

Propuesta para el desarrollo de un sistema de metadatos web para la administración pública*

MANUEL BLÁZQUEZ OCHANDO
Universidad Complutense de Madrid, España

INTRODUCCIÓN

El desarrollo y uso de los metadatos en el sector de la administración pública está orientado a la gestión de información, a su tratamiento estadístico, a la organización de servicios, a la preservación de la documentación, a la gestión del acceso y a la privacidad de la información, así como a la identificación y calificación de los fondos.¹ Estas necesidades plantean el diseño de sistemas de metadatos más evolucionados, para adaptarse a la casuística de cada administración pública, según el corte y tradición en la gestión documental.² El objeto de esta investigación es analizar esquemas de metadatos, compararlos, determinar sus ventajas e inconvenientes para describir y representar la documentación de la administración pública, con el objetivo

* Agradecimiento a la Dirección General de Asuntos del Personal Académico (DGAPA, UNAM), proyecto “Buenas prácticas en el acceso a la información gubernamental” (PAPIIT IN 403113).

de argumentar y proponer un nuevo desarrollo de metadatos mejorado.

ANÁLISIS DE METADATOS AGLS, E-EMGDE Y DC QUALIFIED

Se han seleccionado los metadatos Australian Government Locator Service (AGLS),³ e-EMGDE⁴ y Dublin Core Qualified, por considerarse los más importantes en el ámbito de la descripción documental de la e-administración pública y la descripción de recursos Web. El sistema de metadatos AGLS, Australian Government Locator Service, fue diseñado con el propósito de establecer un método normalizado de descripción para la documentación de la administración pública australiana que permitiera mejorar la visibilidad, disponibilidad e interoperabilidad de la información gubernamental y sus servicios. Para ello, los metadatos AGLS implementan aspectos de la Web Semántica,⁵ con los que se identifica el objeto a describir a través del *dominio y rango*, detallando sus características y rasgos por medio de las *propiedades asociadas* al configurar estructuras de triples semánticas a partir de poder realizar, teóricamente, inferencias semánticas. Durante su fase de diseño, los metadatos AGLS implementaron 46 de los 55 metadatos DC Qualified, con lo cual se representó casi la mitad del número total de metadatos de la norma.⁶

El sistema de metadatos e-EMGDE, Esquema de Metadatos para la Gestión del Documento Electrónico, fue desarrollado como instrumento de descripción normalizado, para la documentación de la administración pública española, que posibilitara la gestión de documentos electrónicos y su interoperabilidad entre sistemas de información. El objetivo del esquema de metadatos e-EMGDE es la descripción del do-

cumento electrónico durante su ciclo vital, proporcionando una perspectiva archivística de la Documentación al encontrar, de esta forma, similitudes con las áreas de descripción de las normas ISAD-G. De hecho el esquema de metadatos está basado en una estructura jerárquica de elementos y subelementos que identifican las propiedades y características de los objetos a describir. Los metadatos e-EMGDE se basan en el modelo Australian Government Recordkeeping Metadata Standard (AGRkMS) australiano del que toman metadatos de seguridad, derechos, datos de contacto, idioma, palabras clave, calificación del documento, características técnicas, ubicación y prioridad. En relación con los metadatos AGLS, incluye la jurisdicción, los metadatos de seguridad, las características técnicas, y los derechos de acceso y uso.

Tanto los metadatos AGLS como e-EMGD están diseñados para tipificar entidades, documentos, agentes, actividades, regulaciones en general y relaciones, quedando perfectamente definido en su documentación técnica. Por ejemplo, en e-EMGD, las agrupaciones documentales (grupo de fondos, fondo, serie, agregación, expediente, documento simple), no sólo son compatibles con la documentación de la unidad administrativa, sino también y más importante, con el sistema archivístico español, lo que mejora la gestión y tratamiento de la información durante su ciclo vital. Lo mismo sucede con la tipificación de los agentes (instituciones, sistemas de información, personas físicas, personas jurídicas) que, a modo de autoridades, se identifican como participantes en la edición del documento, hasta su finalización. La identificación de los tipos de actividad, aun siendo menos precisa por no aportar valores específicos, sí permite distinguir entre función marco, función, actividad y acción para determinar, finalmente, una actividad concreta. Este método

permite diferenciar, en las normas e-EMGD, qué agente tiene una función y bajo qué regulación lleva a cabo su actividad.

