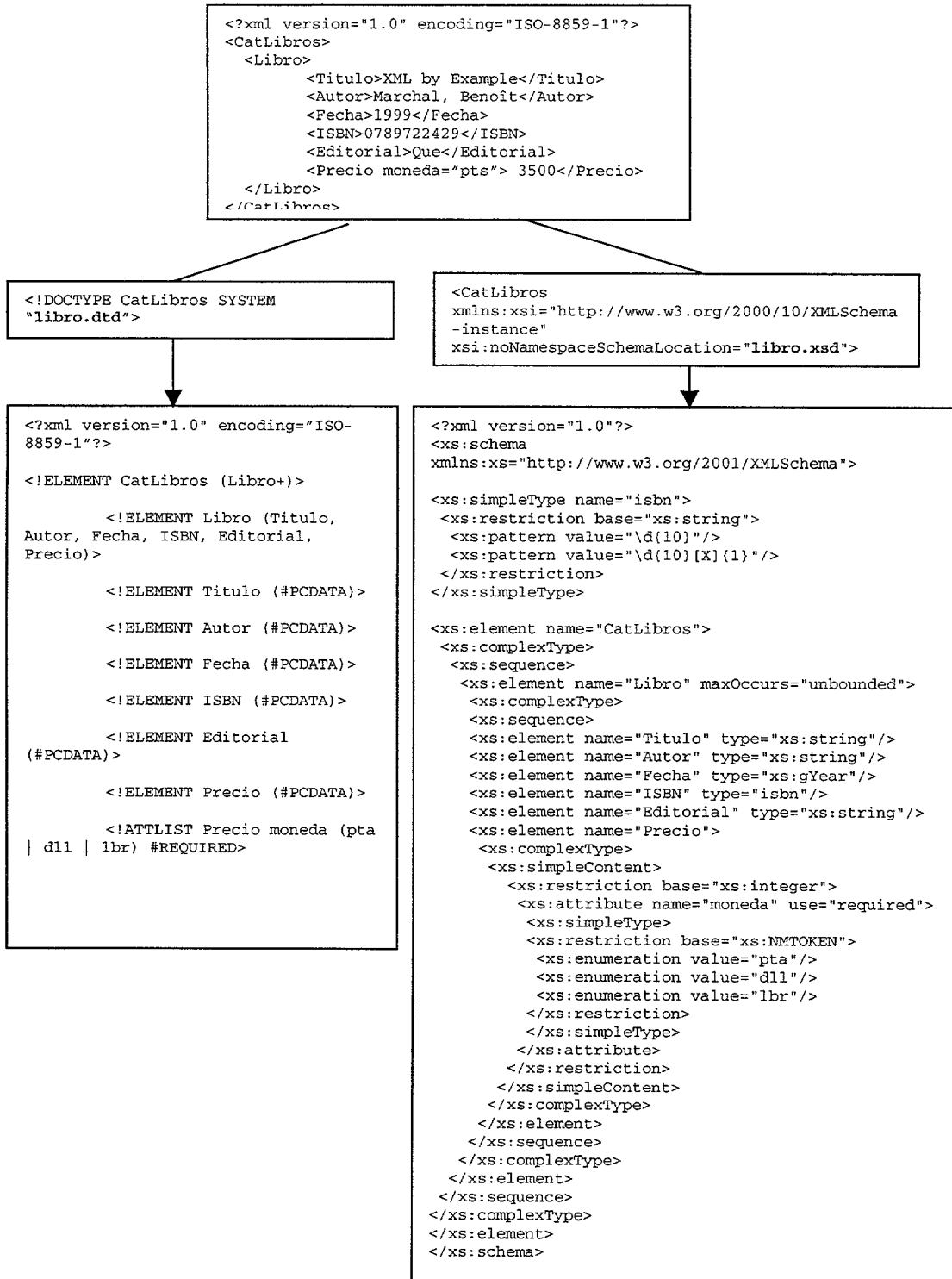


Figura III.4: Documento XML definido mediante el modelo de DTD frente al modelo de Esquema XML.

No deseamos entrar aquí en un examen y descripción exhaustiva de la especificación del modelo de esquema elaborado por el W3C dado que la complejidad del mismo (tan sólo la Parte 2, dedicada a los tipos de datos, tiene una extensión de más de 150 páginas impresas) conllevaría a una excesiva amplitud de este capítulo, amén de ser detallado algunos aspectos de dicha recomendación en el modelo aplicado que se desarrolla en el siguiente capítulo de esta tesis doctoral. Es por ello, que apuntaremos mínimamente algunos de los aspectos más característicos y sobresalientes de la sintaxis y construcción de esquemas XML derivados del modelo aprobado por el W3C²⁰³:

- Todos los ficheros electrónicos que incluyen un esquema XML del W3C llevan la extensión “.xsd”.
- Todo esquema contiene un preámbulo seguido de una o más declaraciones. Dicho preámbulo está compuesto por la instrucción de procesamiento habitual en los documentos XML junto con la etiqueta de inicio `<xs:schema>`, la cual sirve de base para la declaración del *namespace* del modelo de esquema que se está empleando (el del W3C, en este caso). Como es de suponer, el documento XML que define al tendrá como última etiqueta la correspondiente `</xs:schema>` final; por ejemplo:

```
<?xml version="1.0"?>
  <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
    .
    .
  </xs:schema>
```

²⁰³ Además de los recursos ya citados en este apartado, existen otros cuantos documentos de interés para profundizar en el estudio y elaboración de esquemas XML a los cuales remitimos para una mayor profundidad en la aplicación práctica de este complejo estándar; éstos son:

- Kurt Cagle... [et al.]. *Professional XML Schemas*. Birmingham: Wrox Press, 2001.
- Roger L. Costelo. *XML Schemas: Best Practices (A Collectively Develop Set of Schema Design Guidelines)* [documento HTML]. Xfront.com, March 12, 2001. Disponible en <http://www.xfront.com/BestPracticesHomepage.html> (consultado el 28 de marzo de 2001).
- Eric van der Vlist. *Using XML Schemas (Part I)* [documento HTML]. XML.com, November 29, 2000. Disponible en <http://www.xml.com/lpt/a/2000/11/29/schemas/part1.html> (consultado el 20 de diciembre de 2000).
- Eric van der Vlist. *Using XML Schemas (Part II)* [documento HTML]. XML.com, December 13, 2000. Disponible en <http://www.xml.com/lpt/a/2000/11/29/schemas/part1.html> (consultado el 20 de diciembre de 2000).

- La declaración de los elementos se realiza mediante la etiqueta `<xs:element>` y se establece en términos de tipos base, los cuales pueden incluir contenido mixto, tipos de datos o simplemente otros elementos. Los tipos de datos pueden extenderse o restringirse a través de referencias a un tipo base dentro de una jerarquía orientada a objetos. Esta declaración de elemento puede hacer uso de una gran número de atributos para caracterizarlo (*abstract*, *block*, *default*, *final*, *fixed*, *form*, *id*, *minOccurs*, *maxOccurs*, *name*, *nillable*, *ref*, *substitutionGroup*, *type*) donde los dos más comunes suelen ser las referencias al nombre del elemento y al tipo de datos contenido; por ejemplo:

```
<xs:element name="Titulo" type="xs:string">
```

- La declaración de los atributos de un elemento se realiza mediante la etiqueta `<xs:attribute>` y de forma idéntica al proceso anterior para los elementos, incluyendo igualmente una serie de atributos característicos (*default*, *fixed*, *form*, *id*, *name*, *ref*, *type*, *use*), donde, al igual que ocurría en el caso anterior, los más comunes son las referencias al nombre del atributo y al tipo de datos contenido; por ejemplo:

```
<xs:attribute name="clase" type="xs:string"
use="required">
```

- La utilización de un atributo que tenga una lista de valores posibles se realiza desde la base de una restricción del tipo de datos contenido junto con la enumeración de cada uno de los posibles valores admitidos a través del elementos `<xs:enumeration>`; por ejemplo:

```
<xs:attribute name="moneda" use="required">
  <xs:simpleType>
    <xs:restriction base="xs:NMTOKEN">
      <xs:enumeration value="pta"/>
      <xs:enumeration value="dll"/>
      <xs:enumeration value="lbr"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
```

- La frecuencia de aparición de un elemento viene determinada por los valores que pueden tomar los atributos *minOccurs* y *maxOccurs*, expresados estos valores con

números enteros y con la palabra *unbounded* para indicar que puede aparecer un número indeterminado de veces. Si no se incluye información al respecto de alguno de estos atributos el valor por defecto será el de 1; por ejemplo:

```
<xs:element name="Libro" maxOccurs="unbounded">
```

- Cuando un elemento contiene a otros elementos y/o atributos se dice que es de tipo complejo, haciendo uso para ello de la definición de tipo complejo `<xs:complexType>`; por ejemplo:

Un elemento con subelementos:

```
<xs:element name="capitulo">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="sección"
type="xs:string">
        <xs:element name="subsección" type="xs:string">
      </xs:sequence>
    </xs:complexType>
  </xs:element>
```

Un elemento con atributos:

```
<xs:element name="parrafo">
  <xs:complexType>
    <xs:attribute name="clase" type="xs:string">
  </xs:complexType>
</xs:element>
```

- Para determinar el orden de aparición de los subelementos definidos en el modelo de contenido de un elemento se pueden emplear las etiquetas `<xs:choice>`, para la elección de alguno de los subelementos, pero no todos, que puedan aparecer cualquier orden (correspondería al signo de la barra vertical en el modelo de DTD de SGML/XML), `<xs:sequence>`, para la entrada consecutiva en el orden establecido de los subelementos (correspondería al signo de la coma en el modelo de DTD de SGML/XML) y `<xs:all>`, para indicar que todos ellos han de aparecer aunque en cualquier orden (correspondería al signo del *ampersand* de la DTD de SGML -en la DTD de XML no existía esta posibilidad-); por ejemplo:

```
<xs:element name="persona">
  <xs:complexType>
    <xs:all>
      <xs:element name="nombre"
type="xs:string">
      <xs:element name="apellidos" type="xs:string">
    </xs:all>
  </xs:complexType>
</xs:element>
```

- Cuando el modelo de contenido de un elemento ha de ser declarado con un contenido mixto (texto junto con otros subelementos) se hace uso del atributo *mixed* con valor *true* dentro de la etiqueta que define el contenido complejo del elemento; por ejemplo:

```
<xs:element name="capitulo">
  <xs:complexType mixed="true">
    <xs:sequence>
      <xs:element name="sección"
type="xs:string">
      <xs:element name="subsección" type="xs:string">
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

- Los elementos y los atributos pueden ser agrupados en una definición de grupo, bien a través de la etiqueta `<xs:group>` para el primer caso o bien, a través de la etiqueta `<xs:attributeGroup>` para el segundo caso, con el fin de poder ser referenciados en un modelo de contenido de algún elemento; por ejemplo:

```
<xs:attributeGroup name="persona">
  <xs:attribute name="nombre" type="xs:string"
use="required">
  <xs:attribute name="apellidos" type="xs:string"
use="required">
  <xs:attribute name="fechaNac" type="xs:date"
use="optional">
</xs:attributeGroup>
```

En el modelo de contenido del elemento:

```
<xs:element name="usuario">
  </xs:complexType>
  <xs:attributeGroup ref="persona">
  </xs:complexType>
</xs:element>
```

- Si un esquema XML es un documento XML bien formado, es posible incluir comentarios utilizando la sintaxis apropiada para ello (`<!-- .... -->`). Pero, además, se implementa un mecanismo para el establecimiento de anotaciones que han de ser preservadas y analizadas sintácticamente, añadiendo de este modo un mayor contenido semántico a los esquemas. Las anotaciones hacen uso de la etiqueta `<xs:annotation>`, que incluye la información necesaria bien dirigida a los usuarios, utilizando para ello la etiqueta `<xs:documentation>`, o bien para ser procesada por aplicaciones informáticas externas, utilizando para ello la etiqueta `<xs:appInfo>`; por ejemplo:

```
<xs:element name="Autor" type="xs:string">
  <xs:annotation>
    <xs:documentation>apellidos y nombre separados
por una coma</xs:documentation>
  </xs:annotation>
</xs:element>
```

- El modelo de esquema XML proporciona dos grupos básicos de tipos de datos: los tipos de datos primitivos (*primitive datatypes*) y los tipos de datos derivados (*derived datatypes*). Los tipos de datos primitivos son aquellos que nos son definidos basándose en otros tipos de datos; existen *ab initio* (por ejemplo, el tipo de dato *float* es un concepto matemático que no puede ser definido basándose en ningún otro tipo de dato existente). Los tipos de datos derivados son aquellos que están definidos a través de otros tipos previamente establecidos en la Recomendación (por ejemplo, el tipo de dato *integer* es un caso especial del tipo más general *decimal*).

El siguiente gráfico, integrado en la Parte 2 de la especificación, ilustra y enumera perfectamente los posibles tipos de datos (primitivos y derivados) existentes en este modelo:

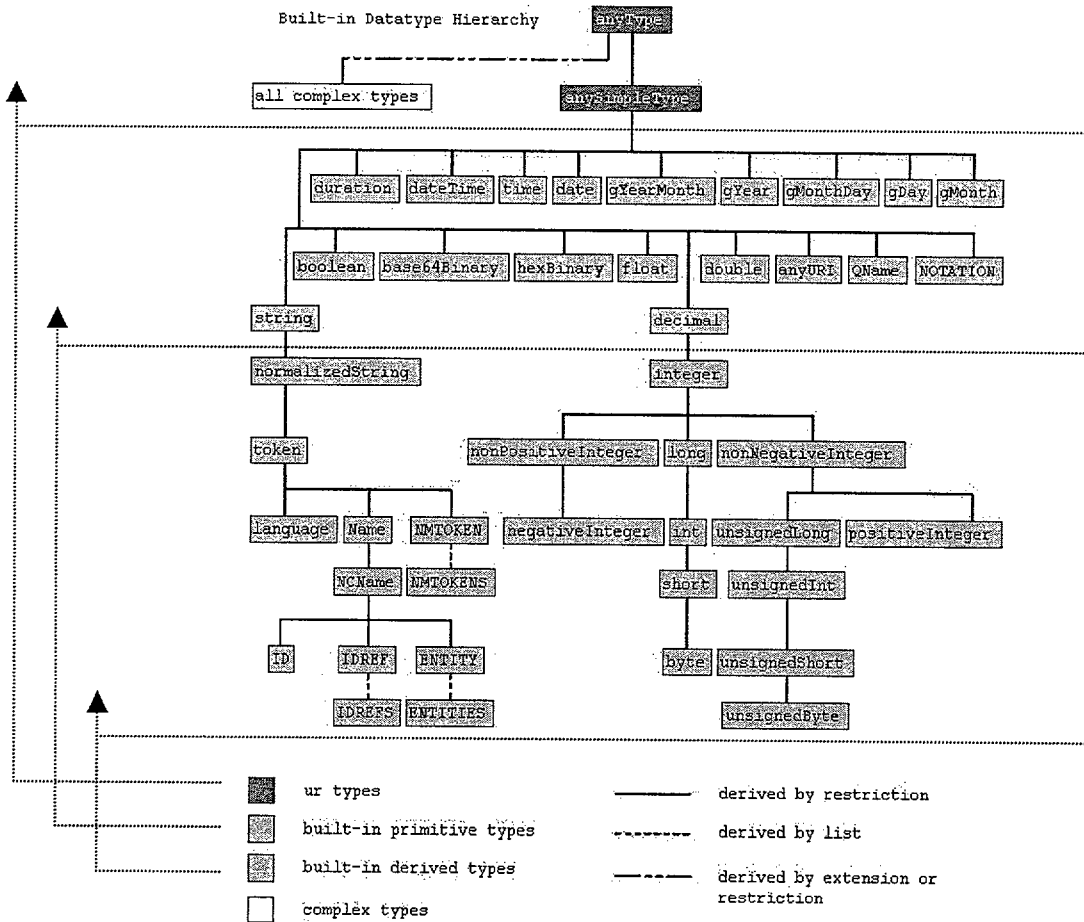


Figura III.5: Tipos de datos en el modelo de Esquema XML.

Fuente: W3C. "XML Schema Part 2: Datatypes". <http://www.w3.org/TR/xmlschema-2/>

Los valores que cada uno de estos tipos de datos (tanto primitivos como derivados) puede tomar se describe brevemente en los siguientes cuadros:

TIPOS DE DATOS PRIMITIVOS

Tipo de dato	Descripción
string	Representa cualquier cadena de caracteres legales en XML
boolean	Un valor booleano (del álgebra de Boole), bien “true” o “false”
float	Concepto normalizado de los números reales correspondiente a una precisión única de un tipo de puntero flotante de 32 bits
decimal	Representa números decimales con una precisión arbitraria
double	Concepto normalizado de los números reales correspondiente a una precisión única de un tipo de puntero flotante de 64 bits
duration	Representa una duración o longitud de tiempo
dateTime	Representa un instante concreto de tiempo
time	Representa un instante de tiempo que se repite todos los días
date	Representa una fecha del calendario
gYearMonth	Representa un mes dentro de un año del calendario gregoriano
gYear	Representa un año del calendario gregoriano
gMonthDay	Representa un día de un determinado mes del calendario gregoriano
gDay	Representa cualquier día del calendario gregoriano
gMonth	Representa cualquier mes del calendario gregoriano
hexBinary	Representa arbitrariamente cualquier dato binario codificado en hexadecimal
base64Binary	Representa arbitrariamente cualquier dato binario codificado en Base64
anyURI	Representa una referencia a un <i>Uniform Resource Identifier</i>
QName	Representa nombres cualificados definidos en la recomendación XML 1.0
NOTATION	Representa el tipo de atributo NOTATION de la recomendación XML 1.0

TIPOS DE DATOS DERIVADOS

Tipo de dato	Descripción	Derivado de
normalizedString	Representa cadenas de caracteres normalizadas con espacios en blanco	string
Token	Representa cadenas “tokenizadas” (cadenas que no contienen caracteres de salto de línea, de tabulación, de secuencias de espacio, etc.)	normalizedString
language	Representa identificadores de lenguaje natural (idiomas)	token

NMTOKEN	Representa el tipo de atributo NMTOKEN de la recomendación XML 1.0	Token
NMTOKENS	Representa el tipo de atributo NMTOKENS de la recomendación XML 1.0	NMTOKEN
Name	Representa nombres de la recomendación XML 1.0	token
NCName	Representa nombres “no-colonizados” de la recomendación XML 1.0	Name
ID	Representa el tipo de atributo ID de la recomendación XML 1.0	NCName
IDREF	Representa el tipo de atributo IDREF de la recomendación XML 1.0	NCName
IDREFS	Representa el tipo de atributo IDREFS de la recomendación XML 1.0	IDREF
ENTITY	Representa el tipo de atributo ENTITY de la recomendación XML 1.0	NCName
ENTITIES	Representa el tipo de atributo ENTITIES de la recomendación XML 1.0	ENTITY
integer	Representa un número entero	decimal
nonPositiveInteger	Representa un número entero que va del cero al menos infinito	integer
negativeInteger	Representa un número entero que va del -1 al menos infinito	nonPositiveInteger
long	Representa un número entero comprendido entre el valor máximo de 9223372036854775807 y un mínimo de -9223372036854775807	integer
int	Representa un número entero comprendido entre el valor máximo de 2147483647 y un mínimo de -2147483648	long
short	Representa un número entero comprendido entre el valor máximo de 32767 y un mínimo de -32768	int
byte	Representa un número entero comprendido entre el valor máximo de 127 y un mínimo de -128	short
nonNegativeInteger	Representa un número entero que va del cero al más infinito	integer
unsignedLong	Representa un número entero que va del cero al 18446744073709551615	nonNegativeInteger
unsignedInt	Representa un número entero que va del cero al 4294967295	unsignedLong
unsignedShort	Representa un número entero que va del cero al 65535	unsignedInt

unsignedByte	Representa un número entero que va del cero al 255	unsignedShort
positiveInteger	Representa un número entero que va del 1 al más infinito	nonNegativeInteger

Finalmente, como ya se explicó con anterioridad, el modelo de esquema XML del W3C permite al usuario crear sus propios tipos de datos haciendo uso para ello de la etiqueta `<xs:simpleType>`. Estos nuevos tipos de datos se definen a través de un tipo de datos ya existente (a éste se le denomina tipo “base”) y especificando valores para uno o más de las facetas opcionales del tipo base (por ejemplo, el tipo base *string* tiene seis facetas opcionales: *pattern*, *enumeration*, *length*, *minLength*, *maxLength*, *whitespace* –con valores: *preserve*, *replace* y *collapse*–); por ejemplo, para definir un tipo de datos nuevo para el elemento ISBN, cuyo valor máximo no valor debe tener exactamente 10 caracteres²⁰⁴, entonces:

```
<xs:simpleType name="isbn">
  <xs:restriction base="xs:string">
    <xs:pattern value="\d{10}" />
    <xs:pattern value="\d{9}[X]{1}" />
  </xs:restriction>
</xs:simpleType>
```

Este nuevo tipo de datos creado puede ser referenciado con posterioridad en el elemento o atributo que haga uso de él, como figura a continuación:

```
<xs:element name="ISBN" type="isbn" />
```

²⁰⁴ Obsérvese que se ha puesto como valor de base para este nuevo tipo de datos el tipo primitivo de “xs:string” (cadena de caracteres) y no el tipo derivado “xs:integer” (números enteros) dado que, como todos sabemos, en el valor de un ISBN pueden aparecer datos de carácter y no sólo números (como es el caso de la x mayúscula al final de la secuencia de números). Por dicho motivo se han creado dos posibles “pattern” (patrones): bien que el ISBN esté compuesto por completo de 10 dígitos o, alternativamente, de 9 dígitos y un último carácter alfabético, siempre fijo en la X. Es posible, igualmente, complicar aún más este modelo de tipo de dato para el ISBN, estableciendo diversos grupos numéricos separados por el guión característico de este código normalizado, pero ello complicaría y alargaría en exceso este apartado de la tesis.

III.4.4. PROGRAMAS ANALIZADORES DE ESQUEMAS XML

Para finalizar este apartado dedicado al modelo de esquema XML del W3C resulta necesario comentar que el número de programas y herramientas informáticas que soportan esta recomendación es aún pequeño. Este hecho se debe en gran medida a que este modelo no ha sido establecido como Recomendación por parte del W3C hasta fechas recientes (mayo de 2001) por lo que, y a fecha de redacción de esta tesis doctoral, la mayoría de las herramientas que trabajan con este modelo de esquema no ofrecen todavía soporte para la versión definitiva y sí para versiones anteriores del modelo de esquema XML del W3C (principalmente de la Recomendación Candidata). Algunas de las herramientas más importantes existen hoy en día son las siguientes²⁰⁵:

A. Validadores de esquema XML exclusivamente en línea de comando:

- **XSV (XML Schema Validator)**²⁰⁶: Desarrollado por la Universidad de Edimburgo bajo la dirección de Henry S. Thompson (uno de los editores de la recomendación del W3C) y Richard Tobin, se encuentra aún en su versión beta. El programa validador de esquemas, de acceso gratuito, puede ser utilizado en línea en el servicio creado a tal fin dentro del W3C o bien, descargado para su instalación, existiendo en este caso versiones del programa para plataforma Windows o para plataforma Pitón. La versión beta de este programa trabaja con la versión del XML Schema de 16 de marzo de 2001,

²⁰⁵ Este listado está basado en el artículo de Eric van de Vlist. *W3C XML Schema Tools Guide* [documento HTML]. XML.com, December 13, 2000. Disponible en <http://www.xml.com/lpt/a/2000/12/13/schematools.html> (consultado el 20 de diciembre de 2000). Pero dado que este artículo resulta ya algo antiguo y desfasado recomendamos que para una completa actualización de la información relativa a este tema se consulte el apartado relativo a las herramientas informáticas (*tools*) del sitio web oficial del XML Schema W3C en <http://www.w3.org/XML/Schema>. Asimismo, resulta también de gran interés la consulta del apartado correspondiente del sitio XMLSoftware de James Tauber (<http://www.xmlsoftware.org/>).

²⁰⁶ Disponible en <http://www.w3.org/2001/03/webdata/xsv>

disponiendo también de la aplicación denominada XSU (*XML Schema Upgrade*)²⁰⁷ para la transformación automática de los esquemas XML desarrollados bajo la versión de 22 de septiembre de 2000 a la versión más actualizada de 30 de marzo de 2001.

- **Sun Multi-Schema XML Validator**²⁰⁸: Desarrollado recientemente por la compañía norteamericana Sun Microsystems (la versión previa 1 apareció en octubre de 2001), es una herramienta escrita en Java para la validación de documentos XML marcados según múltiples formatos y modelos de esquemas, algunos de los cuales ya vistos en este apartado del capítulo. Lamentablemente, no ofrece un soporte completo al modelo de esquema XML del W3C, centrándose en un subconjunto de la Parte 1 de la especificación. Además de ofrecer soporte de validación para un gran número de modelos de esquemas XML a través de su utilización en modo línea de comando es posible integrarlo como una librería para otras aplicaciones Java más potentes en el tratamiento y validación de documentos XML.

B. Validadores de esquema XML con API de programación:

- **Oracle XML Schema**²⁰⁹: Esta importante compañía de software de bases de datos ha desarrollado su propio programa validador, de descarga gratuita para su uso en investigación y desarrollo (en especial bajo su famoso programa gestor de bases de datos Oracle8i/9i). Este programa validador está disponible para las plataformas Java, C++ y C, permitiendo en todos los casos analizar esquema XML del W3C, validando

²⁰⁷ Disponible en <http://www.w3.org/2001/03/webdata/xsu>

²⁰⁸ Disponible en <http://www.sun.com/software/xml/developers/multischema/>

²⁰⁹ La información oficial relativa a esta herramienta de validación, y para cada una de las plataformas disponibles, se encuentra en las siguientes direcciones:

- Java: http://technet.oracle.com/tech/xml/schema_java/index.htm
- C++: http://technet.oracle.com/tech/xml/schema_cpp/index.htm

las instancias de los documentos que hacen uso de los esquemas y suministrando información sobre la estructura de los datos contenidos en los documentos. Este validador trabaja con la versión del XML Schema de fecha de 24 de octubre de 2000.

- **Xerces**²¹⁰: Desarrollado por la prestigiosa organización sin ánimo de lucro Apache, una de las más activas e importantes en el desarrollo e implantación de aplicaciones XML al entorno de trabajo de la Web a través del proyecto Cocoon, del cual se hablará con profundidad en el apartado IV.2.5.1, relativo a la descripción del servidor Web Apache/Cocoon utilizado en el desarrollo práctico de esta tesis. Creado en un principio para la plataforma Java, existen hoy en día versiones gratuitas para la práctica totalidad de plataformas habituales en el mercado. La última versión de este validador trabaja con la especificación de fecha de 24 de octubre de 2000.

- **MSXML4.0**²¹¹: La poderosa compañía norteamericana de software Microsoft ha dado soporte al modelo de esquema XML del W3C a partir de la versión 4 de su *parser* XML (las versiones anteriores sólo validaban esquemas XDR –*XML-Data Reduced*–, modelo de esquema XML, como ya se comentó con anterioridad, de los primeros en aparecer). Con el lanzamiento de esta herramienta de validación esta compañía posibilita que muchos de sus productos capaces de trabajar con documentos XML (Internet Explorer 5.0 y superiores, Office 2000, SQL Server 2000, etc.) puedan trabajar con la especificación del XSD, además de incluir otras funcionalidades adicionales como las extensiones XPath dentro de este modelo de esquema y la incorporación de dos nuevas interfaces para la gestión de *namespaces*. La versión 4.0 del *parser* de Microsoft da soporte a la Recomendación Propuesta de fecha de 16 de marzo de 2001.

- C: http://technet.oracle.com/tech/xml/schema_cpp/index.htm

²¹⁰ Información disponible en <http://xml.apache.org/xerces-j/index.html>

²¹¹ Información disponible en <http://msdn.microsoft.com/code/sample.asp?url=/msdn-files/027/001/594/msdncompositedoc.xml>

C. Validadores de esquema XML con interfaz gráfica (GUI):

- **XML Spy 4.1**²¹²: El editor gráfico de documentos XML de la compañía comercial Altova para la plataforma Windows da soporte a la práctica totalidad de modelos de esquemas XML desarrollados hasta el momento, incluyendo, lógicamente, la edición y validación de esquemas XML del W3C (en su versión definitiva de recomendación del W3C de 2 de mayo de 2001). Además de trabajar en un entorno gráfico, con las facilidades que ello reporta para los usuarios, este programa ofrece la posibilidad de generar una completa documentación (gráficos y comentarios) del esquema XML elaborado. Igualmente, este programa incluye un generador gráfico de documentos XSLT que hace uso de lo especificado en el esquema definido previamente para facilitar la transformación del documento XML al formato HTML. Se trata de un programa comercial aunque puede descargarse una versión de evaluación de 30 días de validez. Es, sin duda, uno de los mejores y más completos editores de documentos XML. Ésta será la herramienta que fundamentalmente se empleará para el desarrollo práctico final, como se expondrá en la siguiente Parte.
- **XML Authority**²¹³: Editor gráfico para esquemas XML de la compañía comercial Tibco Extensibility está desarrollado en Java pero pensado para trabajar en diversas plataformas (Windows y Unix, principalmente). Al igual que en el anterior caso, permite trabajar con diversos modelos de esquemas XML siendo, asimismo, capaz de generar un documento en el cual se muestra gráficamente la estructura de jerarquía de los elementos existentes en el esquema XML. Esta aplicación, aunque independiente en su funcionalidad, se suele integrar en un paquete informático mayor denominado Turbo XML²¹⁴, el cual incluye además de esta aplicación otras orientadas a la edición de instancias de documentos XML (*XML Instance*) así como una consola para el control y

²¹² Información disponible en <http://www.xmlspy.com/>

²¹³ Información disponible en http://www.extensibility.com/solutions/xml_authority/index.htm

gestión de sitios XML (*XML Console*). En todos estos casos, se trata de software comercial con versiones de evaluación por un periodo de 30 días. Este importante producto informático trabaja ya con la versión final de la recomendación de esquema XML del W3C, así como con otros modelos de esquemas descritos con anterioridad (DTD, XDR, SOX y BizTalk).

²¹⁴ Información disponible en http://www.extensibility.com/solutions/turbo_xml/index.htm

CAPÍTULO III.5

DEFINICIÓN Y ESTABLECIMIENTO DE METADATOS EN LOS DOCUMENTOS XML

III.5.1. LOS METADATOS EN EL ENTORNO DOCUMENTAL DE LA WEB

El término **metadatos**, tan de moda en nuestros días, es descrito habitualmente, y de forma ciertamente simplista, como “datos sobre los datos” o “información que describe datos”. Pero más allá de esta sencilla definición, el concepto de metadatos (o también denominado a veces bajo el apelativo de “metainformación”) conlleva asociadas otras cualidades que le hacen ser en la actualidad uno de los focos de mayor atención en el desarrollo sistemas de almacenamiento, búsqueda e intercambio de información en el espacio electrónico de la *World Wide Web*, como se explicará a lo largo de este punto.

En el contexto de las bibliotecas, archivos y centros de documentación tradicionales la labor de informar sobre los datos contenidos en los documentos impresos ha venido llevándose a cabo a través de la aplicación de las diversas técnicas catalográficas existentes. Así, en este contexto, metadatos y catalogación tienen una significación semántica idéntica, siendo señalado por autores como Caplan, Miller y Smits, pues en ambos casos se está hablando de mecanismos que nos ayudan a la localización y selección de fuentes de información²¹⁵. Pero con la incorporación de las tecnologías ópticas e informáticas a este entorno profesional y social de difusión de la información y el conocimiento para la creación y desarrollo de colecciones digitales, las llamadas *bibliotecas digitales*, junto con la

²¹⁵ A este respecto P. Caplan, uno de los primeros autores en hablar del término “metadatos”, afirma que un registro de catálogo son metadatos, lo mismo que una cabecera TEI o cualquier otra forma de descripción, pero prefiriendo emplear este término frente a la palabra “catalogación” dado que es un término neutro que cubre todas las bases y supuestos posibles. Miller señala igualmente que los metadatos son datos sobre los datos y que, por tanto, proporcionan información básica sobre el autor, título, fecha de creación, etc., del mismo modo que lo hace una ficha catalográfica de una biblioteca. De idéntica forma se manifiesta Smits al indicar que una primera instancia, los metadatos proporcionan información de forma similar a como lo hacen las descripciones bibliográficas en los catálogos de las bibliotecas. Para una mayor información sobre esta idea, véase, respectivamente, Priscilla Caplan. “You Call It Corn, We Call It Syntax-Independent Metadata for Documents-Like Objects” [documento HTML]. *The Public Access Computer Systems Review*, v. 6, n° 4, 1995. Disponible en <http://info.lib.uh.edu/pr/v6/n4/cap16n4.html> (consultado el 5 de octubre de 2000), Paul Miller. “Metadata for the Masses” [documento HTML]. *Ariadne*, n° 5, September 1996. Disponible en

aparición y expansión del fenómeno de la WWW, como espacio electrónico de masas para la publicación, la difusión y la consulta de información electrónica de todo tipo, hace que dicho paralelismo no resulte tan claro ni exacto.

En el primer caso mencionado, la digitalización de documentos impresos o, y de forma más precisa, la elaboración directa de documentos electrónicos *ex novo* en el espacio de gestión documental de múltiples organizaciones de todo tipo hizo necesaria la búsqueda de mecanismos de descripción normalizada de contenidos capaces de dar una respuesta satisfactoria a la hora de localizar y recuperar dichos documentos electrónicos en este espacio finito de publicación que constituían las redes informáticas corporativas. Este mismo hecho se extiende a las bibliotecas digitales pues, como señala V. Ortiz-Repiso, con el empleo de la tecnología digital en estos espacios se ven alteradas la propia organización de los objetos (de objetos físicos se pasa a objetos lógicos), el acceso a la colección (de un acceso físico y único se pasa a un acceso lógico y múltiple) y el entorno de la meta-información de la biblioteca (el objeto digital permite el uso de esa misma tecnología para extraer la información contenida)²¹⁶.

Si el empleo de lenguajes de marcado descriptivo, derivados en un primer momento y en su gran mayoría del metalenguaje SGML, resultaban útiles para la descripción y definición formal de las estructuras lógicas de construcción de documentos electrónicos de igual forma estos lenguajes de marcado podrían ser empleados para la descripción normalizada del contenido semántico de dichos documentos. Esta idea fue la que recogieron aplicaciones (o DTDs) SGML tan importantes como TEI (*Text Encoding Initiative*), donde, como se explicó en el apartado III.2.3.2 de esta tesis, cada documento TEI lleva una cabecera conteniendo toda la información bibliográfica necesaria para describir el texto

<http://www.ariadne.ac.uk/issue5/metadata-masses/intro.html> (consultado el 5 de octubre de 2000) y Jan Smits. "Metadata: An Introduction". *Cataloging & Classification Quarterly*, v. 27, n° 3/4, 1999, p. 305.

²¹⁶ Virginia Ortiz-Repiso Jiménez. "Nuevas perspectivas para la catalogación: metadatos versus MARC". *Revista Española de Documentación Científica*, v. 22, n° 2, 1999, p. 199.

marcado, o EAD (*Encoded Archival Description*) para la descripción de estructuras y datos de los materiales documentales propios de los archivos²¹⁷.

Pero, sin duda, ha sido en el contexto de publicación del espacio electrónico de la WWW de la red Internet donde el término de *metadatos* ha tomado toda su razón de ser debido a la extrema importancia que adquieren en este entorno cuestiones referentes a la identificación, descripción y localización de los documentos Web²¹⁸. En este mega-espacio virtual repleto de objetos electrónicos escasamente descritos lo metadatos, como señala el propio Berners-Lee, están creados para el consumo de los ordenadores y no de las personas, con la finalidad de que éstos interpreten una serie de informaciones descriptivas sobre los objetos y recursos web²¹⁹. La Web está llena de innumerables documentos electrónicos sobre cualquier temática o disciplina del saber humano. Se hace, por tanto, evidente la necesidad de elaborar potentes mecanismos de descripción de contenidos para que las herramientas informáticas de búsqueda de información (*search engine*, en la terminología anglosajona) que sean capaces de localizar de forma precisa y exacta la información demandada por el usuario.

²¹⁷ Una fuente de gran interés para conocer el modo en el que estas aplicaciones SGML han venido haciendo uso de los metadatos se encuentra en la documentación generada dentro del proyecto BIBLINK de la agencia británica UKOLN (*UK Office for Library and Information Networking*). Este importante proyecto, desarrollado entre abril de 1996 y febrero de 2000 y financiado con fondos procedentes del Programa de Aplicaciones Telemáticas de la Comisión Europea, tuvo como objetivo principal el establecimiento de relaciones entre las agencias bibliográficas nacionales y los editores europeos para el estudio del mejor método para la normalización en la descripción bibliográfica de diversos materiales con el fin de confeccionar catálogos y otras fuentes secundarias de información de interés para ambos sectores. Para llevar a cabo este objetivo se hizo, pues, necesario estudiar algunos de los estándares existentes en aquel momento para la elaboración y descripción de documentos electrónicos basados en el metalenguaje SGML. Recomendamos, pues, la consulta del documento de Lou Burnard, Richard Light. *Three SGML metadata formats: TEI, EAD, and CIMI: A Study for BIBLINK Work Package 1.1* [documento HTML]. Oxford: UKOLN Metadata Group, rev. May 14, 1998. Disponible en <http://hosted.ukoln.ac.uk/biblink/wp1/sgml/> (consultado el 17 de febrero de 2001).

²¹⁸ Un interesante artículo de divulgación sobre las tentativas que han venido surgiendo para la descripción (tanto externa como interna) y la localización de los documentos electrónicos publicados este servicio de Internet, lo constituye el elaborado por Eva M^a Méndez Rodríguez, José Antonio Merlo Vega. "Localización, identificación de documentos web: tentativas hacia la normalización". En: *Jornadas Españolas de Documentación* (7^o. 2000. Bilbao). Bilbao: Universidad Carlos III de Madrid, 2000, pp. 221-231.

²¹⁹ Tim Berners-Lee. *Metadata Architecture. Documents, Metadata, and Links* [documento HTML]. W3C, February 6, 1997. Disponible en <http://www.w3.org/DesignIssues/Metadata.html> (consultado el 10 de diciembre de 2000).

Y desde la óptica más cercana a los bibliotecarios y documentalistas, la cual atañe a esta tesis doctoral, es necesario destacar las palabras de M. Baca, al señalar que “para que los metadatos sean útiles en la construcción de un catálogo distribuido de la Web, es esencial que tanto la estructura como la sintaxis de la información se ajusten a estándares de amplia aceptación, de modo que la semántica de los registros del catálogo pueda ser correctamente interpretada por los distintos ordenadores que integran la red del catálogo distribuido”²²⁰. En este contexto resultan imprescindibles, como señala R. Swick (representante del grupo de trabajo sobre metadatos del W3C)²²¹, los metadatos para la descripción de los contenidos existentes en este espacio electrónico de comunicación.

Dos han sido los principales desarrollos llevados a cabo por el grupo de trabajo de la W3C encargado del estudio y aplicación de metadatos hasta la aparición del metalenguaje XML²²²: el sistema PICS y el elemento META en los documentos HTML.

PICS (*Platform for Internet Content Selection*)²²³ o la plataforma para la selección de contenidos en Internet fue creado con la finalidad de permitir distribución electrónica sencilla y fiable entre los ordenadores conectados a Internet de reseñas de los documentos electrónicos existentes en aquella incipiente web. PICS surgió en su inicio, como así lo atestigua P. Resnick, primer presidente de este grupo de trabajo del W3C y uno de los autores principales de la especificación, como un sistema que “facultara a los padres y a los

²²⁰ *Introducción a los Metadatos: vías a la información digital*. Murtha Baca (ed.). Los Ángeles: Getty Information Institute, 1999, p. 33.

²²¹ Toda la actividad que ha venido llevando a cabo el W3C en el campo de los metadatos puede ser consultada en la dirección <http://www.w3.org/Metadata/Activity.html>

²²² Ralph Swick es el representante de este importante grupo de trabajo de la W3C donde, como se explicará posteriormente, se ha apostado fuertemente por el desarrollo del estándar RDF como plataforma universal para la descripción de la metainformación de los documentos electrónicos. Toda la actividad de este grupo de trabajo puede ser consultada en la dirección <http://www.w3.org/Metadata/Activity.html>

²²³ Toda la información oficial relativa al desarrollo PICS del W3C puede ser consultada en la dirección <http://www.w3.org/PICS/>

educadores el bloqueo del acceso a materiales juzgados impropios para los niños”²²⁴. Este sistema utiliza un conjunto de etiquetas para la descripción de los documentos incluidos en la red (principalmente descriptivas tanto del contenido literario o científico como el contenido moral del documento o nodo web), confeccionadas bien por el propio autor del documento o bien, por otros organismos o instituciones dedicadas a tal fin, y orientadas al procesamiento por parte de programas especiales más que a su lectura por parte de las personas. Estas etiquetas pueden adjuntarse a los objetos que describen (principalmente documentos HTML para la Web) como fichero electrónicos independientes que apuntan y describen a dichos objeto o imbuidos dentro de la cabecera de los documentos HTML. Sirva como ejemplo el incluido dentro de la propia especificación de las etiquetas PICS²²⁵:

