

## El uso de metadatos en la administración electrónica española: los retos de la interoperabilidad

Tony Hernández-Pérez\*, David Rodríguez-Mateos\*, Bonifacio Martín-Galán\*,  
María Antonia García-Moreno\*\*

**Resumen:** Se analiza el concepto de interoperabilidad, las distintas iniciativas sobre interoperabilidad adoptadas por la Unión Europea y se hace un estudio sobre el uso de metadatos en la administración electrónica española en los distintos ministerios. Se analizaron más de 42.000 páginas web distintas, más de 260.000 etiquetas meta de 18 ministerios distintos y el uso que se hace de cada una de estas etiquetas.

Se observa que en la administración central española no existe ninguna política sobre descripción de recursos electrónicos ni sobre interoperabilidad. No se hace ni un buen uso de metadatos ni de vocabularios controlados e incluso los mismos valores son expresados de muy distinta forma por los ministerios. Solo el Ministerio de Cultura hace un uso parcial de etiquetas Dublin Core. Un número muy elevado de las páginas presentan errores básicos en el uso de etiquetas meta.

**Palabras clave:** metadatos, interoperabilidad, administración electrónica, Dublin Core, España.

### *Use of metadata in Spanish electronic e-government: the challenges of interoperability*

**Abstract:** *This work analyses the concept of interoperability and the various interoperability initiatives adopted by the EU, and examines the use of metadata in e-government in different Spanish ministries. Over 42,000 web pages involving more than 260,000 meta-tags produced by 18 ministries of the Spanish government were examined, and the use made of each of these meta-tags analysed.*

*The central Spanish government was found to have no policy regarding the description of electronic resources or interoperability. Poor use was made of metadata and established vocabularies. Even the same subjects were expressed in different ways. Only the Ministry of Culture made partial use of Dublin Core tags. A very high number of the web pages examined showed basic mistakes in the use of meta-tags.*

**Keywords:** *metadata, Interoperability, e-government, Dublin Core, Spain.*

---

\* Universidad Carlos III de Madrid.

\*\* Universidad Complutense de Madrid. Correo-e: [antonio.hernandez@uc3m.es](mailto:antonio.hernandez@uc3m.es); [david.rodriguez@uc3m.es](mailto:david.rodriguez@uc3m.es); [bonifacio.martin@uc3m.es](mailto:bonifacio.martin@uc3m.es); [mariaant@ccinf.ucm.es](mailto:mariaant@ccinf.ucm.es)

Recibido: 21-4-09; 2.ª versión: 11-6-09.

## 1. Introducción

La Decisión 2004/387/CE (Union Europea, 2004) define el término interoperabilidad como la capacidad de los sistemas de tecnologías de la información y las comunicaciones, y de los procesos empresariales a los que apoyan, de intercambiar datos y posibilitar la puesta en común de información y conocimientos. Entre las Medidas Horizontales en apoyo de los Proyectos de Interés Común, la decisión cita expresamente el «marco de metadatos para aplicaciones paneuropeas» y la «especificación de vocabularios XML, esquemas y entregables».

El Marco Europeo de Interoperabilidad V.1 distingue tres dimensiones de interoperabilidad:

1. **Organizativa**, que contempla la modelización de los procesos y la colaboración entre las administraciones. Este aspecto de la interoperabilidad tiene que ver con la definición de los objetivos de negocio, la modelización de los procesos de negocio y con lograr la colaboración de las administraciones que deseen intercambiar información y que pueden tener diferentes estructuras y procesos internos.
2. **Técnica**, que cubre los asuntos técnicos de enlazar ordenadores y servicios. Incluye aspectos claves como interfaces abiertos, servicios de interconexión, integración de datos y middleware, intercambio y presentación de datos y servicios de accesibilidad y seguridad.
3. **Semántica**, que contempla no sólo que los recursos de información puedan estar conectados, sino que también la información pueda ser interpretable de forma automática y consecuentemente reutilizable por aplicaciones informáticas que no intervinieron en su creación. La interoperabilidad semántica permite a los sistemas combinar la información recibida con otros recursos de información y procesarlos de manera que tengan significado. La interoperabilidad semántica es, por eso, un prerequisite para el interfaz final multilingüe de prestación de servicios al usuario.

Algunos autores (Acar et al., 2009, Kubicek, 2009) añaden un cuarto nivel, el de interoperabilidad sintáctica, relacionada con el procesamiento de los datos que se reciben. Peristeras (2008) hace una revisión bastante completa sobre el estado de la cuestión en el área de mejora de la interoperabilidad mediante la utilización de los modelos comunes de datos y de las ontologías.

La interoperabilidad de los servicios de administración electrónica, basados en normas y en especificaciones e interfaces abiertas, se ha convertido en una tarea transversal de vital importancia para la UE en el nuevo marco estratégico i2010. De hecho, el objetivo de la Comunicación de la Comisión al Consejo y al Parlamento Europeo sobre Interoperabilidad (Comisión Europea, 2006) es concienciar a los Estados miembros sobre los trabajos que han de emprenderse para alcanzar la interoperabilidad a nivel paneuropeo de los servicios de administración electrónica.

Este artículo se centrará principalmente en aspectos relacionados con la interoperabilidad técnica y semántica. Y es que el cambio de presentar simplemente información en un sitio web a que los programas de ordenadores puedan intercambiarla, combinarla con otros recursos de información y, consecuentemente, procesarla de forma que tenga significado, requiere acuerdos sobre una amplia variedad de temas que relacionen el contexto en el que la información es creada y utilizada.

### **1.1. El programa IDABC**

La Decisión 2004/387/EC es la base legal del Programa IDABC, para el período 2005-2009. El programa IDABC (Interoperable Delivery of Pan European e-government Services to Public Administrations, Business and Citizens) (<http://ec.europa.eu/idabc>) de la Comisión Europea aborda los retos de la interoperabilidad para la prestación de servicios paneuropeos de administración electrónica al sector público, las empresas y los ciudadanos. La decisión IDABC (2004/387/EC) se basó en el plan de acción e-Europe y se ajusta a lo establecido en la iniciativa i2010. Esta decisión establece el marco de los principios y normas sobre estándares abiertos e interfaces para la aplicación de la interoperabilidad entre sistemas, aplicaciones, procesos de negocio y los agentes que generan o utilizan los servicios de administración electrónica.

Los programas predecesores de IDABC, IDA (Interchange of Data between Administrations)(1995-1998)<sup>1</sup> e IDA II (1999-2004)<sup>2</sup> ya sentaron las bases para una orientación estratégica sobre interoperabilidad semántica. IDABC tiene como objetivos específicos la identificación, promoción y desarrollo de servicios paneuropeos de administración electrónica dirigidos a ciudadanos, empresas y administraciones (Proyectos de Interés Común), así como de las infraestructuras y servicios necesarios (Medidas Horizontales) para el despliegue de los primeros.

El papel de la IDABC, por tanto, no se limita a la producción de directrices (Amutio Gómez M. A. 2006), sino que también implica la planificación e implementación de las infraestructuras de apoyo a la interoperabilidad. Entre otras tareas, la IDABC debe promover la reutilización y la armonización de los formatos de datos y de los valores semánticos dentro de la Unión Europea. Esto puede lograrse ofreciendo instrumentos que ayuden a lograr dicha interoperabilidad (por ejemplo, esquemas XML, tesauros, taxonomías) que puedan ser utilizados en el intercambio de datos de los proyectos transfronterizos.

---

<sup>1</sup> Decisión (95/468/CE) del Consejo de la Unión Europea, de 6 de noviembre de 1995

<sup>2</sup> Decisiones 1719/1999/CE y 1720/1999/CE, conocidas como Decisiones IDA II (Intercambio de Datos entre Administraciones), y sus enmiendas 2045/2002/CE y 2046/2002/CE, todas ellas del Consejo y del Parlamento Europeo.

## 1.2. Marco Europeo de Interoperabilidad

El Marco Europeo de Interoperabilidad para Servicios Paneuropeos de E-Gobierno (MEI o EIF v 1.0, por sus siglas en inglés), publicado en 2004, proporcionó una guía de alto nivel para las instituciones europeas y sus estados miembros y está sirviendo como paraguas para un conjunto federado de orientaciones de interoperabilidad que deberá desarrollar el IDABC y que ya está haciendo a través del Centro Europeo para la Interoperabilidad Semántica (<http://semic.eu>), antes llamado el centro XML (XML Clearinghouse). Actualmente está disponible una revisión del MEI o EIF 2.0 (<http://ec.europa.eu/idabc/en/document/7728>) que previsiblemente sea aprobada antes de final de 2009.