Observando la *Tabla 1*, comparativa de los metadatos AGLS, e-EMGD y DC Qualified, pueden apreciarse algunas diferencias cuantitativas. Por ejemplo, el número total de metadatos es ligeramente superior en los metadatos e-EMGD a las AGLS, y se superan o alcanzan, en ambos casos, los 100 metadatos. DC Qualified, en su extensión terminológica DCTERMS, alcanza los 55 metadatos, lo que supone la mitad de los campos de descripción del sistema AGLS y e-EMGD. Por otra parte, ningún sistema de metadatos tiene especificado en su normativa de uso el análisis multinivel, según el grado de exhaustividad. Esto significa que su aplicación dependerá de la obligatoriedad de los metadatos y las políticas de aplicación de las instituciones, sin que exista un método que normalice la extensión de las descripciones y, por tanto, una recomendación en el método de descripción. Otro aspecto reseñable es la interoperabilidad nacional e internacional de los metadatos. Por ejemplo DC Qualified puede ser aplicado para una descripción de recursos Web o documentación administrativa, en sentido general, en cualquier país del mundo. Pero AGLS y e-EMGD son esquemas de metadatos específicos para un país, una normativa y una casuística administrativa muy concreta, acotando, con ello, sus límites de aplicación. Esto resta interoperabilidad internacional para ser compartidos en administraciones de nivel superior. Si bien AGLS puede ser implementada por cualquier país de la Commonwealth, el esquema e-EMGD está escrito íntegramente en español y no podría ser compartido por la Unión Europea, con lo que se reduce su ámbito de aplicación a los países hispanohablantes.

Propuesta para el desarrollo de un sistema de metadatos web...

Tabla 1
Comparativa de metadatos AGLS, e-EMGDE y DC Qualified

	AGLS	e-EMGDE	DC Qualified
Número total de metadatos	100	117	55
Análisis multinivel según grado de exhaustividad	No	No	No
Análisis según obligatoriedad	Sí	Sí	Sí
Interoperabilidad internacional	No	No	Sí
Interoperabilidad nacional	Sí	Sí	Sí
Tipos de contenidos descritos	Documentos, Autoridades institucionales, Sujetos productores	Documentos, Agentes, Actividades, Regulaciones, Relaciones	Documentos, Recursos Web
Basados en DC Qualified	Sí	No	-
Basados en AGRkMS	-	Sí	-
Basadas en ISAD-G	No	Sí	-
Metadatos obligatorios	3	16	-
Metadatos condicionales	3	59	-
Metadatos opcionales	89	11	-
Metadatos recomendados	5	-	-
Metadatos de identificación	Códigos de referencia, título, títulos atribuidos, fechas, niveles de descripción, agrupaciones documentales, identificación de fondos y expedientes, características físicas de tamaño, dimensiones, volumen		
	20	16	17
Metadatos de descripción física	Características físicas de tamaño, dimensiones, volumen de la unidad de descripción		
	3	7	3
Metadatos de contexto	Instituciones relacionadas con el desarrollo de la tramitación de los documentos de la administración, participantes, historia institucional, reseña biográfica, historia archivística, forma de ingreso, adquisición		
	3	0	7

El acceso a la información gubernamental...

Tabla 1 (cont.)