```
(PICS-1.1 "http://www.gcf.org/v2.5"
  by "John Doe"
  labels on "1994.11.05T08:15-0500"
    until "1995.12.31T23:59-0000"
    for "http://w3.org/PICS/Overview.html"
    ratings (suds 0.5 density 0 color/hue 1)
    for "http://w3.org/PICS/Underview.html"
    by "Jane Doe"
    ratings (subject 2 density 1 color/hue 1))
```

Por otra parte, el lenguaje de marcado HTML prevé desde la versión oficial 3.2 un sistema para la asignación de metadatos que caracterizan a las páginas web: la inclusión del **elemento META**, elemento vacío que se sitúa en la cabecera de las páginas web²²⁶. Como

²²⁴ Paul Resnick. “La selección de información en Internet”. *Investigación y Ciencia*, n° 248, mayo 1997, p. 54.

²²⁵ World Wide Web Consortium. *PICS Label Distribution, Label Syntax and Communication Protocols: W3C Recommendation* [documento HTML]. Jim Millar (ed.); Tim Krauskopf... [et al.]. W3C, October 31, 1996. Disponible en <http://www.w3.org/TR/REC-PICS-labels-961031.htm> (consultado el 8 de noviembre de 2000).

²²⁶ Existe otro elemento vacío que se sitúa dentro de la cabecera de los documentos HTML que transmite, en cierta forma, información sobre los datos contenidos en el documento. Se trata del elemento LINK empleado para establecer las relaciones que se dan entre dicho documento y otro y otros documentos relacionados (como es, por ejemplo, el caso de la asociación con esquemas de metadatos o con hojas de estilo externas). En este caso, dicho elemento informa más sobre cuestiones estructurales del conjunto de documentos que

señala el propio D. Raggett, autor de la especificación del HTML 3.2, el elemento META puede ser empleado para incluir pares de nombres y valores que permitan una descripción de las propiedades del documento (autor, resumen del contenido, palabras clave, etc.)²²⁷. Este elemento emplea principalmente dos atributos para llevar a cabo esta función descriptiva, el atributo NAME, el cual precisa el nombre de la propiedad descriptiva empleada, y el atributo CONTENT, que desarrolla el valor de la propiedad reseñada. Las diversas especificaciones surgidas hasta la fecha del lenguaje HTML, aunque han mejoraron notablemente las capacidades de la etiqueta META (añadiendo, por ejemplo, el atributo SCHEMA para permitir establecer tipo de metadatos más exactos y el atributo PROFILE para la localización del perfil de dichos metadatos), no han establecido formalmente cuáles habían de ser los nombres de las propiedades descriptivas a situar como valor del atributo NAME. Por ejemplo, la especificación del HTML4 establece tanto al definir la estructura global de estos documentos como en lo señalado en el Apéndice B que resulta recomendable incluir al menos las palabras claves y la descripción del recurso a través de dicha etiqueta o, por el contrario, hacer uso del marco RDF y el vocabulario Dublín Core (de los cuáles se hablará con posterioridad en este punto) para, de este modo, facilitar la búsqueda y recuperación de documentos por parte de los motores de búsqueda en la WWW.

Siendo esto así, diversas compañías y usuarios particulares han ido formalizando a través de la práctica cotidiana en la elaboración de documentos HTML una serie de nombres para dicho atributo. Los nombres más habituales que se pueden encontrar en las páginas web son los de *author*, *keywords*, *description*, *copyright*, *date*, *distribution*, *robots* o *language*, además de

sobre el significado mismo de éstos, haciendo uso para ello del atributo *rel* (relación) en combinación con el atributo *href* (referencia de hipertexto con el nombre del documento relacionado). Para establecer estas relaciones entre un documento y otro u otros del conjunto documental el atributo *rel* puede tomar valores tales como *alternate* (para designar versiones sustitutivas del documento actual), *start* (para indicar el primer documento de una colección de ellos), *next* (documento siguiente), *prev* (documento anterior), *contents* (documento que hace las funciones de una tabla de contenidos o sumario), *index* (documento que contiene el índice relativo al documento actual o conjunto de documentos), *glossary* (documento que proporciona un glosario de términos relativos al documento actual o al conjunto de documentos), *chapter* (documento que figura como un capítulo de una colección de documentos), *section* (documento que figura como una sección), *subsection* (documento que figura como una subsección), etc.

²²⁷ W3C. "HTML 3.2 Reference Specification...". *Op. cit.*, <http://www.w3.org/TR/REC-html32>

otros muchos más específicos desarrollados por entidades, compañías u organismos de diversos sectores industriales y científicos²²⁸. Sirva como ejemplo sencillo el siguiente código incluido en la especificación HTML4²²⁹:

```
<HEAD profile="http://www.acme.com/profiles/core">
  <TITLE>How to complete Memorando cover sheets</TITLE>
  <META name="author" content="John Doe">
  <META name="copyright" content="&copy; 1997 Acme
  Corp.">
  <META name="keywords" content="corporate, guidelines,
  cataloguing">
  <META name="date" content="1994-11-06T08:49:37+00:00">
</HEAD>
```

Aunque no deseamos entrar aquí en el tema de los sistemas y herramientas informáticas para la búsqueda y recuperación de la información en Internet²³⁰, dada la amplitud y complejidad de este tema y, principalmente, por escaparse a los propósitos de esta tesis doctoral, sí es necesario comentar que uno de los grandes fracasos del servicio *World Wide Web* se encuentra en este aspecto fundamental; esto es, las enormes dificultades con las que se encuentra el usuario de este servicio para la localización y recuperación de información

²²⁸ Un par de interesantes artículos divulgativos en formato electrónico sobre el empleo de las etiquetas META en los documentos HTML son el elaborado por Danny Sullivan. *How To Use HTML Meta Tags* [documento HTML]. SearchEngineWatch.com, rev. December 15, 2000. Disponible en <http://searchenginewatch.com/webmaster/meta.html> (consultado el 13 de junio de 2001), en especial por la inclusión de un gran número de referencias a otros muchos recursos de interés en la Web sobre este tema, y el de Christopher Simmons. *How to Optimize Your Web Pages for META Tags* [documento HTML]. Send2Press.com, 2000. Disponible en http://www.send2press.com/PRnetwire/articles/cs_meta-tags.shtml (consultado el 13 de junio de 2001).

²²⁹ W3C. "HTML 4 Specification...". *Op. cit.*, <http://www.w3.org/TR/html4>

²³⁰ Existe una abundante literatura científica al respecto de este tema, tanto en formato impreso como en formato electrónico. Siendo así, resulta realmente difícil hacer un extracto de aquellas obras que pueden ser recomendadas para una ampliación de contenidos en este apartado. Aún así, no nos parece aventurado destacar como obras de partida un par de aportaciones científico-literarias aparecidas en nuestro entorno académico y de investigación dado el correcto carácter divulgativo del mismo y la perfecta conjunción entre descripción y rigor científico. De este modo, es de destacar el artículo de David Green. "The evolution of Web searching". *Online Information Review*, v. 24, n° 2, 2000, pp. 124-137 así como el capítulo de Antonio Hernández Pérez. "La búsqueda y recuperación de información en Internet". En: Mercedes Caridad Sebastián (coord.). *La Sociedad de la Información: Política, Tecnología e Industria de los Contenidos*. Madrid: Centro de Estudios Ramón Areces: Universidad Carlos III de Madrid, 1999, pp. 213-240.

precisa y relevante materia de su interés. Muchos son los factores desencadenantes de esta frustración tan generalizada; a saber: la no inclusión de las etiquetas META por parte de la gran mayoría de los redactores de documentos HTML, la inclusión de información errónea en dichas etiquetas o, en otros casos, deliberadamente insertada con fines puramente publicitarios o comerciales (los famosos *spam*), el desconocimiento de gran parte de los usuarios de los diversos mecanismos existentes en muchos motores de búsqueda para el refinamiento o búsqueda avanzada de información electrónica en la Web; el bloqueo de ciertos servidores web al acceso a sus documentos electrónicos por parte de los robots o *spiders* que constantemente indagan la Web; muchos tan sólo indexan la información contenida en documentos HTML despreciando otros formatos de fichero electrónico existentes en la Web (documentos PDF, documentos PostScript, documentos MSWord, etc.)²³¹; la utilización por parte de muchos motores de búsqueda de información de algoritmos matemáticos poco potentes y precisos²³² (por ejemplo, muchos de ellos no hacen uso de la información encontrada dentro de las etiquetas META para establecer el *ranking* de relevancia de los documentos obtenidos²³³), y un largo etcétera más de factores que han hecho pensar a muchos especialistas en la materia que las herramientas

²³¹ Este aspecto, así como otros relacionados con los elementos que son analizados por los motores de búsqueda más populares en la WWW es detallado en el artículo electrónico *Search Engine Features For Webmasters* [documento HTML]. SearchEngineWatch.com, rev. March 15, 2001. Disponible en <http://www.searchenginewatch.com/webmaster/features.html> (consultado el 25 de mayo de 2001).

²³² Un interesante documento electrónico sobre el uso y funcionamiento de los motores de búsqueda en relación a todos estos aspectos técnicos de indexación de documentos y ponderación de los resultados obtenidos lo constituye el elaborado por el servicio de informática de la Universidad de Cambridge, véase en *The University's site-wide web search engine* [documento HTML]. Cambridge: University of Cambridge Computing Service, September 2000. Disponible en <http://www.cam.ac.uk/CS/web-search/> (consultado el 19 de enero de 2001).

²³³ Uno de los trabajos de mayor interés aparecidos en este año sobre los métodos empleados por los motores de búsqueda más populares en la Internet para el establecimiento del *ranking* de relevancia de documentos obtenidos, y en referencia directa al uso de las etiquetas META para llevar a cabo este proceso de baremación, lo constituye el elaborado por Wendy Lucas. "Search Engines, Relevancy, and the World Wide Web". En: Amita Goyal Chin (coord.). *Text Databases & Document Management: Theory & Practice*. Hersey [etc.]: Idea Group, 2001, pp. 23-51.

informáticas de búsqueda y recuperación de información en documentos HTML para la Web han tocado techo²³⁴.

Otras muchas iniciativas han venido surgiendo desde entonces para la descripción normalizada de la metainformación contenida en los documentos electrónicos existentes en el espacio virtual de la Web²³⁵. Dejando al margen aquellos proyectos provenientes del mundo más cercano a las actividades catalográficas propias de las bibliotecas tradicionales, como fueron la adaptación del formato MARC junto con el empleo del protocolo Z39.50 para el intercambio entre ordenadores de registros electrónicos²³⁶, cabe destacar como

²³⁴ Véase al respecto de este asunto las opiniones pronunciadas por C. Morris en su tan referenciado artículo electrónico sobre la utilidad de los actuales buscadores de información en Internet. Charlie Morris. *Are search engines dead?* [documento HTML]. Web Developer's Virtual Library, June 26, 2000. Disponible en http://www.wdvl.com/Internet/Dead_SearchEngines/ (consultado el 13 de junio de 2001).

²³⁵ Para una mayor profundización sobre los desarrollos que han venido surgiendo desde la aparición de la WWW recomendamos la consulta y lectura del, aunque ya algo antiguo y desfasado, del excelente artículo divulgativo de Judith R. Ahronheim. "Descriptive Metadata: Emerging Standards". *The Journal of Academic Librarianship*, v. 24, n° 5, 1998, pp. 395-403.

²³⁶ Sin entrar en demasiados detalles respecto a esta cuestión, sí es necesario señalar que han existido diversas propuestas desde entornos clásicos del mundo de la biblioteconomía para la utilización de las reglas de descripción de materiales impresos en bibliotecas y centros de documentación (agrupadas bajo el nombre de ISBD –*International Standard Book Description*–) así como la adaptación de los diversos formatos de codificación electrónica para este tipo de materiales (las diversas versiones nacionales del formato MARC –*Machine Readable Cataloguing*–), bien a través de la simple incorporación de un nuevo campo para señalar el URL del documento electrónico en Internet a describir (campo 856) o, de forma más compleja, a través de la generación de DTDs de SGML/XML de los diversos formatos MARC existentes. En el primer caso, la Biblioteca del Congreso de los Estados Unidos, mantenedora de este formato, dispone en su sitio web de abundante información para el correcto uso y aplicación del campo 856 para la descripción catalográfica de recursos electrónicos existentes en la Web, como es el caso, por ejemplo, del documento *Guidelines for the Use of Field 856* [documento HTML]. Washington DC: Network Development and MARC Standards Office, Library of Congress, rev. August 1999. Disponible en <http://lcweb.loc.gov/marc/856guide.html> (consultado el 24 de enero de 2001). En el segundo caso, la citada institución bibliotecaria ha venido trabajando desde mediados de la pasada década junto con otras instituciones e investigadores del mundo de las bibliotecas en la adaptación del formato MARC al lenguaje SGML y, con posterioridad, al lenguaje XML, desarrollando diversas DTDs para dicho formato de codificación e intercambio electrónico de registros catalográficos. Toda la información al respecto se encuentra disponible en las páginas web de la Biblioteca del Congreso norteamericano en la dirección <http://lcweb.loc.gov/marc/marcshtml.html>, siendo recomendable para profundizar en este aspecto la lectura del documento *MARC DTDs Document Type Definitions: Background and Development* [documento HTML]. Washington DC: Network Development and MARC Standards Office, Library of Congress, May 22, 1988. Disponible en <http://lcweb.loc.gov/marc/marcdtd/marcdtdback.html> (consultado el 24 de enero de 2001). Algo similar ha sucedido con el protocolo Z39.50 (véase información al respecto de este estándar en <http://lcweb.loc.gov/z3950/agency/>), surgido en un principio como modelo universal para el intercambio de registros bibliográficos electrónicos codificados en MARC, y sobre el cual

formato de metainformación de propósito general el estándar del Dublin Core²³⁷ y, como marco o plataforma de soporte para éste u otros sistemas de descripción normalizada de metadatos, la aplicación RDF.

El **Dublin Core Metadata Initiative (DCMI)**²³⁸ es, sin duda, la iniciativa para el establecimiento de metadatos más divulgada y aceptada entre la comunidad de especialistas dedicados a la descripción normalizada de información electrónica en el espacio Web, debido, en gran medida, a su carácter abierto y ser su ámbito de aplicación general²³⁹.

Esta importante iniciativa tiene sus orígenes en 1995 bajo un encuentro de trabajo desarrollado en Dublin, Ohio, y organizado por la prestigiosa organización OCLC (*Online Computer Library Center*), el NCSA (*National Center for Supercomputing Applications*) y la

han surgido diversas tentativas como modelo válido para el intercambio de todo tipo de documentos electrónicos existentes en el intangible mundo de la WWW. Para una mayor información sobre todos estos temas, véase la ponencia de María José Gimeno Montoro, José Manuel Barrueco Cruz, Cristina García Testal. "Catalogación de recursos electrónicos accesibles en Internet: revisión de propuestas para una normativa". En: *Jornadas Españolas de Documentación* (6ª. 1998. Valencia). Valencia: FESABID, 1998, pp. 369-376.

²³⁷ Además del Dublin Core, analizado a continuación, surgieron otras muchas iniciativas de interés tanto de propósito general como para la descripción de materiales de contenidos específicos. Sin entrar en un examen exhaustivo de ellos, dado que este tema escapa a los propósitos de esta tesis doctoral (además de estar siendo perfectamente cubierto todo ello por otras tesis doctorales que se están llevando a cabo dentro del propio Departamento de Biblioteconomía y Documentación de esta Universidad), tan sólo señalar la importancia que han tenido otros desarrollos en determinados campos específicos de actuación, como son, por ejemplo, los del FGDC (*The Federal Geographic Data Committee*) para la descripción de metainformación geográfica, cuya información se encuentra disponible en <http://www.fgdc.gov/Metadata/>, o el GILS (*Government Information Locator Service*) para la descripción de la metainformación de carácter gubernamental, cuya información se encuentra disponible en <http://www.gils.net>

²³⁸ El *Dublin Core Metadata Initiative* es, como se señala textualmente en la información oficial existe en su sitio web, "una organización sin ánimo de lucro dedicada a la promoción y la adopción generalizada de estándares de metadatos así como al desarrollo de vocabularios de metadatos para la descripción de recursos que permita la creación de sistemas inteligentes de recuperación de información". Toda la información oficial relativa a este importante desarrollo se encuentra localizada en la dirección <http://dublincore.org/>

²³⁹ Además de los documentos normativos del DCMI que se irán citando en el desarrollo de este punto, resulta de especial utilidad para los usuarios noveles en el uso y empleo del sistema del Dublin Core el documento electrónico elaborado por D. Hillmann, con carácter normativo de Recomendación oficial de este organismo. En él se explican de forma breve y clara algunas cuestiones sobre el significado del término "metadatos", el alcance del modelo del Dublin Core además de mostrarse a través de diversos ejemplos y documentos relacionados el modo correcto para su utilización dentro de los documentos HTML y XML/RDF. Véase, Diane Hillmann. *Using Dublin Core* [documento HTML]. Dublin Core Metadata Initiative, April 12, 2001. Disponible en <http://dublincore.org/documents/2001/04/12/usageguide/> (consultado el 18 de junio de 2001).

compañía de software informático Softquad. Estas instituciones invitaron a un selecto grupo de bibliotecarios, archiveros, informáticos y representantes del colectivo de empresas suministradoras de acceso a Internet para encontrar entre todos un conjunto de 13 elementos descriptivos (*DC elements*) que fueran capaces de caracterizar a los documentos existentes en la Web. Desde entonces, se han venido produciendo hasta un total de 8 seminarios o encuentros de trabajo por diversas ciudades de todo el mundo y se ha ampliado el conjunto inicial hasta un total de 15 elementos. Su éxito ha llegado a ser de tal importancia y magnitud que, como señalan S. L. Weibel, director del DCMI, y T. Koch, constituye hoy en día el estándar *de facto* para la asignación de metadatos en la Web²⁴⁰.

Cada uno de los 15 elementos incluidos en la actual **versión 1.1 del Dublin Core Metadata Element Set** (DCMES)²⁴¹ es definido según un conjunto de diez atributos procedentes del estándar ISO/IEC 11179²⁴² para la descripción de elementos de datos (*Name, Identifier, Version, Registration Authority, Language, Definition, Obligation, Datatype, Maximum Occurrence* y *Comment*). Los 15 elementos que conforman el DCMES pueden ser vistos, y en palabras de Weibel y Miller, “como el bloque semántico común de los metadatos para la Web”²⁴³. Éstos se organizan, como decíamos, en 15 categorías o elementos los cuales pueden ser representados haciendo uso de múltiples y variadas sintaxis (principalmente SGML y XML), permitiendo describir las características más generales que caracterizan a los documentos electrónicos en la Web. Este posibilita, por tanto, una búsqueda y recuperación de información más efectiva en este espacio virtual de publicación. Estos elementos son los siguientes: *Title* (el título asignado al recurso), *Creator*

²⁴⁰ Stuart L. Weibel, Traugott Koch. “The Dublin Core Metadata Initiative: Mission, Current Activities, and Future Directions” [documento HTML]. *D-Lib Magazine*, v. 6, n° 2, December 2000. Disponible en <http://www.dlib.org/dlib/december00/weibel/12weibel.html> (consultado el 23 de enero de 2001).

²⁴¹ Dublin Core Metadata Initiative. *Dublin Core Metadata Element Set, Versión 1.1: Reference Description* [documento HTML]. DCMI, July 2, 1999. Disponible en <http://dublincore.org/documents/1999/07/02/dces/> (consultado el 25 de febrero de 2001).

²⁴² Este estándar ISO se encuentra disponible a texto completo en <http://www.sdct.itl.nist.gov/~ftp/x318/other/Standards/iso11179/>

(persona o entidad que es responsable del contenido del recurso), *Subject* (palabra o palabras clave que describen la materia del recurso), *Description* (frases descriptivas del contenido del recurso), *Publisher* (persona, entidad o servicio responsable de la difusión y publicación del recurso), *Contributor* (persona o entidad que ha contribuido en la realización de los contenidos del recurso), *Date* (fecha asociada a un evento en el ciclo de vida del recurso, especialmente el de su creación), *Type* (la naturaleza o género del contenido del recurso), *Format* (la manifestación física o digital del recurso), *Identifier* (una referencia sin ambigüedades a la identificación del recurso, principalmente será un URI), *Source* (una referencia al recurso general del cual depende la parte señalada), *Language* (la lengua en la que está redactado el contenido del recurso), *Relation* (una referencia a un recurso relacionado), *Coverage* (la extensión o cobertura del contenido del recurso, normalmente será una referencia a una localización geográfica o a un periodo temporal), *Rights* (información sobre la persona o entidad que ostenta los derechos de propiedad intelectual del recurso).

Junto a estas grandes categorías, Dublin Core permite la utilización de una serie de calificadores, los **Dublin Core Qualifiers**²⁴⁴, que matizan o acotan el contenido incluido dentro de algunos de los elementos DC. Existen dos clases de calificadores:

- Acotaciones del elemento (*Element Refinements*): estos calificadores permiten acotar el contenido de un elemento de forma más específica.
- Codificación de esquemas (*Encoding schemas*): estos calificadores expresan el esquema normalizado que se ha empleado para la codificación de los datos contenidos dentro del elemento permitiendo, por tanto, que sean correctamente interpretados por la aplicación informática que ha de procesarlos.

²⁴³ Stuart Weibel, Eric Miller. *An Introduction to Dublin Core* [documento HTML]. XML.com, October 25, 2000. Disponible en <http://www.xml.com/lpt/a/2000/10/25/dublincore/index.html> (consultado el 20 de diciembre de 2000).

²⁴⁴ Dublin Core Metadata Initiative. *Dublin Core Qualifiers* [documento HTML]. DCMI, July 11, 2000. Disponible en <http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/> (consultado el 25 de febrero de 2001).

Por ejemplo, el elemento *Date* del DCMES tiene como acotaciones al mismo los calificadores de *Created*, *Valid*, *Available*, *Issued* y *Modified* siendo posible codificar estos datos, bien según lo establecido por el DCMI para la entrada de datos temporales²⁴⁵ o, alternativamente, con lo establecido por el W3C para la codificación de fechas e intervalos temporales (basado en la norma ISO 8601)²⁴⁶.

Lamentablemente, y a pesar de los grandes beneficios que este sistema de asignación de metadatos reporta para la búsqueda y recuperación de información, no es una práctica generalizada la inclusión de los elementos DC en los documentos HTML que pueblan el territorio de la WWW²⁴⁷. Tan sólo aquellas instituciones nacionales o internacionales más directamente involucradas con los asuntos de la normalización y la estructuración de la información electrónica en Internet tienden a emplear este sistema de asignación de metadatos dentro de las cabeceras de sus documentos HTML. Sirva como magnífico ejemplo de empleo la estructura de metadatos incluida dentro de una de las páginas web publicadas por la importante institución española RedIRIS:

²⁴⁵ Véase el sistema *DCMI Period* en <http://dublincore.org/documents/dcmi-period/>

²⁴⁶ Véase el sistema *W3C-DTF* en <http://www.w3.org/TR/NOTE-datetime/>

²⁴⁷ Un interesante y más que citado estudio al respecto de este tema fue el elaborado en 1999 por S. Lawrence y C. L. Giles en el cual demostraban que tan sólo un 34% de todos los documentos HTML existentes en la Web incluían las meta-etiquetas de “keywords” y de “description”, descendiendo hasta un 0’3% aquellos documentos que hacían uso del estándar Dublin Core para la asignación de metadatos. Para una mayor información al respecto, véase Steve Lawrence, C. Lee Giles. “Accessibility of Information on the Web”. *Nature*, v. 400, n° 6.740, 1999, p. 108. De igual modo, pero circunscrito al ámbito de los documentos HTML publicados en la Web por las bibliotecas españolas resulta, de interés el estudio reciente realizado J. A. Merlo y A. Sorli en el cual se demostraba que tan sólo un 7% de éstas empleaban las cuatro meta-etiquetas básicas (bien propias de HTML o de DC) para la descripción de la información contenida. Para una más amplia información, véase José Antonio Merlo Vega, Ángela Sorli Rojo. “El uso de metainformación en los webs de las bibliotecas españolas”. En: *Jornadas Españolas de Documentación* (7°. 2000. Bilbao). Bilbao: Universidad Carlos III de Madrid, 2000, p. 158.

```
<HEAD>
<TITLE>RedIRIS - Red española de I+D</TITLE>

<!-- Meta caracteres -->
<meta name="robots"          content="noindex, nofollow">
<meta name="keywords"        content="rediris; iris;
jornadas; i+d">
<meta name="description"      content="RedIRIS es la red
académica y de investigación nacional, patrocinada por el
Plan Nacional de I+D+I y gestionada por el Consejo Superior
de Investigaciones Científicas">

<!-- Dublin Core -->
<META NAME="DC.Title"         CONTENT="RedIRIS - Servicio de
Directorio de RedIRIS">
<META NAME="DC.Creator"       CONTENT="dirManager -
infodir@rediris.es">
<META NAME="DC.Subject"       CONTENT="rediris; iris;
jornadas; i+d">
<META NAME="DC.Description"   CONTENT="RedIRIS es la red
académica y de investigación nacional, patrocinada por el
Plan Nacional de I+D+I y gestionada por el Consejo Superior
de Investigaciones Científicas">
<META NAME="DC.Publisher"     CONTENT="CSIC/RedIRIS. España">
<META NAME="DC.Date"          CONTENT="2000-10-20">
<META NAME="DC.Language"      CONTENT="es">
<META NAME="DC.Rights"        CONTENT="http://www.rediris.es/copy.es.html">
.
.
</HEAD>
```

Figura III.6: Código de la cabecera del documento HTML de inicio del sitio web de RedIRIS.

Fuente: *RedIRIS*. <http://www.rediris.es/index.html>

Por todo lo dicho hasta el momento, y antes de introducirnos en el estándar RDF de XML, es fácil entender que en el contexto de la WWW de producción de documentos electrónicos el término “metadatos” exceda, por tanto, el ámbito de la actividad catalográfica y los catálogos de bibliotecas pues, como señala acertadamente S. Gradmann, “en general [los metadatos] no están creados por catalogadores profesionales, están pensados para ser producidos de manera más eficiente que los registros catalográficos, cubren un tipo concreto de material (los recursos electrónicos) y [...] la relación entre los

metadatos y los recursos referenciados difiere substancialmente de la relación entre un registro catalográfico y el libro que tiene la biblioteca”²⁴⁸. Así, aunque entre un registro DC y un registro catalográfico puedan existir similitudes semánticas el contexto general de producción y el uso de ambas productos difieren sustancialmente. Empero, el papel que pueden (y deben) desempeñar los actuales profesionales de la biblioteconomía y la documentación en este nuevo contexto de producción puede resultar de especial valor pues la experiencia y dominio profesional adquirido durante siglos en cuanto a la consistencia en el control de autoridades y la autenticación de los datos resulta un factor determinante para la actividad de la asignación de metadatos a los actuales documentos electrónicos.

²⁴⁸ Stefan Gradmann. “Catalogación Versus Metadata: Vino Viejo en Odres Nuevos?” [documento HTML]. En: *IFLA General Conference* (64th. 1998. Amsterdam). IFLA, 1998. Disponible en <http://www.ifla.org/IV/ifla64/007-126s.htm> (consultado el 28 de marzo de 2001).

III.5.2. UTILIZACIÓN DE LOS ESQUEMAS XML Y RDF COMO FUENTE DE METADATOS

La marca descriptiva es en sí misma, como asevera R. Cover, una forma de “metadatos” pues nos dice lo que el elemento es a través del nombre elegido y cómo los objetos informativos existentes dentro del documento se estructuran en un conjunto coherente y jerarquizado de bloques²⁴⁹. Por tanto, un factor crítico en cualquier lenguaje de marcado descriptivo resulta ser el delicado proceso de selección de los nombres de los elementos y de los atributos, así como los posibles valores que estos pueden tomar, tarea ésta, como ya se ha venido exponiendo en esta tesis doctoral, llevada a cabo por profesionales y expertos en la materia a tratar.

XML, como metalenguaje que es para la construcción de lenguajes descriptivos, surgió con esta filosofía pues ofrecía al usuario, al igual que ocurría con su predecesor SGML, poder establecer aquellos nombres para las marcas que mejor se adecuasen al contenido semántico de cada uno de los objetos del documento a tratar; esto es, crear sus propios vocabularios de marcado de documentos a través de los mecanismos existentes para ello en XML (la DTD o el esquema XML). Bien, esto es así en XML siempre y cuando dichas marcas sean interpretadas por una persona o un colectivo éstas capaces de predecir la información que se encuentra entre una etiqueta de inicio y otra final (o en los valores de los atributos de una etiqueta vacía). Por ejemplo, un autor francófono que desee marcar el título de un documento a buen seguro establecerá como nombre del elemento la palabra *titre* pero, claro está, el contenido semántico de ese objeto marcado sólo será entendido por aquellas personas de su misma lengua o con conocimientos sobre la misma (imaginemos lo que supondría para la mayoría de nosotros un texto con marcas en ruso, japonés o austro-húngaro). Lo mismo se puede decir de un documento cuyo contenido científico sea altamente especializado (física nuclear, microbiología, astrofísica, etc.) donde, a buen

²⁴⁹ Robin Cover. *XML and Semantic Transparency* [documento HTML]. OASIS, rev. November 24, 1998. Disponible en <http://www.oasis-open.org/cover/xmlAndSemantics.html> (consultado el 11 de noviembre de 2000).

seguro, la mayoría de las marcas establecidas para marcar los documentos electrónicos sólo podrían ser interpretadas por aquellos especialistas en cada una de estas disciplinas científicas.

Pero aún hay más. XML fue diseñado en un principio para la Web y en este contexto el principal consumidor de información electrónica no son las personas y sí, como señalábamos con anterioridad, los ordenadores y las aplicaciones informáticas. Para XML, así como para las aplicaciones informáticas que lo han de procesar, resulta igual de absurdo o sin sentido que marquemos el título de un documento con las etiquetas `<titulo></titulo>`, `<title></title>` o `<titre></titre>` o que se hayan empleado estas etiquetas para marcar el autor de dicho documento; tan sólo le interesa que las reglas sintácticas establecidas se hayan cumplido. XML, al igual que SGML, sólo gobierna sobre la sintaxis, no existiendo, como así es señalado por el propio R. Cover, mecanismos formales que sean capaces de dar soporte a declaraciones restrictivas sobre la semántica de los elementos establecidos²⁵⁰.

De este modo, y si pensamos en la búsqueda y recuperación de información en un entorno de redes informáticas la aplicación del marcado XML no tendría demasiados problemas si se estuviese funcionando en un entorno finito de producción y difusión documental (por ejemplo, en una Intranet corporativa) pues aquí sería relativamente sencillo que todos los documentos producidos estuviesen marcados según una o varias DTDs/esquemas determinadas. De este modo, tanto el usuario como cualquier aplicación informática sabría que para buscar el título, el autor o la fecha de un documento tan sólo debería extraer el contenido existente entre las etiquetas de inicio y final de aquellos nombres de los elementos consensuados y establecidos de antemano por la organización. Pero, claro está, en un contexto de redes informáticas más amplio como es el caso de la Internet no existe un consenso universal, una mega-DTD redactada en un lenguaje planetariamente aceptado (un esperanto para la red), que posibilite a una aplicación informática (en este caso, un motor de búsqueda de tipo genérico) recuperar todos los títulos, autores, materias, fechas de elaboración o cualquier otro dato de interés de todos

²⁵⁰ R. Cover. *Ibid.*, <http://www.oasis-open.org/cover/xmlAndSemantics.html>

los documentos XML existentes en este espacio electrónico. Esta ausencia de semántica en el marcado es, entre otros muchos, uno de los principales motivos de fracaso de la búsqueda y recuperación de información en Internet de documentos XML²⁵¹.

Para dar solución a este complejo problema, algunos especialistas vienen trabajando dentro de la W3C desde hace unos cuantos años en la denominada **Semantic Web** (SW)²⁵².

El concepto de una web semántica fue introducido en 1998 por Tim Berners-Lee a través de una serie de escritos en el seno del W3C con la finalidad de crear un verdadero sistema global de información capaz de ser interpretado correctamente tanto por las personas como por las aplicaciones informáticas que la han de procesar. Pero no sería hasta la conferencia ofrecida en el congreso del XML 2000 organizado por la GCA en Washington DC, y debido a las críticas que tanto este autor como el concepto inventado estaban recibiendo²⁵³, cuando Berners-Lee expusiera oficialmente su idea de una Web semántica²⁵⁴.

²⁵¹ En un interesante artículo del sitio Web "SearchTools" se analizan una serie de cuestiones que, en opinión de los autores, deberían asentarse para que realmente XML fuese ese estándar tan deseado que diese solución a los problemas de búsqueda y recuperación de información en la Internet. Entre estas cuestiones a solucionar se citan la limitación de los actuales navegadores para interpretar correctamente documentos XML, la falta de normalización de estructuras documentales para las diversas industrias, las limitaciones de los actuales motores de búsqueda para interpretar correctamente la información existente en las etiquetas XML, la pobreza de los actuales sistemas de interrogación de dichos sistemas de búsqueda, la falta de normalización de los documentos a través del uso de los metadatos, etc. Para una mayor información al respecto, véase *XML and Search* [documento HTML]. SearchTools.com, [sin fecha]. Disponible en <http://www.searchtools.com/info/xml.html> (consultado el 19 de noviembre de 2000).

²⁵² Toda la información y actividad oficial relativa a este importante grupo de trabajo (*W3C Semantic Web Activity*), coordinado por Eric Miller, se encuentra disponible en la dirección <http://www.w3.org/2001/sw/>. De igual forma, y desde fechas recientes, existe un "portal" en Internet, creado y mantenido por Stefan Decker, dedicado exclusivamente a suministrar información relevante sobre todo lo concerniente a la "Web Semántica", accesible en la dirección <http://www.semanticweb.org/>

²⁵³ A este respecto señala E. Dumbill en un artículo divulgativo sobre el desarrollo de la Web semántica por parte del W3C que algunos investigadores del entorno XML criticaban este desarrollo por ser bastante ambiguo y estar mal definido, llegando a denominar a este conjunto de ideas bajo el irónico apelativo de "Pedantic Web" debido a que sobre la base de RDF (tecnología bajo la que se sustenta la idea de una Web semántica para el W3C) emplea el sistema de URIs inventado por Berners-Lee para la Web en sus orígenes. Para una mayor información al respecto, véase Ed Dumbill. *The Semantic Web: A Primer* [documento HTML]. XML.com, November 1, 2000. Disponible en <http://www.xml.com/lpt/a/2000/11/01/semanticweb/index.html> (consultado el 30 de noviembre de 2000).

²⁵⁴ Las transparencias empleadas por Berners-Lee en este congreso se encuentran en Tim Berners-Lee. "Semantic Web on XML" [documento HTML]. En: *XML 2000* (Washington DC. 2000). Disponible en

La idea base de Berners-Lee para este desarrollo está en el aserto de que la mayor parte de la información que se encuentra disponible en la actual Web está diseñada para el consumo humano, conteniendo dicha información unas estructuras de datos poco evidentes o propicias para ser interpretadas y analizadas correctamente por los robots y aplicaciones informáticas que recorren la Web²⁵⁵. Este autor y sus colaboradores comentan en un escrito elaborado recientemente, que la web semántica no implica una web distinta de la actual sino, más bien, una extensión de la misma, donde la información ofrecida esté definida y estructurada de forma adecuada para que los datos contenidos sean fácilmente procesables y sin ambigüedades por los ordenadores y sus aplicaciones informáticas y, por ende, donde la gente pueda trabajar verdaderamente de forma cooperativa²⁵⁶. Así, y según estos autores, lo que la web semántica preconiza es la proposición de un lenguaje que sea capaz de expresar tanto los datos como las reglas que los construyen y que posibilite establecer reglas desde cualquier sistema de representación del conocimiento para ser exportado dentro de la Web.

Estos lenguajes son denominados en el contexto de la inteligencia artificial aplicada a la investigación de la Web bajo el nombre de *ontologías*²⁵⁷, donde una ontología resulta ser,

<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide1-0.html> (consultado el 12 de marzo de 2001). De igual forma, un artículo de E. Dumbill resume lo más significativo de la exposición realizada por Berners-Lee en este congreso internacional, por lo que recomendamos su lectura para una mayor información al respecto. Véase, Edd Dumbill. *Berners-Lee and the Semantic Web Vision* [documento HTML]. XML.com, December 2000. Disponible en <http://www.xml.com/lpt/a/2000/12/xml2000/timbl.html> (consultado el 28 de marzo de 2001).

²⁵⁵ Tim Berners-Lee. *Semantic Web Road map* [documento HTML]. W3C, September 1998, rev. October 14, 1998. Disponible en <http://www.w3.org/DesignIssues/Semantic.html> (consultado el 20 de enero de 2001).

²⁵⁶ Tim Berners-Lee, James Hendler, Ora Lassila. "The Semantic Web" [documento HTML]. *Scientific American*, n° 501, May 2001. Disponible en <http://www.sciam.com/2001/0501issue/0501/berners-lee.html> (consultado el 18 de junio de 2001).

²⁵⁷ El concepto de "ontología" no siempre resulta fácil de definir en este contexto y sujeto, como señala Tom Gruber, a numerosas polémicas y controversias. Se trata de un término heredado del campo de la filosofía (metafísica) el cual hace referencia al ser en general y de sus propiedades trascendentales. Pero como señala este mismo autor, una "ontología" aplicada al campo de la gestión del conocimiento es una "especificación explícita de una conceptualización" o, dicho de otro modo, una descripción formal de conceptos y relaciones que pueden existir para un agente o una comunidad de agentes. Véase, Tom Gruber. *What is an Ontology?* [documento HTML]. Stanford, CA: Knowledge System Laboratory, Stanford University, [s. f.]. Disponible en <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html> (consultado el 8 de mayo de 2001). Nicola Guarino se manifiesta a este respecto de forma más directa al comentar que una ontología está constituida

como señalan Berners-Lee y sus colaboradores, un documento o fichero que define formalmente las relaciones existentes entre los términos, similar a las taxonomías que definen clases y subclases de objetos y las relaciones entre ellos²⁵⁸.

Dos han sido las tecnologías que han entrado en juego para llevar a cabo esta compleja y polémica idea: el metalenguaje XML y, principalmente²⁵⁹, la aplicación RDF.

por un vocabulario específico utilizado para describir una cierta realidad más un conjunto de presunciones explícitas relativas al “significado intencionado” de las palabras del vocabulario. Véase, Nicola Guarino. “Formal Ontology and Information Systems” [documento PDF]. En: *Formal Ontology in Information Systems* (1º. 1998. Trento). Disponible en <http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/FOIS98.pdf> (consultado el 10 de mayo de 2001). En cualquier caso, y para profundizar en el tema de las “ontologías” aplicadas al campo de la Web Semántica recomendamos la consulta de los documentos referenciados en el apartado dedicado a tal fin del anteriormente citado portal temático sobre la materia y disponible en la dirección <http://www.semanticweb.org/knowmarkup.html>

²⁵⁸ T. Berners-Lee, J. Hendler, O. Lassila. *Ibid.*, <http://www.sciam.com/2001/0501issue/0501/berners-lee.html>

²⁵⁹ Y decimos que principalmente dado que existen otras aplicaciones derivadas del metalenguaje XML que están desempeñando un papel importante en esta labor de descripción semántica de la Web: XTM y MDL. La más destacada de estas otras aplicaciones XML alternativas al modelo RDF del W3C es el desarrollo denominado **XML Topic Maps** (XTM). Los mapas temáticos o de materia (*topic maps*) aplicados a la descripción documental fueron establecidos de forma normativa por la ISO en 2000 a través del estándar internacional ISO/IEC 13250 como una aplicación de SGML (documento disponible en formato electrónico en <http://www.y12.doe.gov/sgml/sc34/0129.pdf>). En este contexto, un mapa temático es definido como una representación multidimensional de información utilizada para describir y navegar a través de objetos de información. De forma general, la información estructural transportada por los *topic maps* incluyen, por un lado, grupos de objetos de información accesibles a través de una dirección alrededor de una serie de materias (las denominadas “ocurrencias”) y, de otra parte, relaciones entre dichas materias (las denominadas “asociaciones”). Debido a la perfecta adecuación de este modelo al espacio de la Web y ser XML la nueva *lingua franca* de este espacio para la comunicación entre documentos electrónicos y las bases de datos, rápidamente se empezó a adaptar este desarrollo al metalenguaje XML. Un grupo de instituciones e investigadores fundaron una organización independiente sin ánimo de lucro para llevar a cabo este fin, dando lugar a la organización conocida como TopicMaps.Org (<http://www.topicmaps.org/>). Esta organización ha venido trabajando desde entonces en la creación de una especificación del XTM, siendo presentada la versión 1.0 en diciembre del 2000 en la conferencia del XML 2000, celebrada en Washington D.C. La última revisión de esta especificación es de marzo de 2001, disponible en TopicMaps.Org. *XML Topic Maps* (XTM). *TopicMaps.Org Specification* [documento HTML]. Steve Pepper, Graham Moore (eds.). TopicMaps.Org, March 2, 2001. Disponible en <http://www.topicmaps.org/xtm/1.0/xtm1-20010302-2.html> (consultado el 16 de junio de 2001). El Anexo G de esta especificación se encuentra enlazado con otro documento normativo para la transformación de los *topic maps* conformes a la norma ISO 13250 ajustada a la sintaxis XML. Para una mayor información sobre el origen y evolución del XTM recomendamos la consulta de Michel Biezunski, Steven R. Newcomb. “XML Topic Maps: Finding Aids for the Web” [documento PDF]. *IEEE Multimedia*, v. 8, n° 2, April-June 2001. Disponible en <http://www.coverpages.org/newcomb-ieee-mm-topicmaps-article.pdf> (consultado el 7 de julio de 2001). En la actualidad, existen algunas líneas de investigación muy prometedoras en las cuales ponen en relación el modelo de los *topic maps* con el modelo RDF así como con el modelo de esquema de XML, con el fin de obtener el mayor potencial posible para una efectiva descripción semántica de los recursos existentes en el espacio electrónico de la Web. Sirvan como ejemplos válidos a este respecto los dos siguientes documentos presentados al congreso del XML Europe 2001: la comunicación de

XML, como ya sabemos, posibilita al usuario la creación de sus propias marcas para describir las piezas estructurales del documento, siendo RDF el desarrollo que permite asignar un significado a dicha estructura y sus marcas. Como señala Tim Bray, uno de los padres de XML, este metalenguaje no es capaz por sí solo de representar de forma adecuada los metadatos de un documento debido a los problemas de escalabilidad y de interoperabilidad de la metainformación en un entorno de trabajo adecuado para la Web²⁶⁰.

Es más, el propio Berners-Lee suele designar a la versión 1.0 de XML como algo pesado, lento y bajo estricta observación²⁶¹, diferenciándolo claramente del modelo propuesto por RDF para la tarea de definición de metainformación de los documentos que han de ser expuestos en la Web. Así, para este autor, dejando al margen las cuestiones idiomáticas de las marcas antes expuestas, el modelo propuesto por RDF resulta notablemente más ventajoso que el modelo de XML para la búsqueda y recuperación de información en la Web. Para apoyar este aserto, este autor comenta que la complejidad en la interrogación de los documentos XML viene causada por las muchas formas en las que lógicamente se puede estructurar o “mapear” una misma información frente al modo único y normalizado

Graham D. Moore. “RDF and TopicMaps. An Exercise in Covergence” [documento PDF]. En: *XML Europe 2001 Conference* (2001. Berlín). Disponible en <http://www.gca.org/papers/xml europe/papers/s23-3.pdf> (consultado el 7 de agosto de 2001) y la comunicación de Martin Bryan. “How can XML Schemas Enhance Topic Maps?” [documento HTML]. En: *XML Europe 2001 Conference* (Berlín. 2001). Disponible en <http://www.gca.org/papers/xml europe2001/papers/html/s31-4.html> (consultado el 7 de agosto de 2001). En el caso del **Meaning Definition Language** (MDL), se trata de una aplicación que permite definir y establecer las relaciones entre el significado de los datos contenidos en los documentos XML y su estructura, haciendo para ello uso de expresiones procedentes del estándar XPath del W3C. El borrador de esta especificación se encuentra disponible en <http://www.charteris.com/mdl/> En cualquier caso y para una mayor profundidad en todos estas aplicaciones para la definición semántica de los documentos XML, remitimos a la consulta de la obra Kal Ahmed... [et al]. *Professional XML Meta Data*. Birmingham: Wrox Press, 2001.

²⁶⁰ Tim Bray. *What is RDF?* [documento HTML]. XML.com, January 24, 2001. Disponible en <http://www.xml.com/2001/01/24/rdf.html> (consultado el 1 de febrero de 2001).

²⁶¹ Estas palabras son recogidas en un artículo de E. Dumbill, el cual señala expresamente que este autor gusta de designar a este metalenguaje como “fast, low and under the radar”. Para una mayor información, véase Edd Dumbill. *Building the Semantic Web* [documento HTML]. XML.com, March 7, 2001. Disponible en <http://www.xml.com/lpt/a/2001/03/07/buildingsw.html> (consultado el 28 de marzo de 2001).

en el que lo hace el desarrollo RDF; esto es, un mismo árbol RDF resulta de múltiples árboles XML²⁶².

RDF (*Resource Description Framework*) es, como ya se señaló, una aplicación de XML que proporciona un marco de trabajo para la descripción y el intercambio de metadatos que han de ser servidos, en principio, en la Web. Adicionalmente, RDF proporciona, como señala E. Miller, “un medio para la publicación de vocabularios, entendibles tanto por las personas como por los ordenadores, diseñados para posibilitar la reutilización y expansión de metadatos semánticos entre comunidades de información dispares”²⁶³.

Como ya se comentó al introducir esta tecnología en el apartado III.2.3, RDF se articula funcionalmente en tres grandes bloques: un modelo de datos para la representación de nombres de propiedades y sus valores, una sintaxis (XML) para la redacción y transporte de los metadatos de modo que se vea asegurada la interoperabilidad e independencia entre servidores y clientes web y, finalmente, los esquemas RDF como colecciones de información sobre las clases y subclases de nodos RDF, donde se incluyen las propiedades de los mismos y sus relaciones. Así, como es señalado por E. Méndez, la esencia de RDF se encuentra en el hecho de ser un “modelo formal” para la representación de las propiedades de la metainformación de un documento así como en los valores de dichas propiedades²⁶⁴.

El modelo de datos, núcleo de la tecnología RDF, se articula sobre las siguientes reglas:

- Recursos: es cualquier objeto que pueda tener un URI, incluyendo tanto la totalidad de las páginas Web existentes así como elementos individuales de un documento XML.
- Propiedades: es un recurso que tiene un nombre y que puede ser utilizado como una propiedad (por ejemplo, “Autor”, “Título”, etc.).

²⁶² Tim Berners-Lee. *Why RDF model is different from the XML model* [documento HTML]. W3C, September 1998, rev. November 14, 1998. Disponible en <http://www.w3.org/DesignIssues/RDF-XML.html> (consultado el 24 de noviembre de 2000).

²⁶³ Eric Miller. “An Introduction to the Resource Description Framework”. *Bulletin of the American Society for Information Science*, v. 25, n° 1, 1998, p. 15.

- Descripciones: es la combinación de un recurso, una propiedad y un valor. Estas partes son conocidas como el “sujeto”, el “predicado” y el “objeto” de una declaración.

El modelo propuesto por RDF de asociaciones entre los metadatos y el recurso a describir es similar al que se producía en los documentos HTML, esto es, los metadatos pueden ser incrustados dentro del propio recurso, pueden estar asociados al recurso, pueden estar contenidos en un fichero independiente al recurso o, por último, el recurso puede estar contenido junto con los metadatos. Pero, como acertadamente señala R. Heery, algunas de estos métodos de asociación no son posibles para determinados tipos de recursos: por ejemplo, no es posible incrustar metadatos dentro de un fichero GIF o JPEG de una imagen digital²⁶⁵.

En el plano más pragmático, el uso de RDF conlleva el empleo de otras muchas tecnologías XML asociadas (XML namespaces, XLink, XSLT, etc.). El caso del empleo de XSLT parece obvio pues, y en palabras del propio Berners-Lee recogidas por E. Dumbill, en muchos casos la extracción de los datos que alimentan los ficheros RDF se realiza de forma automática a través de procesos de transformación de formatos realizados desde fuentes de documentos XML²⁶⁶.

En el caso de los *XML namespaces*, RDF, como acertadamente señala S. St. Laurent, fue la primera recomendación del W3C que hizo uso de los nombres locales XML²⁶⁷. Y es absolutamente lógico pues para cualquier descripción del contenido de un recurso web resulta imprescindible diferenciar, al menos, los nombres de los elementos y atributos empleados tanto por la aplicación RDF como de los empleados por el documento electrónico descrito (el vocabulario propio). Y este señalamiento hacia los vocabularios de metadatos empleados en los *namespaces* se realiza mediante el empleo de URIs (de forma

²⁶⁴ Eva M^a Méndez Rodríguez. “RDF: un modelo formal de metadatos flexible para las bibliotecas digitales del próximo milenio”. En: *Jornades Catalanes de Documentació* (7^a. 1999. Barcelona). Barcelona: Col·legi Oficial de Bibliotecaris-Documents de Catalunya, 1999, p. 490.

²⁶⁵ Rachel Heery. “What is... RDF?” [documento HTML]. *Ariadne*, n° 14, July 1998. Disponible en <http://www.ariadne.ac.uk/issue14/what-is/intro.html> (consultado el 12 de noviembre de 1999).

²⁶⁶ E. Dumbill. “Berners-Lee and the... “ *Op. cit.*, <http://www.xml.com/lpt/a/2000/12/xml2000/timbl.html>

real, a través de URLs) posibilitando con estos vocabularios, además, como señala D. Brickley, la construcción de gráficos que expresen cosas de forma simple y normalizada²⁶⁸. Así, por ejemplo, la declaración “el autor del recurso <http://www.bib.uc3m.es/~bmartin/xml/index.xml> es Bonifacio Martín y lleva por título ‘Página principal de Boni’ ” puede ser expresado en RDF de forma directa del siguiente modo:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:s="http://www.bib.uc3m.es/~bmartin/schema/"
  <rdf:Description
    about="http://www.bib.uc3m.es/~bmartin/index.xml">
    <s:Autor>Bonifacio Martín</s:Autor>
    <s:Titulo>Página Principal de Boni</s:Titulo>
  </rdf:Description>
</rdf:RDF>
```