Un marco de interoperabilidad es el instrumento que define los principios, políticas, directrices, especificaciones técnicas, recomendaciones y orientaciones para los servicios de administración electrónica, a fin de garantizar la interoperabilidad y la accesibilidad entre los sistemas de las Administraciones Públicas, entre sectores administrativos y con los ciudadanos (Ministerio de Administraciones Públicas, 2009).

Las bases sobre la interoperabilidad europea ya se han establecido en el documento (European Interoperability Framework for pan-European eGovernment Services [Marco europeo de interoperabilidad (EIF-MEI) para servicios paneuropeos de administración electrónica, <http://www.csi.map.es/csi/pg3315.html>], un documento de trabajo del programa IDA y que se complementa con el documento Architecture Guidelines (AG, <http://ec.europa.eu/idabc/en/document/3485/5585>). Este Marco Europeo de Interoperabilidad (MEI) se construye sobre la base de siete principios:

1. Accesibilidad.
2. Multilingüismo.
3. Seguridad.
4. Privacidad Subsidiariedad.
5. Uso de estándares abiertos.
6. Valoración de los beneficios del software de fuentes abiertas.
7. Uso de soluciones multilaterales.

La Architecture Guidelines (AG) en el área de interoperabilidad de contenidos (semántica) ha desarrollado el Model Requirements for the Management of Electronic Records (MoReq, <http://ec.europa.eu/idabc/en/document/2303/5644>), una guía para la gestión de registros electrónicos y el Managing Information Resources for e-Government (MIReG, <http://ec.europa.eu/idabc/en/document/3615/5585>), un modelo de metadatos basado en Dublin Core para la información gubernamental en aplicaciones paneuropeas, con control de vocabulario, ontologías y topics maps. MIReG define una propuesta de conjunto de metadatos y el marco de aplicación para su uso efectivo.

Las características claves del modelo MIREG están basadas en las recomendaciones de la UE en cuanto a lo que se requiere para que una especificación y sus documentos pueden ser consideradas como un estándar abierto son:

- a) Que el estándar sea publicado y que el documento con las especificaciones del estándar esté disponible de forma gratuita o con un coste mínimo. El estándar debe permitir que se copie, distribuya o utilice sin ningún tipo de tarifa o con una tarifa mínima. Que la propiedad intelectual del estándar, por ejemplo, patentes posibles de todo o de parte, estén disponibles y de forma irrevocable libres de derechos. En este caso, el modelo tiene una fácil implementación en XML, tal y como recomienda la UE y es de uso libre.
- b) Que el estándar sea adoptado y que sea mantenido en el tiempo por una organización sin ánimo de lucro y que las decisiones sobre su desarrollo se basen en un procedimiento de toma de decisiones a disposición de todas las partes interesadas (consenso o decisión por mayoría, etc.). En este caso está basado en el conjunto de elementos definidos en la ISO Dublin Core, Desarrollado en la norma ISO 15836-2003, y mantenido por la Dublin Core Metadata Initiative (DCMI, <http://dublincore.org/>).
- c) Que admita ampliaciones del conjunto de metadatos de acuerdo a las necesidades nacionales y refinamientos de los elementos del conjunto.

La solución a la interoperabilidad semántica es una actividad que seguramente debe realizarse por sectores de actividad, teniendo en cuenta a los distintos actores, la vida de los acontecimientos o los episodios de negocio a los que sirve. Sin embargo, también es necesario identificar a nivel paneuropeo un conjunto común de datos (el núcleo de elementos de datos de la administración electrónica tales como identificadores nacionales básicos de empresas, ciudadanos y administraciones).

Esto requerirá la implementación de infraestructuras tanto técnicas como organizativas y un lenguaje simple que permita la descripción de los significados y estructuras de los datos subyacentes, por ejemplo, un lenguaje de marcas. En el contexto de las tecnologías actuales ese lenguaje es XML. Sin embargo, XML no garantiza por sí mismo interoperabilidad semántica. Esto se alcanza a través de iniciativas para desarrollar una semántica común sobre la base de XML. La introducción de esquemas XML y artefactos relacionados (por ejemplo, metadatos, ontologías, etc) es lo que hace posible integrar servicios que fueron desarrollados con diferentes vocabularios y con diferentes perspectivas sobre los datos.

### **1.3. Iniciativas sobre interoperabilidad en Europa**

En Europa, una de las iniciativas más avanzadas en este ámbito es la llevada a cabo por el gobierno del Reino Unido. De hecho sus iniciativas e-GIF (e-Go-

vernment Interoperability Framework, Version 6.1, de 18 de marzo de 2005, <http://www.govtalk.gov.uk/schemasstandards/egif.asp>) y e-GMS (e-Government Metadata Standard Version 3.1, de 29 de agosto de 2006, <http://www.govtalk.gov.uk/schemasstandards/metadata.asp>) están siendo consideradas por el programa de Intercambio de Datos (IDA-IDAC) de la UE como base para los estándares europeos. E incluso disponen de lo que llaman el IPSV (Integrated Public Sector Vocabulary) <http://www.esd.org.uk/standards/ipsv/1.00/ipsv.pdf>, una iniciativa para crear un vocabulario global gestionado centralmente para ser utilizado por todo el espectro de las organizaciones públicas inglesas y que funciona como un esquema codificado para el elemento tema (subject) de la versión del estándar de metadatos de la administración electrónica (e-GMS). El uso de un vocabulario controlado como IPSV facilita que la gente pueda encontrar más fácilmente recursos de información en los sistemas de información del gobierno pero también ayuda a que los programas informáticos puedan intercambiar datos, interoperar.

El caso inglés no es único, Bélgica también ha desarrollado su eGovernment Interoperability Framework (BELGIF, <http://www.belgif.be>), lo mismo que Estonia, en donde se aprobó el IT Interoperability Framework v2.0 ([http://www.riso.ee/en/files/framework\\_2005.pdf](http://www.riso.ee/en/files/framework_2005.pdf)) y se creó una capa de intercambio de datos llamada X-Road (<http://www.riso.ee/en/information-policy/projects/x-road>). En Grecia, el Greek e-Government Interoperability Framework (<http://www.e-gif.gov.gr/English/tabid/63/language/en-US/Default.aspx>) está en fase de desarrollo. En Alemania, las Standards and Architectures for eGovernment Applications» (SAGA, [http://www.cio.bund.de/DE/Standards/SAGA/saga\\_node.html](http://www.cio.bund.de/DE/Standards/SAGA/saga_node.html)) recogen la importancia de la interoperabilidad. En Francia, en marzo de 2009, está en proceso de aprobación el Référentiel Général d'Interopérabilité (RGI, <http://www.references.modernisation.gouv.fr/rgi-interoperabilite>) Y también en Suecia se ha establecido la Verva (Swedish Administrative Development Agency, <http://www.verva.se/>) y aprobado un marco de interoperabilidad. Aunque casi todos los países recogen la importancia de una u otra manera tanto Dinamarca como Irlanda y el Reino Unido son países de referencia en el campo de la interoperabilidad.

En España, el Ministerio de Administraciones Públicas ha puesto en marcha el Centro de Transferencia de Tecnologías (<http://www.ctt.map.es>), un portal para la difusión y reutilización de soluciones técnicas (servicios, desarrollos, infraestructuras, normativa, etc) de interés para el desarrollo de la administración electrónica. Uno de los proyectos es precisamente la creación del Esquema Nacional de Interoperabilidad (ENI). Este proyecto tiene por finalidad el desarrollo del ENI, previsto en el artículo 42 de la ley 11/2007, de 22 de Junio, de acceso electrónico de los ciudadanos a los servicios públicos. El ENI comprende los criterios y recomendaciones en materia de seguridad, conservación y normalización de la información, de los formatos y las aplicaciones a ser tenidos en cuenta por la Administraciones Públicas para la toma de decisiones tecnológicas, a fin de garantizar la interoperabilidad y la accesibilidad entre los sistemas de las Administraciones Públicas, entre sectores administrativos y con los ciudadanos, en el acceso electrónico a los servicios públicos. El grupo encabezado por Tambouris

(2007) hace un repaso muy completo sobre el estado de la cuestión de la interoperabilidad a escala local y regional.