	AGLS	e-EMGDE	DC Qualified
Metadatos de autoridades	Descripción de personas físicas, jurídicas, entidades públicas, privadas, familias, organizaciones gubernamentales, no gubernamentales		
	17	4	2
Metadatos de contenido y estructura	Alcance, contenido, valoración o valores del documento (histórico, legal, administrativo, probatorio), selección de materiales y documentos, calendarios de transferencias, historia de la tramitación, ingreso, transferencia, trazabilidad		
	2	11	5
Metadatos de trazabilidad	Gestión interinstitucional de los documentos, tramitación, jurisdicción, bases legales del proceso administrativo, seguridad		
	14	13	0
Metadatos de condiciones de acceso y uso	Condiciones de acceso, uso, reproducción, idioma, lengua del documento, requisitos técnicos de reproducción, instrumentos de descripción, conservación		
	11	29	8
Metadatos de documentación asociada	Versiones de la documentación, documentación relacionada, unidades de descripción relacionadas, relación con documentación original		
	34	0	13
Metadatos de control	Códigos de control, referencias de identificación unívoca del recurso, documento, entidad o autoridad		
	1	4	1
Metadatos bibliográficos	Descripción de documentación bibliográfica, título, mención de responsabilidad, edición, publicación, serie, notas, referenciación bibliográfica, cita bibliográfica		
	3	0	3

Sobre la cuantificación de los metadatos según el sector o ámbito de aplicación, pueden clasificarse en parte, según áreas de descripción de las ISAD-G, tales como *identificación, contexto, contenido y estructura, condiciones de acceso y uso, documentación asociada, control*; y según áreas de descripción más precisas, se analizan los metadatos dedi-

cados a *características físicas, autoridades, trazabilidad y metadatos bibliográficos*. Según se detalla en la *Figura 1*, el número de metadatos de identificación resulta similar tanto en AGLS, como en DC Qualified y e-EMGDE. Las diferencias se encuentran en apartados como documentación asociada, en la que AGLS consta de una gran capacidad de interrelación entre recursos, debido a sus principios semánticos con los que se configura un total de 34 metadatos especializados. Sin embargo, los metadatos e-EMGDE ofrecen mejor cobertura en el apartado de condiciones de acceso y uso, conteniendo 29 metadatos especializados. Ello indica que prima la declaración de propiedades de reproducción, acceso, seguridad y casuística legal de protección del documento. Se contabilizan pocos metadatos de tipo bibliográfico y control, tanto en AGLS como en e-EMGDE, lo que significa que existe una reducida capacidad para abordar el tratamiento de tipos documentales, no administrativos. Tanto AGLS como e-EMGDE son esquemas de metadatos diseñados para utilizar los metadatos de identificación de forma polivalente, tanto para describir documentos como instituciones, personas, agentes y actividades, por lo que los campos de autoridades, contexto y descripción física, en muchos casos, están altamente integrados en los grupos de identificación, contenido y estructura.

Analizando con más detenimiento los metadatos AGLS, cabe destacar el empleo de prefijos denotativos de los metadatos y su contexto de aplicación. El metadato genérico viene precedido de la denominación AGLSTERMS;⁷ el uso del prefijo AVAILTERMS declara la disponibilidad de la documentación y sus factores de acceso y uso; ADMINTERMS determina los datos del operador o gestor de la documentación, y finalmente, AGENTTERMS identifica las autoridades o agentes de la tramitación del documento. Esta clasificación interna

de los metadatos ayuda a distinguir su empleo, codificación manual y automática en sistemas de información, así como su recuperación e interpretación. Obsérvese que, a pesar de que cada metadato sugiere conceptos realmente complejos, la extensión del término resulta muy reducida en comparación con los metadatos e-EMGDE. De esta forma, aun mezclándose los metadatos, su uso y empleo es perfectamente distinguible a simple vista.

Los metadatos e-EMGDE se caracterizan por estar escritos en español y disponer de denominaciones más elaboradas, extensas y complejas que dependen del esquema jerárquico de elementos y subelementos que, por ejemplo, no rige el funcionamiento de las AGLS. De hecho, existen metadatos que superan los 50 caracteres de extensión, lo que supone una dificultad añadida, cuando se pretende codificar con el esquema e-EMGDE frente a denominaciones más sencillas. A pesar de todo, e-EMGDE completa en mayor medida el apartado de trazabilidad de la documentación, que supone el control de la normativa incoativa, la jurisdicción de las entidades y agentes, y el traslado e incluso la transferencia de la documentación entre instituciones, lo que favorece la gestión archivística de la documentación digital.