En este caso, se ha conseguido normalizar en cierto grado la descripción de la autoría del documento XML pero tan sólo es significativo para nuestro sistema dado que el vocabulario empleado (los nombres de los elementos definidos en nuestra DTD) para marcar el documento XML no se ajusta a ninguno de los vocabularios creados a escala internacional para la descripción de metainformación. Así, RDF no establece en sí mismo ningún vocabulario predefinido para el establecimiento de los metadatos, esperando que aparezcan, como señala J. Boye, vocabularios normalizados a nivel internacional para facilitar, de este modo, el intercambio, recuperación y reutilización de los metadatos incluidos en los documentos electrónicos²⁶⁹. De este modo, podemos hacer uso, por ejemplo, de los elementos del Dublin Core (aunque es posible combinar dentro de un

²⁶⁷ S. St. Laurent, R. Biggar. *Op. cit.*, p. 215.

²⁶⁸ Dan Brickley. *Understanding RDF* [documento HTML]. Institute for Learning and Research Technology, University of Bristol, 1999, rev. January 22, 2001. Disponible en <http://ilrt.org/discovery/2001/01/understanding-rdf/> (consultado el 31 de enero de 2001).

²⁶⁹ Janus Boye. *RDF – What's in it for us?* [documento HTML]. IRT.org, May 16, 1998. Disponible en <http://www.irt.org/articles/js086/index.htm> (consultado el 7 de septiembre de 1999).

mismo documento RDF elementos procedentes de más de un vocabulario a través, como señala R. Ianella, de la indicación de sus *namespaces* respectivos²⁷⁰) para facilitar dichas tareas. Así, y empleando el anterior ejemplo, éste quedaría de forma básica de la siguiente manera:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  <rdf:Description
    about="http://www.bib.uc3m.es/~bmartin/index.xml">
    <dc:creator>Bonifacio Martín</dc:creator>
    <dc:title>Página Principal de Boni</dc:title>
  </rdf:Description>
</rdf:RDF>
```

RDF es algo más que una simple aplicación de XML, pues al igual que este metalenguaje contempla mecanismos para que cada comunidad pueda desarrollar de forma independiente vocabularios propios de lenguajes de metadatos. Las descripciones de estos vocabularios se realiza a través del modelo de esquema RDF (*RDF Schema*) capaz, por tanto, de establecer y formalizar las relaciones y restricciones que se dan entre los elementos del lenguaje definido. Este modelo de esquema para el desarrollo de vocabularios de metadatos y expresión de las relaciones que se establecen entre los términos (que en el entorno de RDF gustan denominar, como ya se comentó con anterioridad, “ontologías”) es similar tanto al modelo de DTD como, y sobre todo, al modelo de esquema de XML pero, y a diferencia de éstos, contempla su propia sintaxis y ciertas peculiaridades²⁷¹. Por ejemplo, como señala E. Dumbill, los esquemas RDF no definen una restricciones sintácticas como lo hacen las DTDs y los esquemas XML pero en cambio definen clases (a través de la etiqueta <rdfs:Class>) y propiedades (a través de la etiqueta <rdfs:Property>) así como las interrelaciones que se dan entre ellas y las posibles

²⁷⁰ Renato Ianella. *An Idiot's Guide to the Resource Description Framework* [documento HTML]. Queensland: DSTC, University of Queensland, September 3, 1998, rev. January 25, 1999. Disponible en <http://www.dstc.edu.au/Research/Projects/rdf/RDF-Idiot.html> (consultado el 24 de enero de 2001).

restricciones que se pudieran establecer a través de su pertenencia a un determinado dominio (a través de la etiqueta `<rdfs:domain>`) y con un rango específico (a través de la etiqueta `<rdfs:range>`): operan directamente al nivel de modelo de datos más que al nivel sintáctico²⁷².

Así, y sin entrar en un detallado examen de los esquemas RDF dado que ello excede a la intencionalidad de esta tesis doctoral, para el modelo de esquema de RDF los conjuntos de información son entendidos como “nodos” (*nodes*) y son creados para representar un determinado documento marcado y su entorno. Cada nodo es un objeto, y cada uno de éstos es una instancia de una clase definida en el esquema. El conjunto de estos nodos forma un “bosque” (*grove*), y el conjunto de cada uno de estos bosques a través de múltiples documentos conforma un “hiper-bosque” (*hypergrove*).

Se observa, pues, que existe un aparente conflicto entre dos de los desarrollos del propio W3C, esto es, entre el modelo de esquema XML y el modelo de esquema RDF, pues en ambos casos pueden ser utilizados para construir taxonomías. Este hecho está generando en la actualidad no pocas polémicas pues no siempre es fácil discernir cuál es el mejor método para la creación de dichas taxonomías a través de XML, si bien hacer uso de las clases dentro de un documento de esquema RDF o, por el contrario, emplear las capacidades de representación estructural que posibilita el esquema XML. Así, algunos autores, como es el caso de John McClure, se decantan abiertamente por el modelo de esquema de RDF esgrimiendo que resulta más natural y cercano al modelo de diseño y técnicas de programación orientado a objetos representar una taxonomía como clases²⁷³.

²⁷¹ S. St. Laurent, R. Biggar. *Op. cit.*, p. 221.

²⁷² E. Dumbill. “The Semantic Web...” *Op. cit.*,
<http://www.xml.com/lpt/a/2000/11/01/semanticweb/index.html>

²⁷³ John McClure. *XML and RDF Schemas: Technical Notes* [documento HTML]. LegalXML, November 1, 1999. Disponible en
http://www.legalxml.org/DocumentReopsitory/UnofficialNotes/Clear/UN_10005_1999_11_01.htm
(consultado el 5 de junio de 2001).

Sin embargo, otros autores, como Jane Hunter (del prestigioso DSTC²⁷⁴ de la Universidad de Queensland en Australia) y Carl Lagoze (del no menos prestigioso *Digital Library Research Group*²⁷⁵ de la Universidad de Cornell en Estados Unidos) establecieron en una comunicación presentada al décimo congreso internacional de la WWW que ambos modelos pueden ser complementarios entre sí, aportando cada uno de ellos los beneficios inherentes a los mismos dentro de un desarrollo común de metadatos para la WWW²⁷⁶. Así, los esquemas RDF son potentes mecanismos para el establecimiento de ricas definiciones semánticas pero están ciertamente limitados en el soporte que ofrecen para el establecimiento de restricciones de uso local (por ejemplo, en el plano estructural, en el número de ocurrencias posibles y en el establecimiento de tipos de datos específicos). Por el contrario, los esquemas XML proporcionan potentes mecanismos para el establecimiento de complejas relaciones estructurales, para la delimitación en el número de ocurrencias posibles para los datos de cada elemento y en las restricciones que los propios datos han de tener en sus valores pero proporcionan un escaso soporte para la definición del conocimiento semántico de los elementos contemplados, necesario para facilitar un mapeado flexible y dinámico entre diferentes dominios de metadatos.

²⁷⁴ El *Distributed Systems Technology Centre* (DSTC) (<http://www.dstc.edu.au/AboutDSTC/index.html>) es un organismo sin ánimo de lucro financiado por el programa gubernamental australiano para la cooperación entre centros de investigación ubicado dentro de la Universidad de Queensland. En él se integran más de veinte organizaciones de diversa índole y condición con el fin de desarrollar una infraestructura tecnológica de calidad dentro de este país. Aunque este organismo investiga y trabaja con un amplio espectro de tecnologías diversas para la organización y gestión de la información electrónica, uno de sus pilares tecnológicos básicos se sustenta en el uso de tecnologías XML para la transferencia de información electrónica entre los diversos actores (gobierno e industria, principalmente) que integran la compleja maraña política, social y económica de este país. En el caso concreto del campo de los metadatos, además del desarrollo de algunos proyectos de aplicación de gran interés, esta organización participa activamente dentro del grupo de trabajo del W3C relativo al modelo RDF. Para una mayor información, véase la información suministrada al respecto en la dirección <http://www.dstc.edu.au/Research/Projects/rdf/index.html>

²⁷⁵ Este prestigioso grupo de investigación adscrito al Departamento de Informática de la Universidad de Cornell en Ithaca, Nueva York ha venido investigando y trabajando en el establecimiento de las llamadas bibliotecas digitales (de hecho, y entre otras muchas cosas, son uno de los co-editores de la prestigiosa revista *D-Lib Magazine*), convirtiéndose en un centro de excelencia en este tema. Para una mayor información al respecto, véase la información suministrada en su sitio web oficial en <http://www.cs.cornell.edu/cdlrg/>

²⁷⁶ Jane Hunter, Carl Lagoze. *Combining RDF and XML Schemas to Enhance Interoperability Between Metadata Application Profiles* [documento HTML] Queensland: DSTC, University of Queensland, November 2000.

Esta integración de esquemas se puede hacer, según estos autores, de dos maneras distintas; a saber:

1. **Incrustando localmente las semánticas del esquema RDF dentro de las anotaciones del esquema XML:** Este método conlleva la incorporación local de las definiciones de *Class*, *subClassOf*, *Property* y *subPropertyOf* del esquema RDF dentro del fichero que contiene el esquema XML a través del elemento de anotación (*annotation*), y de forma más precisa, dentro del subelemento de información para el procesamiento por parte de las aplicaciones informáticas (*appinfo*). Por ejemplo, imaginemos que deseamos establecer un elemento XML cuyo identificador genérico sea “Autor” y cuyo contenido sea una cadena de caracteres. Para asignarle una semántica a dicho elemento podemos incrustar una propiedad RDF que lo equipare o lo traduzca al elemento “Creator” del Dublin Core; así:

```
<?xml version="1.0"?>
<xs:schema
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
>

[...]

<xs:simpleType name="Autor">
  <xs:annotation>
    <xs:appinfo>
      <rdf:Property ID="Autor">
        <rdfs:subPropertyOf

          rdf:resource="http://purl.org/dc/elements/1.1/#Creato
r"/>

        </rdf:Property>
      </xs:appinfo>
    </xs:annotation>
    <xs:restriction base="xs:string"/>
  </xs:simpleType>
```

[...]

</xs:schema>

2. Enlazando externamente definiciones del esquema RDF a un esquema XML:

En este caso se hace uso de la tecnología XLink para el establecimiento de relaciones hipertextuales entre documentos XML y, de forma más concreta, el empleo del atributo del tipo *xl:arcrole* definido en un esquema XML en el *namespace* de XLink. Este atributo, añadido a una declaración de tipo de dato simple o complejo, puede asignar un valor al dato declarado que corresponda con una propiedad RDF. Así, por ejemplo, para el mismo caso anterior en el esquema XML el “Autor” quedaría definido del siguiente modo:

```
<?xml version="1.0"?>
<xs:schema
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  targetNamespace="http://purl.org/dc/elements/1.1/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:xl="http://www.w3.org/2001/xlink-ns#"
>

[...]

<xs:complexType name="Autor">
  <xs:simpleContent>
    <xs:extension base="xs:string">
      <xs:attribute name="arcrole"
type="xl:arcrole"/>
    </xs:extension>
  </xs:simpleContent>
</xs:complexType>

[...]

</xs:schema>
```

Y en la instancia del documento XML sería, por tanto, donde se haría referencia al elemento correspondiente del esquema RDF del Dublin Core, del modo siguiente:

```
<Autor arcrole="http://purl.org/dc/elements/1.1/ dcmes.rdf#creator">  
Bonifacio Martín Galán  
</Autor>
```

Cada uno de estos dos modelos posibles de integración de esquemas XML y RDF tiene sus ventajas y sus inconvenientes, como así es señalado por los propios autores. En el primer caso, se obtiene el beneficio de la integración de las definiciones semánticas y las restricciones sintácticas dentro de un mismo fichero pero, por el contrario, resulta un método menos flexible que el segundo y, sobre todo, las clases y las propiedades RDF definidas explícitamente dentro de las anotaciones del esquema XML son establecidas de forma local, sólo para el conjunto de documentos XML que se adscriban a dicho esquema. Este hecho conlleva el gran inconveniente de no poder ser reutilizadas o señaladas por otros esquemas dado que los elementos no han sido nombrados de forma global para su acceso. En el segundo caso, se consigue solucionar el problema de una definición externa y global pero, por el contrario, las semánticas RDF sólo son especificadas en el momento de crear la instancia del documento XML y no en el propio esquema XML, que es lo más lógico.

En cualquier caso, esta integración de ambos modelos de esquema no está, según los anteriormente citados autores, carente de problemas siendo el principal de ellos la actual inexistencia de programas de validación de esquemas que sean capaces de validar conjuntamente ambos esquemas así como el chequeo de la consistencia entre ambos.

Todos estos problemas y sus posibles soluciones alternativas para la asignación de un verdadero significado a los elementos empleados para marcar los documentos XML que han de ser servidos e intercambiados en la Web es uno de los principales caballos de batalla de todos aquellos investigadores que actualmente están trabajando en la construcción de la llamada “Web semántica” a través de la definición de metadatos. Esta tesis doctoral no pretende dar solución a tan complejo tema pues, como ya se ha comentado con anterioridad, ello excede a la temática y pretensiones de la misma. Además, otros investigadores de nuestro Departamento están trabajando profundamente en estos temas por lo que remitimos a las conclusiones obtenidas en sus tesis doctorales y las

publicaciones científicas surgidas con dicha temática Sin embargo, y dado que en el desarrollo práctico que se lleva a cabo en esta tesis doctoral los documentos jurisprudenciales emanados del Tribunal Constitucional han de ser servidos en el espacio electrónico de Internet y, por tanto, éstos han de estar preparados para ser recuperados por los futuros buscadores generales de información electrónica en formato XML, este tema se volverá a retomar posteriormente en el apartado IV.2.4.2 del desarrollo práctico de esta tesis.

Finalmente señalar algo que a estas alturas ya es obvio: uno de los vocabularios para la definición de metadatos en el entorno documental de trabajo XML/RDF que más proyección está teniendo es, como cabría esperar, el del Dublin Core²⁷⁷ al ser éste un modelo candidato de descripción de metainformación de carácter general de recursos existentes en la WWW y, más recientemente, de aplicación en entornos restringidos de intranets corporativas²⁷⁸. El principal inconveniente en la aplicación del vocabulario del DC

²⁷⁷ El empleo del Dublin Core en los documentos XML/RDF no fue mínimamente normalizado hasta mediados del año 2000 con la presentación de un borrador de trabajo por parte de Dave Beckett y otros para su discusión entre los investigadores en la materia (disponible en <http://dublincore.org/documents/2000/07/14/dcmes-xml/>). La última versión actualizada de este documento, aún en fase de Recomendación propuesta, se encuentra en Dave Beckett, Eric Miller, Dan Brickley. *An XML Encoding of Simple Dublin Core Metadata* [documento HTML]. Dublin Core Metadata Initiative, April 11, 2001. Disponible en <http://dublincore.org/documents/2001/04/11/dcmes-xml/> (consultado el 10 de junio de 2001). En el Apéndice A de dicho documento se halla la propuesta de DTD XML para el conjunto de elementos definidos por el modelo Dublin Core. A pesar de que el modelo de metadatos Dublin Core no está aún definido como un modelo de esquema RDF ello no significa que no se hayan realizado diversos intentos desde tiempo atrás para la asignación de un modelo de esquema XML para dicho vocabulario. Así, en mayo de 1999, Andrew Layman presentó a esta organización una propuesta de descripción de las etiquetas del Dublin Core desarrollado siguiendo el modelo de esquema XML-Data (el único con cierto desarrollo, peso e implantación en aquellos años) elaborado por él mismo y otros autores dentro de la compañía Microsoft. Véase, Andrew Layman. *Dublin Core Tags Applied to XML-Data Schemas for Purpose of Description and Categorization* [documento HTML]. Dublin Core Metadata Initiative, May 20, 1999. Disponible en <http://dublincore.org/documents/1999/05/20/dc-xml-data-schemas/> (consultado el 2 de mayo de 2001).

²⁷⁸ Sin entrar en detalle en este tema, sí es necesario comentar que existen algunos autores que preconizan el valor e importancia de los modelos de metadatos RDF aplicados al entorno de producción documental dentro de las intranets corporativas, tanto o más que al entorno de producción de la red Internet (debido a las dificultades reales de construir una verdadera Web semántica). Así, y citando uno de los últimos artículos aparecidos al respecto, Lisa D. Murphy apuesta decididamente por la adopción del vocabulario de metadatos del Dublin Core al entorno de gestión de documentos electrónicos dentro de las organizaciones pero proponiendo la ampliación del conjunto de elementos DC a otros nuevos que reflejen adecuadamente las

en el entorno descriptivo del modelo RDF está en el hecho de su gran generalidad y, por tanto, escasa especificidad (o *granularity*, como gustan decir los especialistas anglosajones en este campo) lo que conlleva que en la mayoría de los proyectos en los que se ha hecho uso de estas tecnologías se hayan tenido que “extender” con la adición de nuevos elementos *ad hoc* que maticen al elemento DC general²⁷⁹.

Pero, y de forma paralela, se están desarrollando otros muchos esquemas RDF para dominios de metadatos mucho más específicos que a buen seguro competirán entre ellos por alcanzar una situación de predominio en este agitado entorno de normalización y que en el campo de la búsqueda y recuperación de información basada en XML/RDF en el espacio de la Web podría hacer evolucionar a las denominadas *subject-gateways* hacia potentes servicios de información temática o sectorial²⁸⁰.

necesidades de control y gestión interna de los documentos electrónicos (elementos como “affiliation”, “authorization”, “contact”, “security level”, “status”, etc.). Para una mayor información al respecto, véase Lisa D. Murphy. “Addressing the Metadata Gap: Ad Hoc Digital Documents in Organizations”. En: Amita Goyal Chin (coord.). *Text Databases & Document Management: Theory & Practice*. Hersey [etc.]: Idea Group, 2001, pp. 52-77. De forma similar, Eva Mª Méndez defiende el modelo de metadatos formulado por RDF para la consistencia de la información electrónica circulante en un entorno de Intranet corporativa a través de desarrollos de esquemas de metadatos junto con la creación de tesauros electrónicos que permitan un eficiente control del vocabulario empleado dentro de la organización. Para una mayor información al respecto, véase Eva Mª Méndez Rodríguez. “Metadatos y Tesauros: aplicación de XML/RDF a los sistemas de organización del conocimiento en Intranets”. En: *Jornadas Españolas de Documentación* (7ª. 2000. Bilbao). Bilbao: Universidad del País Vasco, 2000, pp. 211-219.

²⁷⁹ Por citar un ejemplo relevante, destacar un interesante proyecto de la *Grainger Engineering Library Information Center* de la Universidad de Illinois en el cual se hizo uso de las tecnologías XML y RDF para el desarrollo de su modelo de biblioteca digital en el apartado de la difusión de su colección de publicaciones periódicas y seriadas. En este caso, el vocabulario empleado para asignar la semántica a los elementos descritos de interés para la búsqueda y recuperación documental fue el Dublin Core. Pero debido a la gran generalidad de sus elementos, se tuvo que aumentar el nivel de profundidad o especificidad a través de un conjunto de nuevos elementos. Por ejemplo, al elemento DC que describe la fuente (*Source*) se incluyeron otros subelementos referentes a la publicación de la publicación seriada, como fueron el título de la revista, su abreviatura, el volumen, el número, la página de comienzo del artículo y su página de finalización del mismo. Para una mayor información, véase Timothy W. Cole, William H. Mischo, Robert Ferrer, Thomas G. Habing. *Using XML, XSLT, and CSS in a Digital Library* [documento HTML]. En: *ASIS 2000 Annual Meeting* (2000. Chicago). Disponible en http://dlib.grainger.uiuc.edu/publications/dlibtestsuite/asis/asis2000/cole_xml.htm (consultado el 29 de noviembre de 2000).

²⁸⁰ Sin entrar en detalle sobre este tema dado que escapa a la intencionalidad de esta tesis doctoral, un *subject-gateway* es, y en palabras de J. Kirriemuir y otros colaboradores, un servicio que permite la localización y el acceso de forma rápida y sencilla a recursos electrónicos dentro de una determinada materia de interés. En el terreno de la WWW este hecho se traduce en aquellas páginas o servicios de información que contienen un gran número de enlaces a otros recursos de la red pero con el común denominador de adscribirse todos ellos a una determinada materia o disciplina del saber humano (los denominados comúnmente directorios o

No escapa a esta fiebre en el empleo de las tecnologías XML el campo de actuación del mundo jurídico, en el cual, como veremos en el próximo apartado de este capítulo, se están llevando a cabo un gran número de interesantes proyectos y propuestas de normalización tanto para los documentos legislativos como para los documentos jurisprudenciales, materia de interés de esta tesis doctoral.

buscadores temáticos). Son múltiples los recursos de este tipo que han venido surgiendo en la última década en la Web basados en el lenguaje HTML, existiendo, como es lógico suponer, ejemplos perfectos para cada una de las materias y disciplinas científicas del saber humano (recomendamos para ello la consulta de Traugott Koch. "Quality-controlled subject gateways: definitions, typologies, empirical overview". *Online Information Review*, v. 24, n° 1, 2000, pp. 24-34). Según Kirriemuir *et al.*, "La principal contribución de RDF puede estar en el campo de la navegación cruzada más que en la actual búsqueda de información relacionada [...] RDF promete suministrar un muy necesitado mecanismo de normalización que sería soportado por los servicios de navegación cruzada de recursos altamente organizados [...] Si las colecciones disponibles en cada uno de estos servicios de navegación fueran creadas mediante una descripción RDF sería posible la construcción de *vistas* jerarquizadas de los servicios distribuidos, ofreciendo una interfaz de usuario organizado a través de una clasificación de materias más que a través de una localización física del recurso". Véase, John Kirriemuir, Dan Brickley, Susan Welsh, Jon Knight, Martin Hamilton. "Cross-Searching Subject Gateways: The Query Routing and Forward Knowledge Approach" [documento HTML]. *D-Lib Magazine*, January 1998. Disponible en <http://www.dlib.org/dlib/january98/01kirriemuir.html> (consultado el 12 de febrero de 2001).

CAPÍTULO III.6



XML APLICADO A LA DOCUMENTACIÓN JURÍDICA

III.6.1. INTRODUCCIÓN

Hagamos una pequeña recapitulación de lo expuesto hasta este momento en relación con el tratamiento y difusión de la información de valor jurídico para situarnos, de este modo, en el punto exacto de nuestra disertación científica.

En la anterior Parte de la tesis se habló extensamente de la importancia del metalenguaje SGML aplicado al tratamiento de la información jurídica, exponiéndose aspectos tales como las características que presenta la documentación jurídica para su procesamiento a través de este metalenguaje, las áreas de aplicación y los principales obstáculos encontrados para dicho tratamiento. De igual modo, se mostraron algunos ejemplos significativos del empleo por parte de diferentes instituciones y organismos, públicos y privados, nacionales e internacionales y la importante labor de aplicación y promoción de estos estándares llevada a cabo por algunos juristas de prestigio (volvemos a recordar aquí el importante papel desempeñado por Alan Asay de la *Utah Administrative Office of the Courts*).

Debido a las grandes similitudes existentes entre los metalenguajes SGML y XML, como se ha venido explicando a lo largo de los anteriores capítulos, todo lo expuesto en dicha Parte es perfectamente aplicable a este nuevo entorno tecnológico, por lo que evitaremos, de este modo, volver a repetir cuestiones ya analizadas con anterioridad.

La evolución lógica de dicho tratamiento electrónico de la documentación jurídica ha ido viniendo hasta nuestros días a través de dos procesos tecnológicos íntimamente relacionados entre sí y ya analizados en diferentes apartados de esta tesis doctoral: por un lado, se ha venido profundizando y mejorando los sistemas informáticos para el tratamiento de la información electrónica de carácter jurídico, en especial en el ámbito de la edición, la gestión e intercambio de información electrónica emanada de los tribunales de justicia, y, por otro, con la aparición y expansión de la red Internet y sus servicios, la necesidad de publicar, difundir, intercambiar, localizar y recuperar dicha información jurídica en este espacio electrónico de comunicación ha venido siendo una constante en todos los países del mundo.

Remontándonos aún más atrás, es necesario recordar que en el apartado I.2.2 se introdujeron algunos aspectos concernientes a los sistemas informáticos para la gestión

electrónica de los documentos jurisprudenciales (los denominados *Case Management Systems*), destacando de manera especial dentro de estos sistemas todo lo concerniente a la generación, almacenamiento y difusión de documentos electrónicos emanados de los tribunales de justicia, concepto éste correspondiente al denominado en el mundo anglosajón de aplicación como *Electronic Court Filing*, ya explicado en dicho apartado. Por otro lado, en el capítulo I.3 se analizaron igualmente algunos aspectos relacionados con la publicación y difusión en Internet de información jurídica, haciendo especial hincapié en las implicaciones legales, políticas, sociales y económicas que este hecho conllevaba, no exento de cierta polémica, exponiendo, de igual modo, algunos de los ejemplos más significativos de difusión de información jurídica en este espacio electrónico de redes de comunicación interconectadas.

En el caso concreto de la búsqueda y recuperación de información en el mega-espacio electrónico formado por la Web se ha analizado igualmente y de forma genérica en el apartado del presente capítulo dedicado al establecimiento de metadatos para los documentos que pueblan dicho espacio las grandes limitaciones que los actuales servicios de información de la Web, basados en el lenguaje de marcado HTML, se encuentran tanto para elaborar complejas sentencias de búsquedas así como para la obtención de resultados pertinentes.

Bien, con todo ello expuesto, ¿dónde entra, por tanto, el metalenguaje XML en esta doble orientación tecnológica en el tratamiento de la información jurídica? A estas alturas del desarrollo expositivo que se ha venido manteniendo aquí, parece lógico suponer que XML viene a desempeñar un papel fundamental en esta doble orientación pues viene a dar respuesta y solución a ambas circunstancias, como ha sido señalado por diversos especialistas en los campos de la judicatura y la docencia universitaria a lo largo de estos últimos años y ratificado en numerosos proyectos de investigación. Analicemos, pues, algunas de las opiniones más señaladas al respecto para posteriormente entrar a describir algunos de los proyectos de mayor interés.

Desde el primer momento de la creación del metalenguaje de marcado XML por parte del W3C la comunidad científica y profesional del campo jurídico ha vislumbrado este

desarrollo como una pieza fundamental para el correcto tratamiento de la información de valor jurídico emanada de las distintas instituciones del estado. Y este hecho es así pues, como ya se ha explicado reiteradamente en anteriores capítulos, XML recoge todos beneficios y experiencias aportados por su predecesor, el metalenguaje SGML, para la generación, tratamiento y gestión de documentos electrónicos estructurados pero añade el enorme beneficio de ser servido en un espacio electrónico tan importante en nuestros días como es la World Wide Web.

Así fue contemplado en aquellos momentos por una veintena de abogados y administradores de tribunales de justicia cuando en febrero de 1998 participaron en un seminario “virtual” (en línea) sobre nuevas aplicaciones tecnológicas en el campo de las transacciones de información electrónica en los tribunales de justicia norteamericanos bajo el título de *Electronic Filing of Court Records: Virtual Courthouse or Virtual Courier?*²⁸¹. Este seminario fue patrocinado por el *Counsel Connect*, un servicio de la American Lawyer Media²⁸². Durante el desarrollo de este seminario, el jurista John Messing, de la compañía *Law on Line, Inc.*²⁸³, sugirió que la tecnología XML podría ser empleada de forma satisfactoria para el desarrollo de estándares de aplicación a los documentos jurídicos utilizados en los procesos electrónicos dentro de los tribunales de justicia²⁸⁴. Este investigador atribuyó esta idea sobre el empleo del XML a otro importante jurista, Winchel “Todd” Vincent de la Escuela de Derecho de la Universidad del estado de Georgia, con el cual había mantenido conversaciones anteriores al respecto de este tema.

²⁸¹ Toda la información relativa a este importante evento se encuentra disponible en la dirección <http://home.pacbell.net/webright/personal/CCEFiling/>

²⁸² Empresa mediática norteamericana de gran peso e importancia en el país circunscrita en el campo de la información jurídica. Esta compañía edita más de 25 revistas y periódicos jurídicos de ámbito nacional y regional, promocionando igualmente diversos actos, seminarios y conferencias en la aplicación de las tecnologías informáticas al campo del derecho. Para una mayor información sobre esta empresa, véase <http://www.americanlawyermedia.com/>

²⁸³ <http://www.law-on-line.com/>

²⁸⁴ El texto electrónico original enviado a este seminario por este investigador se encuentra disponible en la dirección <http://home.pacbell.net/webright/personal/CCEFiling/xml.htm>

El *College of Law at Georgia State University* de los Estados Unidos ocupa, como veremos con posterioridad, un papel predominante en esos primeros años de desarrollo y aplicación del estándar XML a la documentación judicial. En un interesante proyecto comenzado a finales del año 1997 para la mejora de los sistemas de procesamiento e intercambio de documentos electrónicos judiciales, del que se hablará en el siguiente apartado de este capítulo, se estudiaron y analizaron los principales formatos de documentos electrónicos existentes en el mercado capaces de dar una respuesta satisfactoria a los objetivos planteados en este proyecto. En dicho informe se resaltó la importancia y validez del formato propuesto por XML frente a otros el RTF, el PDF o el mismísimo SGML para llevar a cabo el desarrollo del proyecto comenzado²⁸⁵.

Pero, sin duda, otras instituciones de mayor calado y alcance nacional e internacional han venido analizando igualmente la oportunidad y validez de este nuevo metalenguaje de marcado de textos electrónicos aplicado a los documentos jurídicos. Destacar en este momento dos de estas instituciones, que serán igualmente tratadas con mayor profundidad en el próximo punto de este apartado, radicadas en los Estados Unidos: el *National Center for State Courts* (NCSC) y la *Washington State Bar Association* (WSBA).

En la comunicación presentada por James E. McMillan (a la sazón, director del NCSC) en la sexta conferencia nacional sobre tecnología aplicada a los tribunales de justicia (CTC6) desarrollada en septiembre de 1999, se analizaba profusamente las características y peculiaridades de los textos jurisprudenciales y las dificultades existentes para difundir estos textos en la red Internet mediante el lenguaje de marcas HTML (principalmente, la flexibilidad inadecuada de este formato para la presentación de información en las múltiples maneras que son requeridas por los editores de información en línea y, de forma especial, las limitaciones para la búsqueda de información precisa dentro de dichos documentos). McMillan concluye su comunicación realizando una apuesta fuerte por el uso de XML para

²⁸⁵ *What is the Best Format for E-CT-Filing?* [documento HTML]. Atlanta, GA: College of Law, Georgia State University, [1998?]. Disponible en <http://gsulaw.gsu.edu/gsuecp/CourtFilings/DocumentFormat/DocumentFormatContents.htm> (consultado el 14 de agosto de 2000).

el tratamiento de los documentos electrónicos de carácter jurídico, los cuales se convierten así en “documentos inteligentes”²⁸⁶.

La *National Association for Court Management* (NACM) *Joint Technology Committee* en colaboración con la *Conference of State Court Administrators* (COSCA), organismos dependientes de la NCSC y de los cuales también hablaremos con posterioridad, elaboraron en el año 2000 una propuesta de estándar para un sistema de fichero electrónico basado en el lenguaje XML y de aplicación a los tribunales de justicia de Norteamérica. Aunque esta propuesta de normalización será analizada más adelante, sí resulta de interés en este momento resaltar algunos de los párrafos ahí contenidos, justificativos de la aplicación de la tecnología XML al entorno de producción documental de carácter jurídico. Así, en el punto tercero de este documento se dice textualmente que²⁸⁷:

XML proporciona una gran oportunidad a las personas y entidades integrantes dentro de un sistema judicial para el intercambio de información entre diferentes sistemas informáticos y bases de datos.

[...] Para los tribunales y los juristas, es posible introducir información dentro de los sistemas automatizados de gestión de casos judiciales y de oficinas legales sin necesidad de que el personal y los secretarios tengan que introducir esta información dentro del ordenador si alguien ya lo hizo con anterioridad.

[...] El uso de XML, soportado por un único estándar consensuado, permitiría la transmisión de datos a múltiples despachos de abogados, agencias y tribunales de justicia -todos ellos utilizando sus propios sistemas de gestión automatizada- sin pérdidas de información. Pero para que esto suceda, los tribunales han de adoptar, y requiere que los juristas y otras entidades con los que el sistema ha de interactuar electrónicamente también lo adopten, un único estándar XML para la transmisión de datos de tipo jurídico.

En términos similares se expresa Roger Winters, del Tribunal Superior del Condado de King en Seattle y miembro del Comité para las Comunicaciones Electrónicas de la WSBA,

²⁸⁶ James E. McMillan. “XML for Courts” [documento HTML]. En: *National Court Technology Conference* (CTC6) (6ª. 1996. Williamsburg). Williamsburg, VA: National Center for State Courts, September 1999. Disponible en <http://www.ncsc.dni.us/NCSC/TIS/TIS99/CTC6/CTC6%20-%20801%20-%20Session.htm> (consultado el 15 de septiembre de 2000).

²⁸⁷ National Center for State Courts. *Proposed LegalXML Court Filing Standard 1.0* [documento PDF]. Williamsburg, VA: COSCA, NACM, 2000. Disponible en <http://www.ncsc.dni.us/NCSC/CTP/HTDocs/pdfdocs/BodyAppendAB.pdf> (consultado el 12 de febrero de 2001).

en dos escritos elaborados a mediados de 1999. En el primero de ellos²⁸⁸, de carácter más generalista, y tras realizar una exposición de los beneficios aportados por los documentos electrónicos para la gestión de la información de carácter jurídico, se exponen las ventajas e inconvenientes de los formatos PDF, HTML y XML, decantándose por este último formato dada la idoneidad del mismo para el tratamiento de datos y documentos electrónicos de carácter jurídico (formateado flexible, ocultación de etiquetas a través de plantillas o formularios, identificación de los datos por su contenido semántico, establecimiento de un estándar principal con variaciones locales, filtrado de los datos por el receptor, soporte de las principales compañías de software, sistema no propietario, etc.). En el segundo de ellos²⁸⁹, más extenso aunque de carácter más específico al circunscribirse al ámbito de actuación de los tribunales de justicia, se expone y justifica la tecnología XML aplicada a este tipo concreto de documentos en dónde la búsqueda de un estándar (una DTD) en el ámbito estatal y, mejor aún, nacional para este tipo de documentos constituirá el principal caballo de batalla de la mayoría de las instituciones y organizaciones de carácter jurídico en los próximos años. Finalmente, este autor señala como el estado de Washington ha creado un comité formado por jueces, administradores, empleados de justicia, miembros de la WSBA y otras personalidades afines a este terreno de actuación, estableciendo una serie de recomendaciones para llevar a cabo el desarrollo de dicho estándar. Estas recomendaciones son:

- Es absolutamente deseable y necesario establecer un desarrollo cooperativo y coordinado de todas las instituciones que estén trabajando en la búsqueda de un estándar XML para el tratamiento de la documentación jurídica, estableciéndose cuál es la mejor iniciativa de las existentes hasta el momento para tomarla como punto central de desarrollo.

²⁸⁸ Roger Winters. *Electronic Legal Documents; an Overview* [documento PPT]. Washington DC: Washington State Bar Association, Electronic Communications Committee, July 13, 1999. Disponible en <http://www.wsba.org/c/ec2/xml/slides/WSBAEC2onXML.ppt> (consultado el 14 de agosto de 2000).

²⁸⁹ Roger Winters. *XML and Electronic Filing Issues for Courts* [documento HTML]. Washington DC: Washington State Bar Association, Electronic Communications Committee, October 12, 1999. Disponible en <http://www.wsba.org/c/ec2/xml/1999/eficourt.htm> (consultado el 14 de agosto de 2000).

- En el caso de trabajar dichas instituciones en paralelo, cada uno de estos esfuerzos debería contribuir al establecimiento de un único estándar de alcance nacional, evitándose, de este modo, una competitividad nada deseable entre estas instituciones.
- El estándar que se desarrolle ha de ser abierto y, por tanto, no propietario.
- Es necesario que el desarrollo de dicho estándar venga acompañado de material didáctico que facilite la implementación práctica por parte de cualquier juzgado y tribunal de la nación.
- Este estándar debe mostrar claramente las identidades estructurales, las funciones y las relaciones de todos los elementos que se dan en un expediente judicial.
- En este desarrollo debería primar la sencillez, centrándose en los datos de relevancia contenidos en los documentos judiciales así como en otros documentos relacionados de carácter jurídico.
- El estándar debe dar soporte a diversas aplicaciones, hojas de estilo y otros productos relacionados de tal modo que cubra perfectamente las posibles variaciones locales existentes en la terminología, los procedimientos, las instituciones y, en definitiva, la cultura de cada estado.
- El estándar debería ser elaborado con cierta rapidez dado que muchos tribunales de justicia están presionando fuertemente para obtener un sistema eficiente de gestión electrónica de documentos.
- El estándar debe englobar el tratamiento de cualquier tipo de material electrónico documental existente en los tribunales de justicia, tales como documentos elaborados a través de procesadores de texto, documentos digitalizados así como otros documentos no basados en el marcado del texto.

Y de hecho esta labor de aplicación de la tecnología y búsqueda de un estándar adecuado para el tratamiento de la documentación judicial se está llevando a cabo, como así lo atestigua R. Plotkin, jurista norteamericano especializado en aplicaciones tecnológicas al campo jurídico y miembro de la prestigiosa organización Legal XML. En uno de sus artículos dedicados a analizar la evolución sufrida hasta nuestros días por los sistemas de edición electrónica de documentos judiciales, este autor reconoce que la capacidad

inherente a este metalenguaje de marcas para especificar el significado del texto proporciona un gran número de beneficios a la hora de tratar documentos electrónicos judiciales pues, de este modo, se pueden identificar dentro de los mismos, por ejemplo, a las partes, el número de registro, el tribunal, etc. A partir de ahí, cualquier sistema automatizado de gestión de casos puede extraer la información marcada en dichos documentos y agregarla a los ficheros que componen el expediente judicial sin que exista intervención humana en ello. Por este importante motivo, este autor apunta que el incremento en el número de personas e instituciones que están trabajando en el desarrollo de estándares para este tipo de sistemas está siendo espectacular en todo el mundo, teniendo la mayoría de ellos al metalenguaje XML como base de trabajo común y a la organización internacional Legal XML, de la cual se hablará profusamente más adelante, como foco principal de la labor de coordinación y estandarización de dichos desarrollos basados en XML²⁹⁰.

Igualmente destacables resultan las opiniones vertidas al respecto por el *Standing Committee on Technology and Information Systems de la American Bar Association* (ABA), asociación de la cual se hablará con posterioridad, al señalar en un artículo reciente que "El poder real de XML está en su habilidad para conectar las diversas comunidades de nuestra profesión dentro de un mismo espacio. XML proporciona los cimientos sobre los que se asientan los sistemas de gestión de casos judiciales, los sistemas de gestión del tribunal, los sistemas de intercambio de ficheros electrónicos y otras tecnologías relacionadas [...]. XML promete incrementar espectacularmente el modo en el que los juristas ejercen su profesión así como la forma en la que se intercambian y comparten los datos y la información de carácter jurídico"²⁹¹.

Circunscribiéndonos al ámbito de actuación jurídica en países europeos, resulta especialmente reseñable el informe emitido en 1999 por la *Commission de l'Informatique, des*

²⁹⁰ Robert Plotkin. "Electronic Court Filing: Past, Present, and Future" [documento HTML]. *Boston Bar Journal*, May/June 2000. Disponible en <http://www.bostonbar.org/bbj/electronic.htm> (consultado el 12 de enero de 2001).

²⁹¹ "Spotlight on XML". *SCOTIS Technology Newsletter*, v. 1, n° 6, June/July 2000. Disponible en <http://www.abanet.org/scotis/vol1no6.html> (consultado el 22 de marzo de 2001).

Réseaux et de la Communication Electronique (COMI) bajo requerimiento del Ministerio de Justicia francés con la finalidad de modernizar el sistema judicial en este país, en el cual se dice expresamente que uno de los temas de estudio e investigación más necesitados está en "los lenguajes interactivos (como el XML), como referencia para el intercambio de datos en el campo judicial."²⁹².

Por último y para no alargar en exceso este punto introductorio dedicado a exponer algunas de las opiniones más destacables sobre la aplicación del metalenguaje XML a la información jurídica resulta ciertamente interesante la consulta de las opiniones vertidas en la *Conference on Congressional Organizations' Application of XML*²⁹³ celebrada en Washington DC el 24 de abril de 2001 y organizada por la anteriormente citada organización y la *House Committee on Administration*, y resumidas en el artículo de A. Kotok²⁹⁴. Aunque en este caso dicho congreso se circunscribe en el campo de actuación de la aplicación del XML a la actividad legislativa y, por tanto, escapa a la materia de interés de esta tesis doctoral, sí es importante señalar, como así lo pone de manifiesto Kotok, los diversos proyectos que están llevando a cabo en este campo, donde el peso y la justificación de la mayoría de éstos recae en la necesidad de suministrar a los ciudadanos información de carácter jurídico a través de la Web.

En el apartado siguiente se expondrán las principales experiencias que se están llevando a cabo en todo el mundo en la aplicación de la tecnología XML al tratamiento y difusión de información de carácter jurídico, encontrándonos con algunos proyectos ya consolidados y perfectamente operativos así como otros aún en fase de desarrollo.

²⁹² Ministère de la Justice. "Dossier: Le ministère de la Justice à l'heure des nouvelles technologies de l'information" [documento HTML]. *Courrier de la Chancellerie*, n° 46, août 1999. Disponible en <http://www.justice.gouv.fr/chancell/cc46ntic.htm> (consultado en 28 de noviembre de 2000).

²⁹³ Toda la información oficial relativa a esta importante conferencia se encuentra disponible en la dirección <http://www.legalxml.org/coax.htm>

III.6.2. PRINCIPALES EXPERIENCIAS

En el capítulo anterior dedicado al metalenguaje SGML se expusieron de modo genérico algunas de las experiencias y proyectos más interesantes que se han llevado a cabo en diversos países del mundo en el ámbito de la actividad jurídica. En dicho apartado se mostraron estas experiencias sin hacer una distinción clara entre los proyectos orientados al tratamiento de los documentos electrónicos de carácter legislativo y aquellos otros orientados al tratamiento de documentos electrónicos de carácter judicial pues en ese momento del discurso científico de esta tesis doctoral resultaba esclarecedor señalar dichas aplicaciones en el ámbito global del mundo jurídico. Sin embargo, dado que el tema central de desarrollo de esta tesis doctoral está circunscrito al tratamiento de los documentos electrónicos jurisprudenciales mediante tecnologías derivadas del metalenguaje XML parece oportuno centrarse en señalar y describir aquellos proyectos e iniciativas nacionales e internacionales que exclusivamente tratan este tipo de información, dejando las iniciativas XML que se están desarrollando en la actualidad en el campo de la documentación electrónica legislativa a otras tesis doctorales que, como ya hemos señalado en varias ocasiones, se están elaborando actualmente en el seno de nuestro Departamento universitario.

III.6.2.1. ESTADOS UNIDOS

Es, sin duda, en los Estados Unidos donde de forma más clara y decidida se han emprendido la mayor parte de los proyectos de aplicación de las tecnologías XML a la documentación gestionada en los tribunales de justicia. Tanto las diferentes

²⁹⁴ Alan Koto. *Can XML Help Write the Law?* [documento HTML]. XML.com, May 9, 2001. Disponible en <http://www.xml.com/lpt/a/2001/05/09/legalxml.html> (consultado el 21 de mayo de 2001).

administraciones gubernamentales federales como estatales con competencias en la materia, así como los propios tribunales de justicia de este país han apostado firmemente por el estudio, investigación y aplicación de las tecnologías XML al tratamiento de dicha documentación, entrando en una clara competición entre diferentes proyectos de envergadura para la definición y establecimiento de un estándar XML adecuado para el tratamiento de la información judicial. La mayoría de los proyectos que se han venido sucediendo en este país se han enmarcado dentro de un proceso mayor y más global de mejora de los sistemas automatizados para el tratamiento y gestión de los expedientes judiciales, donde, como ya se ha venido exponiendo en capítulos precedentes, la búsqueda de un formato adecuado de documento electrónico para las transacciones de información entre las partes intervinientes en un proceso judicial ha constituido la piedra angular de muchos de estos sistemas.

Los primeros proyectos de investigación aplicada que se emprendieron en este país coincidieron, como ya se ha señalado con anterioridad, con el propio lanzamiento del metalenguaje XML en el año 1998. Este hecho parece lógico pues muchas de las instituciones aquí señaladas habían venido experimentando con las tecnologías derivadas del metalenguaje SGML y con el modelo EDI para el intercambio de datos y documentos jurídicos así como, y de forma especial en los últimos años, con el lenguaje HTML y la publicación de textos jurisprudenciales en el espacio electrónico de la WWW. La utilización de XML era, por tanto, la continuación lógica de dichos proyectos de investigación aplicada.

Es importante señalar igualmente que estos primeros proyectos realizados en los Estados Unidos se han venido circunscribiendo a centros judiciales tecnológicamente muy activos (tribunales de justicia de ámbito local, comarcal o estatal) o en instituciones universitarias de prestigio donde el entusiasmo de unas cuantas personas por el estudio y aplicación de la informática jurídica documental ha conseguido que sea ésta una materia o disciplina de estudio de gran interés por parte de sus facultades y escuelas de derecho. A partir de ahí, esta ilusión por la utilización de las tecnologías derivadas del metalenguaje XML ha contagiado a otros centros e instituciones gubernamentales con competencias en esta materia, así como a ciertas asociaciones y organizaciones profesionales del mundo del

derecho, que provocó con posterioridad que desde ahí se emprendiesen o canalizasen muchos de los actuales esfuerzos. Empecemos, pues, señalando algunos de los más importantes proyectos de aplicación de las tecnologías XML a la documentación judicial emprendidos en focos locales de este país.

Uno de los primeros proyectos norteamericanos emprendidos fue el llevado a cabo a comienzos de 1997 por el **College of Law at Georgia State University**²⁹⁵ en colaboración con la **Escuela de Negocios de J. Marck Robinson**²⁹⁶ de dicha Universidad y financiado por la Georgia Courts Automation Commission. La finalidad inicial de este proyecto, conocido con el nombre de **Electronic Court Filing Project (E-CT-Filing)**²⁹⁷ y bajo la dirección del anteriormente citado Winchel "Todd" Vincent, era la de analizar los diferentes sistemas de tratamiento electrónico existentes para la mejora de los procesos de gestión e intercambio de documentos electrónicos entre los tribunales de justicia de dicho estado norteamericano siendo el metalenguaje XML, como se expuso en la introducción de este punto, el formato de codificación de documentos electrónicos más destacado por este grupo de investigación. A finales de 1997 este grupo inicial convocó a veinticinco miembros de la comunidad jurídica del estado de Georgia para mostrar los primeros resultados del trabajo emprendido y, sobre todo, para concienciar a esta comunidad de la importancia en el desarrollo de estándares comunes para los tribunales de justicia de dicho estado. En 1998, con el metalenguaje XML ya reconocido por la W3C como estándar oficial, este grupo de trabajo se une con el grupo de investigadores de la Universidad de Utah (los cuales venían trabajando en el proyecto *Utah Electronic Law & Commerce Partnership (UELCP)*²⁹⁸) y con el investigador Gabe Wachob del portal jurídico FindLaw²⁹⁹ para la

²⁹⁵ <http://law.gsu.edu/>

²⁹⁶ <http://www.cba.gsu.edu/robinson/home.asp>

²⁹⁷ Toda la información oficial relativa a dicho proyecto se encuentra disponible en la dirección <http://gsulaw.gsu.edu/gsucfp/>

²⁹⁸ <http://www.uelcp.org/default.cgi>

colaboración científica entre todas estas instituciones y la revitalización del primitivo grupo de trabajo sobre XML creado un año antes en la Universidad de Utah³⁰⁰. El nuevo grupo de trabajo de la Universidad del estado de Georgia y las otras dos instituciones pone en marcha en noviembre de 1998 un servicio de listas de distribución de correo electrónico para el debate en línea entre los investigadores del país interesados en la materia (en aquellos momentos participaban unos 18 miembros). Este primitivo grupo de discusión y trabajo ha venido siendo considerado como el verdadero inicio de la organización internacional Legal XML. A este grupo de trabajo se fueron incorporando rápidamente otros miembros (en marzo de 1999 se había llegado a los 45 miembros y en el verano de ese mismo año se llegó hasta los 60) incrementándose, consecuentemente, el tráfico de mensajes electrónicos circulantes en la lista de distribución. Desde este grupo de trabajo de la Universidad de Georgia se veía necesario mantener algún encuentro presencial entre todos los miembros participantes en la lista de distribución para tratar de darle a este grupo una mejor estructura operativa y establecer las diversas competencias en la dirección y gobierno del mismo. Este hecho se conseguiría con posterioridad gracias a la incorporación de otras importantes instituciones, entre las que destacó el NCSC, citado con anterioridad y analizado con mayor profundidad posteriormente³⁰¹.

²⁹⁹ Se trata éste de un clásico portal web dedicado íntegramente a la localización de recursos existentes en Internet de interés jurídico. El sitio web de FindLaw se estructura de forma similar a los directorios o listas temáticas de recuperación de información (siendo su más claro exponente el buscador Yahoo!), existiendo un primer nivel estructural según el tipo de usuario que realiza la consulta (profesionales de la actividad jurídica, estudiantes, sectores comerciales o de negocios, recursos humanos y público en general). A partir de ahí, el directorio se va subdividiendo jerárquicamente en otros tantos apartados. Este sitio web dispone igualmente de un motor de búsqueda para la localización directa del recurso o recursos materia de nuestro interés. FindLaw es accesible en la dirección <http://www.findlaw.com/>

³⁰⁰ La dirección electrónica de este primitivo pero importante grupo de trabajo liderado por Brent Isrealson de la Universidad de Utah se encuentra en <http://www.uelcp.org/xml.cgi> Este grupo de trabajo se creó con la filosofía y los objetivos iniciales del debate intelectual y el desarrollo de una o varias DTDs XML normalizadas para la edición e intercambio de documentos electrónicos jurídicos. Desafortunadamente, este grupo de trabajo se quedó en la simple discusión, no llegando a desarrollar ningún resultado real.

³⁰¹ Tan sólo apuntar aquí que en la actualidad toda la información relativa a las investigaciones y trabajos que se están llevando a cabo en el estado de Georgia sobre la aplicación del metalenguaje XML a la información jurídica de los tribunales de dicho estado se encuentra centralizada en el grupo de trabajo creado a tal fin en el seno de la organización LegalXML. Para una más amplia información, véase la página web relativa a dicho grupo de trabajo en la dirección <http://www.legalxml.org/Georgia/>

Independientemente a la labor emprendida por las anteriores instituciones y personas, resulta igualmente encomiable la labor llevada a cabo durante esos años por Richard Himes (a la sazón, otro brillante tecnólogo amigo directo de Alan Asay) dentro del **United State District Court, District of New Mexico**³⁰². Este importante tribunal federal del distrito de Nuevo Méjico constituye en la actualidad, sin lugar a dudas, uno de los tribunales de justicia más avanzados tecnológicamente de los Estados Unidos e, igualmente, de todo el mundo³⁰³. En el caso concreto del tratamiento de documentos electrónicos de carácter judicial a través de la tecnología XML resulta ser este tribunal otro de los centros iniciales y punteros en esta actividad. Así, en la búsqueda incesante de métodos para la mejora y desarrollo de iniciativas tecnológicas de aplicación judicial, su investigador principal, R. Himes, ha venido trabajando desde mediados de 1998 en el desarrollo conocido como **Extensible Markup Language Court Interface (XCI)**³⁰⁴ con la finalidad de alentar el desarrollo e investigación de estándares independientes para los sistemas de edición y transmisión electrónica de documentos judiciales. Como se indica en la propia página web informativa de este desarrollo, el XCI está basado en la premisa de que los sistemas cerrados y propietarios no son admisibles para el colectivo profesional jurídico (los litigantes, abogados y para los propios tribunales) dado el elevado precio que éstos alcanzan en el mercado y el encasillamiento al cual se ven sometidos los tribunales en el empleo de tecnologías que son difíciles de soportar cuando existen otras alternativas tecnológicas más prometedoras ya disponibles, caso del XML. Con este desarrollo, toda la comunicación

³⁰² <http://www.nmcourt.fed.us/dcdocs/>

³⁰³ Resulta fascinante leer la información suministrada al respecto de este tema por el propio tribunal pues el nivel de tecnificación judicial es impresionante. Así, las diversas salas de este tribunal cuentan con un sistema electrónico de transcripción en tiempo real, sistema de audio y vídeo conferencia, cámara de vídeo digital para la presentación y proyección de documentos y objetos en tres dimensiones, pizarra digital para el señalamiento manuscrito de dibujos o pruebas grafológicas, monitores de 15 pulgadas tanto para los jueces, abogados y miembros del jurado, sistema de escucha y traducción simultánea por infrarojos, etc. Para una más amplia información, véase http://www.nmcourt.fed.us/dcdocs/court_technology.html

³⁰⁴ A pesar de que la información relativa a este importante desarrollo se debería encontrar disponible en la dirección electrónica de su sitio oficial (<http://www.nmcourt.fed.us/xci/xcihome.htm>), algunos documentos electrónicos de importancia han sido eliminados de este servidor Web. Por ello, recomendamos que para consultar uno de los mejores documentos electrónicos explicativos del funcionamiento del XCI sea acceda a

documental que se produce entre este tribunal de justicia y los diversos interlocutores se realiza en un formato electrónico normalizado a través de XML.

En junio de 1999, y dado el enorme interés suscitado por este desarrollo en la comunidad judicial norteamericana, la Administrative Office of the US Courts³⁰⁵ solicitó a este equipo de investigación una demostración operativa del sistema implementado con el fin de promocionar entre los tribunales de justicia de titularidad federal el uso de la tecnología XML y la búsqueda común de un estándar adecuado para todos ellos. Esta demostración práctica fue desarrollada por el propio Rich Himes bajo la dirección de Robert M. March, administrativo del citado tribunal de justicia³⁰⁶. Este desarrollo constaba de tres piezas básicas: una cubierta XML para la estructura de los datos contenidos en los mensajes, software servidor para la traducción desde el formato HTML de la Web a y desde XML, y software del servidor XCI para el procesamiento y generación de los mensajes. El sistema empleaba, además del propio XML, otras tecnologías informáticas abiertas, como era el caso del sistema operativo Linux, el servidor web Apache, Java Servlets y otros programas informáticos complementarios para garantizar la seguridad en dichas transacciones de información electrónica (SSL, firmas y certificados digitales, etc.).

Sirva como ejemplo ilustrativo de este complejo sistema la tipología de mensajes electrónicos marcados con XML para la comunicación entre este tribunal de justicia y los abogados y procuradores intervinientes en un caso judicial. Dichos mensajes eran divididos, de forma general, en dos clases distintas: los que llegaban al tribunal (mensajes de súplica o petición) y los que emitía esta institución (mensajes de réplica o respuesta). El formato general de cada uno de estos mensajes era el siguiente³⁰⁷:

través de la copia local existente en el servidor Web de la organización OASIS, en la dirección <http://www.oasis-open.org/cover/xcispec19990402.html>

³⁰⁵ <http://www.uscourts.gov/>

³⁰⁶ Toda la información técnica relativa a esta demostración práctica se encuentra disponible en la dirección <http://www.nmcourt.fed.us/xci/xcidemo.htm>

³⁰⁷ Ejemplos obtenidos y traducidos del documento anteriormente citado sobre el XCI disponible en la dirección <http://www.oasis-open.org/cover/xcispec19990402.html>

Mensaje de súplica	Mensaje de réplica
<pre><request type="filing" sealed="yes / no" message_id="id del mensaje"> <court> <type> tipo de tribunal </type> <name> nombre del tribunal</name> <location> dirección del tribunal</location> </court> <case> <id> identificación del caso en el formato del tribunal </id> <year> año del caso </year> <type> tipo de caso</type> <number> número de secuencia del caso </number> <title> título del caso </title> </case> <document> <type> tipo de documento </type> <docket_template> plantilla de entrada del legajo </docket_template> <options> opciones </options> <signed> <file attributes> fichero del documento </file> <signature attributes> firma del abogado </signature> otras firmas requeridas </signed> </document> <attachment> <name> archivo adjunto o nombre expuesto </name> <signed> <file attributes> fichero adjunto </file> <signature attributes> firma del abogado </signature> otras firmas requeridas </signed> </attachment> otras ficheros adjuntos requeridos <certificate attributes> certificación del abogado </certificate> otras certificaciones requeridas </request></pre>	<pre><response type="filing" status="ok / error / partial / delayed / duplicate" sealed="yes / no" message_id="id del mensaje"> <message> respuesta textual al mensaje </message> <document status="ok / error"> <file_stamp> <court> <type> tipo de tribunal </type> <name> nombre del tribunal </name> <location> dirección del tribunal </location> </court> <case> <id> identificación del caso en el formato del tribunal </id> <year> año del caso </year> <type> tipo de caso </type> <number> número de secuencia del caso </number> <title> título del caso </title> </case> <type> tipo de documento </type> <description> texto generado del legajo </description> <docket_number> número de legajo </docket_number> <date_filed> fecha de archivado </date_filed> <time_filed> hora de archivado </time_filed> <court_signature attributes> firma del tribunal </court_signature> </file_stamp> </document> <attachment status="ok / error"> <name> nombre del archivo adjunto </name> <court_signature attributes> firma del tribunal </court_signature> </attachment> otras ficheros adjuntos <court_response_document> documento de respuesta del tribunal</court_response_document> <certificate attributes> certificación del tribunal </certificate> </response></pre>

Figura III.7: Marcado XML en mensaje de súplica y en mensaje de réplica según el lenguaje XCI.

En la actualidad, este desarrollo se encuentra integrado dentro de un sistema mayor de gestión automatizada del tribunal de justicia de Nuevo Méjico, denominado *Advanced Court*

Engineering (ACE)³⁰⁸, desarrollado íntegramente por el departamento de sistemas de información de dicho tribunal.

Cambiando de estado norteamericano, resultan especialmente destacables todas las iniciativas que se han venido produciendo en California, algunas de ellas emprendidas como iniciativa del sector privado o comercial.

Rich Himes, tras la experiencia adquirida en el tribunal del distrito federal de Nuevo Méjico, abandona este centro y marcha al California para incorporarse a la compañía @Court, Inc.³⁰⁹ Es aquí donde desarrolla la iniciativa conocida bajo el nombre de **Open XML Court Interface** (OXCI). Esta nueva experiencia de R. Himes, continuación lógica del anterior desarrollo XCI y, por tanto, de similares características, fue puesta en marcha en marzo de 2000 con la intención de desarrollar un estándar abierto para un sistema de intercambio de documentos electrónicos de interés judicial dentro de los tribunales del estado de California. Este proyecto, a diferencia de otros muchos adscritos al campo jurídico, no fue inscrito oficialmente dentro de la organización Legal XML para su desarrollo prefiriendo su autor que dicho desarrollo fuese debatido dentro de la *SourceForge*, organización dedicada al desarrollo y promoción de estándares abiertos³¹⁰. Este hecho no ha implicado que la organización Legal XML no se haya interesado por su parte en este proyecto, siendo finalmente admitido como un *working draft* para su estudio y debate entre

³⁰⁸ Toda la información oficial relativa a este desarrollo, así como un gran número de documentos y manuales de uso, puede ser consultada en la dirección <http://www.nmcourt.fed.us/dcdocs/ace.html>

³⁰⁹ Esta empresa radicada en la ciudad de Alameda, California, está especializada en el suministro de servicios de soporte de pleitos a través de Internet para abogados y tribunales de justicia. Todas las soluciones tecnológicas que ofrece esta empresa para la transmisión y gestión de información judicial tienen como núcleo central al metalenguaje XML. Para una mayor información, véase en la dirección <http://www.atcourt.com/>

³¹⁰ A pesar de la importancia de este proyecto el eco producido no ha sido elevado debiéndose, tal vez, al adscribir el proyecto a esta organización, con muy buenas intenciones pero de escaso peso en el panorama normativo. En cualquier caso, véase la información suministrada al respecto de este proyecto en la citada organización, en la dirección <http://sourceforge.net/projects/oxci/>