#### **1.4. El uso de metadatos como factor de interoperabilidad**

Un seminario, celebrado en febrero de 2009 (<http://www.cen.eu/cenorm/businessdomains/businessdomains/iss/workshops/wsegovshare.asp>), sobre descubrimiento y acceso a los recursos de la administración electrónica, organizado por el Comité Europeo de Normalización (CEN) reconocía que el descubrimiento y el acceso a los servicios de la administración electrónica es a menudo difícil, a veces, por una simple ausencia de clasificación, de una correcta terminología o por descripciones inexistentes de los servicios, pero también por problemas de lenguaje o de otras habilidades semióticas.

Las conclusiones del seminario resumían los obstáculos para ampliar el acceso y la reutilización de los recursos de la administración electrónica en:

1. La falta de una solución amplia y fácil de usar de esquemas normalizados de metadatos para su descripción.
2. La falta de una terminología administrativa fácilmente accesible para usar en esas descripciones (el caso del IPSV inglés mencionado antes puede considerarse una rareza).
3. Falta de conciencia sobre la diversidad cultural incluyendo las competencias lingüísticas de los usuarios.

Un importante número de gobiernos en todo el mundo reconocen la necesidad de establecer estándares de metadatos como un ingrediente integral y fundamental de su marco de interoperabilidad. El conjunto de elementos Dublin Core ha sido utilizado como base en diferentes administraciones nacionales. El valor de los modelos de datos genéricos deriva de su reusabilidad y de su conformidad a descripciones conceptuales normalizadas previstas entre diferentes dominios. Y de hecho varios estados miembros de la UE han obligado el uso de DC como el estándar de metadatos en sus administraciones públicas o han desarrollado sus conjuntos de metadatos en sus administraciones públicas con variaciones de DC.

Alasem (2009) hace un repaso de las iniciativas de metadatos de la administración electrónica de diversos gobiernos basadas en Dublin Core, centrándose en Reino Unido, Irlanda, Canadá, Australia y Nueva Zelanda. En ese artículo se propone un plan de trabajo para el desarrollo de esquemas de metadatos en la administración electrónica. El plan se divide en cuatro fases: 1) el establecimiento de un grupo de trabajo. 2) La identificación de los requisitos de los proveedores y productores de recursos gubernamentales y de los usuarios de la administración electrónica. 3) El estudio de los sitios web gubernamentales. 4) La determinación de los elementos de metadatos apropiados.

En España se está desarrollando el Esquema Nacional de Interoperabilidad. En nuestro estudio pretendemos justamente analizar el uso que los sitios web

del gobierno español están haciendo, si es que lo hacen, de los metadatos como mecanismo de descripción de los recursos web con el fin de determinar si las entidades gubernamentales españolas han hecho algún esfuerzo previo a la adopción del ENI para adoptar algún esquema, definiciones o vocabularios específicos para la descripción de recursos de la administración electrónica.

## 2. Metodología y análisis de resultados

Con el fin de detectar el uso de metadatos y/o vocabularios que se utilizan en la administración electrónica española se seleccionó una muestra de 18 dominios web, los de los ministerios del gobierno central, más el de la Presidencia del Gobierno española. La evolución que experimentan los distintos ministerios en los diferentes períodos de gobierno hizo que al final fuesen considerados 31 sitios web.

Por ejemplo, existen ministerios que funden competencias procedentes de otros ministerios, es el caso del sitio web del Ministerio de Medio Ambiente, Rural y Marino que cuenta con tres dominios distintos de su sitio web principal (marm.es, el actual; mma.es, del anterior Ministerio de Medio Ambiente; y mapa.es, del anterior Ministerio de Agricultura, Pesca y Alimentación); ministerios con versiones en distintos idiomas de un mismo sitio web, como el caso del Ministerio de Cultura (www.mcu.es; gl.www.mcu.es o en.www.mcu.es) o ministerios muy grandes con varios sitios web principales, como el caso del Ministerio de Industria, Turismo y Comercio (mityc.es o comercio.es). Se dejaron fuera tanto las administraciones públicas locales y regionales así como los organismos autónomos como puertos, aeropuertos y otras instituciones públicas.

La muestra elegida se construyó recorriendo con un robot un mínimo de 2.000 páginas web de cada uno de los 18 sitios seleccionados, entre el 11 y el 15 de marzo de 2009. En total se analizaron 2.874.034 enlaces de los que tras una depuración por tipo de recursos se extrajeron 42.436 páginas web distintas. De los enlaces analizados se excluyeron ficheros de formato multimedia, propietarios y/o comprimidos por lo que el número final de páginas recogidas en algunos de los sitios fue inferior a las 2.000 recuperadas.

Aunque los ficheros pdf, doc, mp3, jpg, swf, etc., pueden llevar metadatos, los tipos de metadatos de estos recursos no responden generalmente a los recomendados para la administración electrónica, por lo que la muestra se limitó fundamentalmente a páginas en formato html, htm, xml, php, asp y sus diferentes variantes. De las 42.436 páginas, 35.109, el 82,73% que serán el objeto de nuestro análisis, poseían etiquetas de metadatos de algún tipo.

Como veremos más adelante, lo que en principio parecía un dato prometedor pronto se revelaría como un espejismo puesto que muchos de esas etiquetas no tenían contenido o su contenido no se ajusta a los metadatos que se utilizan en la administración electrónica.

De las 35.109 páginas se extrajeron 268.457 etiquetas meta en 106 atributos distintos. Por tanto, la media de etiquetas meta por página o url distinto analiza-



do fue de 7,59 (Tabla I) mientras que la media de atributos distintos por url fue de 17,5, lo que es lógico puesto que en una etiqueta meta pueden colocarse varios atributos. Un ejemplo puede verse en el siguiente ejemplo de una etiqueta con tres atributos.

```
<meta name="DCTERMS.created" scheme="ISO8601" content="2008-07-04">
```

**TABLA I**  
*Dominios analizados*

Totales por sitio			
Sitio	Nº Meta	Nº URL	Media
www.mcu.es	42.229	2.061	20,49
www.mpr.es	2.967	257	11,54
www.la-moncloa.es	2.543	242	10,51
www.fomento.es	37.422	3.728	10,04
www.mde.es	21.431	2.352	9,11
www.mviv.es	28.758	3.246	8,86
www.msc.es	1.893	235	8,06
www.mtas.es	15.896	1.994	7,97
www.meh.es	10.211	1.204	8,48
www.marm.es	45.318	5.578	8,12
web.micinn.es	12.222	2.034	6,01
www.mityc.es	11.364	1.938	5,86
www.mepsyd.es	9.065	1.784	5,08
www.mir.es	8.875	1.769	5,02
www.map.es	7.624	1.718	4,44
www.maec.es	2.620	792	3,31
www.mjusticia.es	7.605	3.955	1,92
www.migualdad.es	414	222	1,86
<b>TOTALES</b>	<b>268.457</b>	<b>35.109</b>	<b>7,59</b>

Al observar el conjunto de etiquetas META empleadas por sitio, se observan algunos datos que conviene poner en su contexto. Por ejemplo, destacan los 47 atributos diferentes que ofrece el Ministerio de Vivienda (un total de 28.758 eti-

quetas META para 3.246 páginas). No obstante, un análisis detallado muestra que al menos un 58,7% de dichas etiquetas, están vacías o contienen caracteres como «,» o «...». Descontadas las mismas, la media descendería desde las 8,86 hasta las 3,65 etiquetas por página. También es relevante el caso del Ministerio de Cultura. De las 42.229 etiquetas META que se encontraron en este Ministerio, al menos 10.816 tienen un *name* con un valor, acompañado de un atributo *content* que está vacío, lo cual reduciría la media de 20,49 a 15,24 etiquetas por página.