Desde el punto de vista teórico, los metadatos AGLS y e-EMGDE permitirían la descripción de la documentación administrativa, pero desde un punto de vista pragmático, se plantean diversos interrogantes. ¿Resulta sencillo identificar el tipo de objeto, elemento o agente que hipotéticamente se está describiendo con cada esquema de metadatos? En e-EMGDE, está comprobado que los valores ya están predefinidos, pero no es así en el caso de AGLS. ¿Es posible codificar de manera sencilla y manualmente los metadatos? En e-EMGDE resulta difícil reseñar cada metadato; por ello, su puesta en práctica resulta complicada. ¿Cómo saber que cada metadato

se refiere exactamente al documento y no al agente que lo está gestionando? En ambos casos, ello requiere la selección de los metadatos y su correcta combinación, que dependerá del centro catalogador que los aplique, sin que exista una ficha predeterminada para cada supuesto. ¿Cómo se podrían señalar las distintas autoridades, objetos y elementos que figuran en el contenido de los expedientes? En los metadatos e-EMGDE, pueden emplearse los metadatos de entidad relacionada y puntos de acceso, pero no termina de quedar claro si éstos se refieren al sujeto productor o la entidad de destino. En AGLS, también queda patente el mismo problema ya que, al igual que los metadatos e-EMGDE, depende de la existencia de un registro que catalogue cada autoridad y éste sea vinculado por los metadatos del registro principal del documento o expediente administrativo. Ello permite formular otra pregunta, ¿la descripción de la trazabilidad de los trámites administrativos con un expediente debe quedar reflejada en las descripciones de varios metadatos, o bien, a través de datos vinculados o linked data? El modelo de descripción híbrido permite un uso dual, describiendo autoridades y a la vez vinculándolas a su ficha meta-informativa, lo cual plantea un problema de normalización y probablemente de duplicación. ¿Podría simplificarse el modelo de metadatos para la administración pública y mantener el nivel de descripción, mejorando el orden y comprensión de la trazabilidad?

PROPUESTA DE METADATOS WEB PARA LA ADMINISTRACIÓN PÚBLICA

A partir de estas reflexiones, se plantea una propuesta de metadatos web para la administración pública, que implica:

El acceso a la información gubernamental...

1) determinar qué documentos, autoridades, sujetos productores, agentes, funciones, servicios, relaciones y actividades deben ser descritas; 2) qué nivel de exhaustividad deberá tener la descripción: elaborar metadatos adecuados al nivel de profundidad, que faciliten la interoperabilidad entre organismos e instituciones nacionales e internacionales; 3) diseñar programas de edición y codificación asistida para la descripción y análisis de la documentación administrativa; esto es, el desarrollo de editores y sistemas de publicación de metadatos; 4) diseñar los sistemas de recuperación de información para aprovechar el sistema de metadatos propuesto.

Qué elementos y objetos deben ser descritos

Coincidiendo con los esquemas de metadatos AGLS y e-EMGDE, es necesario establecer una serie de elementos, objetos y agentes que claramente estén definidos y puedan ser descritos con garantías y de acuerdo a un método normalizado. Se propone que el nuevo sistema de metadatos incluya prefijos para identificar a qué elemento se le aplica la descripción correspondiente. De esta forma es posible determinar que la descripción corresponde a una agrupación documental, a una actividad o a varios tipos de elementos en la misma descripción (*Tabla 2*).

Tabla 2

Identificación principal de la documentación, tipos de agrupaciones documentales, agentes, actividades, marcos legales y trámites administrativos

Prefijo metadato	Denominación	Valores
DOCGROUP.	Agrupación documental	Fondo, subfondo, serie, subserie, sección, subsección, expediente simple, expediente compuesto, caja, legajo...
<code><meta name='DOCGROUP:identifier' type='Expediente simple' code='Código normalizado del expediente' title='Título del expediente' value='http://www.direccion-uri-del-recurso.es/'/></code>		

Tabla 2 (cont.)