los profesionales adscritos a dicha organización³¹¹. OXCI es una metodología para el procesamiento de documentos XML donde, en el plano arquitectónico, se entremezclan diversos componentes para formar un conjunto homogéneo de aplicación con una fuerte orientación a objetos, definidos éstos como interfaces de aplicación de programa (API). El desarrollo OXCI define el software intermedio encargado de recoger los mensajes electrónicos procedentes de una o más agencias, de autenticar al emisor del mismo, de validar que dichos mensajes se ajustan al formato establecido por el tribunal, de convertir y enviar estos datos a los sistemas automatizados para la gestión de casos judiciales así como el propio para la gestión de documentos y, por último, de enviar las respuestas oportunas a las agencias emisoras de los mensajes. El siguiente diagrama, contenido en el borrador de trabajo del desarrollo OXCI, ilustra perfectamente el *modus operandi* de este sistema:

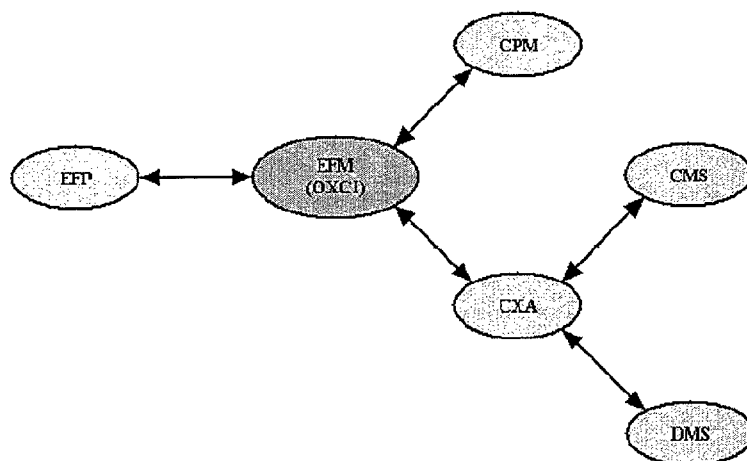


Figura III.8: Diagrama de conexión de los diversos módulos integrados en el modelo del OXCI.

Fuente: R. Himex. "Open XML Court Interface (OXCI)".

<http://www.legalxml.org/California/files/OXCIArchitecture2.pdf>

³¹¹ Richard Himes. *Open XML Court Interface (OXCI). Working Draft* [documento PDF]. LegalXML, October 15, 2000. Disponible en <http://www.legalxml.org/California/files/OXCIArchitecture2.pdf> (consultado el 16 de febrero de 2001).

El EFP (*Electronic Filing Providers*) representa a las agencias externas a los tribunales que son capaces de suministrar documentos electrónicos normalizados según el estándar en desarrollo de la Legal XML. El EFM (*Electronic Filing Manager*) es el programa de software definido por OXCI para la intermediación entre todos los componentes. El CPM (*Court Policy Manager*) es el gestor de la política establecida por el tribunal empleado para efectuar un cierto nivel de validación sobre los documentos electrónicos, que han de estar adaptados a las reglas establecidas en el ámbito local por el tribunal. El CXA (*Court XML Adapter*) es el adaptador que define la interfaz para la manipulación de documentos específicos del tribunal así como otras solicitudes. El CMS (*Court Management System*) y el DMS (*Document Management System*) son los sistemas locales para la gestión automatizada de los expedientes y documentos judiciales.

Mención igualmente destacada dentro del estado de California resulta la labor emprendida por el **Orange County Superior Court**³¹². En este caso, se trata de una iniciativa conjunta del sector público con el sector privado pues en el ambicioso proyecto llevado a cabo por este tribunal ha venido participando activamente además del citado tribunal, la Orange County Bar Association (OCBA)³¹³ y la poderosa compañía comercial norteamericana West Group³¹⁴, especializada en la publicación y suministro de información jurídica con alto valor añadido a la comunidad profesional de este país. El desarrollo comercial obtenido del proyecto emprendido a principio de los años noventa lleva por nombre WestFile³¹⁵. Como señala Alan Slater, oficial jefe ejecutivo del citado tribunal se trata éste del primer proyecto en el que verdaderamente se integran aspectos tales como el intercambio de documentos electrónicos a través de la Web, la gestión de expedientes judiciales y los sistemas de gestión de documentos utilizando para ello las tecnologías

³¹² <http://courts.oc.ca.gov/>

³¹³ <http://www.ocbar.org/>

³¹⁴ <http://www.westgroup.com/>

³¹⁵ <http://www.westfile.com/>

XML³¹⁶. Esta iniciativa empezó implantándose en la división de derecho de familia del citado tribunal de justicia para extenderse con posterioridad a las divisiones de derecho civil y probatorio³¹⁷.

Pero no son éstas las únicas iniciativas de interés que se han producido en el estado de California pues nos encontramos, sin duda, ante uno de los estados de Norteamérica donde las instituciones gubernamentales de ámbito estatal con competencias en este materia han resultado ser más activas y productivas en el fomento de la investigación aplicada del lenguaje XML al tratamiento de los documentos judiciales. Así, la Administrative Office of the Courts (AOC)³¹⁸ se configura como la institución central canilizadoradora e impulsora de la mayoría de los proyectos que se abordan en este estado. Este centro tiene como misión principal la de potenciar la excelencia de los tribunales de este estado para beneficio de sus ciudadanos. Entre las doce divisiones en las que se estructura la AOC resulta destacable la División de Servicios de Información, la cual coordina las diversas iniciativas tecnológicas emprendidas a lo largo del estado, gestiona de manera centralizada dichos esfuerzos y optimiza el alcance y la accesibilidad de la información emanada de los tribunales de justicia del estado. Una de las múltiples iniciativas emprendidas por esta institución hace referencia a la búsqueda de estándares comunes relativos a los diferentes elementos que entran en juego en el desarrollo de verdaderos tribunales electrónicos de justicia. Este importante proyecto es conocido bajo el nombre de **California Electronic Filing Technical Standards** (CEFTS)³¹⁹ y donde la tecnología XML alcanza un papel principal en la búsqueda de este estándar. El proyecto cubre un gran número de áreas o apartados tratados, cada uno de los cuales contempla su propia especificación (algunas requeridas,

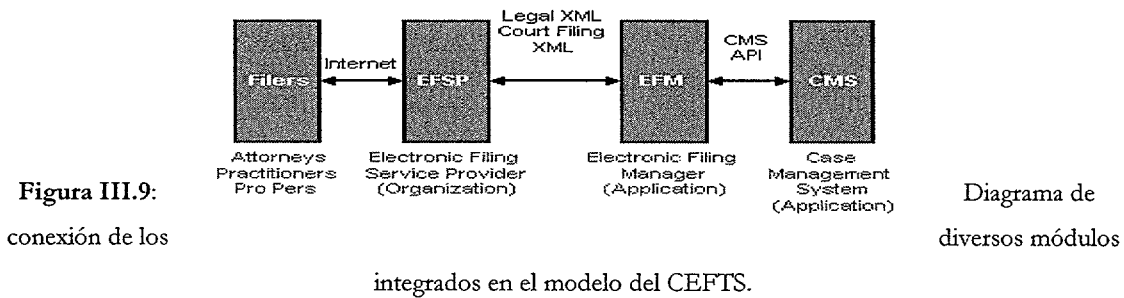
³¹⁶ Alan Slater. "Letters to the Editor: Seeing Red Over Orange" [documento HTML]. *Law Technology News*, October 2000. Disponible en http://www.lawtechnews.com/october00/letters_to_the_editor_p06.html (consultado el 25 de enero de 2001).

³¹⁷ Para una más completa información sobre la implantación y desarrollo de esta iniciativa, véase la información suministrada al respecto por la propia compañía comercial West Group en la dirección <http://www.westfile.com/News/OCBAwestfile.htm>

³¹⁸ <http://www.courtinfo.ca.gov/courtadmin/aoc/>

otras recomendadas y en algunos casos, opcionales)³²⁰, tales como la definición de los formatos documentales aceptados por estos tribunales (PDF, TIFF, JPEG, GIF, etc.), la interacción con las bases de datos jurisprudenciales, las firmas electrónicas y el cifrado de los documentos, el sistema de gestión automatizada de los expedientes judiciales, servicios electrónicos de noticias, mecanismos para el pago electrónico de las tasas, y un largo etcétera de elementos de interés lo que convierten a esta iniciativa en una de las más destacadas tanto dentro de los Estados Unidos como fuera de las fronteras de este país.

El modelo conceptual de este sistema es ciertamente similar al propuesto por R. Himes para el desarrollo del OXCI, donde existe una interacción entre los diversos módulos que integran el sistema, como queda reflejado de forma esquemática en el siguiente gráfico:



Fuente: AOC. "California Electronic Filing Technical Standards Project".

<http://www.courtinfo.ca.gov/reference/documents/efilingstand.pdf>

Este modelo define cuatro elementos principales para el sistema de tratamiento e intercambio de información electrónica judicial de ámbito estatal (para el estado de California), pero sin renunciar a una posible ampliación consensuada a escala nacional:

³¹⁹ La información oficial relativa a este importante proyecto se encuentra disponible en la dirección <http://www.courtinfo.ca.gov/reference/cefts.htm>

³²⁰ El documento en cuestión en el cual quedan reflejados cada una de estas áreas de aplicación se encuentra en su versión 6 a fecha de 9 de marzo de 2001. Véase Administrative Office of the Courts. *Electronic Filing Technical Standards Project: Technical Standards. Version 6* [documento PDF]. San Francisco, CA: AOC, March 9, 2001. Disponible en <http://www.courtinfo.ca.gov/reference/documents/efilingstand.pdf> (consultado el 25 de junio de 2001).

1. Suministradores de ficheros (*Filers*): abogados particulares, bufetes de abogados, litigantes, agencias del estado y del condado y cualquiera otra institución o persona que tenga que intercambiar documentos electrónicos con un tribunal de justicia.
2. Proveedores de servicios de tratamiento de documentos electrónico (EFSP): éstos son empresas que proporcionan servicios de edición, tratamiento y transmisión de documentos electrónicos de carácter judicial para sus clientes (*filers*). En muchos casos, se tratarán de empresas comerciales pero indudablemente pueden existir organizaciones sin ánimo de lucro que se dediquen a dicha labor e incluso los propios tribunales.
3. Software de gestión de documentos electrónicos judiciales (EFM): se trata de una aplicación informática que acepta documentos XML desde un EFSP, analiza estos documentos, traslada los datos contenidos al siguiente módulo de gestión de expedientes (CMS), salva dichos documentos si es que el CMS no está equipado para ello y devuelve al EFSP la respuesta generada por el CMS en formato XML.
4. Sistemas de gestión de casos judiciales (CMS): son las aplicaciones empleadas en los tribunales para la gestión electrónica de los expedientes que se generan en el desarrollo de un caso judicial.

Por todo ello, como se observa en el propio gráfico, el objetivo fundamental de este proyecto está en el establecimiento de especificaciones técnicas que permitan establecer de forma normalizada, en principio para todos los tribunales de justicia de este estado, un sistema automatizado que agilice dichas transmisiones de información entre las diversas partes interesadas de forma ágil, consistente y válida. Es importante comentar nuevamente que entre los formatos de documentos electrónicos anejos al envoltorio XML que son aceptados por los tribunales de justicia del estado de California se encuentran aquellos que son aceptados de forma universal (PDF y TIFF, principalmente). Sin embargo, esta especificación contempla que igualmente estos documentos electrónicos podrán venir en un futuro codificados en origen en formato XML aunque ello está limitado al desarrollo y definición futura de tipos documentales (a través de las correspondientes DTDs o esquemas XML) de ámbito nacional, con las lógicas adaptaciones a la casuística de los

tribunales de cada estado. En cualquier caso, esta especificación remite a lo dispuesto al respecto de este tema dentro del estándar XML en desarrollo de mayor alcance internacional: el *XML Court Filing* de la organización Legal XML, del cual se hablará con posterioridad.

Señalaremos igualmente para las iniciativas emprendidas en el estado de California que la organización Legal XML mantiene un grupo de trabajo específico para los desarrollos llevados a cabo en este estado; el **California Workgroup**³²¹, al cual remitimos para una mayor información sobre la importante labor que se viene realizando durante los últimos años.

Destacaremos por último dentro del apartado de las iniciativas emprendidas a escala estatal dentro de los Estados Unidos la labor que viene realizando desde hace ya varios años el **Grupo de Estudio XML** de la **Washington State Bar Association**³²², radicada en la ciudad de Seattle. Como se indica expresamente en sus páginas web, la misión de este grupo de estudio recae principalmente en los siguientes puntos:

El estudio y aplicación de la tecnología XML como herramienta de aplicación para la construcción de documentos digitales de carácter jurídico.

La constitución y desarrollo de un organismo público y organizado para la comprensión del XML como herramienta de utilidad para la comunidad jurídica del estado de Washington.

La promoción e implementación de los desarrollos emprendidos como método de acceso libre y abierto a los documentos electrónicos jurídicos.

Este grupo de estudio ha venido exponiendo sus líneas de trabajo en diversos congresos y encuentros organizados al respecto de este tema, así como en diversos documentos y presentaciones electrónicas realizados por algunos de sus miembros destacados, caso de

³²¹ Véanse en la dirección <http://www.legalxml.org/California/> los diversos documentos que se han venido generando fruto de la intensa actividad llevada a cabo por este grupo de trabajo.

³²² Toda la información oficial relativa a este activo grupo de estudio se encuentra disponible en la dirección <http://www.wsba.org/c/ec2/xml/home.htm>

Roger Winters, del cual ya hemos referenciado con anterioridad algún documento suyo en este capítulo.

Existen, sin embargo, otras organizaciones en los Estados Unidos de carácter supraestatal que están tratando de llevar a cabo la coordinación de todas las iniciativas desarrollados en los diversos estados norteamericanos con la finalidad de obtener un estándar que sea a la vez común para todos los tribunales de justicia de la nación pero con la suficiente flexibilidad para adecuarse a la variaciones locales que se pudieran producir. Entre este tipo de organizaciones son destacables el National Center for State Courts (NCSC), la Conference of State Court Administrator (COSCA), la National Association for Court Management (NACM) y la American Bar Association (ABA).

El **National Center for State Courts** (NCSC)³²³ es una organización sin ánimo de lucro fundada en 1971 con el propósito de mejorar los servicios que prestan los tribunales de justicia estatales en los Estados Unidos. Para llevar a cabo esta misión esta organización se estructura en diversos comités y divisiones, cada una de ellas tratando un aspecto destacable en el desarrollo de misión. una de esas importantes divisiones la constituye la *Technology Division*³²⁴ dentro de la cual destaca sobremanera el *Court Technology Laboratory*³²⁵. Una de las líneas principales de investigación de este laboratorio se sitúa en el estudio y aplicación del metalenguaje XML a la producción, gestión y transmisión de documentos electrónicos judiciales.

En el caso que nos ocupa, ya se expuso en el apartado II.4.1 la importante labor que esta organización había llevado a cabo en el estudio y aplicación en los tribunales de justicia norteamericanos de la aplicación conocida por el nombre de JEDDI (*Judicial Document and Data Interchange*) para el intercambio de datos y documentos electrónicos entre estos tribunales y las partes intervinientes en un proceso judicial. En esa búsqueda constante de

³²³ <http://www.ncsconline.org/>

³²⁴ <http://www.ncsc.dni.us/NCSC/CTP/index.html>

una tecnología que satisfaga los requisitos demandados por los profesionales de la judicatura, esta organización apostó fuertemente por el nuevo metalenguaje XML en detrimento del anterior desarrollo, por todas las razones anteriormente citadas. Tal ha sido la apuesta de esta organización por el estudio y aplicación de la tecnología XML que ya a principios del año 1999 fue lanzada una propuesta de estandarización basada en esta tecnología para su estudio y debate entre la comunidad profesional con una doble intencionalidad: por un lado, se pretendía elaborar un diccionario que normalizase la terminología empleada en el marcado de los documentos XML de carácter judicial y, como extensión de esta propuesta, la definición inicial de una DTD XML para el marcado de dichos documentos. Este borrador, cuyo título es el de *Concepts for a Judicial XML Namespace & Data Tag Dictionary*³²⁶, fue elaborado por James E. McMillan, de la NCSC, y por Stephanie Rondenell, del Alpha Consulting Group, Inc., con el patrocinio de la todo poderosa empresa editorial de material jurídico Lexis-Nexis³²⁷. Esta DTD propuesta, aunque de gran interés por el esfuerzo de normalización y de aplicación a todos los tribunales de justicia de los distintos estados norteamericanos, contemplaba una definición

³²⁵ <http://ctl.ncsc.dni.us/>

³²⁶ James E. McMillan, Stephanie Rondenell. *Concepts for a Judicial XML Namespace & Data Tag Dictionary. Draft 1.0* [documento HTML]. NCSC, May, 1999. Disponible en <http://ctl.ncsc.dni.us/xmlconcept2.htm> (consultado el 14 de agosto de 2000). El diccionario de etiqueta de marcado fue elaborado, como se indica en el propio documento, por Stephanie Rondenell a través de la recopilación de diversos diccionarios de datos jurídicos existentes en más de una docena de sistemas integrados de gestión de casos judiciales. Todos ellos fueron compilados, reestructurados y almacenados en tres documentos independientes en formato de hoja de cálculo de MS Excel 97/2000: Criminal, Civil y Juvenil. El fichero comprimido con estas tres hojas de cálculo se encuentra disponible en la dirección <http://ctl.ncsc.dni.us/bbsfiles/xml/crtdata.zip>

³²⁷ Aunque no se va analizar aquí el peso y la importancia de esta potente e influyente casa editorial norteamericana LEXIS-NEXIS (<http://www.lexis-nexis.com/>), sí resulta importante aclarar que el papel desempeñado por ésta en el desarrollo y aplicación de estándares basados en XML para el tratamiento de los documentos jurídicos, legislativos y jurisprudenciales, ha sido fundamental, apoyando desde el sector privado un gran número de iniciativas y proyectos de investigación en este campo. Sirva como ejemplo, la conferencia pronunciada por Don Steiner en 1999 para la Southern Ohio Chapter de la prestigiosa *American Society for Information Science* (ASIS) sobre el uso de los lenguajes de marcado de texto, y en especial XML, para el tratamiento de la información jurídica (información disponible en <http://www.asis.org/Chapters/soasis/events/19990513.html>). En la actualidad, esta compañía hace uso de la tecnología XML en todos los servicios que proporciona a la comunidad jurídica a través de la WWW, apostando en este caso por la aplicación y uso del software comercial Microsoft SQL Server 2000 para el almacenamiento y gestión de sus bases de datos XML. Véase dicha información en la dirección <http://www.microsoft.com/servers/evaluation/casestudies/lexnex.asp>

estructural y de contenido ciertamente inadecuada debido, y según nuestro propio parecer, a la inclusión de extensos y ambiguos modelos de contenido para cada uno de los elementos incluidos en la misma y la falta de desarrollo de algunos apartados (caso, por ejemplo, de la cabecera del registro). Así, por ejemplo, el elemento raíz o elemento de documento, denominado *courtdocument*, contemplaba más de veinte elementos hijos, todos ellos definidos con una frecuencia de aparición de cero a infinito (utilización del signo del asterisco en la definición del modelo de contenido). En cualquier caso y para una mayor aproximación a esta DTD recomendamos la consulta del documento original elaborado por los autores antes citados, incluido como Apéndice 2 de esta tesis doctoral, recogiendo íntegramente la definición de tipo documental formulada.

Señalar igualmente aquí, que la labor emprendida por la NCSC en pos de establecer líneas comunes de investigación entre los diversos grupos estatales que estaban debatiendo e investigando sobre las aplicaciones del metalenguaje XML a la documentación judicial dio como fruto en aquellos años el desarrollo de diversos encuentros, seminarios y conferencias: el 15 de septiembre de 1999, bajo el seno de la sexta conferencia nacional sobre tecnologías de aplicación en los tribunales de justicia (CTC6)³²⁸, se reunirían en Los Ángeles representantes de la NCSC, del New México State Courts y de la COSCA y la NACM; un par de días después, el 17 de septiembre de ese año, la NCSC patrocinaría el primer encuentro presencial entre los miembros asociados al grupo de debate y trabajo del Legal XML. A partir de esa fecha se han venido manteniendo con regularidad estos encuentros y congresos entre los investigadores integrantes de las distintas organizaciones e instituciones de importancia y competencias en esta materia.

La **Conference of State Court Administrators (COSCA)**³²⁹ es una institución vinculada en la actualidad a la NCSC pero fundada antes que ésta (en el año 1953) con la finalidad de mejorar los sistemas de gestión de los tribunales estatales de justicia. La misión principal de

³²⁸ Toda la información oficial relativa a este importante encuentro se halla disponible en la dirección <http://www.ncsc.dni.us/ctc6.htm>

³²⁹ <http://cosca.ncsc.dni.us/>

esta organización radica, como se indica textualmente en sus páginas electrónicas de información, en la "promoción de un foro nacional para ayudar a los administradores de los tribunales estatales en el desarrollo de un sistema judicial más justo, efectivo y eficiente."³³⁰ Para llevar a cabo esta importante misión, dicha organización, en colaboración con otras de similares propósitos, viene trabajando activamente en el desarrollo de políticas, principios y estándares apropiados para la administración de dichos sistemas judiciales.

La **National Association for Court Management (NACM)**³³¹, como asociación de ámbito nacional que es contempla unos propósitos de mayor alcance al intentar ser un foco de potenciación y mejora de la gestión de los tribunales de justicia norteamericanos tanto a nivel federal, estatal como local. Esta asociación también incide fuertemente en la potenciación de los aspectos tecnológicos de estos tribunales, asesorando a sus miembros asociados sobre diversos aspectos del uso de la informática y los sistemas de gestión automatizada en dichos centros. Otra importante línea de actuación de esta asociación se sitúa en la potenciación y mejora de los servicios que posibilitan a los ciudadanos un acceso libre y gratuito a la información generada en los tribunales de justicia norteamericanos, promocionando y participando activamente en diversos proyectos donde las tecnologías relacionadas con la red Internet facilitan dicho propósito.

Pero, ajustándonos al caso que nos ocupa, la importancia de estas dos últimas organizaciones viene derivada por haber constituido el primer intento verdaderamente importante para la creación de un organismo de ámbito nacional para el estudio y aplicación de las tecnologías XML en el campo de la información electrónica judicial. Este grupo de investigación, con nombre formal de **COSCA/NACM Joint Technology, XML Standards Committee, Electronic Filing Subcommittee** pero conocido de forma

³³⁰ Información obtenida en la dirección http://cosca.ncsc.dni.us/about_cosca.html

³³¹ <http://www.nacmnet.org/>

abreviada por el nombre de **Court XML**³³², surge tras una reunión de trabajo llevada a cabo a finales de agosto de 1999 en la ciudad de Santa Fe, Nuevo Méjico, en la cual el Joint Technology Committee (JTC) de la COSCA y la NACM deciden crear un subcomité para el estudio y desarrollo de estándares para el tratamiento de ficheros electrónicos en los tribunales de justicia norteamericanos. El foco principal de actuación estaría en la búsqueda y creación de un estándar de aplicación nacional basado en la tecnología XML, el cual debería permitir realizar las oportunas transferencias de ficheros electrónicos a través de la red Internet entre las diversas partes interesadas de forma unificada y efectiva.

Desde noviembre de 1999 este subcomité ha venido patrocinando un gran número de encuentros y reuniones de trabajo entre diversos especialistas y profesionales en la materia para llevar a cabo este cometido, siendo especialmente destacable la reunión mantenida en diciembre de 1999 en la cual se decidió unánimemente la adscripción como socio a la organización Legal XML para el desarrollo conjunto de estándares XML de aplicación en los tribunales de justicia³³³. En la reunión de marzo de 2000 desarrollada en la ciudad de Atlanta se presentó al público la propuesta de estándar en la que habían estado trabajando el JTC y el grupo de trabajo de la Legal XML, siendo aprobada ésta en la siguiente reunión realizada en la ciudad de Kansas a finales de dicho mes³³⁴.

Este documento normativo, con fecha de 24 de marzo de 2000, constituye en la actualidad el estándar XML *de facto* (una DTD) para el tratamiento de los ficheros electrónicos que se han de emplear en los sistemas de transmisión de información

³³² <http://www.courtxml.org/>

³³³ Toda la información relativa a los encuentros y reuniones de trabajo que este subcomité ha venido manteniendo se encuentra disponible en la dirección <http://www.courtxml.org/xml/JjntXmlStandard.nsf/516c7664fda1528a862565ec00504473/2dd9169668cab7e1862567e1006231b8?OpenDocument>

³³⁴ Para una mayor información sobre el origen y evolución de la organización CourtXML así como otros detalles de importancia sobre el estándar XML propuesto para el tratamiento de los ficheros electrónicos judiciales servidos a través de la red Internet, resulta recomendable la consulta de la información suministrada por Robin Cover en la página web dedicada a este particular. Véase, <http://www.oasis-open.org/cover/courtXML.html>

electrónica entre las diversas partes intervinientes en los procesos judiciales³³⁵. Este documento está pensado como vehículo normalizador para la descripción estructural y semántica de la información contenida en los ficheros electrónicos de los nuevos tribunales de justicia que hacen uso de las tecnologías derivadas de Internet³³⁶.

Esta aplicación XML o DTD es ciertamente compleja y extensa donde se tiene como elemento raíz o elemento de documento al identificador genérico de *LegalEnvelope* (pudiéndose traducir por algo similar a envoltorio o cubierta jurídica), identificativo del tipo de mensaje XML que va a ser transmitido. En un nivel inferior de la estructura jerárquica de esta DTD se encuentran diversos elementos que van a identificar aspectos tales como un identificador normalizado para el mensaje a transmitir, el emisor, el destinatario principal, las copias a otros destinatarios, la fecha y hora de creación del mensaje, método empleado para validar la integridad del mensaje y el elemento *Legal* como elemento genérico para la entrada de todos los datos, información y ficheros adjuntos que van a ser transmitidos en esta transacción electrónica de documentos. Este primer nivel en la DTD queda representado por el siguiente gráfico:

³³⁵ Marty Halvorson, Richard Himes. *XML Standards Development Project. Electronic Court Filing Proposed Standard* [documento PDF]. Rolan Yoshinaga (ed.). CourtXML; Legal XML, March 24, 2000. Disponible en [http://www.courtxml.org/xml/JjntXmlStandard.nsf/516c7664fda1528a862565ec00504473/9cbe71496cde5a14862567e100533da8/\\$FILE/LegalXMLCourtFilingProposedStandard.PDF](http://www.courtxml.org/xml/JjntXmlStandard.nsf/516c7664fda1528a862565ec00504473/9cbe71496cde5a14862567e100533da8/$FILE/LegalXMLCourtFilingProposedStandard.PDF) (consultado 25 de noviembre de 2000). Este mismo documento (con ligeras modificaciones debido a la inclusión de una serie de gráficos estructurales de los elementos descritos en la DTD) se encuentra disponible de forma oficial en el sitio de la organización Legal XML, en la dirección http://www.legalxml.org/DocumentRepository/ProposedStandards/Clear/PS_10001/PS_10001_2000_07_24.htm (consultado el 23 de marzo de 2001).

³³⁶ Resulta imprescindible volver a aclarar aquí que este estándar XML no normaliza en sí documentos que se emplean habitualmente en los tribunales de justicia y que conforman los expedientes de los casos judiciales. En sí se trata, como decíamos, de una DTD para generar y marcar información electrónica codificada mediante XML que ha de ser intercambiada a través de Internet entre los tribunales de justicia y las partes implicadas (abogados, procuradores, partes implicadas, etc.), constituyendo, por tanto, pieza esencial de los actuales sistemas de los *tribunales electrónicos*. De hecho, como ya se ha explicado con anterioridad, estos mensajes electrónicos codificados en XML contemplan la posibilidad de llevar anejos a los mismos documentos electrónicos probatorios o de otra índole en los diversos formatos de codificación admitidos (PDF, TIFF, etc.). La evolución lógica en estos sistemas es que todo el entramado documental esté definido mediante XML, siendo éste el caso que se expondrá en el apartado final del capítulo.

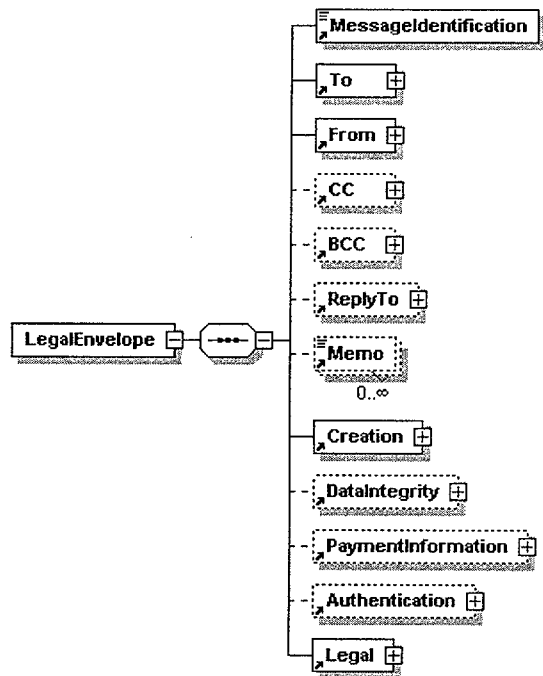


Figura III.10: Diagrama particular de estructuración del elemento `LegalEnvelope` del Court XML.

A partir de aquí, cada uno de estos elementos se va subdividiendo en otros tantos con la finalidad de establecer un exacto control informativo de todos los datos necesarios para la realización e identificación del mensaje a transmitir. Resulta especialmente complejo y extenso el apartado dedicada al elemento *Legal*, donde simplemente la información empleada para describir al documentos principal transmitido contiene más de 15 subelementos. Sirva como ejemplo de la complejidad de esta DTD la definición del elemento *Person* (elemento que aparece en diversas partes de la DTD), expresada gráficamente del siguiente modo:

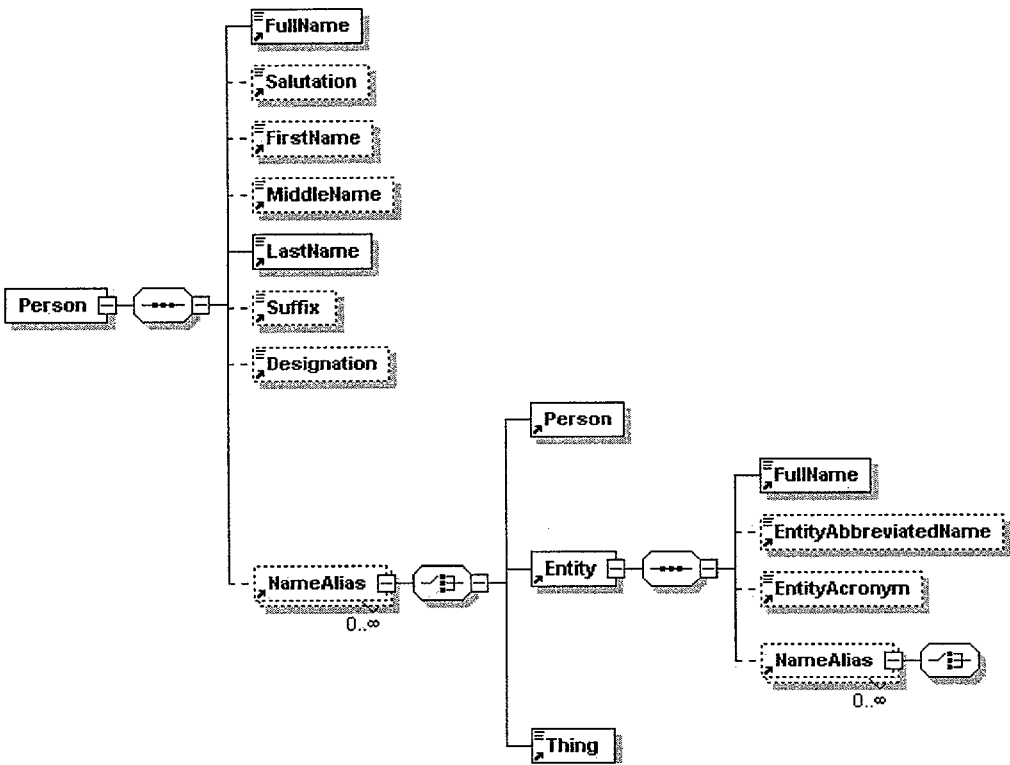


Figura III.11: Diagrama particular de estructuración del elemento Person del Court XML.

Es importante señalar igualmente que este estándar fue patrocinado por el anteriormente citado *Electronic Court Filing Project* de la Georgia State University y por la compañía JusticeLink, Inc.³³⁷. Se trata esta última de una de las mayores empresas de los Estados Unidos dedicadas al suministro en tiempo real de soluciones para el procesamiento y suministro electrónico de documentos legales entre los tribunales y las partes encausadas y que de forma más decidida ha apostado por la tecnología XML para sus desarrollos comerciales.

Antes de dar paso a la descripción de la organización más importante en la actualidad en la promoción y desarrollo de estándares XML de aplicación al entorno jurídico, el Legal

³³⁷ <http://www.justicelink.com/>

XML, apuntar mínimamente la igualmente importante labor llevada a cabo por la **American Bar Association** (ABA)³³⁸. Se trata ésta de una de las asociaciones profesionales con mayor número de afiliados en todo el mundo (más de 400.000 miembros inscritos), dedicada entre otras cosas a la certificación y acreditación de las escuelas y facultades de derecho en los Estados Unidos, la formación continua de los juristas, a la promoción de la información de carácter legal, la realización de programas de ayuda laboral para los abogados y jueces afiliados y la promoción de iniciativas para la mejora de los sistemas jurídicos de cara al ciudadano. Dentro de esta asociación destaca la labor emprendida por James Keane en la promoción del estándar XML al organizar en 1999 el primero de los encuentros conocidos bajo el nombre de *XFiles!* Estos encuentros fueron los primeros que se desarrollaron bajo el prisma de poner en contacto a los juristas con las empresas informáticas más punteras en este campo para la discusión y potencial de los estándares XML. En la actualidad, esta asociación mantiene dentro de su *Legal Technology Resource Center*³³⁹ un apartado dedicado a la investigación y promoción entre sus asociados de las nuevas tecnologías de la información aplicadas al entorno de producción documental de los tribunales de justicia.

Finalmente, comentar en este apartado algunos aspectos de interés de la anteriormente anunciada organización de ámbito internacional **Legal XML**³⁴⁰. Se trata de una organización sin ánimo de lucro fundada, como ya se ha comentado con anterioridad, en noviembre de 1998 en la cual se integran personas procedentes de todo tipo de instituciones relacionadas con el campo jurídico (empresas, organizaciones sin ánimo de lucro, agencias gubernamentales, instituciones académicas, etc.) con el objetivo de desarrollar estándares abiertos y no propietarios basados en el metalenguaje XML y de

³³⁸ Para una más amplia información, véase la información suministrada en la dirección <http://www.abanet.org/>

³³⁹ <http://www.lawtechnology.org/>

³⁴⁰ <http://www.legalxml.org/>

aplicación a los documentos jurídicos y material relacionado³⁴¹. En la actualidad cuenta con más de 650 miembros procedentes de diversos países del planeta (mayoritariamente, de los Estados Unidos³⁴²), siendo sus patrocinadores principales el macro-portal jurídico FindLaw y el ya citado Georgia State University Electronic Court Filing Project.

Esta activa e importante organización internacional además de fomentar el debate y el intercambio de ideas a través del uso de las tecnologías de Internet (especialmente, haciendo uso de las listas de distribución de correo electrónico) fomenta de igual modo, como ya se ha visto con anterioridad, el debate presencial entre los investigadores en la materia a través de la puesta en marcha de congresos anuales. Como ya se comentó anteriormente, el primero de ellos tuvo lugar en septiembre de 1999 en la ciudad de Los Ángeles (California); el segundo de ellos se produjo en marzo de 2000 en Atlanta (Georgia) en las dependencias de la universidad del citado estado; y el tercero, y último hasta la fecha,

³⁴¹ La apasionante historia del origen y evolución de esta organización (que, en definitiva, es en gran medida la historia del XML aplicado a la documentación judicial en los Estados Unidos) no resulta fácil localizarla en la actualidad. A mediados del año 2000, el documento *HTML Legal XML Overview*, elaborado por W. "Todd" Vincent, se hallaba ubicado dentro del servidor web de esta organización en la dirección <http://www.legalxml.org/Information/LegalXMLOverview.asp> (documento del cual tengo copia impresa). Pero a mediados del año 2000 la dirección de este página fue modificada (dejando de estar operativa el anterior URL), situando sus gestores el enlace ahora en la dirección <http://www.legalxml.org/DocumentRepository/LegalXMLOverview.asp>, aunque no es posible encontrar el citado documento bajo dicho URL (en principio, dentro del sitio web del Legal XML resulta a fecha de hoy imposible acceder al mencionado documento). En cualquier caso, el mismo texto de este documento (aunque bajo otro título) fue publicado en el año 2000 en la revista que publica la *Administrative Office of Georgia Courts* (<http://www.georgiacourts.org/>). Véase, por tanto, Winchel "Todd" Vincent. "History of Legal XML" [documento PDF]. *Web Currents*, v. 1, n° 4, December 2000, pp. 2-4, 6. Disponible en <http://www.georgiacourts.org/aoc/publications/webv14.pdf> (consultado el 17 de junio de 2001).

³⁴² En una nota general sobre esta organización ubicada en las páginas web de R. Cover (pues dicha información, al igual que ha sucedido con el anterior documento, no se encuentra presente en el sitio web del Legal XML) se expone que entre las instituciones que se integran dentro del Legal XML un 50% de ellas pertenecen al sector privado, un 25% son agencias gubernamentales y el restante 25% corresponde a instituciones académicas y otras organizaciones profesionales sin ánimo de lucro (destacando entre estas últimas el *Joint Technology Committee of COSCA/NACM*, la *National Court Reporters Association*, *SEARCH*, la *California Administrative Office of Courts* o la organización alemana del *LEXML*). En cuanto a su distribución geográfica, un 25% no están radicadas en los Estados Unidos, siendo mayoritarias dentro de este porcentaje las instituciones de países como Australia, Canadá y algunos países europeos. Para una mayor información, véase la información suministrada en la dirección <http://xml.coverpages.org/legalXML-overview20000927.html> así como la actualización a fecha de 13 de agosto de 2001 de la información sobre el *Legal XML Working Group* en la dirección <http://xml.coverpages.org/legalXMLWG.html>

ha sido desarrollado conjuntamente con el séptimo de la NCSC (el CTC7) en la ciudad de Baltimore (Maryland)³⁴³.

Para llevar a cabo la importante misión propuesta, esta organización se estructura tanto horizontal como verticalmente en diversos subdominios. Así, entre los subdominios verticales se encuentran algunos como los de *Court Filing*, *Transcripts*, *Judicial Decisions*, *Public Law* (leyes, proyectos de ley, estatutos, etc.), *Private Law* (contratos, testamentos, etc.) y *Publications* (libros jurídicos, revistas de derecho, etc.) y entre los subdominios horizontales podemos encontrar algunos como los *Citations*, *General Vocabulary* (nombres, direcciones, etc.) y *Logical Document Structure* (elementos iniciales, tablas, referencias, párrafos, firmas o metodología para la definición de estructuras documentales). Estos subdominios horizontales se pueden encontrar, obviamente, en diversos subdominios verticales por lo que la coordinación entre todos estos subdominios verticales y horizontales es un factor de especial importancia dentro de esta organización.

Estos dominios y subdominios dan lugar a los grupos de trabajo (*Legal XML Workgroups*)³⁴⁴, siendo éstos los verdaderos centros operativos de esta organización. Los grupos de trabajo tienen como misión el desarrollo de diversas especificaciones que han de definir los estándares técnicos XML en el ámbito jurídico. Al igual que ocurre con otras organizaciones de características similares que han venido surgiendo a lo largo de la historia de la red Internet (y en especial con la IETF y el W3C), al frente de cada uno de estos grupos de trabajo se sitúa uno o dos directores o responsables del mismo, fomentando el diálogo y el debate a través de las listas de discusión vía correo electrónico existentes para cada uno de ellos. Existen cuatro grandes grupos o categorías son: los relacionados con las cuestiones administrativas (*Administrative*), los relacionados con la coordinación de todos los grupos y sectores implicados en esta organización (*Coordinating*), los de ámbito geográfico o jurisdiccional (*Jurisdictional*) y aquellos en los que se debaten temas

³⁴³ Toda la información oficial relativa a los diversos eventos y conferencias que la organización Legal XML ha venido desarrollando a lo largo de su historia se encuentra disponible en la dirección <http://www.legalxml.org/activity/index.htm>

³⁴⁴ La información oficial relativa a los grupos de trabajo existentes en la organización Legal XML se encuentra disponible en la dirección <http://www.legalxml.org/workgroups.htm>

propriadamente jurídicos (*Substantive*). Dentro de esta última categoría se encuentran grupos de trabajo tales como los de *Appellate Court Filing*, *Citations*, *Dictionary*, *Integrated Justice*, *Legislation*, *Signatures* o, el que es materia de nuestro interés, el del *Court Filing*.

Aunque algunos de estos grupos de trabajo resultan de gran interés, como es el caso del grupo dedicado a establecer un diccionario de metadatos de interés jurídico³⁴⁵, para los intereses de esta tesis doctoral nos centraremos exclusivamente en el denominado **Court Filing Workgroup**³⁴⁶. Codirigido por John Greacen (autor del directorio web de información jurídica *New México State Judiciary*³⁴⁷) y Moira Rowley (de la compañía norteamericana suministradora de soluciones informáticas *STC*³⁴⁸), se especializa en el debate sobre todos aquellos aspectos relacionados con los documentos y los formatos de intercambio de información electrónica de interés para las aplicaciones informáticas propias de los tribunales de justicia. Dado que dentro de este grupo de trabajo se integran gran parte de las organizaciones e instituciones que se han venido detallando en este último apartado del capítulo, resulta lógico que el desarrollo normativo de las principales aplicaciones XML en las que dichas instituciones venían trabajando pase ahora a manos de este grupo de trabajo. Y de hecho así es en lo tocante a la aplicación XML relativa al *Electronic Court Filing*, elaborada en un primer momento bajo los auspicios de la Court XML y coordinada en la actualidad por este grupo de trabajo del Legal XML. De aquella DTD

³⁴⁵ Y decimos que resulta de especial interés pues, como se explicará más adelante, la rama europea de la organización Legal XML está trabajando activamente en este mismo campo, siendo uno de los pilares de su actividad. En el caso del *Dictionary Workgroup* (<http://www.legalxml.org/dictionary/>), y hasta el momento de redactar esta tesis doctoral, este grupo tan sólo ha planteado una declaración de buenas intenciones en el desarrollo de un diccionario de metadatos de interés jurídico en lengua inglesa. Así, y de forma más específica, este grupo de trabajo pretende elaborar diversos productos; a saber: un documento en el que se especifiquen los requisitos y las interfaces para el diccionario de metadatos, un diccionario de términos jurídicos y de negocios, un software de código abierto para el mantenimiento de este diccionario y, finalmente, el mapeado de las entradas del diccionario a aplicaciones XML. Para una más amplia información, véase John McClure, John Messing, Jerome Cook. *Dictionary Workgroup Charter* [documento HTML]. LegalXML, May 15, 2001. Disponible en http://www.legalxml.org/dictionary/CH_100XX_2001_05_15.xhtml (consultado el 21 de junio de 2001).

³⁴⁶ La información oficial de este grupo de trabajo se encuentra disponible en la dirección <http://www.legalxml.org/CourtFiling/>

³⁴⁷ <http://www.nmcourts.com/>

³⁴⁸ <http://www.stc.com/>

desarrollada inicialmente se mantuvo en la versión del Legal XML el texto íntegro de la misma, siendo únicamente alterada y mejorada la presentación de la especificación con la inclusión de una serie de gráficos relativos a la estructura jerárquica del modelo de contenido de los elementos y subelementos que se integran dentro de esta DTD³⁴⁹. A esta versión se la conoce como el **LegalXML Court Filing 1.0 standard**. A partir de ahí, se ha seguido trabajando y depurando algunos aspectos de esta DTD para alcanzar una versión definitiva y estable y, por tanto, un estándar de aceptación internacional para este tipo de documentos electrónicos dentro de los tribunales de justicia. La última versión del estándar propuesto, documento fechado a 4 de enero de 2001³⁵⁰, en el cual se incluyen entres sus apéndices una serie de ejemplos de uso muy ilustrativos. El texto íntegro de esta DTD viene expuesto en el Apéndice 3.

Además de esta importante aplicación basada en XML, este grupo de trabajo ha venido desarrollando otras especificaciones complementarias a la misma. Este es el caso, por ejemplo, del *EFM-CMS Interface Requirements*³⁵¹, actualmente en su versión 7, en el cual este grupo de trabajo en colaboración con el grupo de trabajo de California han establecido los requisitos básicos para el desarrollo de una interfaz de programa de carácter universal para la interacción entre los módulos de aplicación del EFM (*Electronic Filing Manager*) y el CMS (*Case Management System*) para el intercambio de datos judiciales en un entorno electrónico de actuación. Pero, sin duda, la especificación más interesante para los propósitos de esta tesis doctoral la constituye el denominado XML Court Document 1.0 Proposed

³⁴⁹ Como ya se comentó en la nota 336, la autoría de esta DTD corresponde a Marty Halvorson y Richard Himes y dentro del sitio web del Legal XML se encuentra ubicada en la dirección http://www.legalxml.org/DocumentRepository/ProposedStandards/Clear/PS_10001/PS_10001_2000_07_24.htm

³⁵⁰ Marty Halvorson, Richard Himes. *XML Standards Development Project. Electronic Court Filing Proposed Standard* [documento DOC]. Winchel "Todd" Vincent (ed.). Legal XML, January 4, 2001. Disponible http://www.legalxml.org/DocumentRepository/PROPOSEDSTANDARDS/CFPROPOSEDSTANDAR2001_01_04.DOC (consultado el 24 de mayo de 2001).

³⁵¹ Mark Yuan, Steve Spohn. *EFM-CMS Interface Requirements* [documento PDF]. Roger Winters (ed.). Legal XML, July 22, 2001. Disponible en http://www.legalxml.org/California/files/EFM-CMS_APIv7.pdf (consultado el 13 de agosto de 2001).

Standard³⁵², que, como veremos a continuación, es la evolución lógica del estándar *Court Filing 1.0*. A esta nueva propuesta, por tanto, se la considera formalmente como la versión 2.0 de la anterior especificación. Veamos algunos aspectos interesantes sobre la historia y aplicación de este estándar, dada su importancia y trascendencia en el entorno judicial.

La versión 1.0 del estándar del *XML Court Filing* nació con un grave problema: el estándar sólo se centraba en la información necesaria para transmitir electrónicamente al tribunal de justicia un envoltorio documental e, igualmente, establecía los mecanismos para el envío de la confirmación de dicha transmisión por parte del tribunal. Este estándar no especificaba, por tanto, el tipo de elementos de datos o etiquetas necesarias para marcar el contenido informativo del documento a archivar. Así, el documento podía ser un fichero de texto llano, un URL apuntando al documento, una imagen digital, un documento marcado en PDF o un documento cualquier otro formato binario. Resultaba, pues, necesario establecer aquellos elementos y sus correspondientes etiquetas para marcar claramente los contenidos de los documentos utilizados en estas transacciones de información electrónica. Sería, por tanto, inevitable tratar de obtener una nueva versión del estándar ya aprobado, es decir, llegar a definir la versión 2.0 del mismo. Para ello se acordaron una serie de consultas y debates entre los profesionales adscritos a este grupo de trabajo para el establecimiento de una serie de requisitos generales y básicos para esta nueva versión del estándar.

A mediados de enero de 2001, se elaboró un documento normativo con las reflexiones y aportaciones recogidas, el *Legal XML General Requirements for Court 2.0 XML Document*³⁵³. Como se señala en este documento, el objetivo principal de esta segunda versión del

³⁵² Mohyeddin Abdulazziz, Rolly Chambers, John Messing. *Legal XML Proposed Standard: XML Standards Development Project – XML Court Document 1.0 Draft Standard* [documento HTML]. Legal XML, June 21, 2001. Disponible en [http://www.legalxml.org/appeals/courtdocument/Legal%20XML%20Court%20Document%201.0%20Draft%20DTD\(rev\).htm](http://www.legalxml.org/appeals/courtdocument/Legal%20XML%20Court%20Document%201.0%20Draft%20DTD(rev).htm) (consultado el 23 de agosto de 2001). Éste y otros documentos relacionados con el mismo (documentos informativos, ficheros de la DTD con la versión de cada fecha de revisión, diagramas de las mismas y una serie de ejemplos prácticos) se encuentran reunidos y accesibles desde en la dirección <http://www.legalxml.org/appeals/courtfilindocument.htm>

³⁵³ Mohyeddin Abdulazziz, Rolly Chambers, John Messing. *LegalXML General Requirements for Court 2.0 XML Document* [documento HTML]. LegalXML, January 15, 2001. Disponible en <http://www.legalxml.org/appeals/courtfilindocument/requirements.htm> (consultado el 25 de febrero de 2001).

estándar estaría en la definición de los elementos de datos XML y sus correspondientes etiquetas capaces de marcar y describir el contenido existente en los documentos electrónicos judiciales así como los metadatos asociados a estos documentos y otros documentos no XML, como las imágenes asociadas a los documentos judiciales. De este modo, y según el propio texto aparecido, con esta nueva versión del estándar “los autores de los documentos, además de otras muchas cosas, serán capaces de proporcionar automáticamente información actualizada necesaria para su agenda de actuación, para los temas financieros, para la indexación, información de contacto así como para otras aplicaciones apropiadas”. En cuanto a los requisitos a cumplir, se establecieron en este documento 4 puntos principales, algunos de ellos con ciertas subdivisiones para su perfecta definición. En líneas generales, se venía a decir que esta nueva versión del estándar debería ajustarse perfectamente a las diversas tecnologías XML existentes; debería ser lo suficientemente simple y comprensible para que las aplicaciones informáticas que hubieran de trabajar con la misma; debería de establecer una estructura básica para estos documentos según una serie de características requeridas (título del caso, organización de los contenidos del cuerpo del documento, firmas electrónicas, ficheros adjuntos, etiquetas para los metadatos, etc.); y, por último, la posibilidad de ampliar o extender el modelo de contenido del cuerpo del documento, se deberían contemplar otras características adicionales como compatibilidad con la versión anterior, la encriptación de los documentos, etc.

Fruto de estas reflexiones surgen nuevos documentos informativos que tratan de matizar algunos aspectos de esta nueva versión del estándar, tal es el caso del documento de ayuda elaborado por este grupo de trabajo en el cual se añaden y se anotan para su perfecto entendimiento por parte de los investigadores y profesionales de la justicia este nuevo estándar propuesto³⁵⁴.

³⁵⁴ Nos estamos refiriendo aquí al documento Mohyeddin Abdulazziz, Rolly Chambers, John Messing. *Amended and Annotated 1.0 Proposed Court Document Standard (formerly, version 2.0 of the Court Filing Standard)* [documento PDF]. LegalXML, March 2001. Disponible en <http://www.legalxml.org/appeals/courtfilingsdocument/requirementsAnnotatedMarch2001.pdf> (consultado el 16 de junio de 2001).

En la actualidad, como ya se ha comentado con anterioridad, el último borrador del estándar *XML Court Document 1.0* (aunque formalmente establecido como la versión 2.0 del *XML Court Filing*) está fechado el 21 de junio de 2001.

Esta definición de tipo documental, cuyo texto íntegro aparece reflejado en el Apéndice 4 de esta tesis doctoral, resulta ciertamente amplia, elaborada y, consiguientemente, compleja, con la inclusión de algunas entidades paramétricas internas y externas que definen los modelos de contenido de algunos elementos que la hacen aún más compleja, hecho éste que nos impide realizar una explicación minuciosa de la misma pues ello alargaría en extremo la extensión de este capítulo. Así, nos centraremos aquí en ofrecer una explicación general de los principales elementos de la misma junto con la exposición de algunos gráficos ilustrativos del modelo de contenido de dichos elementos³⁵⁵.

Como señalábamos con anterioridad, dos son las entidades paramétricas principales que se incluyen dentro de esta DTD y que remiten a modelos de contenidos definidos externamente.

La primera de ellas es la correspondiente al elemento *BasicActor*, utilizado en diversas partes de la DTD para marcar la información esencial de las personas u organizaciones involucradas en el caso judicial tratado. El modelo de contenido de este elemento es una reducción del modelo que se había definido para el elemento *Actor* del estándar *Court Filing 1.0*. Así, además de una serie de atributos propios del elemento *BasicActor* (además de una serie de códigos identificativos para establecer referencias entre documentos –*ID*, *IDReference*– existen otros como el *ActorIdentifier*, el *ActorType* –con los valores posibles de *Person* u *Organization*– y el *RoleType* –con los valores posibles de *Attorney*, *Party*, *Witness* o *JudicialOfficer*–), su modelo de contenido incorpora a dos entidades paramétricas internas

³⁵⁵ Existe, en cualquier caso, un documento electrónico dentro del sitio web del Legal XML que presenta de manera clara y perfecta los diagramas correspondientes al modelo de contenido de cada uno de los elementos que se integran en esta DTD (en su versión del 3 de junio de 2001), por lo que remitimos al lector a la consulta del mismo para obtener una visión más directa de dicha definición de tipo documental. Este documento se encuentra disponible en la dirección <http://www.legalxml.org/appeals/CourtFilingDocument/June032001/courtfillingdocumentDiagram.htm>

para definir los datos relativos al nombre del actor (entidad paramétrica *Name.content*)³⁵⁶ y a la función desempeñada por el mismo (entidad paramétrica *Role.content*) junto a los elementos opcionales y no repetibles *BasicAddress*, *Phone*, *Fax*, *Email* y *ActorInformation*. A su vez, los modelos de contenido de algunos de estos subelementos han sido igualmente definidos a través de entidades paramétricas internas, como es el caso del elemento *BasicAddress*, o como entidades paramétricas externas, como es el caso del elemento *ActorInformation* (incluido dentro de la gran entidad paramétrica externa *SimpleActorInformation.module*). La información relativa a los elementos *telephone*, *fax* y *email* viene establecida como datos de carácter (PCDATA) para introducir el dato correspondiente, junto con una serie de atributos para detallar aún más esta información (en todos ellos se repite el atributo *Type*, con un valor posible entre los de *Business*, *Home* y *Unknown*, y el atributo *Status* para introducir algún texto significativo sobre dicho elemento). Algo similar ocurre con la segunda entidad paramétrica interna, correspondiente al elemento *BasicAddress*, la cual contiene datos de localización relativos a un *BasicActor*. Al igual que en el caso del elemento *BasicActor*, se trata de una reducción del elemento *Address* de la DTD del *Court Filing 1.0*. En su modelo de contenido sólo figura como subelemento obligatorio el relativo al *State* siendo, por tanto, el resto opcionales (*AddressLine* —que, a su vez, incluye a los subelementos *OrganizationName*, *Street* y *POBox*— *City*, *County*, *PostalCode* y *Country*). La información relativa La representación gráfica del modelo de contenido del elemento *BasicActor* quedaría, por tanto, de la siguiente forma:

³⁵⁶ Sirva como ejemplo de la complejidad y precisión con la que ha sido diseñada esta DTD que simplemente para definir el nombre del actor o actores participantes en un caso judicial su entidad paramétrica refleja un modelo de contenido ciertamente complejo con un gran número de subelementos, quedando formalmente definida del siguiente modo:

```
<!ENTITY % Name.content “((FullName | AbbreviatedName | AliasName) | (Salutacion?,  
GivenName?, MiddleName?, FamilyName, Suffix?))”>
```

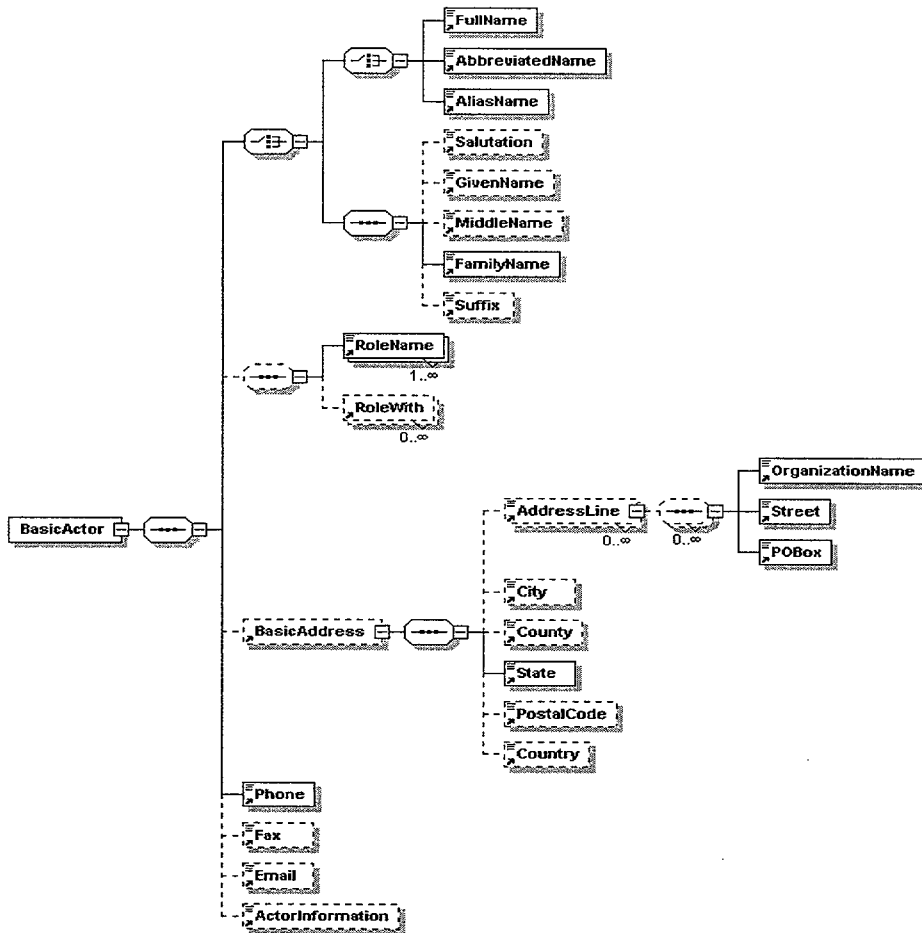


Figura III.12: Diagrama particular de estructuración del elemento BasicPerson del Court XML.

Esta DTD establece como elemento raíz de la misma al elemento *Legal* el cual contiene como único elemento posible, obligatorio y repetible, al elemento *CourtDocument*. Éste, a su vez, contiene a otro elemento obligatorio y único que representa la información contenida en el documento principal o base de la transmisión electrónica, el elemento *MainDocument*. Opcionalmente, puede existir otro elemento denominado *ElectronicSignature*, el cual contendrá toda la información relativa a las firmas electrónicas empleadas para firmar y, por tanto, autenticar el documento del tribunal. Por último, el *CourtDocument* contempla varios atributos, entre los cuales destaca el denominado *FilingType*, tomando como valor uno de los ítem de la lista de valores posibles relativos al área del derecho se adscribe el documento (*AdministrativeLaw*, *Appeal*, *Bankruptcy*, *Civil*, *Criminal*, *DomesticRelations*, *Juvenile*,

Probate, SmallClaims, Traffic o *Arbitration*). El elemento *MainDocument*, o documento principal, contiene una serie de elementos hijos que describen las partes básicas del documento electrónico transmitido. Estos elementos son: *DocumentMetadata*, *CaseCaption*, *DocumentBody*, *DocumentSigner* (opcional y repetible), *ProofOfService* (opcional) y *Attachment* (opcional y repetible). Estos niveles quedan gráficamente reflejados en el siguiente diagrama:

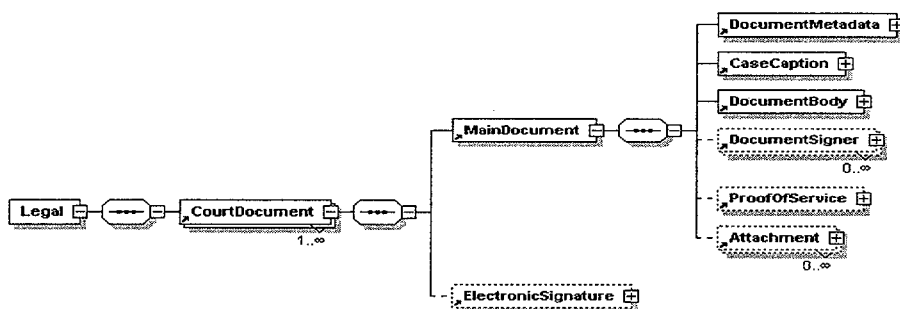


Figura III.13: Diagrama particular de estructuración del elemento Legal del Court XML.

El elemento *DocumentMetadata* describe la metainformación relativa al documento electrónico principal, donde la mayoría de los elementos hijos están basados en el conjunto de elementos de metadatos definidos por el Dublin Core en su versión 1.1. Así, aquí nos encontramos con una serie de elementos obligatorios, como son *Creator* (la persona u organización responsable principal de la redacción del contenido del documento), *Filer* (la persona u organización que realiza la presentación o transferencia del documento, definida mediante la incorporación del elemento *BasicActor* anteriormente explicado), *Identifier* (código identificador del documento), *DocumentType* (describe el tipo de documento con la inclusión de un texto y un atributo opcional para la asignación de un código indentificativo) y *Format* (la manifestación física o digital del documento, normalmente será el tipo MIME text/xml), junto con otros opcionales y no repetibles, como son *Description* (resumen descriptivo del contenido del documento), *Subject* (materia o materias descriptivas del contenido del documento, siendo éstas palabras clave, frases o códigos de clasificación separados por una coma y un espacio en blanco), *Coverage* (al igual que en el Dublin Core, este elemento describe el ámbito jurisdiccional del documento), *Reference* (una referencia a

un recurso mencionado en el documento, empleando para ello el subelemento *Link* o el subelemento *Image*), *BasicDocumentInformation* (otra información adicional indicada a través de otros subelementos, como el *DocumentStatus* -si el documento es un borrador y es su versión final-, *Privacy* -código del nivel de privacidad-, *Matter* -identificador de la materia- y *CauseOfAction* -identificador de la causa de actuación-) y *FilingInformation* (definido su modelo de contenido a través de la otra gran entidad paramétrica externa *SimpleFilingInformation.module*, en la cual se hace referencia a aspectos tales como el pago de tasas, información relativa al caso, etc.). Destacar aquí que el elemento *Link* viene definido a través de una entidad paramétrica interna y hace uso de los atributos más comunes definidos en el estándar asociado XLink en su versión de 1999 (*xlink:type*, *xlink:href*, *xlink:role*, *xlink:title*, *xlink:show* y *xlink:actuate*). La representación gráfica del modelo de contenido dedicado a los metadatos sería la siguiente:

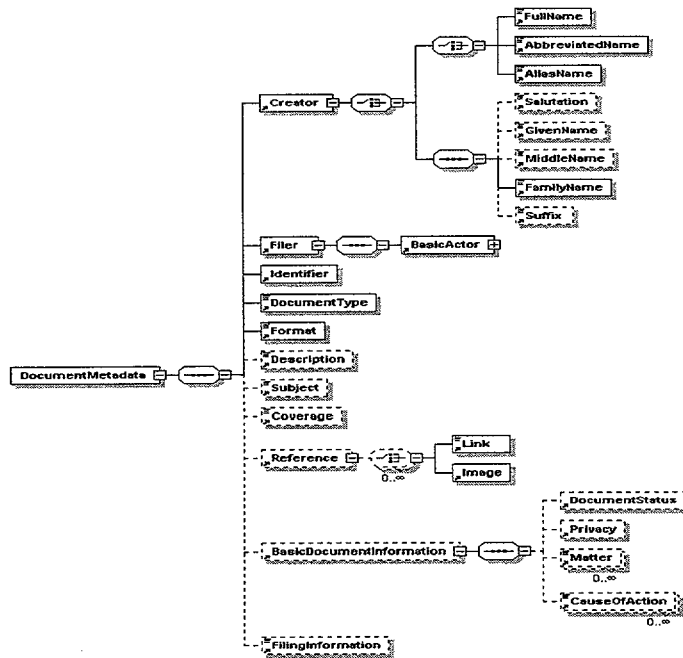


Figura III.14: Diagrama particular de estructuración del elemento DocumentMetadata del Court XML.

El elemento *CaseCaption*, definido igualmente a través de una entidad paramétrica, describe la información encontrada en los títulos de los documentos de alegación (*pleading*)

o de seguimiento judicial (*discovery*) relativa al nombre, el tipo y la localización del tribunal en el cual el caso está pendiente de resolución, el número de caso y el nombre de las partes. Al igual que en otros casos, este elemento hace uso de las entidades paramétricas internas *BasicActor* y *BasicAddress*. Su representación gráfica sería la siguiente:

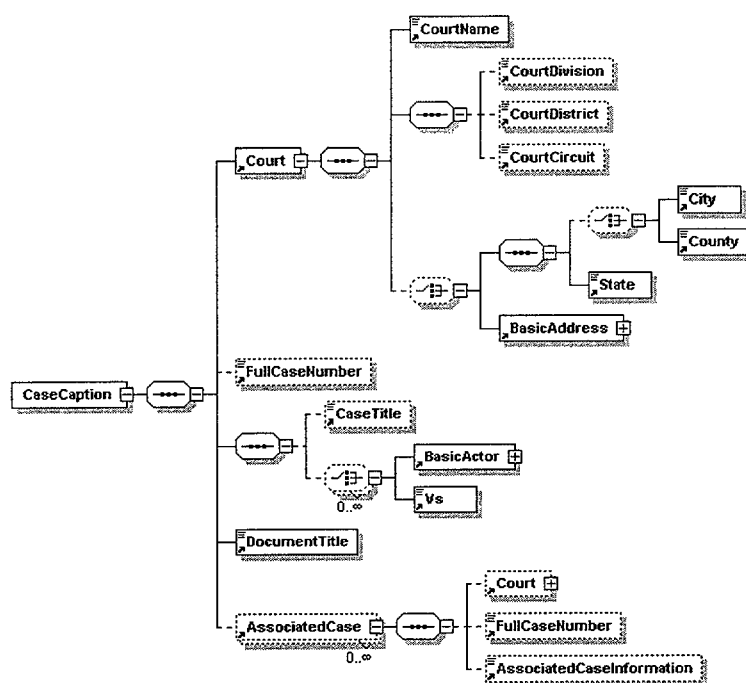


Figura III.15: Diagrama particular de estructuración del elemento CaseCaption del Court XML.

El elemento *DocumentBody* contempla los párrafos y las agrupaciones de éstos que conforman el cuerpo del documento, como se muestra en el siguiente diagrama:

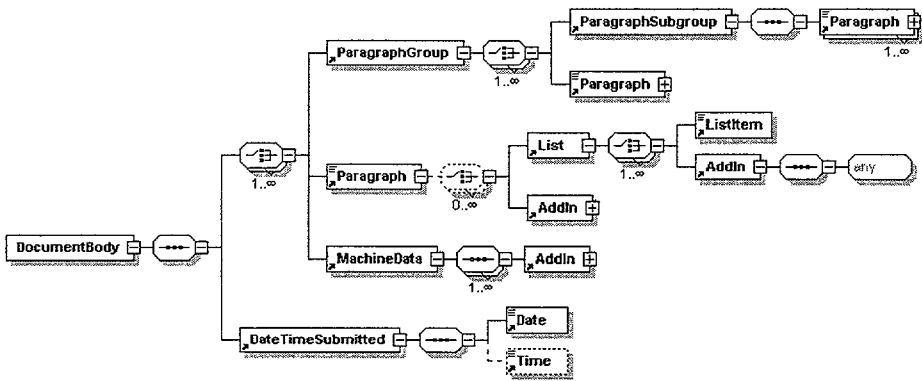


Figura III.16: Diagrama particular de estructuración del elemento `DocumentBody` del Court XML.

El elemento *DocumentSigner* es el que define la información correspondiente a los datos de la firma y de contacto de la persona o personas que suscriben el documento de súplica o de seguimiento a través de la utilización del elemento hijo *Signatory*. Opcionalmente, se puede hacer uso igualmente del elemento hijo *Notarization* para dar forma con valor notarial a un tipo especial de firma para actos notariales, declaraciones juradas y cuestiones similares. Para el correcto control del actor de estas informaciones se hace uso igualmente del elemento *BasicActor*, anteriormente descrito. Gráficamente quedaría representado del siguiente modo:

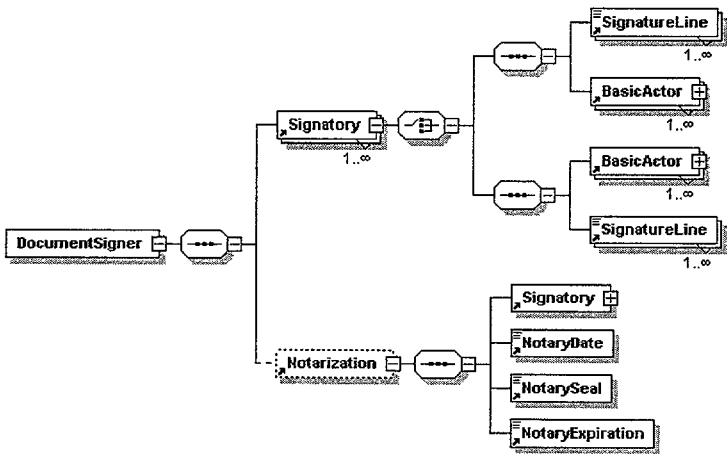


Figura III.17: Diagrama particular de estructuración del elemento `DocumentSigner` del Court XML.

El elemento opcional *ProofOfService* contempla aquella información necesaria para un certificado de servicio. No se trata aquí de la certificación habitual empleadas en las firmas digitales sino, más bien, de un párrafo certificando que este tipo de documento electrónico judicial ha sido entregado. Este elemento, además de la información relativa a la persona o institución que ha llevado a cabo esta labor, incluye otros subelementos relativos a la fecha y hora de entrega, en el formato establecido por la norma ISO 8601. Su representación gráfica sería la siguiente:

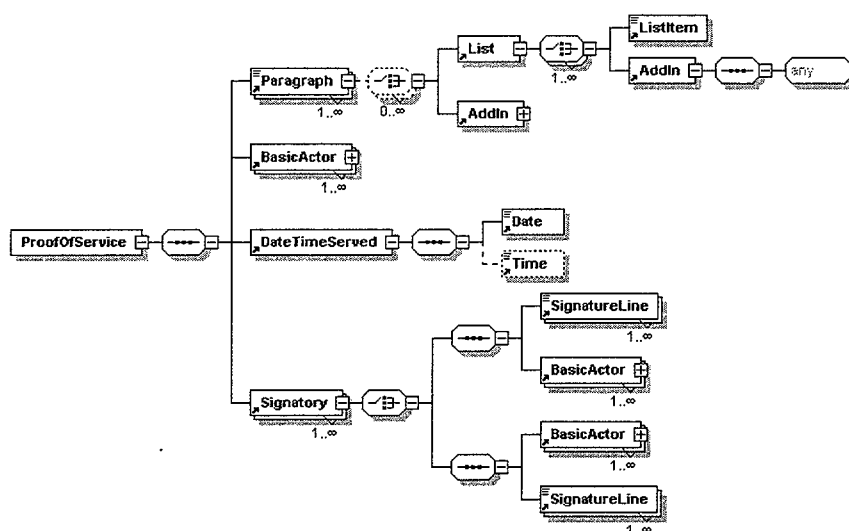
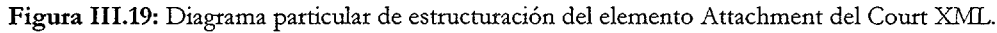
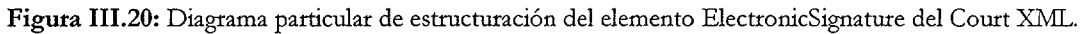


Figura III.18: Diagrama particular de estructuración del elemento *ProofOfService* del Court XML.

El elemento *Attachment* encierra toda la información relativa a los ficheros adjuntos al documento XML enviado. Este elemento, además de una serie de atributos propios, contempla un elemento hijo denominado *AttachmentContent*. En éste, a su vez, se definen una serie de atributos para establecer el tipo MIME del fichero electrónico adjunto, la codificación base del mismo y la referencia hipertextual al mismo, y otro subelemento para indicar el título del documento adjunto. Su representación gráfica sería la siguiente:



gráfica sería la siguiente:



III.6.2.2. EUROPA Y OTROS PAÍSES DEL MUNDO

La aplicación de las tecnologías de la información y la computación al procedimiento judicial dentro de los países europeos se encuentra, a fecha de hoy, muy lejos de lo que sucede en los Estados Unidos, máxime cuando hablamos del empleo de las tecnologías derivadas del metalenguaje XML. Es cierto que existen algunas honrosas excepciones a este aserto, como son los casos de Austria, Finlandia, Reino Unido y Alemania donde el grado de tecnificación de sus tribunales de justicia alcanza cotas ciertamente significativas. No ocurre esto mismo cuando hablamos de la aplicación de las tecnologías XML en instituciones europeas de carácter supranacional, como es el caso de los diferentes

organismos de la Unión Europea, que, como veremos más adelante, sí que se está haciendo una apuesta firme y decidida por el empleo de dichas tecnologías.

Es cierto que en una primera fase de aplicación de las nuevas tecnologías de la información y la comunicación, los países europeos y, de forma muy especial las organizaciones surpranacionales en las cuales se integran, han venido preocupándose más por conseguir una armonización legislativa entre todos ellos así como en la actuación y procedimientos informativos de traducción y publicación de la documentación generada desde los diversos organismos de dichas instituciones.

El capítulo III.4 estuvo dedicado a las experiencias existentes en todo el mundo en la aplicación del metalenguaje SGML a la información y documentación de carácter jurídico, pudiéndose comprobar este hecho anunciado: la mayoría, por no decir todos, de los proyectos europeos de investigación aplicada de lenguajes de marcado a documentos jurídicos correspondían al tratamiento y gestión de documentos legislativos. Este era el caso de proyectos europeos de amplio propósito y gran envergadura como el *proyecto Raske*, de la Universidad finlandesa de Jyväskylä, y el *proyecto Corpus Legis*, de la Universidad sueco de Estocolmo, ampliamente descritos con anterioridad.

Con este panorama europeo resulta, pues, lógico que en estos últimos años se haya seguido con una línea continuista en la investigación y aplicación de los nuevos lenguajes de marcado de textos electrónicos, caso del XML y tecnologías asociadas, a la documentación jurídica, a excepción de lo sucedido dentro del Tribunal de Justicia de las Comunidades Europeas, como se expondrá con posterioridad. Así, algunos proyectos de gran interés auspiciados y soportados económicamente por la Unión Europea han venido insistiendo en el tratamiento de la información electrónica de carácter legislativo mediante el uso de dichas tecnologías.

Sin entrar en un profundo análisis de dichos proyectos, al ser ésta materia de otra tesis doctoral que se están llevando a cabo en nuestro Departamento, sí resulta necesario describir mínimamente estos desarrollos al ser de los pocos ejemplos existentes en el panorama jurídico europeo en la aplicación del metalenguaje XML.

Este es el caso del **Proyecto EULEGIS**³⁵⁷ (*European User Views to Legislative Information in Structured Form*), proyecto convocado a iniciativa del Parlamento de Finlandia y financiado por el *Telematics Application Programme* de la **Dirección General XIII de la Comisión Europea**. Este proyecto se puso en marcha en 1998 y en él se integraron un gran número de empresas e instituciones de otros países europeos como el Reino Unido, Bélgica, Portugal, Noruega, Suecia, Luxemburgo y España (a través de la sociedad INDRA SSI S.A.) con el objetivo central de la búsqueda de nuevas formas y métodos para la recuperación de información legislativa en Europa. Partiendo de la experiencia y resultados obtenidos en el ya mencionado proyecto RASKE del parlamento finlandés, se trataba ahora de buscar y ofrecer un único y consistente interfaz al usuario para la recuperación de información legislativa generada en cualquiera de los países de la Unión Europea tratando, de este modo, de superar las enormes barreras que existían (y aún existen en la actualidad) para dicho acceso debido a la enorme pléyade de repositorios y bases de datos jurídicas existentes en Europa, con formatos dispares y en un amplio número de lenguas. Esta gran heterogeneidad en dichas bases de datos jurídicas hacía que la información contenida estuviese estructurada, organizada y clasificada de múltiples formas, con contenidos en diferentes idiomas y con técnicas de recuperación ciertamente dispares entre ellas. El modelo propuesto se basaba en la utilización de la tecnología SGML/XML y en la idea de asignación de metadatos a la información tratada. Así, en el núcleo del sistema implementado se encuentra una base de datos relacional sobre metadatos, definida su

³⁵⁷ La dirección oficial en Internet de este interesante proyecto era la de <http://www.eulegis.net/>. Y decimos que era pues actualmente no se encuentra, lamentablemente, operativa debido, en gran medida, al estado en el que se encuentra este desarrollo (un prototipo no comercial). Es por ello, que la información relativa al mismo debe ser consultada hoy en día a través de otros documentos referenciales, impresos y electrónicos. Un resumen de este proyecto y de la bibliografía generada hasta el año 2000 se encuentra disponible entre las páginas personales de una de sus principales investigadoras, la ya citada con anterioridad en esta tesis doctoral Airi Salminen, en la dirección <http://www.cs.jyu.fi/~airi/docman.html#eulegis>. Además de lo relacionado en dicha página web resulta ciertamente recomendable la lectura de la comunicación presentada al congreso *XML Europe 2001* por ésta y otros investigadores participantes en este proyecto; véase Virpi Lyytikäinen, Pasi T. Tiitinen, Airi Salminen. "XML Metadata for Accessing Heterogeneous Legal Databases" [documento HTML]. En: *XML Europe 2001* (Berlín. May 2001). Disponible en <http://www.gca.org/papers/xml europe2001/papers/html/s27-4.html> (consultado el 25 de junio de 2001), así como el artículo de estos mismos autores en Virpi Lyytikäinen, Pasi T. Tiitinen, Airi Salminen. "Supporting Access to Information Created in Interorganizational Processes". En: Amita Goyal Chin (ed.).

estructura mediante una DTD de XML, y donde este lenguaje es empleado para establecer el formato de intercambio de datos entre los módulos que conforman el sistema. De igual forma, XML fue el formato utilizado para el intercambio de datos entre las personas encargadas de la recopilación de la información necesaria para alimentar el sistema. El modelo de metadatos empleado en este proyecto se basaba sobre todo en el empleo de los mapas de materias (*topic maps*) más que en la utilización de tecnologías XML específicas de metadatos, como es el desarrollo RDF. Así, el sistema contemplaba cuatro módulos esenciales puesto en relación unos con otros: el DATABASE, el cual describe la base de datos legislativa; el PROCESS, el cual describe el proceso de generación de la legislación en cada uno de los sistemas nacionales tratados; el ACTOR, el cual describe las organizaciones y las personas involucradas en la generación de la legislación correspondiente a cada uno de esos sistemas legislativos nacionales; y el DOCUMENT, el cual describe los tipos de documentos generados en el proceso legislativo dentro de cada uno de esos sistemas. Así, la base de datos relacional de los metadatos contiene una descripción de DATABASE para cada uno de los sistemas legislativos tratados, así como un módulo de PROCESS, otro de ACTORS y otro de DOCUMENTS. Cada uno de estos módulos, como decíamos anteriormente, estaba definido con su correspondiente DTD de XML. El resultado final proporcionaba una visualización gráfica de todos estos módulos, con lo que el usuario podía ir navegando desde lo general a lo particular a través de estos mapas visuales o ir directamente a la base de datos legislativa deseada (de alguno de los países integrantes en el proyecto a la propia base de datos CELEX de la Unión Europea) y lanzar, por tanto, la búsqueda de información a través de un formulario unificado. Los resultados obtenidos de la base de datos eran transformados al formato HTML a través de la utilización de la tecnología XSLT.

El **Parlamento Europeo** en Estrasburgo es otra de las instituciones europeas que de forma más decidida ha apostado por la tecnología XML para el tratamiento de los documentos de carácter jurídico. Desde esta institución supranacional se ha venido

potenciando en los últimos años la necesidad de crear mecanismos tecnológicos comunes entre todos los parlamentos europeos que permitan, por un lado, incrementar el diálogo democrático entre los ciudadanos y los parlamentarios que los representan y, por el otro, que agilicen y mejoren la gestión de la ingente cantidad de información y de documentación que se produce con la aplicación de estas nuevas tecnologías informáticas y de comunicación. Si todos los parlamentos democráticos europeos mantienen una serie de características comunes en su funcionamiento, estructura, actuaciones legislativas y procesos de difusión de sus actos resulta lógico pensar que en el ámbito de actuación comunitario se traten de crear mecanismos que posibiliten una mejor accesibilidad, comprensión e intercambio de la información producida, haciendo uso para ello de las nuevas tecnologías de Internet, y en especial del servicio de la WWW. Partiendo de esta idea base, el Parlamento Europeo, a través del *Task Force Information and Documentation Management*, lanzó en abril de 2000 un documento³⁵⁸ para su debate entre los parlamentarios y otras partes interesadas en el cual se establecía la necesidad de desarrollar un lenguaje de marcado común para la información y los documentos parlamentarios; a esta iniciativa se la conoce bajo el nombre de **ParlML**³⁵⁹. Este grupo de trabajo entiende que, como así es expresado en el propio documento, la gran diversidad de lenguas (tanto humanas como informáticas) que se dan en los distintos sistemas parlamentarios de la Unión Europea acarrea serios problemas para hacer efectivos los principios de intercambio e de integración de información antes mencionados. Se entiende, igualmente, que hoy en día estas diferencias pueden ser en gran medida salvadas con empleo de las tecnologías XML pues, aunque no todas ellas son aún estándares terminados, ofrecen la suficiente estabilidad y confianza como para ser adoptadas por este organismo. De este modo y a través de esta llamada, el Parlamento Europeo invita a todos los interesados a participar con sus ideas y

³⁵⁸ European Parliament. *A common vocabulary for Parliamentary language. Call for Expressions of Interest: Development of a markup language for Parliamentary texts and information* [documento PDF]. Brussels: European Parliament, Task Force Information and Documentation Management, April 26, 2000. TFDM(2000)0014. Disponible en [http://www.europarl.eu.int/docman/texts/TFDM\(2000\)0014EN0.pdf](http://www.europarl.eu.int/docman/texts/TFDM(2000)0014EN0.pdf) (consultado el 24 de mayo de 2001).

³⁵⁹ Toda la información oficial relativa a esta iniciativa del Parlamento Europeo se encuentra disponible dentro de su sitio web oficial en la dirección <http://www.europarl.eu.int/docman/ParlML/>

contribuciones a este proyecto de investigación a través del foro electrónico creado a tal fin³⁶⁰.

Pero ha sido, sin duda, en el campo de la traducción y la publicación donde la Unión Europea y sus organismos dependientes más han invertido en el aspecto tecnológico del tratamiento de la información generada. Y es lógico que esto sea así pues en una organización en la que se han de traducir y publicar los documentos generados a las once lenguas oficiales existentes³⁶¹ en dicha organización supone una labor ardua y compleja, necesitada de cualquier ayuda informática para una mayor agilidad de trabajo en dichos procesos.

Al amparo de esta imperiosa necesidad, la Unión Europea ha venido desarrollando varios proyectos de gran interés en el campo de la asistencia tecnología a la labor de traducción en todos los organismos dependientes de dicha institución, siendo el proyecto más emblemático en el campo de la administración de justicia europea el denominado proyecto GTi, del cual hablaremos con posterioridad.

Comentaremos antes que, como ya se introdujo en el apartado II.4.2, la **Oficina de Publicaciones Oficiales de la Unión Europea** ha sido uno de los centros más activos en esa labor de traducción, edición y publicación de los documentos oficiales generados por las diferentes instituciones integrantes en esta organización supranacional. Si en un principio esta oficina venía desarrollando su labor editorial haciendo uso de las tecnologías SGML, como ya se explicó en su momento, en los últimos años ha ido variando esta política de trabajo hacia el uso de las tecnologías XML³⁶².

³⁶⁰ El acceso a los documentos ya existentes dentro de este foro de discusión es libre y gratuito, encontrándose disponibles para su consulta en la dirección <http://www1.europarl.eu.int/forum/docman/dispatch.cgi/ParlML> aunque el acceso completo al resto de documentos internos y a los mensajes que se producen en los debates electrónicos requiere de un registro previo.

³⁶¹ En realidad son doce lenguas oficiales: el español, el danés, el alemán, el inglés, el francés, el griego, el italiano, el holandés, el portugués, el finlandés, el sueco y el irlandés, aunque en la práctica esta última lengua jamás se ha utilizado.

³⁶² En septiembre de 1999 se presentó al Parlamento Europeo una propuesta de sistema editorial global para todas los organismos dependientes de la Unión Europea basado en el uso de las tecnologías XML. Esta

En el caso concreto de la aplicación de las tecnologías XML al campo de la justicia en el seno de la Unión Europea lo podemos encontrar, por tanto, en la aplicación del **Proyecto GTi** (*Generic Text interface*)³⁶³ a la traducción de los documentos emanados del **Tribunal de Justicia de las Comunidades Europeas** (incluyendo aquí, obviamente, el Tribunal de Primera Instancia)³⁶⁴. Este importante proyecto se pone en marcha en 1997 como prototipo de desarrollo de un sistema de generación automática de proyectos de traducción de documentación judicial, en aquel primer momento circunscrito exclusivamente a la traducción de las sentencias emanadas del Tribunal. Este prototipo, denominado por aquel entonces *Generic Translation interface*, funcionaba en un entorno de producción de WordPerfect para DOS. El empleo de este prototipo para otros tipos documentales, como las conclusiones o los resúmenes hizo que a finales de 1998 se viese la posibilidad de extender estas capacidades para la creación y traducción de todo tipo de documentos jurisprudenciales, rebautizándose en esos momentos a su actual nombre de *Generic Text interface*. De este modo, y de forma reciente, se ha venido investigando y trabajando en este

ambiciosa idea tecnológica, tras la cual estaba la compañía internacional *Infotechnique* (<http://www.infotechnique.com>), contemplaba el tratamiento y actuación en los tres ámbitos habituales de trabajo documental dentro de la Unión; esto es: la creación de los documentos, la traducción de los mismos a las diferentes lenguas oficiales y la publicación de los mismos en diferentes medios y soportes. En todos los casos, en el núcleo central de esta idea se situaba el metalenguaje XML apoyándose en otras tecnologías asociadas, como el protocolo HTTP para la publicación en la web, el sistema Unicode para la codificación de los caracteres en las diversas lenguas oficiales, los estándares XSL/XSLT para la asociación y/o transformación de formatos, los estándares X-Link y X-Pointer para el establecimiento de relaciones hipertextuales entre los documentos, el empleo del procesador de texto WordPerfect 8 SGML para la generación de los documentos electrónicos, etc. Fruto de esta idea, los diferentes organismos de la Unión Europea han empezado a trabajar sobre esta base siendo, lógicamente, la Oficina de Publicaciones Oficiales el organismo que con mayor decisión ha empezado a hacer uso de todas estas tecnologías. El documento de presentación de este sistema editorial basado en el uso de la tecnología XML se encuentra disponible en http://www1.europarl.eu.int/forum/comedi/dispatch.cgi/Infodocs/docProfile/100097/d1991008125521/No/XML_GTI.pdf

³⁶³ Natalie Maurice, Directora de los servicios de traducción del Tribunal de Justicia de las Comunidades Europeas, presentó en el seno de la *Conferencia para una Infraestructura Terminológica en Europa*, desarrollada en París entre los días 13 al 15 de marzo de 2000, una interesante comunicación en la cual se explicaba con todo detalle el desarrollo y puesta en marcha de este importante proyecto. Recomendamos, pues, para una mayor información sobre el mismo la consulta de Natalie Maurice. "Intégration de ressources documentaires et linguistiques dans un processus de production multilingue de documents. Le projet GTi (Generic Text interface)" [documento HTML]. En: *Conférence pour une Infraestructura Terminologique en Europe* (2000. París). Disponible en <http://www.unilat.org/dtil/etis/actasTDCnet/maurice.htm> (consultado el 25 de febrero de 2001).

proyecto para cubrir todo el proceso de producción y de gestión de los documentos emanados del Tribunal de Justicia de las Comunidades Europeas, tanto los internos (la jurisprudencia) como los externos (las diferentes piezas que se integran dentro del procedimiento judicial). Así, el sistema desarrollado proporciona ayuda para la redacción de los documentos, para la inclusión de citas a otras sentencias o normas jurídicas, para la redacción de los fundamentos jurídicos, para la traducción de todos estos documentos y, finalmente, para la comunicación y publicación de los mismos en los diferentes medios y espacios de difusión, tanto electrónicos como impresos. Todo este complejo entramado tecnológico se sustenta bajo una arquitectura de trabajo cliente-servidor y donde el lenguaje XML es empleado como formato para el marcado, el almacenamiento y el intercambio de estos documentos electrónicos. El tratamiento de todos estos textos se realiza mediante el procesador WordPerfect 9 XML. Se aplican, igualmente, otras tecnologías actuales en la Web, como es el caso de los lenguajes de programación Java y JavaScript, para la creación de la interfaz entre el tratamiento de los textos y los diferentes módulos del sistema GTi, así como para la interfaz del navegador web empleado para la búsqueda y recuperación de información entre las diversas fuentes documentales de consulta (bases de datos procedimentales, terminológicos, fraseológicos, y textuales). Todo el sistema de gestión de bases de datos relacionales recae sobre el programa Oracle 8i.

Sin embargo, y a pesar de tan importante desarrollo, el Tribunal de Justicia de las Comunidades Europeas, a diferencia de lo que está sucediendo en los tribunales de justicia norteamericanos descritos con anterioridad, no ha contemplado todavía la posibilidad de que toda la actividad relativa a un procedimiento judicial sea iniciada a través de los actuales medios electrónicos y el empleo de la tecnología XML. Así, el procedimiento ante este tribunal de justicia debe iniciarse del modo clásico, es decir, mediante un escrito de recurso dirigido a la Secretaría de dicho tribunal para que, de este modo, comiencen los trámites y actuaciones correspondientes.

³⁶⁴ Toda la información oficial relativa al Tribunal de Justicia y al Tribunal de Justicia de Primera Instancia de las Comunidades Europeas se encuentra disponible en la dirección <http://www.curia.eu.int/>

Dentro del ámbito de actuación jurídica de los países europeos cabe destacar igualmente la labor emprendida en Alemania y otros países vecinos para la constitución de un grupo de investigación y trabajo de similares características al ya analizado Legal XML; es decir, la versión o rama europea de esta organización internacional. Se trata del grupo denominado **LeXML**, con tres nodos principales en la actualidad: en Alemania³⁶⁵, en Austria³⁶⁶ y en Holanda³⁶⁷. Este importante grupo europeo de investigación y aplicación de las tecnologías XML a la documentación jurídica surge prácticamente al mismo tiempo que su *hermano mayor*, el Legal XML³⁶⁸, y en la cual se integra como un grupo de trabajo. Este grupo de trabajo europeo persigue la misma finalidad que la organización internacional pero tratando de explorar y adaptar los estándares desarrollados a lenguas no inglesas. Como se indica en la propia página dedicada a los objetivos de esta organización europea, el LeXML es “un punto de coordinación y de fuerza de trabajo para el desarrollo de estructuras normalizadas, vocabularios y herramientas para el intercambio de datos. LeXML trata de alcanzar sus objetivos a través del desarrollo de una modelo global de datos jurídicos e, igualmente, a través del desarrollo de un programa de código abierto para la gestión de oficinas jurídicas y bufetes de abogados, el cual hable y entienda el XML”³⁶⁹. La actividad principal de este grupo europeo se centra en la actualidad en el desarrollo de un diccionario de metadatos (desarrollado bajo el estándar RDF) de términos jurídicos³⁷⁰, siendo

³⁶⁵ <http://www.lexml.de/>

³⁶⁶ <http://www.lexml.at/>

³⁶⁷ Aunque el URL del nodo holandés apunta a un espacio de trabajo de la Facultad de Derecho de la universidad de Leiden en Holanda (<http://law.leiden.edu/xml/>), al activarlo redirecciona su posición al nodo principal de esta organización en Alemania.

³⁶⁸ La estrecha relación existente entre el Legal XML y la representación europea del mismo a través del LeXML fue expuesta Peter Ebenhoch en el congreso de XML Europe 2001. Para una mayor información al respecto de este tema, véase Peter Ebenhoch. “The quest of standardising law. LegalXML and LeXML” [documento HTML]. En: *XML Europe 2001* (2001. Berlin). Disponible en <http://www.gca.org/papers/xml europe2001/papers/html/s27-5.html> (consultado el 3 de julio de 2001).

³⁶⁹ Texto traducido de la dirección http://home.snaufe.de/lexmlde/mission_english.htm

³⁷⁰ El sitio web oficial en el cual se encuentra toda la información disponible sobre la aplicación del estándar RDF a este proyecto se halla en la dirección <http://rdf.lexml.de/>

presentada la versión en lengua germana en septiembre de 2001 en el congreso de la EDV Gerichtstag desarrollado en la ciudad alemana de Saarbrücken³⁷¹. Este diccionario jurídico en RDF se encuentra aún, al igual que el que se está desarrollando en el seno de la organización Legal XML, en fase de desarrollo por lo que la información oficial que se suministra sobre este interesante proyecto no va más allá de una mera declaración de intenciones y buenos propósitos, remitiéndonos, por el momento, a varios documentos anteriores en el que se abordan la necesidad de definir y construir tipos de datos jurídicos a través del uso de la tecnología XML³⁷².

Finalizaremos este apartado dedicado a las aplicaciones jurídicas del metalenguaje XML al continente europeo concluyendo exponiendo los planes de modernización tecnológica de las administraciones públicas de algunos países europeos en los cuales se contempla el uso de dichas tecnologías de marcado de documentos electrónicos.

En primer lugar, resulta ciertamente destacable los esfuerzos encaminados a este fin por el gobierno del **Reino Unido**. En octubre de 2000 el ministro británico de industria y comercio Ian McCartney (actualmente, Ministro de Pensiones y Seguridad Social) anunció³⁷³ el lanzamiento de un plan de modernización tecnológica de las administraciones

³⁷¹ Este importante congreso se desarrolló en la Universidad de Saarland, sita en la ciudad de Saarbrücken al sur-oeste de Alemania. La dirección en Internet en la cual se halla toda la información relativa a este congreso es <http://edvgt.jura.uni-sb.de/index.shtml> La información oficial de la organización LeXML relativa a esta presentación se encuentra <http://home.snafu.de/mmuller/lexmlde/flyeredvgt.htm>

³⁷² Ambos artículos están en alemán. En el primero de ellos, redactado en junio de 2000, se establece de forma general el uso y aplicación del modelo de DTD de XML para la construcción de tipos de datos de carácter jurídico; véase en Murk Muller. *Ein juristisches Datenmodell für XML* [documento HTML]. LeXML, 23.6.2000. Disponible en <http://home.snafu.de/mmuller/lexmlde/datenmodellmm.htm> (consultado el 25 de febrero de 2001). El segundo de ellos, elaborado por el anterior autor en abril de 2001, se centra mucho más en la aplicación concreta del modelo RDF de XML para la construcción de los tipos de datos semánticos capaces de definir la información jurídica contenida; véase Murk Muller. *XML und RDF Dictionary: Austausch juristischer Informationen zwischen Computern* [documento HTML]. LeXML, April 2001. Disponible en <http://home.snafu.de/mmuller/lexmlde/xml-rdf.htm> (consultado el 13 de junio de 2001).

³⁷³ Como señala Robin Cover en un documento electrónico elaborado al respecto del tema, esta presentación se realizó en la QEII Conference Centre, desarrollada en Londres el 11 de octubre de 2000, y en la cual también intervinieron representantes de compañías comerciales del ámbito tecnológico tan importantes como



públicas a través del denominado, y ya visto con anterioridad en el capítulo primero de esta tesis doctoral, *e-Government Interoperability Framework* (e-GIF)³⁷⁴. Como se comenta en la propia información oficial de este importante marco de trabajo, la modernización del gobierno y de sus instituciones pasa por la integración de los sistemas de información existentes en el sector público del país. El e-GIF, por tanto, define los requisitos previos considerados como esenciales para conseguir esta integración tecnológica de los sistemas de información gubernamentales y su disponibilidad por parte del ciudadano a través del servicio de la WWW. Así, el gobierno británico ha considerado que para poder llevar a cabo dicha integración de los datos elaborados por el sector público del país resulta imprescindible el uso de políticas y estándares tecnológicos abiertos, siendo la tecnología XML la más adecuada para llevar a cabo dicha integración³⁷⁵.

De forma más detallada, este ambicioso marco de trabajo señala tres puntos clave de desarrollo y aplicación de la tecnología XML al e-GIF: en primer lugar, el metalenguaje XML y los esquemas XML para la integración de los datos; en segundo lugar, la aplicación de los estándares UML, RDF y XML para la modelización de los datos y el lenguaje descriptivo; y tercero, el empleo de los estándares XSL, DOM y XML para la presentación de los datos. Se pretende así, con la apuesta firme y decidida por parte del gobierno británico en la adopción de las tecnologías de Internet y la WWW y el estándar XML, que todos los servicios gubernamentales sean accesibles por los ciudadanos en un plazo no superior al año 2005. Para poder alcanzar este ambicioso objetivo, el gobierno británico ha