Este problema de las etiquetas con contenido vacío afecta a casi toda la muestra analizada, aunque la incidencia varía de unos sitios a otros. Por ejemplo, el Ministerio de Ciencia, de cuya muestra se recogieron 12.222 atributos, contiene 8.148 atributos *name*, con valores *author*, *description*, *keywords* y *owner*, cuyos respectivos atributos *content* están vacíos. Es decir, que contiene únicamente una media de apenas 2 metadatos válidos por página. Algo similar ocurre con la Secretaría General de Presupuestos (uno de los sitios del Ministerio de Economía), en el que a priori se cuentan 1.336 atributos *name* que contienen a su vez un atributo *content* cuyo valor está vacío. Este dato es especialmente llamativo, dado que supone el 25% de todos los metadatos de esa sección y, a su vez, más de un 10% de todos los metadatos obtenidos de la muestra tomada en el Ministerio de Economía.

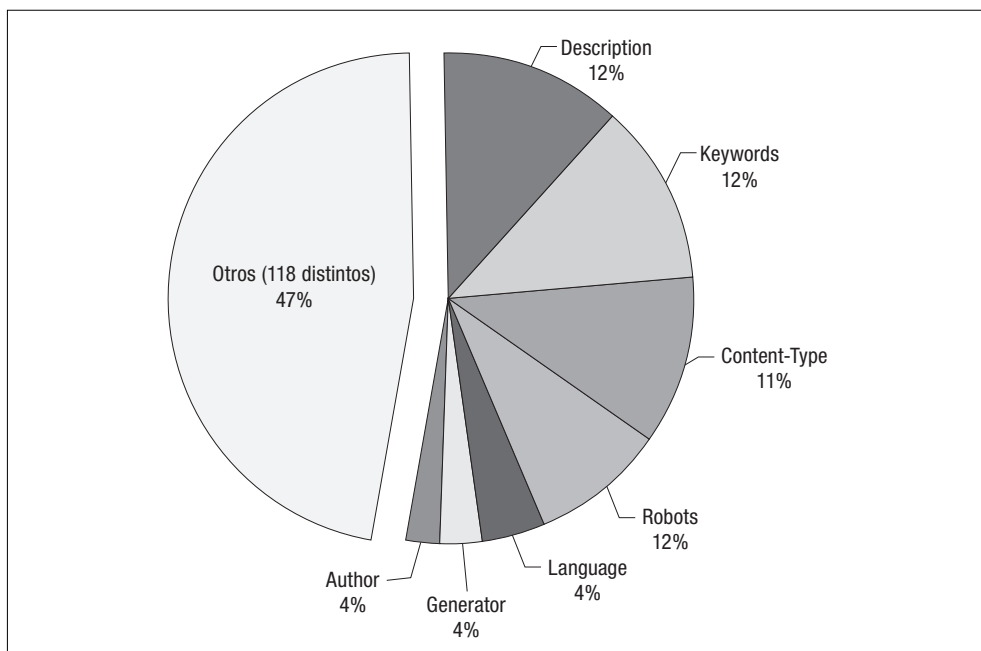
En otros casos, el número de elementos vacíos afecta a algún metadato concreto. Es el caso de los atributos *description* y *keywords* existentes en la sección de Medio Ambiente del Ministerio de Medio Ambiente, Marino y Rural. Un 15% del valor del correspondiente atributo *content*, junto a dichos atributos (712, en total, entre ambos) también están vacíos.

En resumen, los cinco atributos más utilizados fueron *keywords* (33.596 casos), *description* (33.530), *robots* (24.442), *language* (11.990) y *generator* (8.595). Ahora bien, aproximadamente un 15% de los atributos *description* y *keywords*, los más semánticos, en los diferentes sitios que los han utilizado, incluyen un atributo *content* cuyo valor está vacío, por lo que la aparición del atributo no dice absolutamente nada respecto al recurso (Figura 1).

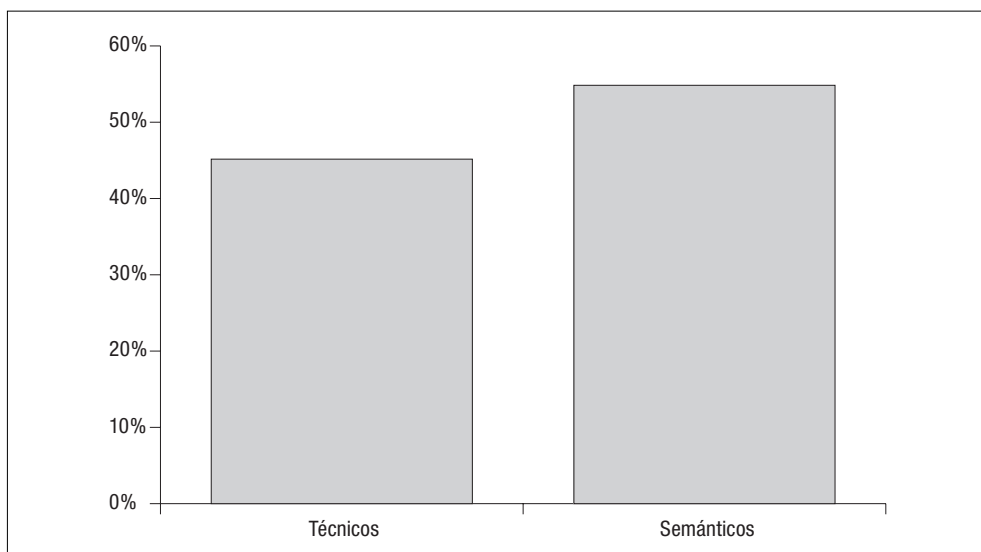
Con el fin de simplificar la presentación de los datos, los atributos fueron agrupados en dos grandes categorías: semánticos y técnico-funcionales (Figura 2).

Entre los semánticos, que supusieron un 54,84% del total, los relacionados con la descripción o representación de los recursos (*keywords*, *classification*, *description*, *title*, *category*, *subject* y sus variantes como *dc.keywords* o *categoría*, etc.); los relacionados con la propiedad, autoría o edición y sus variantes (*publisher*, *author*, *copyright*, *owner*, *editor*, *creator* o *contributor*); y otros comunes como *language*, *date*, *type*, *coverage*, *traducción*, etc. Entre los técnicos y funcionales, que sumaban un 45,16% del total, se analizaron en profundidad los relacionados con la codificación de caracteres, los relacionados con los robots (*expires*, *visit*, *revisit*, *visit-after*, *verify-v1*, *spiders*) y otros como *rating*, *generator*, etc.)

**FIGURA 1**  
*Atributos de metadatos más utilizados*



**FIGURA 2**  
*Típos de metadatos*



## 2.1. Metadatos relacionados con la descripción y/o representación del recurso

Entre los atributos más significativos de carácter semántico empleados para la descripción de contenido, se encuentran los siguientes (Tabla II):

**TABLA II**  
*Metadatos semánticos para describir contenido*

Atributo	Ocurrencias	Páginas (%)
keywords	33.596	95,69
description	33.530	95,50
DC.description	4.519	12,87
title	3.966	11,30
DC.title	2.787	7,94
DC.keywords	2.064	5,88
MCU.area	2.053	5,85
category	1.940	5,53
Subject	992	2,83
DC.Subject	344	0,98
DC.Source	337	0,96
Classification	97	0,28
DC.title.alternative	25	0,07

El valor del atributo name más empleado en las etiquetas META de los documentos analizados es *keywords*, con 33.596 ocurrencias, lo que significa que aparece en casi el 95,69% de las páginas analizadas, si bien, como ya se ha explicado, en el 15,22% de los casos estas etiquetas no tienen ningún valor, aparecen vacíos. De los 18 dominios analizados su uso aparece en todos ellos a excepción del Ministerio del Interior y del Ministerio de Asuntos Exteriores.

Especialmente significativos son los casos del Ministerio de Ciencia e Innovación y del Ministerio de la Vivienda. En el primero de ellos, todas sus páginas web analizadas aparecen con el valor de este metadato vacío. En el segundo, en la mayoría de ellas encontramos elementos como la coma (,), los puntos suspensivos (...) o términos como «mambo, Mambo», haciendo con ello clara alusión al nombre del gestor de contenidos web empleado.

Destaca igualmente el caso de las páginas del Ministerio de Fomento, donde se mezclan páginas con un uso adecuado del metadato *keywords* con otras donde se emplea éste para realizar una descripción somera del contenido de las páginas, haciendo las funciones del metadato *description*, como se verá más

adelante. La repetición del mismo valor en los casos que aparece este metadato es característica del dominio de Comercio del Ministerio de Industria, Comercio y Turismo, en donde encontramos en la práctica totalidad de los casos el texto «INFORMACIÓN ESPECÍFICA DE COMERCIO».