Prefijo metadato	Denominación	Valores
AGENT.	Agente	Institución pública, Institución privada, persona física, persona jurídica, órgano, departamento, negociado, dispositivo, sistema de información, programa informático, base de datos, flujo de trabajo...
<pre><meta name='AGENT.identifier' type='Institución pública' code='Código normalizado de la institución pública' title='Nombre o de la institución pública' value='http://www.direccion-uri-del-agente.es/'/></pre>		
ACTIVITY.	Actividad	Meta, función, actividad, acción, proceso...
<pre><meta name='ACTIVITY.identifier' type='Actividad' code='Código normalizado de la actividad' title='Denominación o descripción breve de la actividad' value='http://www.direccion-uri-de-actividad.es/'/></pre>		
FRAMEWORK.	Marco regulatorio	Constitución, Real Decreto Ley, Real decreto Legislativo, Ley Autonómica, Ley orgánica, Ley ordinaria, Ley marco, Estatuto, Convenio, Tratado Internacional, Contrato...
<pre><meta name='FRAMEWORK.identifier' type='Estatuto' code='Código normalizado del estatuto' title='Título o denominación del estatuto' value='http://www.direccion-uri-oficial-del-estatuto.es/'/></pre>		
TRACE.	Trazabilidad	Incoación, edición, transferencia, expurgo, copia, firma, verificación, supervisión, sellado, compulsado...
<pre><meta name='TRACE.identifier' type='Incoación' activitycode='Código de la actividad' agentcode='Código del agente que lleva a cabo la actividad' title='Denominación del trazo o prueba del proceso realizado sobre el documento, agente u objeto afectado por la actividad' date='fecha normalizada del cambio efectuado en el documento' value='http://www.direccion-uri-de-la-version-del-documento.pdf/'/></pre>		

La identificación del elemento sujeto a descripción se lleva a cabo con el metadato *identifier*. Además, en la codificación se establecen atributos obligatorios para completar la identificación. El atributo *type* especifica el tipo de agrupación documental (DOCGROUP.identifier), agente (AGENT.identifier), actividad (ACTIVITY.identifier), marco regulatorio (FRAMEWORK.identifier) y trazo (TRACE.identifier) que se está identificando. El atributo *code* determina el código normalizado para el elemento que se describe y que actúa como

identificador unívoco del recurso. El atributo *title* contiene la denominación o título del elemento sometido a descripción. El atributo *value* permite aportar la dirección URI del documento, agrupación documental, agente, actividad, marco regulatorio y trazo o prueba del proceso, y sirve, al igual que el atributo *code*, de método de identificación unívoco susceptible de ser utilizado a modo de enlace de datos o linked data. Se establecen atributos especiales, como *activitycode* y *agentcode* para especificar el código de actividad y el código de agente implicados en la trazabilidad o tramitación de una agrupación documental, en concreto de un documento. De esta forma, es posible repetir el metadato *TRACE.identifier* tantas veces como cambios o trámites sufra el expediente administrativo o el documento que se esté analizando, con lo cual quedan reflejadas las actividades, agentes principales y fechas mediante el atributo *date*.