Microsoft, IBM y CISCO, además de un representante de la Comisión Europea. Para una más amplia información, véase en la dirección <http://xml.coverpages.org/egifAnn20001011.html>

³⁷⁴ Toda la información oficial relativa a este importante plan nacional sobre el desarrollo tecnológico en el Reino Unido, que desde mayo de 2001 se encuentra en su versión 2, se encuentra disponible en el sitio web del organismo oficial denominado *Office of the e-Envoy* en la dirección <http://www.citu.gov.uk/publications/frameworks/egif2/execsum.htm>. La versión 2 del documento oficial que da soporte a este importante marco estratégico se encuentra disponible en *E-Government Interoperability Framework* [documento PDF]. London: Office of the e-Envoy, May 2001. Disponible en http://www.e-envoy.gov.uk/publications/frameworks/egif2/e-GIF_ver2.pdf (consultado el 12 de septiembre de 2001). Este documento electrónico se encuentra disponible en otros formatos de codificación (Word, rtf y HTML).

³⁷⁵ Además del sitio web oficial del e-GIF, comentado en la anterior nota, resulta recomendable la consulta de la página web elaborada por Robin Cover relativo a todo lo concerniente a este plan estratégico político en relación con el uso y aplicación de la tecnología XML, en la dirección <http://xml.coverpages.org/egif-UK.html>

tratado de involucrar a todos los organismos del sector público, de la industria y a cualesquiera otras personas u organismos interesados en el debate electrónico de ideas y el desarrollo conjunto de políticas y estándares capaces de obtener dicho objetivo marcado. Con esta idea se puso en marcha la iniciativa denominada *UK GovTalk*³⁷⁶, la cual cubre todos los aspectos relacionados con el e-GIF así como otra iniciativa relacionada con la misma, denominada *e-Government Metadata Framework* (e-GMF)³⁷⁷, relativa a la política gubernamental de aplicación de metadatos a los recursos de información pública. A pesar de que aún queda, como hemos señalado, un largo trayecto hasta que este importante proyecto llegue a su conclusión y, por tanto, todos los servicios gubernamentales utilicen las tecnologías mencionadas y la información contenida en sus bases de datos pueda ser accesible en línea por todos los ciudadanos, ya es posible acceder a una buena parte de los servicios electrónicos del gobierno británico, incluyéndose entre éstos los relativos a los tribunales de justicia de este país.

Francia es sin duda otro de los países europeos que de forma más decidida está apostando por el desarrollo tecnológico e informático de sus administraciones públicas y donde, como veremos a continuación, el desarrollo y aplicación de los estándares XML juega un papel ciertamente importante. El gobierno francés estableció en agosto de 1998 un marco político y estratégico, para impulsar el desarrollo tecnológico en todas las administraciones públicas del estado bajo el nombre de *Mission interministérielle de soutien technique pour le développement des technologies de l'information et de la communication dans l'administration* (MTIC). Este importante programa, bajo la responsabilidad y autoridad del

³⁷⁶ <http://www.govtalk.gov.uk/>

³⁷⁷ Desde las páginas del sitio web de la *Office of the e-Envoy* se remite al ciudadano o investigador interesado en este desarrollo a una dirección específica del *UK GovTalk* (concretamente, a la dirección <http://www.govtalk.gov.uk/egif/meta.html>) ésta no se encuentra aún operativa por lo que resulta, por tanto, recomendable la consulta de la página web de la primera oficina citada para una mayor información al respecto del tema, en <http://www.e-envoy.gov.uk/publications/frameworks/metadata/metadata.htm>. Dentro de este espacio se encuentra, en formato electrónico, el documento oficial en el cual se describen las líneas de actuación de este importante marco de trabajo del gobierno británico. Véase *E-Government Metadata Framework* [documento PDF]. London: Office of the e-Envoy, May 2001. Disponible en <http://www.e-envoy.gov.uk/publications/frameworks/metadata/metadata.pdf> (consultado el 12 de septiembre de 2001).

Primer Ministro del país, dio lugar a mediados del año 2001 a la creación de un organismo gubernamental propio para el desarrollo y seguimiento de este plan estratégico: la denominada *Agence pour les Technologies de l'Information et de la Communication dans l'Administration* (ATICA)³⁷⁸. Este plan se ha venido articulando en tres bloques principales de actuación, a saber: la definición de un marco común de interoperabilidad ministerial, el desarrollo y potenciación de proyectos de trabajo interministerial y, finalmente, la implantación y desarrollo de las tecnologías de la información y la comunicación dentro las administraciones públicas. Para llevar a cabo este último objetivo se establecieron diversas vías de desarrollo, como eran las del ofrecimiento para su debate de posibles áreas de aplicación así como la difusión de los casos más sobresalientes, la estimulación de los denominados *clubes de capacidad* en ciertos temas (servicios en línea, aplicación de XML, etc.), la organización de encuentros, seminarios y congresos sobre sus programas de actuación, la difusión de informes y documentos técnicos sobre las distintas temáticas de interés³⁷⁹, etc. Centrándonos en el caso concreto del estudio, investigación y aplicación que desde esta agencia gubernamental se está realizando de las tecnologías XML resulta muy significativo el apoyo recibido a esta tecnología así como toda la labor emprendida hasta la fecha³⁸⁰.

Además de la existencia de diversos foros electrónicos de debate sobre XML, resulta especialmente destacable el denominado *Club XML*³⁸¹. Surgido al amparo de las decisiones adoptadas por el *Comité Interministériel à la Réforme de l'Etat* en octubre de 2000, relativas a la

³⁷⁸ Toda la información relativa a este importante plan estratégico y político para la implantación y desarrollo de las tecnologías de la información y la comunicación en Francia, así como sobre la agencia gubernamental encargada de su seguimiento, se encuentra disponible en la dirección <http://www.atica.pm.gouv.fr/>

³⁷⁹ Resulta sorprendente la capacidad y el dinamismo que ha venido manteniendo este programa de actuación gubernamental en el aspecto documental pues son numerosos los artículos, informes y notas que diversos autores han venido elaborando sobre multitud de temas relacionados con las nuevas tecnologías de la información y sus aplicaciones al ámbito de la administración pública (algunos de ellos relativos a XML). Todos estos informes se encuentran disponibles en la dirección <http://www.atica.pm.gouv.fr/dossiers/>

³⁸⁰ La página del sitio web de ATICA en la que se relacionan todas las iniciativas y actuaciones emprendidas alrededor de esta tecnología se encuentra disponible en la dirección <http://www.atica.pm.gouv.fr/XML/>

³⁸¹ Toda la información oficial relativa a este importante grupo de investigación y trabajo se encuentra disponible en la dirección <http://www.atica.pm.gouv.fr/XML/clubxml/introduction.shtml>

utilización de las tecnologías XML para el desarrollo de un amplio repertorio de esquemas XML de aplicación en las diversas administraciones públicas, este club se conforma como un elemento fundamental para el debate e intercambio de ideas entre aquellos profesionales e investigadores del país interesados en el desarrollo de dicho repertorio³⁸². Así, y fruto de esta actividad, se han venido poniendo en marcha en estos últimos años unos cuantos proyectos de investigación en el desarrollo de esquemas XML y DTDs aplicados a la información electrónica de diversos organismos públicos del Estado³⁸³.

Además de los importantes desarrollos que se están llevando a cabo en diversos ministerios de este país (como son el Ministerio de Transportes y Vivienda, el Ministerio de Agricultura y Pesca, el Ministerio de Cultura y el Ministerio de Empleo y Solidaridad), resulta destacable por su adscripción jurídica el llevado a cabo por el Consejo Constitucional de Francia.

El **Consejo Constitucional**³⁸⁴ de Francia (de similar función y características al Tribunal Constitucional español) está desarrollando en la actualidad un proyecto de investigación aplicada bajo tecnologías XML, aún en fase de experimentación. Este proyecto se enmarca dentro de otro proyecto global del estado francés denominado *Service Public d'Accès au Droit* (SPAD)³⁸⁵ cuya finalidad principal está en la creación de un servicio público de acceso al derecho que ponga a disposición de todos los ciudadanos, a través de Internet y de forma gratuita, toda la legislación y jurisprudencia emanada de los diversos órganos competentes

³⁸² Resulta destacable el informe elaborado en abril de 2001 dentro del seno de esta agencia gubernamental sobre la importancia del estándar XML y la necesidad que tienen las administraciones públicas del estado francés en el desarrollo de un repertorio de esquemas de acceso público con el doble objetivo de favorecer la interoperabilidad e intercambio electrónico de datos así como el de facilitar la conservación de los documentos electrónicos. Véase en *De XML au répertoire de schémas XML de l'administration* [documento HTML]. Paris: ATICA, 12 avril 2001. Disponible en <http://www.atica.pm.gouv.fr/dossiers/documents/lettreXMLrepertoire.shtml> (consultado el 12 de septiembre de 2001).

³⁸³ Toda la información oficial relativa a estos proyectos de investigación aplicada en el desarrollo de esquemas y DTDs XML se encuentra disponible en la dirección <http://www.atica.pm.gouv.fr/XML/repertoire.shtml>

³⁸⁴ <http://www.conseil-constitutionnel.fr/>

³⁸⁵ La información oficial relativa a este importante proyecto en relación con el Consejo Constitucional se encuentra disponible en la dirección http://www.atica.pm.gouv.fr/XML/repertoire_de_schema/spad.shtml

en la materia. De ese modo, se pretende crear en un futuro próximo una gran base de datos jurídica para el almacenamiento y difusión de toda esta información. Los documentos que han de alimentar dicha base de datos serán tratados mediante XML para facilitar su almacenamiento y difusión en el espacio de la WWW. Y es esto en lo que actualmente está investigando y trabajando el Consejo Constitucional francés, a través del desarrollo de una DTD válida para marcar los documentos jurídicos emanados de esta institución. Existe definida ya una DTD XML³⁸⁶ para hacer posible esta transmisión de documentos jurídicos del Consejo Constitucional para su difusión a través del futuro servicio SPAD, de la cual se adjunta copia de la misma en el Apéndice 5 de esta tesis doctoral.

Es de lamentar que esta DTD XML no venga acompañada de un documento explicativo sobre la definición y el contexto de aplicación de cada uno de los elementos señalados en la misma o, al menos, de algún comentario inserto en la propia especificación de la DTD. Afortunadamente nos encontramos en el sitio web informativo de este proyecto un ejemplo de aplicación de esta DTD a una sentencia del Consejo Constitucional francés que nos facilita realizar una mayor aproximación al entendimiento de la misma³⁸⁷. Por dicho motivo, resulta ciertamente difícil realizar una interpretación exacta del significado de algunos de los elementos contenidos en esta DTD para todas aquellas personas ajenas a la documentación jurídica emanada de este Consejo Constitucional pues, además de los problemas ya destacados, abundan sobremanera elementos opcionales y la entrada de códigos previamente establecidos para determinar el contenido informativo de muchos de

³⁸⁶ El documento electrónico en formato de MS-Word en el cual se encuentra transcrita esta DTD, además de la representación gráfica del modelo de contenido de cada uno de los elementos integrantes de la misma, está disponible en la dirección http://www.atika.pm.gouv.fr/XML/documents/spad_docum.doc

³⁸⁷ El documento XML en cuestión se encuentra disponible en la dirección http://www.atika.pm.gouv.fr/XML/documents/spad_CCO00DC04441.xml. Lamentablemente éste documento XML no puede ser visualizado directamente en el navegador (aunque sí descargado) al estar mal ubicada la referencia a la DTD que ha de validar la instancia de este documento. La sentencia tratada mediante el marcado XML es la correspondiente a la Decisión n° 2000-441 DC, de 20 de diciembre de 2000, relativa a las rectificaciones realizadas por el gobierno en el 2000 sobre la Ley de finanzas. El texto de esta sentencia se encuentra igualmente disponible en formato HTML en el sitio web del propio Consejo Constitucional en la dirección <http://www.conseil-constitutionnel.fr/decision/2000/2000441/index.htm> así como en el portal jurídico en la web del gobierno francés *Legifrance* en la dirección http://www.legifrance.gouv.fr/citoyen/jorf_nor.ow?numjo=CSC0004589S

los elementos existentes en la misma. En cualquier caso, trataremos de dar una explicación lo más exacta posible de la composición de los principales elementos de esta DTD.

La representación gráfica del conjunto de elementos y subelementos definidos en esta DTD es la siguiente:

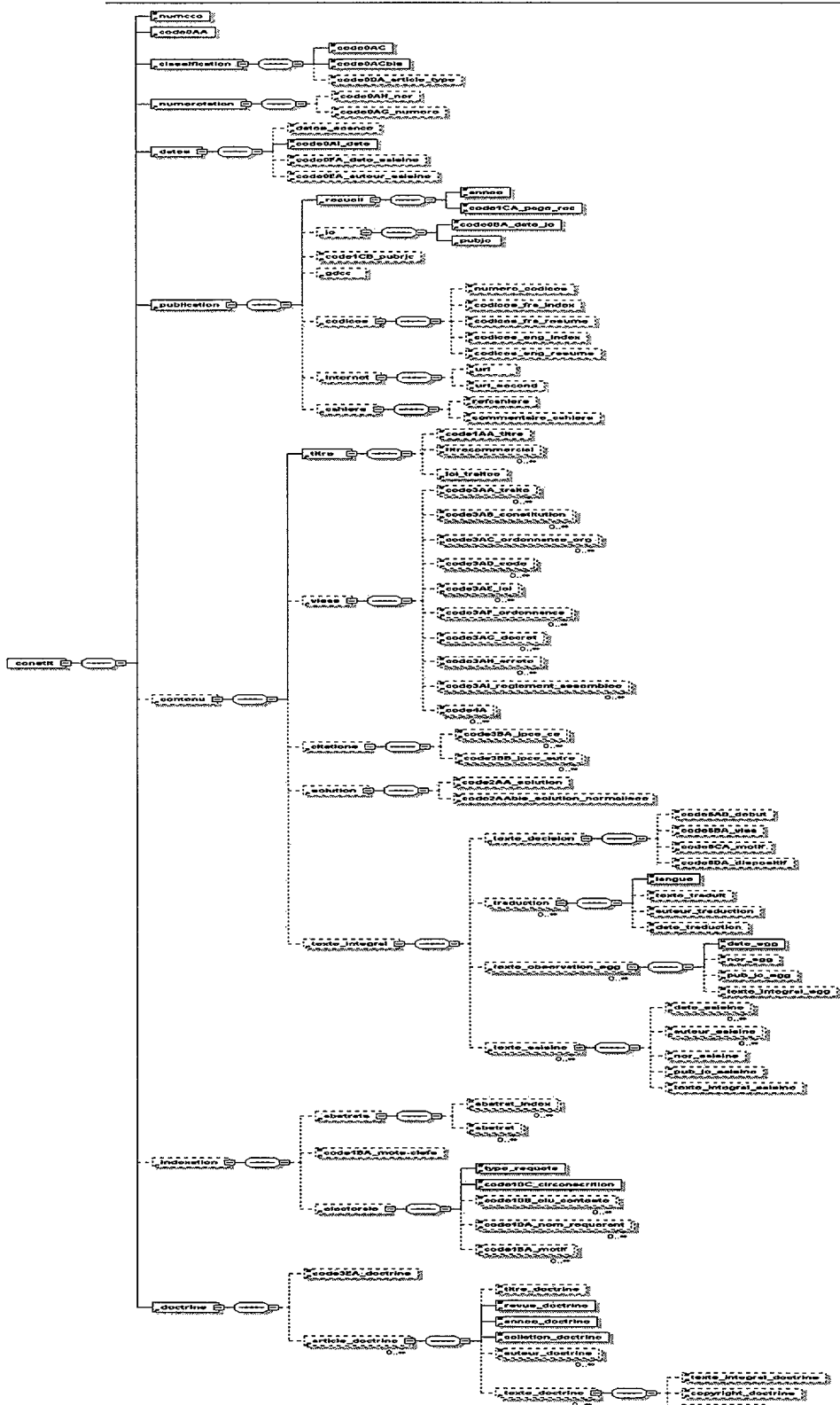


Figura III.21: Diagrama general de estructuración de la DTD definida para el Consejo Constitucional francés.

El elemento de documento o elemento raíz viene constituido por la palabra *constit*, el cual contempla en su modelo de contenido a otros elementos hijos, la mayoría obligatorios y no repetibles (*numcco*, *code0AA*, *classification*, *numerotation*, *dates*, *publication* y *doctrine*) y otros opcionales e igualmente no repetibles (*contenu* e *indexation*).

El primer elemento hijo señalado, el *numcco*, hace referencia al número de control o de registro de este documento por parte del sistema.

El siguiente elemento, el *code0AA*, hace referencia al nombre de la institución emisora del documento (en este caso el propio Consejo Constitucional), definido su contenido a través del valor de su atributo (*emetteur*).

El elemento *classification* define en su modelo de contenido tres elementos hijos (dos de ellos obligatorios, el *code0AC* y el *code0Acbis*, y el otro opcional, el *code0DA_article_type*) para establecer, a través de los valores posibles de un atributo para cada uno de ellos, una serie de códigos clasificatorios del tipo de documento jurídico que se está marcando (por ejemplo, para una decisión de este tribunal se empleará en el primer elemento el valor DC, en el segundo el DC04 y el tercero correspondería al número del artículo y número de párrafo en aplicación de la constitución francesa).

Algo similar ocurre con el elemento *numerotation* pues su modelo de contenido define dos elementos hijos, *code0AH_nor* y *code0AG_numero*, en este caso ambos optativos, para establecer los códigos relativos a la norma jurídica y al número de la misma.

Igualmente son optativos todos los elementos hijos contenidos en el modelo de contenido del elemento *dates*, relativos a las fechas de la sesión y la sentencia resultante así como autores o intervinientes en la consulta, como son *dates_seance*, *code0AI_date*, *code0FA_date_saisine* y *code0EA_auteur_saisine*.

En el caso de los elementos hijos que se establecen para el espacio dedicado a la publicación del documento, definido con el elemento *publication*, tan sólo uno es obligatorio, el elemento *recueil*, destinado a establecer los datos informativos relativos a la obra impresa, apoyándose para ello en la subdivisión de dos elementos hijos, como son *annee* y *code1CA_page_rec*. Los demás elementos hijos para la publicación son, como decíamos, optativos y corresponden a *jo* (fecha y localización en la publicación del boletín oficial del estado), *code1CB_publrc*, *gdcc* (nombre, número y página), *codices* (los índices y los

resúmenes de los códigos jurídicos en francés y en inglés), *internet* (los URLs donde se encuentran disponibles el texto de la sentencia) y *cabiers* (comentarios y referencias a los pliegos).

El elemento opcional destinado a analizar el contenido de la sentencia (*contenu*) contiene como único elemento hijo obligatorio el destinado a reflejar el título (*titre*) de la sentencia, estableciendo para ello otros subelementos para llevar a cabo este control (*code1AA_titre*, *titrecommercial* y *loi_traitee*). Los restantes elementos hijos del contenido de la sentencia analizan aspectos tales como las normas jurídicas referenciadas en la sentencia (*visas*) a través de diversos subelementos correspondientes a la codificación de los títulos de las normas nacionales o comunitarias, las citas (*cites*) explícitas a textos jurídicos, si las hubiera, la resolución adoptada por el Consejo Constitucional (*solution*) definiendo para ello dos subelementos para incluir el texto como fue redactado (*code2AA_solution*) y la forma textual normalizada (*code2Aabis_solution_normalisee*), y, finalmente, el elemento que engloba el texto íntegro de la sentencia (*texte_integral*). Los elementos hijos, y los subelementos de éstos, que se incluyen dentro del elemento relativo al texto íntegro de la sentencia son ciertamente numerosos. Así, nos encontramos en un primer nivel con el texto propiamente dicho de la decisión (*texte_decision*), el cual incluye cuatro subelementos característicos de los bloques principales de la sentencia: preámbulo con la composición de los miembros del consejo (*code6AB_debut*) y el texto con los antecedentes jurídicos empleados (*code6BA_visa*), las motivaciones o fundamentos jurídicos esgrimidos en el acto (*code6CA_motif*) y, finalmente, el fallo de la decisión adoptada (*code6DA_dispositif*). En ese mismo primer nivel encontramos al elemento *traduction* con los distintos subelementos correspondientes para señalar el idioma (*langue*), el texto traducido (*texte_traduit*), el autor (*auteur_traduction*) y la fecha de la traducción (*date_traduction*), si la hubiere. Igualmente se reflejan en ese nivel las observaciones o anotaciones realizadas por el gobierno a dicho texto (*texte_observation_sgg*), a través de los correspondientes subelementos de fecha (*date_sgg*), el código de la norma comentada (*nor_sgg*), la publicación de estos comentarios en el diario oficial (*pub_jo_sgg*) y el texto íntegro de los mismos (*texte_integral_sgg*). Finalmente, en ese mismo nivel se contempla el texto de la consulta (*texte_saisine*) realizada por este Consejo a los senadores y a los diputados al respecto del tema tratado, utilizando para ello los subelementos relativos

a la fecha de la consulta (*date_saisine*), la autoría de la misma (*auteur_saisine*), la norma tratada (*nor_saisine*), la publicaciones de este texto en el diario oficial (*pub_jo_saisine*) y, finalmente, el texto íntegro de la consulta (*texte_integral_saisine*).

Otro de los elementos opcionales del elemento raíz es el dedicado a la indexación (*indexation*) del documento en el cual se emplean como elementos hijos del mismo el resumen (*abstrats*) y otros relativos a legislación electoral (*electorale*).

Finalmente, el elemento raíz constit contempla como elemento obligatorio al elemento *doctrine*, en el cual se va a incluir toda la información relativa a indicar las referencias doctrinales empleadas en esta decisión. Así, el primer elemento hijo que nos encontramos (*code3EA_doctrine*) está dedicado a aglutinar de forma general todos los datos relativos a las obras doctrinales empleadas. Es en el elemento hijo *article_doctrine* donde se hace un desglose pormenorizado de cada una de las referencias doctrinales existentes, estableciendo diversos subelementos para el título del artículo (*titre_doctrine*), el título de la revista (*revue_doctrine*), año de publicación (*annee_doctrine*), volumen, número y páginas de ubicación (*collation_doctrine*), el autor del artículo (*auteur_doctrine*) y, si fuera menester, el texto íntegro del artículo doctrinal (*texte_doctrine*), conteniendo aquí otros subelementos relativos al texto propiamente dicho (*texte_integral_doctrine*), los derechos de reproducción (*copyright_doctrine*) y, si existiera, la dirección en Internet de dicho texto (*url_doctrine*).

Canadá, y más concretamente el **Gobierno de Québec**, es otro de los ejemplos más ilustrativos en el empleo de las tecnologías XML aplicadas al tratamiento y gestión de la información procedente de las administraciones públicas. En 1998 el gobierno Québec lanzó a través del Ministerio de Cultura y las Comunicaciones y del Ministerio de la Administración y la Función Pública un ambicioso plan político y estratégico de desarrollo tecnológico de las administraciones públicas denominado *Autoroute de l'information*³⁸⁸. Este importante plan, definido formal y oficialmente en el documento titulado *Politique québécoise*

³⁸⁸ En la actualidad, es la *Direction générale de l'autoroute de l'information*, dependiente del Ministerio de Cultura y de las Comunicaciones, en cooperación con el Ministerio de Estado y la Administración y la Función Pública, sobre quienes recaen la responsabilidad de llevar a cabo este programa. En cualquier caso, para una mayor

de l'autoroute de l'information³⁸⁹, tenía el objetivo principal de incrementar la investigación y las inversiones económicas por parte del gobierno de esta región autónoma canadiense en pos de alcanzar en pocos años el ideal internacionalmente planteado de una auténtica “sociedad de la información”, en la cual los ciudadanos de esta región hiciesen uso frecuente de las tecnologías de la información, y de forma muy especial de la red Internet, en la nueva sociedad y economía que se está fraguando.

Pero sin duda, y para los intereses de esta tesis doctoral, resulta de especial importancia lo reflejado en el Capítulo 5 del documento que define este plan político. En este capítulo, cuyo título es el de *Rapprocher L'Etat du citoyen et des entreprises*, se establecía la importancia capital de la información como recurso estratégico para el Estado y, consiguientemente, la necesidad en la renovación de los servicios que presta la administración pública a los ciudadanos y a las empresas haciendo uso para ello de las nuevas tecnologías de la información y la comunicación. De este modo, se establecía la necesidad de crear un plan específico de actuación conjunta para las diversas administraciones públicas del estado de Québec. Así, en abril de 1998 y desde la Secretaría del *Conseil du trésor* del Ministerio de Estado y Administración y de la Función Pública, se lanza el plan denominado *Inforoute gouvernementale*³⁹⁰, definido formalmente en el documento *L'inforoute gouvernementale: pour mieux servir les citoyens et les entreprises*³⁹¹, en el cual se ilustra la contribución de la administración pública a la política quebecuá de l'autoroute de l'information. En este nuevo plan se indica

información al respecto del tema, el sitio web oficial de este importante plan de desarrollo tecnológico se encuentra en la dirección <http://www.autoroute.gouv.qc.ca/>

³⁸⁹ *Politique québécoise de l'autoroute de l'information* [documento PDF]. Québec: Ministère de la Culture et des Communications, Ministère de l'Administration et à la Fonction publique, 1998. Disponible en <http://www.autoroute.gouv.qc.ca/politique/pdf/politiq.pdf> (consultado el 25 de junio de 2001).

³⁹⁰ Toda la información oficial relativa a este importante plan de actuación de la administración pública del estado de Québec se encuentra disponible en la dirección <http://www.tresor.gouv.ca/inforoute/inforoute.htm> El plan de actuación se divide en cinco grandes apartados, los cuales generan sus propios proyectos de investigación aplicada y documentación relacionada con los mismos; éstos son: puesta en marcha de la infopista gubernamental, seguridad e infraestructuras en la administración pública, servicios públicos en línea, infraestructuras y servicios comunes y, finalmente, la gestión de los recursos de información.

³⁹¹ *L'inforoute gouvernementale: pour mieux servir les citoyens et les entreprises* [documento PDF]. Québec: Secrétariat du Conseil du trésor, Avril 1998. Disponible en <http://www.tresor.gouv.qc.ca/doc/acrobat/gouverne.pdf> (consultado el 25 de junio de 2001).

expresamente que la evolución de están sufriendo las tecnologías de la información y las comunicaciones hacen factible el intercambio, la distribución y la reutilización de la información gubernamental en conformidad con las normas internacionales. Así, estas nuevas tecnologías han de servir para establecer infraestructuras de base y servicios comunes a la vez que se obtenga una simplificación en los procesos administrativos y mayor coherencia en las presentaciones de los servicios públicos suministrados a los ciudadanos y a las empresas.

Si bien es cierto que a lo largo del documento que define formalmente este plan se hacen frecuentes referencias a la tecnología de Internet como el vehículo más adecuado para llevar a cabo esta política de acercamiento y acceso a la información generada por la administración pública de Québec tanto a los ciudadanos como a las empresas, en éste no se hace ninguna referencia directa al empleo de los lenguajes de marcado de texto para el tratamiento de dicha información, quedándose en un simple y genérico apunte sobre el empleo de normas estándar al hablar de la arquitectura general que ha de tener este desarrollo³⁹².

Fruto de esta búsqueda del estándar adecuado para el tratamiento de la información electrónica de la administración pública, se ha venido trabajando a lo largo de estos últimos años en el desarrollo de esta idea así como en la creación de la infraestructura tecnológica para llevar a cabo tan ambicioso plan, siendo el interés existente hacia las tecnologías XML uno de los ejes fundamentales de las actuales actuaciones de la administración de Québec. Tan es así que en fechas recientes se ha presentado ante el Secretariado del Consejo del Tesoro un informe de investigación sobre la aplicación del XML para el tratamiento de la información electrónica que es generada, y que ha de ser servida, por la administración pública del gobierno de Québec. Este extenso e interesante informe, cuyo título exacto es

³⁹² En este informe se dice textualmente al respecto de este tema, dentro del apartado dedicado a la arquitectura general de la infopista gubernamental, lo siguiente: “definir, de forma conjunta entre los ministerios y organismos, las orientaciones, los estándares y los servicios relativos a los principales componentes de la infraestructura gubernamental y generar las oportunas guías para su organización, su implantación y su utilización [...]”. Véase, *Ibid.*, p. 47, <http://www.tresor.gouv.qc.ca/doc/acrobat/gouverne.pdf>

el de *XML en route au gouvernement du Québec*³⁹³, ha sido elaborado por el Grupo Departamental de Investigación en Estructuras Documentales (GDRS) de la Escuela de Biblioteconomía y Ciencias de la Información de la Universidad de Montreal, dirigido por el profesor Y. Marcoux, los cuales ya fueron citados en el capítulo segundo de esta tesis doctoral al hablar de los desarrollos llevados a cabo en Canadá mediante la utilización del metalenguaje SGML aplicado a la documentación jurídica.

Este trabajo se enmarca dentro del ámbito de investigación y trabajo del gobierno de Québec denominado *Chantier en ingénierie documentaire*³⁹⁴, proyecto llevado a cabo de forma conjunta por diferentes ministerios y organismos del gobierno de Québec (la Biblioteca Nacional, los archivos nacionales y el Grupo de responsables de gestión documental de dicho gobierno, principalmente) entre los años 1997 a 1999 con el objetivo principal de establecer una serie de directrices generales para la gestión integral del conjunto de documentos impresos y electrónicos gubernamentales. En algunos de los documentos generados a lo largo de esta importante experiencia se hacía mención a la necesidad en la utilización del formato XML para el tratamiento de la información y, consiguientemente, el desarrollo de esquemas para la definición de los tipos documentales existentes en la administración pública del gobierno de Québec. Y es aquí, por tanto, donde enlaza y continúa este nuevo proyecto.

Este informe establece diferentes campos de aplicación de la tecnología XML a la documentación del Gobierno de Québec; a saber:

1. Metadatos: Establecidos éstos por la Chantier para los tres tipos generales de documentos (de referencia, de transacción y de mensajes de correo) y que se han de modelizar e integrar dentro de los formularios normalizados.

³⁹³ Yves Marcoux, Diane Girard, Anne Hakier, Chirstian Rémillard, Isabelle Spina. *XML en route au gouvernement du Québec* [documento PDF]. Québec: Secrétariat du Conseil du trésor, 4 février 2001. Disponible en <http://www.autoroute.gouv.qc.ca/publica/xml.pdf> (consultado el 12 de septiembre de 2001).

³⁹⁴ La información al respecto de este desarrollo se encuentra disponible dentro del sitio web de la *Autoroute de l'information* del gobierno de Québec en la dirección <http://www.autoroute.gouv.qc.ca/dossiers/ingenierie.htm>

2. Esquemas de clasificación: La generación de esquemas y cuadros de clasificación basados en XML que se relacionen y acomoden a la estructura de un tesoro propio de la actividad gubernamental.
3. Cuadro de autorizaciones: De igual modo, se trata de establecer una formalización del concepto de autorización dentro del entorno XML, desarrollando la correspondiente DTD o esquema XML que de soporte a este apartado.
4. Etiquetas de la signatura numérica: Como complemento al cuadro de autorizaciones se ha de normalizar y formalizar la signatura numérica de los documentos XML.
5. Formularios electrónicos: Se trata de un aspecto complejo pues aquí se trata de establecer conceptualmente y de modelizar los formularios electrónicos dentro de un contexto de registro referencial de esquemas XML y en relación con los procesos de concepción y de desarrollo de aplicaciones informáticas.
6. Registro referencial de esquemas XML: Análisis de los diferentes modelos y propuestas de esquemas que han venido surgiendo desde la comunidad XML en cuanto a su potencial dentro del contexto de la gestión documental del gobierno de Québec.
7. Circuitos de producción documental: Las técnicas derivadas de los flujos de trabajo (*workflow*) ocupan un lugar destacado en los problemas analizados dentro del gobierno. Así, resulta necesario analizar y disponer de las herramientas de gestión automática de trabajo así como profundizar en la descripción de los procesos de tareas y de mantenimiento para la mejora en las formas y métodos de interoperabilidad.
8. Herramientas de tratamiento XML: El estudio y análisis de las herramientas XML existentes en el mercado para el tratamiento y gestión de aplicaciones documentales es otro de los pilares fundamentales de este proyecto. En este contexto, resulta pues imprescindible analizar editores de esquemas, editores de instancias de documentos, sistemas para la gestión de bases de datos, servidores XML para la Web, etc.
9. Biblioteca de esquemas: Al igual que sucedía para el caso del gobierno francés, antes analizado, el desarrollo de un completo repertorio o biblioteca de esquemas XML para los tipos documentales manejados por la administración pública en su relación

con los ciudadanos y empresas resulta ser una de las tareas primordiales de este proyecto.

De este modo, y a lo largo de las más de 150 páginas que incluye este informe se analizan y detallan en profundidad cada uno de estos puntos señalados. Pero, sin duda, la parte más relevante del mismo corresponde al planteamiento y desarrollo de esquemas XML donde, y a través del establecimiento de un protocolo para la modelización de estos esquemas para el desarrollo de formularios electrónicos, se realiza el trabajo más importante e intenso. El desarrollo de esquemas de clasificación para llevar a cabo la conversión del tesoro de la actividad gubernamental de Québec en XML ha dado como fruto la generación de una DTD basada en el modelo propuesto de la organización OASIS, institución ya comentada con anterioridad en esta tesis doctoral, para la representación de las distintas relaciones asociativas, jerárquicas y de equivalencia existentes en dicho tesoro³⁹⁵. Para el caso de los formularios electrónicos a través de la Web se pretende trasladar los tradicionales formularios impresos procedentes de los distintos ministerios y organismos públicos del gobierno de Québec³⁹⁶ dentro de un contexto de registro referencial XML en relación con los procesos de concepción y desarrollo de aplicaciones informáticas.

Así, y fruto de estas reflexiones, se pretende elaborar un protocolo riguroso de redacción de esquemas para dichos formularios electrónicos. Los formularios que ya han sido definidos (modelizados) con la ayuda de dicho protocolo corresponden a los siguientes organismos públicos: Inspección General de Instituciones financieras (formularios

³⁹⁵ Este desarrollo se encuentra ubicado en dos espacios de uno de los servidores web del gobierno de Québec. En el primero de ellos, en la dirección <http://rr.gouv.qc.ca/r/gouv-qc>, se encuentra el tesoro de la actividad gubernamental (el fichero XML se encuentra concretamente en <http://rr.gouv.qc.ca/r/gouv-qc/thesaurus-activite-gouvernementale.1.xml>) y en el segundo, en la dirección <http://rr.gouv.qc.ca/r/gouv-qc/grds>, se encuentra el esquema de clasificación correspondiente a la estructura de espacios nominativos del registro referencial de esquemas XML (el fichero XML se encuentra concretamente en http://rr.gouv.qc.ca/r/gouv-qc/grds/espaces-nominaux_scheme-clasification.1.xml).

³⁹⁶ El gobierno de Québec mantiene un más que interesante servicio público de acceso a través de la Web a los programas y servicios proporcionados por dicho gobierno. Accediendo al mismo a través de un índice alfabético es posible localizar, además de la información relativa para la tramitación de cualquier asunto en relación con la materia tratada, los formularios oficiales en formato electrónico, normalmente en formato

correspondientes para los estatutos de constitución), el Ministerio de Agricultura (los correspondientes para los minoristas en alimentación y restaurantes), el Ministerio de Medio Ambiente (certificados para el almacenamiento y distribución de agua así como para el tratamiento de aguas residuales), Gestión Urbanística de Québec (los correspondientes para la licencia de empresas de construcción) y el Ministerio de Pensiones y Seguridad Social (los correspondientes para la solicitud de pensión)³⁹⁷. Se ha establecido que para la presentación en la Web de estos formularios electrónicos ha de hacerse uso del estándar asociado XForms, ya analizado con anterioridad en su apartado correspondiente, capaces de reflejar los modelos de contenido definidos en los esquemas XML y de reducir la dependencia de la utilización de *scripts* en los actuales formularios HTML.