En el otro extremo, resalta el buen uso que hacen de este metadato las páginas web del Ministerio de Cultura, donde un alto porcentaje (más del 80%) de ellas incluyen información al respecto, siendo ésta correcta y precisa (aunque sorprende que se haya utilizado las mayúsculas para la introducción de las palabras clave en todos los casos existentes). Igualmente es destacable la labor realizada en las web del Ministerio de Educación, Política Social y Deporte, el Ministerio de Industria, o el Ministerio de Defensa con una relación correcta entre la cantidad de términos empleados y la adecuación de los mismos.

El metadato *description*, uno de los más habitualmente empleados en la web para realizar una breve descripción del contenido de la página web en la que se encuentra, es, como ya se señaló anteriormente, el segundo en número de apariciones de nuestro estudio. Con 33.530 ocurrencias aparece en el 95,50% de las páginas analizadas, en prácticamente todas las ocasiones en las que aparece el elemento *keywords*. En este caso la cifra de valores vacíos encontrados ha sido algo superior al anterior, obteniendo un porcentaje del 17,38% de los casos. Al igual que sucedía antes, todos los valores que se han encontrado para este atributo en el Ministerio de Ciencia e Innovación aparecen vacíos. De forma casi similar, en el Ministerio de Administraciones Públicas se ha encontrado un número muy reducido de casos (el 1,6%) en los que se ha asignado valor a este metadato.

Entre los casos donde su uso no es el adecuado podemos destacar el caso del Ministerio de la Vivienda donde, de nuevo, nos encontramos con una ausencia casi total de valores correctos para este atributo: aparecen, una vez más, los puntos suspensivos, las comas, o información relativa al gestor de contenidos Mambo; o el caso del Ministerio de Justicia, donde los valores para este atributo son prácticamente idénticos a los que encontramos para su atributo *keywords*.

En el caso del Ministerio de Educación, Política Social y Deporte sí que se emplea un texto descriptivo para el valor de este atributo pero éste es el mismo en todos los casos, tratándose de una mera declaración de la función y competencias de dicho ministerio. Esto mismo sucede en las páginas, cuando se cumplimenta este valor, del Ministerio de Medio Ambiente, Medio Rural y Marino, del Ministerio de la Presidencia o de La Moncloa. Por ello, su utilidad es realmente escasa a efectos de recuperabilidad e interoperabilidad.

En el nivel superior de calidad en el uso de este metadato encontramos muchos y buenos ejemplos entre las páginas del Ministerio de Industria, Turismo y Comercio, en el Ministerio de Sanidad y Consumo, en el Ministerio de Economía y Hacienda y, muy especialmente, en el Ministerio de Cultura, aunque también el porcentaje de valores vacíos es algo elevado para este ministerio.

El equivalente a este tipo de metadato en el modelo del Dublin Core aparece reflejado en las etiquetas META como *DC.description*. En este caso, el contenido informativo de este atributo suele ser siempre el mismo para todas las pá-

ginas del sitio en el que aparece, por lo que su valor descriptivo tiene poco sentido.

De este abuso en la repetición de los textos descriptivos no se libra ni el Ministerio de Cultura, pues, a pesar de ser, como hemos venido repitiendo, uno de los mejores en cuanto a la calidad de la descripción, en sus páginas de la Agenda Cultural se emplea siempre el mismo texto.

Las etiquetas *title*, *DC.title* y alguna otra no solo son utilizadas escasamente sino que a menudo no responden a un título descriptivo del recurso. Así, por ejemplo, en el Ministerio de Cultura se puede encontrar el título «Agenda Cultural» en más de mil ocasiones o en otros ministerios «procedimientos administrativos» en más de 100 ocasiones o «index» o «Plan de Calidad para el Sistema Nacional de Salud» en casi otras tantas.

Un último metadato dentro de este grupo es *subject*, propio del lenguaje HTML pero habitualmente poco utilizado al no ser tenido en cuenta, en principio, ni por los navegadores web ni por los programas de recuperación de información en Internet. Si se emplea, su contenido debería hacer referencia a la materia genérica o principal tratada en la página web. Su utilización ha sido prácticamente nula, no alcanzando ni el 3% del total de etiquetas META.

El equivalente de *subject* en el modelo del Dublin Core aparece consignado en las etiquetas META con *DC.subject*, cuya utilización es escasa respecto del conjunto de etiquetas META del DC, no llegando al 1%. Sus valores son los mismos que los contenidos en el metadato *keywords*, siendo el mismo en todas las páginas que los incluyen.

## 2.2. Metadatos relativos a la autoría

Los atributos más significativos relativos a la autoría de las páginas son los siguientes (Tabla III):

**TABLA III**  
*Metadatos semánticos sobre autoría*

Atributo	Ocurrencias	Páginas (%)
author	6.739	19,19
copyright	3.929	11,19
owner	3.816	10,87
DC.editor	2.054	5,85
editor	2.054	5,85
publisher	1.808	5,15
autor	1.005	2,86
DC.creator	717	2,04
DC.publisher	587	1,67
DC.contributors	232	0,66

Los valores *author*, *autor* y *DC.creator* del atributo *name* se han empleado para representar a autores institucionales, es decir, al ministerio u organismo que ha generado la página, o bien, a la empresa que ha participado en la elaboración del sitio. En total, el valor aparece en un 24,09% de las páginas analizadas. El más utilizado es *author*, repartido sobre todo en cinco ministerios (Trabajo, Ciencia y secciones de Economía, Industria y Justicia). No obstante, un tercio de los mismos tenían a continuación un atributo *content* cuyo valor estaba vacío.

*Autor* es un valor utilizado exclusivamente en una de las secciones del Ministerio de Industria, mientras que *DC.creator* ha sido empleado en 9 sitios diferentes.

Otro grupo de metadatos lo componen las etiquetas que indican habitualmente a la institución que asume la responsabilidad de los contenidos. Sería el caso de los valores *DC.editor* y *editor*. Todos los casos de ambas etiquetas se concentran en un único ministerio, el de Cultura, que tiene ambas etiquetas en todas sus páginas, y en cada página, ambos atributos tienen el mismo valor: la sección del ministerio que se hace responsable del contenido de la página.

Un caso similar al anterior es el de las etiquetas *publisher* y *DC.publisher*. El valor *publisher* es usado en la totalidad de las páginas del Ministerio de Interior, así como en unas pocas páginas de Defensa. En cuanto a *DC.publisher*, está más repartido, pero solo se usa realmente en dos secciones de Medio Ambiente y Economía, así como en un 15% de las páginas del Ministerio de Educación. En el caso de *DC.publisher*, todos los casos en los que aparece coincide con un atributo *DC.creator*, y ambos tienen el mismo valor.

El atributo *owner*, que apunta directamente al propietario de las páginas, está repartido únicamente en dos ministerios: Interior (con el nombre del ministerio como valor) y Ciencia (en todos los casos, con valor vacío). Algo similar ocurre con *copyright*, que aparece en la totalidad de las páginas de Trabajo y Educación, y parcialmente en Medio Ambiente. En todos los casos, se remite al ministerio correspondiente o, en su defecto, a alguna unidad del mismo.

Por último, en el caso del atributo *DC.contributors* (que debería ser *DC.contributor*) el valor del respectivo atributo *content* está vacío en todos los casos.

### 2.3. Otros metadatos comunes (language, date, coverage, type)

La descripción sobre el lenguaje empleado es una de las características más comunes de las páginas. Este tipo de descripción se encuentra, descontando las repeticiones de etiquetas (por ejemplo, el caso de «lenguaje» y «traducción», utilizados simultáneamente para describir esta característica por el Ministerio de Cultura), en el 53,46% de las páginas analizadas. Los principales atributos empleados son los siguientes (Tabla IV):

**TABLA IV**  
*Metadatos semánticos sobre lengua*

Atributo	Ocurrencias	Páginas (%)
language	11.990	34,15
content-language	2.927	8,34
lenguaje	1.972	5,62
traducción	1.972	5,62
dc.language	1.577	4,49
idioma	506	1,44

El atributo más utilizado por la mayoría de los ministerios es *language* y, como era de esperar, el valor más común es el español, expresado a través de tres formas principales: «es» 13.604 veces; «spanish» 5.183 veces; y «es-es» 1.075 veces.

Ningún otro valor aparece en más de 300 ocasiones. El inglés aparece como «en» en 170 ocasiones y varias decenas más como «en-gb» y «en-us», el francés aparece como «fr» 247 veces, y el resto de los valores se reparten en las distintas lenguas oficiales del estado (ca o ca-es para catalán, eu ó eu-es para euskera, ga o gl para gallego o va en alguna ocasión aislada para valenciano).

No obstante, aunque los ministerios no parecen introducir la lengua como metadato, encontramos casos de páginas traducidas a otras lenguas oficiales. Pero, en vez de reflejarlo como parte de la información que describe el recurso, lo hacen a través del procesamiento de programas. Esta circunstancia es relativamente fácil de deducir mediante el análisis del url.

Algunos ejemplos de URL con indicación de lengua son:

```

http://ca.www.mcu.es/sitios/cambiarIdioma.do?newLanguage=es
http://ca.www.mcu.es/sitios/cambiarIdioma.do?newLanguage=eu
http://www.fomento.es/MFOM/LANG CASTELLANO/INFORMACION_MFOM/
PRESUPUESTOS_E_INVERSIONES/
http://www.fomento.es/MFOM/LANG FR/default.htm
http://www.la-moncloa.es/IDIOMAS/9/MapaWeb.htm
http://www.mapa.es/app/pnir/Publico/Resultados/Resultados.asp?idioma=es
http://www.mapa.es/app/Zootecnia/Catalogo.aspx?lng=es
    
```

Para mostrar las fechas de creación o publicación de los recursos los distintos ministerios utilizan distintas variantes del atributo *date* (Tabla V).

Tan solo un escaso 19,09% de estas atributos tienen valores reales de fecha mientras que el 80,90% restante contiene valores vacíos o erróneos. *date* y *DC*.



**TABLA V**  
*Metadatos semánticos sobre fecha*

Atributo	Ocurrencias	Páginas (%)
date	2.059	5,86
DC.date	2.153	6,13
DC.date.created	336	0,96
fecha	96	0,27
DCTERMS.created	51	0,15
DC.date.modified	6	0,02

*date* son ampliamente utilizados por el Ministerio de Cultura. «Fecha» es utilizado solo por Presidencia del Gobierno y por el Ministerio de Presidencia.

Los formatos de valores de fecha varían de ministerio a ministerio pero, sin duda, el más utilizado es el de dd/mm/aaaa hh:mm, aunque también el de aaaammdd, y algunos otros como 2009-01-09T23:21:55+01:00.

El atributo *coverage* sirve para indicar el lugar o la cobertura temporal de un recurso y se representa mediante el nombre de un sitio o su localización expresada en forma de coordenadas geográficas o mediante el nombre de un período, una fecha o un rango de fecha. En los ministerios analizados se han encontrado otras etiquetas afines (*dc.coverage*, *dc.coverage.placename*, *geo.placename*, *geo.position*, *geo.region*, *lugar* o *mcu.area*) (Tabla VI).

**TABLA VI**  
*Metadatos semánticos sobre cobertura*

Atributo	Ocurrencias	Páginas (%)
dc.coverage	2.054	5,85
coverage	2.053	5,85
MCU.area	2.053	5,85
dc.coverage.placename	199	0,57
geo.placename	51	0,15
geo.position	51	0,15
geo.region	51	0,15
lugar	55	0,16

*coverage* y *DC.coverage* son utilizados exclusivamente por el Ministerio de Cultura, pero sus valores están vacíos en todos los casos. En este ministerio, cada vez que se utiliza uno de estos dos atributos, se utiliza también el atributo *MCU*.

*area* en el que se toma como valor o bien «Ministerio de Cultura» o bien alguna de sus áreas «Bibliotecas», «Archivos», «Cine y audiovisual», «Cooperación cultural», etcétera.

*DC.coverage.placename* aparece en el Ministerio de Economía, y con contenido vacío en todos los casos. *geo.placename* es utilizado solo por el Ministerio de Medio Ambiente, Rural y Marino, y siempre con el mismo valor, el de la dirección postal del ministerio («Paseo de Infanta Isabel, 1 - 28071, Madrid, España»). Igual ocurre para *geo.position* (40.4089333333, -3.6903611111) y *geo.region* (ES-M). Por su parte, Presidencia del Gobierno y el Ministerio de la Presidencia utiliza el atributo *lugar* con valores de días de la semana (Friday, Tuesday), pero también con valores que podrían ser propios del atributo (como Madrid, Banjul, Tenerife, Congreso de los Diputados).

Finalmente, el atributo *type* se utiliza para describir la naturaleza o el género del recurso, no el formato del fichero, el medio físico o las dimensiones del recurso. El atributo es utilizado en apenas un 4,84% de las páginas analizadas. En prácticamente todos los casos se utiliza con el valor «text». El resto de los casos suelen ser errores en los que se utiliza el valor «imagen». También existen algunos casos en los que se utiliza como atributo *resource-type*, prácticamente todos con valor de «document»; y algunos otros en los que se utiliza el atributo *ximdex\_doc\_type*.

#### 2.4. Atributos funcionales (codificación de caracteres, robots)

La codificación empleada para elaborar la página, en la mayoría de los casos, aparece como un valor del atributo *http-equiv*, que suele llevar el valor «text/html; charset=», seguido del tipo de codificación empleado. El 83,93% de las páginas con metadatos contienen información sobre la codificación que se utiliza. En el 57,51% de las páginas con información sobre este aspecto la codificación se realiza de acuerdo a la norma ISO 8859-1, el 42,21% a UTF-8, y en menos de 100 casos a Windows-1252 o UTF-16.

En cuanto a los valores funcionales del atributo *name*, el más utilizado con diferencia es *robots*. Este valor indica cuál debe ser el comportamiento que debe seguir el robot de un buscador, cuando llega a la página. Por defecto, si no aparece este elemento, se entiende que el robot tiene vía libre para realizar ambas tareas.

Este elemento aparece en un 70% de las páginas. No obstante, aproximadamente un tercio de los mismos son duplicados; es decir, que emplea más de un meta con valor *robots* en una misma página. Estos casos pueden ser: o diferentes valores que son complementarios (por ejemplo, en una misma página, un *robots* con valor «INDEX, FOLLOW», y otro con valor «ALL»); o dos atributos *robots* en una misma página: uno con valores válidos (INDEX, FOLLOW, etc.) y otro, con valores vacíos.

El 89% de estos valores permiten que los *robots* recopilen tanto las páginas como los enlaces de las mismas. Solo un 10% de los valores válidos imponen

que los *robots* no incluyan ni páginas ni enlaces dentro de sus bases de datos: así ocurre principalmente en Asuntos Exteriores, Cultura, parcialmente en Presidencia del Gobierno y, puntualmente, en recursos de otros ministerios.

Por su parte, el valor *expire* indica a los *robots* si el contenido de la página expira en alguna fecha y hora, y cuándo llega ese momento. Apenas un 10% de las páginas incluye este valor, fundamentalmente dentro del elemento *http-equiv*.

Otros valores relacionados son los que hacen referencia a si el *robot* debe visitar o no la página de nuevo; y, en ese caso, cuál es el período pasado el cual han de volver a la página. Es el caso de *revisit* y *revisit-after*. Ambos son utilizados fundamentalmente por Medio Ambiente (con valores «15 days» y «15 day»).

En un 24,48% de las páginas se emplea el atributo *generator*, que sirve para la descripción de la aplicación informática empleada para generar las páginas web. Los valores más importantes detectados son Mambo, Microsoft Visual Studio y XimDex. Asimismo, un 14,14% de las páginas analizadas incluyen el uso del valor CODE-LANGUAGE. En la mayoría de los casos, se han empleado Visual Basic, seguido de C#.

Casi como anécdota, un 40% de las páginas del Ministerio de Medio Ambiente, así como algunas páginas sueltas en otros ministerios, incluyen el valor *rating*, que indica si el contenido de las páginas está restringido a personas a partir de una determinada edad. En todos los casos, el acceso es para todos los públicos.