Es de lamentar la aún inexistencia de desarrollos XML de aplicación a la administración de justicia en el estado autónomo de Québec aunque consideramos que dados los magníficos criterios de trabajo que se han empezado a emprender por parte de este gobierno no ha de tardar mucho el momento en el que aparezcan propuestas al respecto de este tema.

Para finalizar este apartado del capítulo, terminaremos hablando de la importante labor emprendida en **Australia** en el campo del tratamiento de la documentación jurídica mediante la aplicación de las tecnologías XML. Para ello será necesario analizar varias instituciones y organismos de este país así como las investigaciones y desarrollos prácticos que están llevando a cabo; éstas son: la *Law & Justice Foundation of New South Wales*, el

PDF para su impresión por parte del ciudadano o empresa interesada. El URL de este sitio web del gobierno de Québec es <http://www.formulaire.gouv.qc.ca/>

³⁹⁷ La lista completa de esquemas XML desarrollados a este respecto, ciertamente extensa, así como las direcciones donde pueden ser localizados todos estos ficheros XSD, se encuentra ubicada en el Anexo 2 del informe elaborado por el GRDS de la Escuela de Biblioteconomía y de Ciencias de la Información de la Universidad de Montreal. De igual forma, en el Anexo 4 de este documento, se incluyen una serie de imágenes correspondientes a diversos pantallazos de la visualización de uno de estos formularios XML (el correspondiente a la petición de licencia de empresa de construcción, el cual se encuentra en la actualidad en formato PDF para su descarga e impresión por parte de los ciudadanos en la dirección http://www.rbq.gouv.qc.ca/qualifications/documents/guide/Form_del.pdf), tanto para la entrada de datos como del formato de impresión del mismo.

Australian Institute of Judicial Administration, la *Federal Court of Australia* y el grupo de trabajo *Legal XML Australia*, adscrito a la organización internacional Legal XML.

La **Law & Justice Foundation of New South Wales**³⁹⁸ es una organización sin ánimo de lucro financiada y soportada por el gobierno autónomo de Nueva Gales del Sur con el fin de investigar en el campo de la aplicación de las nuevas tecnologías de la información y las comunicaciones al tratamiento y difusión de la documentación de carácter jurídico producido en esta comunidad. Ya en 1997 se creó desde esta fundación un ambicioso proyecto para el desarrollo de estándares capaces de dar una respuesta satisfactoria al tratamiento de la información legislativa y jurisprudencial producida por los distintos organismos públicos con competencias en esas materias. Esta importante iniciativa, conocida bajo el nombre de *Legal Information Standards Council (LISC)*³⁹⁹, ha venido trabajando durante estos años (su actividad ha cesado en el año 2001 para crear un grupo de trabajo de ámbito nacional, como veremos posteriormente) de forma más específica en campos como el diseño y desarrollo de sitios web de contenido jurídico, el establecimiento de normas para conseguir un modelo neutral para la redacción de las citas jurídicas en los distintos organismos y tribunales de este estado y de la nación (principalmente, el Tribunal Supremo de Australia, el Tribunal Supremo de Nueva Gales del Sur, el Instituto Australiano de Administración Judicial y el AustLII), el establecimiento de estándares para el tratamiento de los textos legislativos y otras cuestiones relativas a la accesibilidad y usabilidad por parte de los ciudadanos de los textos jurídicos electrónicos en la Web⁴⁰⁰.

³⁹⁸ Hasta mediados de 2001 esta Fundación llevaba el nombre de Law Foundation of NSW pues su labor se centraba principalmente en el tratamiento electrónico de los textos legislativos y otras temas relacionados con esta materia. Con la importancia adquirida igualmente en todo el mundo el tratamiento electrónico de los documentos jurisprudenciales su nombre cambió al que actualmente lleva. En cualquier caso, toda la información oficial relativa a esta importante fundación del gobierno de Nueva Gales del Sur se encuentra disponible en la dirección <http://www.lawfoundation.net.au/>

³⁹⁹ Toda la información oficial relativa a esta iniciativa se encuentra disponible en la dirección <http://www.lawfoundation.net.au/lisc/>

⁴⁰⁰ Citar como ejemplo representativo de la labor desarrollada por este grupo de trabajo durante los años 1997-2001 las directrices marcadas para el desarrollo de sitios web jurídicos. Estas directrices trataban aspectos tales como la identificación clara de la persona u organismo responsable de la información suministrada, la revisión técnica de los contenidos expuestos por parte de un jurista con experiencia en la materia, la sencillez y claridad en temas como el diseño, la presentación de los contenidos o el ámbito

Otro de los proyectos más importantes que se están llevando a cabo a iniciativa de esta Fundación es el denominado **Online Legal Access Project (OLAP)**⁴⁰¹. Se trata de un proyecto ambicioso y de gran envergadura donde se están desarrollando un gran número de líneas de trabajo y actuación en campos tales como el establecimiento de directrices para el desarrollo de estándares en el tratamiento electrónico de los textos jurídicos, la potenciación de las investigaciones en este campo y la promoción y difusión de las mismas, el desarrollo de sistemas de información jurídica en línea o la promoción y potenciación de estos espacios electrónicos para incrementar su utilización entre los ciudadanos.

Así, y en relación con la materia de interés para esta tesis doctoral, algunas de las directrices e informes elaborados a iniciativa de este proyecto han tenido una repercusión internacional. Por ejemplo, destacar el informe elaborado allá por 1998 por S. Davey y S. Scott⁴⁰² y presentado a la primera Conferencia de Tecnología aplicada a la Justicia (de la cual hablaremos con posterioridad en este apartado), donde se destaca la importancia del desarrollo de estándares para el tratamiento electrónicos de texto jurídicos que han de ser servidos a través de Internet. Estos autores apuntaron una serie de áreas en las que era imprescindible desarrollar estos estándares: para los datos propiamente dichos en el entorno de producción de la legislación y de la jurisprudencia, destacando a los metalenguajes SGML y XML como las tecnologías más adecuadas para llevar a cabo el desarrollo de definiciones de tipos documentales así como el empleo de los metadatos para la descripción de las fuentes secundarias de contenido jurídico; para la normalización en la redacción de las citas a leyes y sentencias en los textos electrónicos de contenido jurídico; para la promoción en la accesibilidad de dichos textos a través de la WWW por parte de

jurisdiccional, el establecimiento de relaciones hipertextuales adecuadas entre las páginas que conforman el sitio web así como los recursos externos relacionados, la forma correcta de citar la legislación y la jurisprudencia reseñada, cuestiones relativas al uso y derechos de propiedad de los materiales expuestos, etc. Para una mayor información, véase *Best practice guidelines for Australian legal websites* [documento HTML]. Sydney: Legal Information Standards Council, Law Foundation of NWS, rev. October 16, 2000. Disponible en <http://www.lawfoundation.net.au/lisc/recommend/bpguide.html> (consultado el 14 de diciembre de 2000).

⁴⁰¹ Toda la información oficial relativa a este proyecto se encuentra disponible en la dirección <http://www.lawfoundation.net.au/olap/>

⁴⁰² Sandra Davey. "Managing the Magic: Standards for Australian Electronic Legal Information" [documento HTML]. En: *Technology for Justice Conference* (1º. 1998. Melbourne). Disponible en <http://www.aija.org.au/conference98/papers/sdavey/standards.html> (consultado el 12 de junio de 2000).

todos los ciudadanos; y, finalmente, para el desarrollo de sitios web de contenido netamente jurídico. Igualmente resulta de gran interés el informe elaborado por esta fundación dedicado al estudio y establecimiento de un esquema de metadatos jurídicos basados principalmente en el modelo del Dublin Core para la descripción de recursos electrónicos de carácter jurídico que permitan una recuperación eficaz de los mismos en Internet por parte de los ciudadanos de este estado⁴⁰³.

Fruto de esta investigación realizada se lanzó en diciembre de 2000 un informe para el establecimiento de estándares de metadatos para el sector judicial de este estado⁴⁰⁴. Esta propuesta de estándar pretendía ser además un manual o guía de utilización para todas aquellas instituciones que desearan hacer uso del modelo de metadatos del Dublin Core para la descripción de recursos electrónicos (normalmente, y en principio, documentos HTML) de contenido judicial que han de ser publicados en Australia. De este modo, el informe establece un conjunto de campos de metadatos, considerados como imprescindibles, para su inclusión o relación con los materiales electrónicos jurídicos. Así, el esquema de metadatos jurídicos propuesto contemplan el uso de algunos de los campos del modelo de esquema del Dublin Core (*DC.Creator*, *DC.Publisher*, *DC.Rights*, *DC.Subject* - en el Apéndice A se establece el listado de materias válidas para este campo-, *DC.Description*, *DC.Language*, *DC.Coverage*, *DC.Coverage.Jurisdiction*, *DC.Coverage.Region*, *DC.Date.Created*, *DC.Date.Modified*, *DC.Type* y *DC.Format*), junto con otros directamente relacionados con este modelo (*Keywords* basado en el *DC.Subject*, y *Description* basado en el *DC.Description*) y otros campos propios definidos por diversos organismos de la administración de justicia (*Justice Sector Metadata Standards* -JSMS- y el *Australian Government Locator Service* -AGLS-) del gobierno australiano (*JSMS.Category* para la naturaleza del recurso referenciado cuyos posibles valores son expuestos en el Apéndice D de este

⁴⁰³ *LawZone: A new way of searching. Improving community access to legal information on the Internet* [documento HTML]. Sydney: NSW, Attorney General's Department, Law Foundation of NSW, December 1999. Disponible en <http://www.lawlink.nsw.gov.au/agd.nsf/pages/lawzone/> (consultado el 19 de octubre de 2000).

⁴⁰⁴ *Justice Sector Metadata Standards: a users guide* [documento HTML]. Sydney: Law Foundation of NSW, December 20, 2000. Disponible en <http://www.lawfoundation.net.au/olap/guidelines/metaintro.html> (consultado el 25 de abril de 2001).

informe, y el *AGLS.availability* para la información relativa a la disponibilidad del documento web referenciado cuyos posibles valores son expuestos en el Apéndice B -códigos normalizados de países por la RFC 1766 de la IETF- y en el Apéndice C -códigos MIME para los tipos de medios referenciados en este campo-) y, finalmente, los propiamente administrativos (*Admin.Creator* para el nombre de la persona y/u organización que ha realizado esta entrada de metadatos, *Admin.Date.Created* para la entrada de la fecha de creación de los metadatos, y *Admin.Date.ValidTo* para establecer la fecha hasta la que es válida la información del recurso).

Este informe se complementa con el desarrollo de una aplicación informática en línea para la generación automática de metadatos en páginas web de contenido jurídico, donde, y a través de la redacción de los campos existentes en el formulario propuesto, el sistema genera automáticamente las correspondientes etiquetas <META> ya cumplimentadas y que han de insertarse en la cabecera de los documentos HTML⁴⁰⁵.

El **Australian Institute of Judicial Administration**⁴⁰⁶ es un instituto de investigación y formación adscrito a la Universidad de Melbourne en el cual se incluyen más de mil miembros entre jueces y magistrados, oficiales de justicia, administrativos, profesores, bibliotecarios y documentalistas y otras personas relacionadas con el mundo de la judicatura, tiene como objetivos principales la investigación en el campo de la administración de justicia y el desarrollo e impartición de programas de formación. En cuanto a la investigación llevada a cabo por este instituto a lo largo de estos últimos años se han venido desarrollando diversos proyectos de investigación en diversos campos, como, por ejemplo, en el establecimiento de líneas de funcionamiento de los tribunales de justicia, la aplicación de las nuevas tecnologías de la información y las comunicaciones a estos espacios, la elaboración de tesauros jurídicos o la promoción de estándares para el tratamiento de los documentos electrónicos de contenido jurídico. En esta labor de

⁴⁰⁵ El acceso a esta interesante aplicación, la cual lleva el sugerente título de *Justice Sector Metadata HTML Generator*, se realiza a través de la dirección <http://www.lawfoundation.net.au/olap/guidelines/metatagform.html>

⁴⁰⁶ Toda la información oficial relativa a esta institución se encuentra disponible en la dirección <http://www.aija.org.au/>

divulgación y promoción de las nuevas tecnologías de la información y las comunicaciones en el campo jurídico resulta destacable la labor llevada a cabo a través de los congresos y conferencias que esta institución ha venido organizando, en especial las denominadas *Technology for Justice*. Hasta la fecha se han realizado dos de estos importantes encuentros, en 1998 y en 2000, y donde algunas de las comunicaciones presentadas han girado en torno a la aplicación de las tecnologías XML para el tratamiento de los datos jurídicos así como los esfuerzos de normalización que se están llevando a cabo por parte de muchas instituciones públicas y organizaciones internacionales del campo de la justicia⁴⁰⁷.

El **Federal Court of Australia**⁴⁰⁸ es, sin duda, una de los escasos tribunales de justicia no ubicados en los Estados Unidos que mayor uso hacen de las nuevas tecnologías de la información y las comunicaciones, siendo incluso posible el intercambio de documentos electrónicos entre las partes. Este importante proyecto de actuación del Tribunal Federal de Australia lleva por nombre *eCourt*⁴⁰⁹ y ya es toda una realidad en la actualidad, teniendo como núcleo tecnológico de la misma el uso de las tecnologías derivadas de Internet y del metalenguaje XML. La iniciativa emprendida cubre un gran número de campos de actuación y desarrollo, como son, por ejemplo, el *eCourt On-Line Forum*, espacio judicial virtual de este tribunal mediante el cual se asiste y gestionan los problemas prejudiciales de las partes implicadas en un proceso judicial y se orienta sobre el modo de proceder para el envío de la documentación necesaria haciendo para ello uso de la red Internet, el *Electronic Filing System*, interfaz electrónica de acceso público para el alojamiento de aplicaciones que han de dar soporte a los documentos electrónicos así como para toda la gestión de pagos y

⁴⁰⁷ Destacar, por ejemplo, la exposición realizada en la sesión del año 2000 por el ya mencionado con anterioridad en esta tesis doctoral Winchel "Todd" Vincent, donde se realizaba un excelente análisis sobre las actuales barreras para el desarrollo de transacciones electrónicas de información entre los tribunales y las partes de un proceso y cómo el desarrollo y promoción de las tecnologías XML aplicadas a este campo llevadas a cabo por la organización Legal XML constituían una solución real a este problema. Véase, Winchel "Todd" Vincent. "Introduction Legal XML" [documento PPT]. En: *Technology for Justice* (2º. Melbourne. October 2000). Disponible en <http://www.aija.org.au/tech2/presentations/vincent/AIJA.2000.10.10.ppt> (consultado el 5 de mayo de 2001).

⁴⁰⁸ <http://www.fedcourt.gov.au/>

tasas a través de tarjeta de crédito, el *Case Management System*, para la gestión automatizada de los expedientes, y documentos integrantes, de los casos judiciales existentes, el *Records and Document Management System*, elemento integrador de los dos sistemas anteriores y otros módulos del sistema general implementado, o, finalmente, el *Information to the Public*, donde el desarrollo de un sitio web de acceso público a las fuentes primarias de recursos jurídicos juega un papel fundamental dentro de esta iniciativa.

Como decíamos anteriormente, este tribunal acepta que la comunicación y entrega de documentos para el desarrollo de un proceso judicial sea completamente “en-línea”. Así, como se informa en la página de acceso al *Federal Court’s “on-line” Electronic Filing*⁴¹⁰, los documentos electrónicos que pueden enviarse a este tribunal pueden, bien ser elaborados directamente ajustándose al modelo impreso oficial o bien, descargarse los oportunos formularios electrónicos (en formato PDF y RTF) existentes para ello en el área correspondiente del sitio web de este tribunal⁴¹¹. Hasta el momento, estos formularios no pueden ser cumplimentados y enviados en línea sino que han de ser descargados, cumplimentados en el ordenador del usuario y enviados posteriormente al tribunal. Finalmente, si alguno de esos documentos deben ser firmados se debe adjuntar la imagen digitalizada del documento con la firma incorporada (imagen digital en formato TIF, GIF o JPG) pues aún este tribunal no ha desarrollado los mecanismos oportunos que garanticen una completa integridad en la firma electrónica de estos documentos.

Finalizar este apartado dedicado a los proyectos y desarrollos que diversas instituciones jurídicas de Australia están llevando a cabo con la labor emprendida hace ya algunos años

⁴⁰⁹ Toda la información oficial relativa a este proyecto de actuación judicial mediante el uso de las tecnologías de la información y las comunicaciones se encuentra disponible en la dirección <http://www.fedcourt.gov.au/ecourt/ecourt.html>

⁴¹⁰ Documento informativo disponible en la dirección <http://www.efiling.fedcourt.gov.au/fedcourt/WelcomeToFedCourtEFS.htm>

⁴¹¹ La información oficial relativa a los formularios oficiales establecidos por este tribunal se encuentran en la página web titulada *FILINGS, FORMS and FEES* en la dirección <http://www.fedcourt.gov.au/fff/fff.html>. De forma concreta el listado alfabético de formularios en formato PDF o RTF para su descarga por parte del interesado se encuentra disponible en la dirección http://www.fedcourt.gov.au/fff/fff_formsallA.html

desde el grupo de trabajo **Legal XML Australia**⁴¹², adscrito a la organización internacional Legal XML, ya analizada con anterioridad en este capítulo de la tesis doctoral. Este grupo de trabajo, coordinado por Allison Stanfield del *eLaw Australia Pty Ltd.*⁴¹³, y por Sue Scott de la *Law & Justice Foundation of NSW*, es en la actualidad el aglutinador y promotor de todos los desarrollos que se están llevando a cabo en Australia en el terreno de la información jurídica y su tratamiento a través de las tecnologías XML. Para llevar a cabo esta promoción y difusión dentro del país, además de la propia página web del grupo de trabajo, se realizan encuentros presenciales en diferentes ciudades del país entre los investigadores y profesionales interesados en este campo de actuación. Además de las actuaciones australianas ya reflejadas aquí, destacar solamente la propuesta de DTD (nota no oficial UN_10013_2000_06_27 de la organización Legal XML)⁴¹⁴ para el tratamiento de procesos judiciales en este país elaborada desde este grupo de trabajo por James Osborn. Esta DTD se encuentra disponible en el Apéndice 6 de esta tesis doctoral (la DTD original ha sido corregida en ciertos errores sintácticos y de definición de modelos detectados para su correcto procesamiento⁴¹⁵).

Se trata de una DTD bastante sencilla que, como señala su propio autor, trata de maximizar su empleo entre la comunidad judicial australianas al no descender excesivamente en la especificidad de los elementos informativos. Así, desde el elemento inicial *AU_LegalCase*, éste contempla en su modelo de contenido sólo tres bloques principales o elementos hijos: uno obligatorio y no repetible, *HeadNote*, otro obligatorio y repetible tantas

⁴¹² Toda la información oficial relativa al grupo de trabajo australiano integrante de la organización Legal XML se encuentra disponible en la dirección <http://www.legalxml.org/Australia/activities.htm>

⁴¹³ <http://www.elaw.net.au/>

⁴¹⁴ James Osborn. *A Proposed Document Type Definition for Judgment* [documento HTML]. Legal XML, June 27, 2000. Disponible en http://www.legalxml.org/Australia/uncopies/UN_10013_2000_06_27.htm (consultado el 14 de noviembre de 2000).

⁴¹⁵ La mayoría de las correcciones que se han realizado hacen referencia a la no utilización de los signos correspondientes para introducir comentarios dentro de la DTD, la inclusión de comentarios dentro de los modelos de contenido de algunos elementos (sintácticamente correcto en la DTD de SGML pero no en la de XML), incorrección sintáctica al definir el modelo de contenido del elemento *Lawyer* y la repetición en la declaración del elemento *Law*, principalmente. En cualquier caso, invitamos al lector a realizar la oportuna

veces como sea necesario, *Judgment*, y otro opcional, *EndNote*. Cada uno de estos elementos se divide a su vez en otros elementos hijos, algunos de los cuales contempla igualmente una mayor profundidad en la información detallada con sus correspondientes hijos, los cuales no explicamos con mayor profundidad para no alargar en exceso este apartado del capítulo.

Así, para detallar la información de las notas de cabecera, este elemento describe la información relativa a la sección de encabezamiento (*HeadingSection*), el tribunal (*Court*), el número de jueces intervinientes (*NumJudges*), los datos relativos a estos jueces (*Judge*), los asistentes letrados (*Appearances*), información relativa a la apelación (*AppealInfo*), las leyes citadas (*LawCited*), la asignación de palabras clave para la descripción de la materia tratada (*Keywords*), el resumen de la sentencia judicial (*SummaryOrder*) y otras información adicionales (*Extra*). En cuanto al elemento del juicio en sí, este elemento refleja los hechos o circunstancias acaecidos (*Facts*), las leyes empleadas (*Law*) y el texto completo de la sentencia (*FullOrder*). El tercer y último elemento hijo del elemento raíz, dedicado a las notas finales, contempla la información del resumen de la sentencia (*SummaryOrder*) y la relativa a la certificación electrónica de todo lo reflejado (*Certification*).

En cualquier caso, y para una mayor profundidad sobre esta DTD, remitimos al Apéndice antes mencionado. Su representación gráfica sería la siguiente:

CONCLUSIONES A LA PARTE

A lo largo de esta tercera Parte de la presente tesis doctoral hemos ido analizando diversos aspectos relacionados con las actuales tecnologías de la información y las comunicaciones, en especial todo lo referente a las tecnologías derivadas del servicio más popular de la red Internet como es la *World Wide Web*, y los lenguajes para el tratamiento de los documentos electrónicos que han de ser servidos en este espacio virtual. De forma concreta y con la exhaustividad necesaria se ha venido igualmente exponiendo la importancia y alcance del metalenguaje XML para la construcción de lenguajes específicos de marcado descriptivo de textos electrónicos de cualquier temática y condición, así como los mecanismos existentes en este metalenguaje para la generación de dichos lenguajes o aplicaciones. De forma más concreta aún, se han expuesto las principales iniciativas y proyectos en todo el mundo, ya concluidos o en fase de realización, para el tratamiento de información y documentación electrónica de carácter jurídico empleados en multitud de tribunales de justicia mediante el uso de la tecnología XML. A modo de resumen, como conclusiones a esta Parte se pueden establecer los siguientes puntos de interés:

- La red Internet constituye hoy en día un espacio muy importante y valioso de comunicación e intercambio de ideas entre multitud de personas y organizaciones de todo tipo repartidas por todo el planeta. Su popularidad y difusión ha venido en estas últimas décadas de la mano del desarrollo y consolidación de su servicio más característico: la *World Wide Web*. Propuesto en su origen como un sistema idóneo basado en las ideas de los hipertextos para la gestión de la información electrónica en un entorno corporativo, su aplicación y extensión al espacio virtual de la Web no ha estado exenta de problemas, tanto de índole social como de carácter técnico.
- Desde el punto de vista técnico, la Web ha venido empleando un lenguaje de marcado para los documentos electrónicos que principalmente son publicados en este espacio, el HTML. Este sencillo lenguaje de marcas derivado del metalenguaje SGML ha venido evolucionando desde su invención a finales de la década de los 80 de forma paralela a la

Web, siendo, por tanto, sus virtudes y defectos similares. HTML se configuraba como un lenguaje de propósito general para el marcado de documentos electrónicos con estructuras muy sencillas lo que facilitaba un rápido aprendizaje y utilización por parte de los usuarios y una relativa simplicidad en la construcción de las herramientas informáticas necesarias para la edición y visualización de dichos documentos. Estos factores, unidos a la necesidad humana de expresarse y difundir ideas, hicieron que este espacio virtual y su lenguaje de construcción de documentos hipertextuales se fuera expandiendo y desarrollando hasta el fenómeno de masas que constituye hoy en día este servicio de Internet. Pero al igual que la sencillez del HTML facilitó en un principio su aceptación y ascenso por parte de la mayoría de la comunidad de usuarios de la Web, en el espacio electrónico que se pretende construir para el siglo XXI, una *Web semántica* basada en los contenidos, esa sencillez no resulta idónea.

- La construcción de una *Web semántica*, de aplicación tanto en el entorno mundial de producción documental que constituye Internet como en el entorno más reducido de las redes informáticas corporativas, exige la construcción de un auténtico sistema global de información capaz de ser interpretado de forma adecuada tanto por los humanos como por las aplicaciones que han de procesar la información circulante. Las grandes limitaciones del modelo ofrecido por el lenguaje HTML para llevar a cabo este cometido (principalmente las derivadas de su escasa capacidad para definir estructuras documentales complejas capaces de marcar con precisión la semántica de los documentos electrónicos susceptibles de ser recuperados eficazmente por los motores de búsqueda en Internet) hicieron necesaria la elaboración de un nuevo sistema de marcado de los documentos que han de poblar este nuevo espacio: el metalenguaje XML.
- XML se configura como un SGML adaptado a la Web, tomando todas las virtudes aportadas y ya demostradas tiempo atrás por este metalenguaje pero desechando todas aquellas partes complejas o ambiguas técnicamente y, por tanto, de difícil implementación en las herramientas informáticas que han de procesarlo. Así, como

metalenguaje que es, permite al usuario construir sus propios lenguajes o vocabularios capaces de describir con precisión la estructura y la semántica de los contenidos de los documentos electrónicos. Además, la ambición de ideas con la que fue creado XML permite que muchos lenguajes específicos contruidos al amparo del mismo (estándares asociados) amplíen las prestaciones y capacidades de los documentos electrónicos en facetas tales como el establecimiento de relaciones hipertextuales, la presentación de los contenidos en diferentes medios y dispositivos de salida, la búsqueda y recuperación documental o la asignación de mecanismos normalizados para la descripción semántica de los contenidos informativos existentes en los documentos electrónicos.

- Como toda innovación tecnológica, XML va creciendo y evolucionando a través de sus errores y aciertos, haciéndose cada vez más complejo. Tal vez esta complejidad surgida en aras de una mayor potencialidad sea uno de los motivos de su relativamente lenta adopción en el entorno de la web, sobre todo en su aplicación a documentos de contenido predominantemente textual: existe un consenso casi unánime entre la comunidad científica acerca de las bondades del XML para el tratamiento de documentos electrónicos pero su aplicación por los usuarios de la Web resulta escasa si lo comparamos con lo sucedido con el HTML. Son las grandes corporaciones empresariales las que están haciendo un mayor uso de las tecnologías XML centrándose principalmente en aquellas aplicaciones orientadas a la transmisión e intercambio de datos (y en especial para el denominado *comercio electrónico*) pues es en este campo donde las compañías informáticas con intereses en la WWW han visto mayores posibilidades de negocio y expansión.
- Sin embargo, el peso que esta tecnología está tomando día a día en el terreno del tratamiento y difusión de documentos electrónicos hace que la denominada *orientación al texto* de XML (y no sólo su aplicación a los datos) esté tomando un fuerte impulso que se aprecia en la construcción de bases de datos alternativas a los modelos relacionales, donde el documento es analizado y descrito mediante el lenguaje de marcas utilizado. Con ello es posible evitar los complejos procesos de traslación de la estructura

contenida en estos documentos, normalmente ambigua en mayor o menor grado, al modelo de relaciones entre tablas y campos de los anteriores sistemas. La indexación automática del texto contenido en estos documentos capacitará al motor de búsqueda para localizar luego la información precisa. Este modelo tan cercano a los profesionales de la Documentación automatizada es conocido en este entorno tecnológico por el nombre de *native XML Databases*, siendo el modelo utilizado para la realización del desarrollo práctico incluido en la cuarta Parte de esta tesis doctoral.

- Dentro del mundo de la tecnología XML tiene un papel fundamental la definición formal de las estructuras documentales para la generación de vocabularios XML específicos. Desde el modelo primitivo de la DTD de SGML, adaptada a los documentos XML, se ha producido en este último periodo un gran salto cualitativo con la incorporación del modelo de esquema XML del Consorcio Web. Con el modelo del XML Schema se rompen las limitaciones y carencias existentes en el viejo modelo de la DTD, siendo posible ahora definir complejas estructuras documentales así como ajustar con precisión los tipos de datos que pueden estar contenidos dentro de los elementos marcados. En cuanto a la descripción semántica de los contenidos del documento, los esquemas XML ya ofrecen por sí mismos una garantía ciertamente notable en un entorno reducido de producción documental pero para su exposición en el espacio constituido por la WWW resulta necesario dotar a dicha semántica de una comprensión y aceptación universal. Se trata pues aquí de la definición y establecimiento de los llamados *metadatos* del documento que en la tecnología XML pueden lograrse haciendo uso de alguno de los vocabularios de aceptación general existentes a tal fin (el modelo propuesto por el Dublin Core resulta ser hoy en día el más aceptado) o definidos por el usuario u organización a través del modelo propuesto por el estándar RDF, y de forma más concreta por su modelo de esquema.
- Finalmente, como adelantábamos, la aplicación de la tecnología XML se extiende en la actualidad al tratamiento de los documentos electrónicos producidos por todo tipo de instituciones. No es una excepción a este aserto, sino todo lo contrario, el uso y

aplicación que del lenguaje XML se está haciendo en la esfera jurídica. Son innumerables los juristas, documentalistas y otros profesionales de prestigio adscritos a instituciones jurídicas de todo el mundo que han venido mostrando su interés y apoyo a la aplicación de esta nueva tecnología al tratamiento de los documentos de carácter jurídico a través de numerosos artículos científicos, conferencias, debates sobre la materia y, de forma más pragmática, con el desarrollo de un gran número de proyectos de investigación aplicada. Todos los proyectos e iniciativas reflejados en el último capítulo de esta tercera parte de la tesis vienen a demostrar que con un mayor o menor grado de éxito son muchos los países de nuestro entorno en los que sus administraciones de justicia están apostando de forma decidida por el uso de la tecnología XML, bien sea para llevar a cabo de forma satisfactoria los procesos de intercambio de datos y documentos entre los tribunales y las partes implicadas en un proceso judicial, o bien para el tratamiento global de toda la información y documentación generada en este proceso y la correspondiente publicación de la jurisprudencia emanada de estos tribunales.

- La necesidad de emplear unos procedimientos y lenguajes de marcado comunes para estos documentos judiciales ha sentado las bases para el establecimiento de proyectos de colaboración tanto a escala nacional, entre instituciones afines dentro de un mismo territorio, como en el ámbito internacional, siendo el máximo exponente de este afán por el desarrollo de estándares de aplicación internacional la organización Legal XML. Si bien el camino que queda aún por recorrer es todavía amplio, los avances y logros que se están produciendo en el seno de esta organización hacen augurar un futuro más que prometedor en el desarrollo de estándares internacionales para el marcado de documentos judiciales, como se está poniendo de manifiesto a través de los distintos textos normativos ya elaborados y difundidos.

En la siguiente parte de esta tesis doctoral se presentará una propuesta para el marcado XML de los documentos jurisprudenciales emanados del Tribunal Constitucional español así como una pequeña maqueta de desarrollo práctico implementado. No se pretende con

ello dar una solución definitiva al tratamiento de este tipo de documentos, ni tan siquiera tratar de hacer norma la estructuración de los contenidos y el vocabulario de marcado propuestos. Se pretende, más bien, demostrar de una forma sencilla y palpable las posibilidades que esta tecnología ofrece en nuestros días para la construcción de bases de datos de información jurídica, tratando así de despertar la curiosidad e interés hacia el uso de las misma por parte del colectivo profesional y académico del mundo jurídico español así como de otros investigadores interesados en la materia.