### 3. Conclusiones

El 82,73% de las 42.436 páginas analizadas contienen etiquetas meta de algún tipo. Lo que puede parecer una alta conciencia respecto a la conveniencia de describir los recursos en la administración electrónica parece más un espejismo que una realidad. En total se analizaron 268.458 atributos META, de ellos casi el 10% de los mismos estaban vacíos y un porcentaje similar contenía datos erróneos como simples signos de puntuación o nombres del gestor de contenidos en atributos cuyos valores deberían contener una descripción o las palabras clave del recurso.

De los datos analizados se deduce que no existe ningún estándar de aplicación o de política consensuada en la administración pública central respecto a la inclusión de metadatos en sus espacios Web de difusión de información. Cada web ministerial analizado hace un uso particular de los posibles atributos a emplear y de sus correspondientes valores, y tampoco existe coherencia ni en el número ni en el tipo de metadatos empleados dentro de un mismo ministerio. Tan solo un tercio de los ministerios aplica los mismos metadatos a todas sus páginas.

Esta falta de armonización alcanza no sólo al uso de metadatos, de esquemas para la normalización de valores, o de palabras claves para designar temas y materias tratadas sino, ni siquiera, para códigos de lengua, de fecha o de locali-

zación geográfica. Así, los ministerios utilizan 125 atributos distintos con la peculiaridad de que algunos emplean en la misma página hasta tres etiquetas para la representación de un mismo contenido. Por ejemplo, se utiliza la etiqueta *keywords*, *dc.keywords*, que no existe como tal, y se dejan vacías para colocar el contenido en alguna otra etiqueta propia.

Del total de metadatos obtenidos en nuestro estudio, algo más de la mitad son de tipo semántico, y el resto, técnicos y funcionales. Entre los metadatos de carácter funcional los más utilizados son *content-type*, aparecen en el 85,12% de las páginas, en la mayoría de los casos con el valor *text/html*, y *robots* y *generator*, con el 69,56% y el 24,48% respectivamente.

Entre los de carácter semántico distinguiremos las etiquetas puras de HTML respecto del uso que se hace de Dublin Core. Entre las de HTML, los tres atributos más utilizados en sus diferentes formas fueron: *keywords*, aparece en el 95,69% de las páginas analizadas, *description*, en el 95,50%, y *language*, en un 34,15%.

No obstante, el uso que se hace de estas etiquetas es muy deficiente. Se puede afirmar que, en prácticamente el 50% de los casos relacionados con *keywords*, la información que aparece está vacía, contiene algún signo de puntuación o información irrelevante como «mambo» (un gestor de contenidos) o «ministerio» o «estadísticas» como palabra clave única. O, peor aún, cientos de páginas son etiquetadas de la misma forma: «información específica de comercio» o «Ministerio de Cultura, información, agenda, actividades culturales, calendarios» y otros por el estilo. Algo parecido ocurre con *description* y sus etiquetas equivalentes. En aproximadamente un 31% de los casos en que aparece la etiqueta está vacía, con algún signo de puntuación o con información irrelevante.

A pesar de la relevancia del modelo Dublin Core (DC) respecto a la interoperabilidad entre los sistemas de información en la administración pública, el número total de etiquetas DC encontradas en el estudio alcanza un triste porcentaje del 8% respecto del total. Además, la mayoría de la metainformación que se emplea mediante este modelo hace uso de unos pocos elementos del conjunto que constituye su núcleo principal (15), encontrando en muy pocos casos el uso de calificadores para delimitar su alcance o la remisión a un esquema para el uso de valores normalizados.

El resto de las 125 etiquetas son utilizadas en menos de un 20% de los casos y a menudo son utilizadas por un solo ministerio. Los datos parecen indicar más una preocupación por lograr una mayor visibilidad ante los motores que por una descripción de los recursos puesto que las tres etiquetas más empleadas (*keywords*, *description* y *robots*) tienen que ver directamente con los motores de búsqueda. Las otras tres etiquetas más utilizadas (*content-type*, *language* y *generator*) son etiquetas que se generan de forma automática.

El ministerio con mayor y mejor uso de metadatos en sus páginas es el de Cultura: tiene una buena variedad de atributos (en torno a 20 metadatos por página), su distribución en las páginas resulta bastante homogénea y es prácticamente el único que incluye metadatos Dublin Core de forma sistemática. Sin

embargo, ni siquiera este ministerio es ajeno al abuso de metadatos con valores repetidos. Contiene, además, varios metadatos que son de uso exclusivo, lo que dificulta su interpretación por parte de los principales robots en Internet.

El Ministerio de Fomento también hace uso de metadatos con una alta coherencia, tanto en cantidad, unas 10 etiquetas meta por página, como en variedad, lo mismo aunque en menor medida que los Ministerios de Defensa, Medio Ambiente, Economía y Sanidad. Todos ellos incluyen al menos los principales atributos analizados en este estudio junto a ciertas particularidades.

Un tercer grupo de ministerios está formado por aquellos que hacen uso de una aceptable cantidad de etiquetas meta en sus páginas pero que, sin embargo, de un análisis detallado de los mismos se deduce que sus valores están vacíos o repetidos en muchos casos, con lo que el número real de metadatos para cada uno de estos ministerios resulta aún más bajo de lo esperado. Es el caso de los Ministerios de Presidencia, Vivienda, Trabajo, Ciencia e Industria, además de la Presidencia del Gobierno.

Por último, existe un grupo de ministerios que, o bien tienen un bajísimo número de metadatos por página, como Administraciones Públicas, Asuntos Exteriores, Justicia e Igualdad, con una cifra igual o inferior a cinco; o bien, hacen un uso muy escaso de etiquetas semánticas, como Educación o Interior; en muchos casos, además, con el mismo valor en todas las páginas. En el caso concreto de Asuntos Exteriores, todos los metadatos que aparecen son exclusivamente de carácter técnico-funcional, sin aportar, por tanto, información alguna de carácter semántico.

A partir de las conclusiones obtenidas, ofrecemos algunas recomendaciones y sugerencias para la mejora de los metadatos y, con ellos, de la interoperabilidad en los sitios web de la Administración española

#### **4. Sugerencias y recomendaciones sobre buenas prácticas en el uso de metadatos en la administración electrónica**

Basándonos en el análisis de la muestra de las páginas analizadas y en las observaciones formuladas en los epígrafes precedentes, podemos llegar a las siguientes sugerencias y recomendaciones para los creadores de páginas y administradores de sitios web de la administración pública. El objetivo de estas sugerencias es mejorar la carga semántica de las páginas para facilitar su interoperabilidad, mejorar su visibilidad y su facilidad de recuperación a través de los motores de búsqueda. Mayor interoperabilidad, mayor visibilidad y mejores posibilidades de recuperación contribuyen al cumplimiento de lo que se establece en la Ley LEY 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público.

1. Tanto las políticas nacionales como europeas e internacionales tienden a primar el uso de estándares abiertos, por lo que la primera sugerencia

sería tender al uso de lenguajes de marcado como html, xhtml o xml así como, en la medida de lo posible, el empleo de formatos de documentos abiertos.

2. Tanto los creadores de páginas web como los administradores de sitios web deberían concienciarse y tomar las medidas necesarias para que ninguna página careciese de título y de unas mínimas etiquetas de metadatos. En tanto en cuanto no se aprueben el Esquema Nacional de Interoperabilidad (ENI) y/o el resto de directrices o normas que la acompañen, se recomienda utilizar las etiquetas meta reflejadas en la norma html o un conjunto o subconjunto de las etiquetas definida en el también estándar abierto de Dublin Core (ISO 15836), hasta ahora el adoptado por la mayoría de las administraciones internacionales más avanzadas en el campo de la administración electrónica. En cualquier caso, el uso de etiquetas meta simples o de etiquetas Dublin Core debería ser siempre consistente respecto a un sitio. Esto es, que se determine el uso de uno u otro pero no una mezcla de ambos.
3. Mientras no se aprueben las normas que lo regulen, cada administración debería definir un conjunto más o menos amplio de las etiquetas meta a introducir de forma casi automática, muy fácil de implementar a través de los gestores de contenido que se están utilizando. La sintaxis de las etiquetas debería respetar las normas internacionales y utilizarlas tal y como se dice en las normas («title», y no «título»; o «keywords», y no «palabras clave») con el fin de facilitar la interoperabilidad entre las máquinas.
4. No obstante, se utilicen etiquetas meta simples del html o etiquetas procedentes de Dublin Core, todos los documentos deberían contener una lista mínima de etiquetas de metadatos que reflejen: el título, el organismo responsable de la información que se publica (creador o autor), la lengua, la fecha y las palabras clave.
5. Según nuestros análisis, la etiqueta *title* como metadato se utiliza en casi el 20% de las páginas (*title* en el 11,30% y *dc.title* en el 7,94%) analizadas. Ésta es una de las etiquetas obligatorias en la práctica totalidad de los estándares de metadatos utilizados, por lo que tanto los creadores de páginas como los administradores de sitios deberían tomar medidas para garantizar que el 100% de las páginas de un sitio contengan siempre esta etiqueta.
6. Un porcentaje importante de las páginas analizadas contienen etiquetas como *author* (19,19%), *autor* (2,86%), *copyright* (11,19%), *dc.editor* (5,85%) *publisher* (5,15%) y algunas otras similares. En realidad, la etiqueta que debería ser obligatoria también en el 100% de las páginas es la de *creator* o *dc.creator*, que indica la unidad administrativa responsable de la información que se publica en la página. Hasta que no se aprueben normas, se recomiendan como buenas prácticas: utilizar tanto el nombre del máximo organismo responsable (un ministerio) como el

de la unidad administrativa encargada (por ejemplo, «Subdirección General de »); evitar el uso del copyright: la Ley 37/2007, de 16 de noviembre ya regula la reutilización de la información del sector público; y evitar el uso de traducciones como «autor» u otras posibles etiquetas como «editor» o «publisher». No estamos queriendo decir que no sean etiquetas válidas o poco importantes sino que, hasta que se aprueben las normas sobre metadatos, los esfuerzos sobre autoría deberían centrarse sobre la etiqueta «creator».

7. La solución adoptada por cada administración para publicar páginas en las distintas lenguas de nuestro país es muy variada, como ya se ha explicado. Casi el 40% de las páginas contienen etiquetas de *language* (34,15%) o de *dc.language* (4,49%) y muchas más contienen algún tipo de referencia a la página. La dispersión sobre los valores del elemento muestra, sin embargo, la dificultad de la normalización y de las dificultades con las que pueden encontrarse los motores de búsqueda para responder de forma adecuada a los usuarios. Nuestra recomendación es hacer de esta etiqueta *language* o *dc.language* una etiqueta obligatoria en todas las páginas. Y normalizar el valor que puede tomar la etiqueta de acuerdo a la norma ISO 639. Un proceso simple puesto que tan solo hay que recordar las lenguas oficiales en nuestra constitución y quizá cuatro o cinco códigos de lengua más.
8. De la misma forma, todos los documentos deberían contener al menos una fecha de documento. La importancia de que la fecha que quede como metadato sea la fecha de creación, de última modificación o cualquier otra fecha, puede variar respecto al organismo productor de la información. Creemos que como mejor práctica el valor de la etiqueta debería ser siempre el de la última actualización del documento, lo que facilitaría la creación automática de la etiqueta. Y debería utilizarse de acuerdo a la norma ISO 8601.
9. Las palabras claves son una de las etiquetas meta más utilizadas en el análisis realizado, presente como *keywords* en un 95'69% y como *dc.keywords* en un 5'88% de las páginas analizadas. El uso de esta etiqueta debería ser prioritario respecto a otras como otras detectadas en el análisis como *subject*, *category*, *classification*, etc. No queremos decir que este tipo de etiquetas no tengan validez o que tengan menos importancia que las anteriores, pero creemos que, hasta que no se aprueben directrices o normas sobre los esquemas o vocabularios a utilizar, son metadatos que requieren de consenso generalmente entre más de una unidad, incluso entre más de una administración o un sector social y/o económico. Y por tanto su utilidad es relativa en tanto en cuanto ese consenso no exista.
10. Existen algunas otras etiquetas que también son recomendables, como *description*, para explicar de forma breve el contenido de la página. Este elemento puede ser clave para los usuarios ya que suelen ser utilizados

por los motores de búsqueda para mostrar las dos primeras líneas que, junto al título, ofrecen a la persona que busca un resumen sobre lo que se puede encontrar al hacer click sobre un enlace de los resultados de un motor de búsqueda.

## 5. Bibliografía

- Acar, S.; Alonso, J. M., y Novak, K. (2009): *Improving Access to Government through Better Use of the Web: C Interest Group Note* [Página web]. 12/5/2009. <http://www.w3.org/TR/egov-improving/> [Consulta: 11 de junio de 2009].
- Alasem, A. (2009): An Overview of e-Government Metadata Standards and Initiatives based on Dublin Core. *Electronic Journal of e-Government*, vol. 7 (1, págs. 1-10. <http://www.ejeg.com/volume-7/vol7-iss1/Alasem.pdf> [Consulta: 2 de abril de 2009].
- Amutio Gómez, M. A. (2004): La construcción de los servicios pan-europeos de administración electrónica: de IDA a IDABC. En: *Tecnimap 2004. e-Cooperación en la Administración Pública*. Murcia: 28 de septiembre - 1 de octubre. [http://www.csi.map.es/csi/tecnicmap/tecnicmap\\_2004/comunicaciones/tema\\_05/5\\_001.pdf](http://www.csi.map.es/csi/tecnicmap/tecnicmap_2004/comunicaciones/tema_05/5_001.pdf) [Consulta: 2 de abril de 2009].
- Amutio Gómez, M. A. (2006): Los servicios paneuropeos de administración electrónica. *Revista de Internet, Derecho y Política* (2), págs. 4-22. <http://www.uoc.edu/idp/2/dt/esp/amutio.pdf> [Consulta: 2 de abril de 2009].
- Benamou, N., Busson, A., y Kavel, A. (2004): Impact of e-Government Interoperability in Local Governments. En: Traunmüller, R. (Ed.), *Electronic Government. Third International Conference, EGOV 2004 Zaragoza, Spain, August 30 - September 3, 2004 Proceedings*. Berlin: Springer, p. 82-87. <http://www.springerlink.com/content/lu46vtu-fbywhahjk> [Consulta: 2 de abril de 2009].
- Comisión Europea (2006): Comunicación de la Comisión al Consejo y al Parlamento Europeo. Interoperabilidad de los servicios paneuropeos de administración electrónica. COM(2006) 45 final. [http://www.csi.map.es/csi/pdf/COM\\_2006\\_45\\_final\\_Interoperabilidad\\_ES.pdf](http://www.csi.map.es/csi/pdf/COM_2006_45_final_Interoperabilidad_ES.pdf) [Consulta: 2 de abril de 2009].
- Kubicek, H.; Cimander, R. (2009): Three dimensions of organizational interoperability Insights from recent studies for improving interoperability frame-works. *European Journal of ePractic*, vol. 6. <http://www.epractice.eu/files/6.1.pdf> [Consulta: 2 de abril de 2009].
- Ministerio de Administraciones Públicas (2008): La Construcción de los servicios paneuropeos de Administración Electrónica. Estado de situación de la integración de los servicios paneuropeos de administración electrónica y actuación de la Administración. <http://www.csi.map.es/csi/pg3315.htm> [Consulta: 2 de abril de 2009].
- Peristeras, V.; Tarabanis, K.; Goudos, S. K. (2008): Model-driven eGovernment interoperability: A review of the state of the art. *Computer Standards & Interfaces*, vol. 31 (4), p. 613-628. doi:10.1016/j.csi.2008.09.034 [Consulta: 2 de abril de 2009].
- Tambouris, E.; Tarabanis, K.; Peristeras, V. y Liotas, N. (2007): *Study on Interoperability at Local and Regional Level. Final Version*. 20 April 2007. <http://ec.europa.eu/idabc/en/document/7038/254> [Consulta: 2 de abril de 2009].



Unión Europea (2004): Decisión 2004/387/CE del Parlamento europeo y del Consejo de 21 de abril de 2004 relativa a la prestación interoperable de servicios paneuropeos de administración electrónica al sector público, las empresas y los ciudadanos (IDABC). Diario Oficial de la Unión Europea, 30 de abril de 2004. <http://ec.europa.eu/idabc/servlets/Doc?id=1892> [Consulta: 2 de abril de 2009].