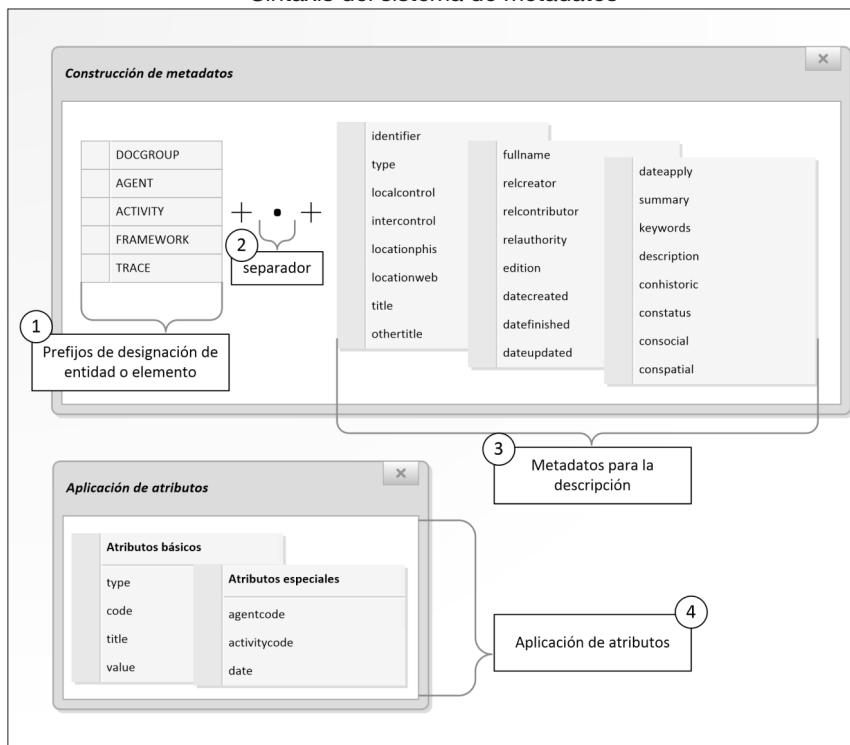
Qué nivel de exhaustividad deberá tener la descripción

El nivel de exhaustividad no siempre es el mismo en todas las unidades administrativas y, por ello, lo normal es permitir cierta flexibilidad en el uso de los metadatos. No obstante, la identificación de los elementos, expuesta en la *Tabla 4*, debería ser obligatoria, ya que se considera la identificación básica de los recursos de la administración. Por tanto, un nivel superior de descripción es la introducción de los metadatos de la *Tabla 5*. En ésta, se exponen 48 metadatos diseñados para combinarse con los prefijos DOCGROUP, AGENT, ACTIVITY, FRAMEWORK y TRACE, de lo que un total de 240 posibilidades de combinación para la descripción (véase método sintáctico de combinación en *Figura 1*). Esta cifra supera el número total de metadatos diseñados en los

esquemas AGLS y e-EMGDE, permitiendo al catalogador elegir los más adecuados para describir cualquier supuesto.

Figura 1

Sintaxis del sistema de metadatos



A la vez, se observa cómo las áreas de descripción, en menor medida tratadas en AGLS e-EMGDE, han sido equilibradas, con lo que se favorece una mayor polivalencia analítica; por ejemplo, los metadatos *localcontrol* e *intercontrol*, que permiten la introducción de códigos de control nacional e internacional. En el ámbito de la identificación de las menciones de responsabilidad y el origen del documento,

El acceso a la información gubernamental...

los metadatos *relauthority*, *relcreator*, *relcontributor*, están destinados a la relación de los datos vinculados, vía linked data, de la autoridad, sujeto productor y colaborador, distinguiéndolos de un simple punto de acceso. La introducción de metadatos de contexto permitirá adaptar mejor la descripción, no sólo al ámbito de los archivos de oficina, sino también históricos. Con respecto al esquema e-EMGDE, se han reducido los metadatos de acceso y uso, obligando al catalogador a referir toda la información en los metadatos *rights*, *accessconditions*, *security*, *signature* y *classaccess*. Esto evita la dispersión de la información y ayuda a controlar con menos campos la información principal de acceso y sus restricciones clave (*Tabla 3*).

Tabla 3
Metadatos para su combinación con los prefijos especificados

Metadato	Atributos	Denominación	Descripción
identifier	type='valor del prefijo', code='código de identificación', title='denominación de la identificación', value='identificador url unívoco', activitycode='código de actividad implicada', agentcode='código de agente implicado'	Identificación del sujeto/objeto	Identificación unívoca del sujeto, objeto que se describirá en el documento electrónico. Implicará el empleo de atributos type, activitycode, agentcode, title, date y value.
type	scheme='esquema del tipo de elemento', value='tipo de elemento definido por el catalogador'	Tipo de elemento	Se define el tipo de sujeto y objeto de acuerdo a los parámetros normalizados o según las necesidades del catalogador. De esta forma el metadato type puede registrar tipos de autoridad no recogidos en las normas y que sin embargo pueden figurar en los contenidos de la documentación administrativa.
localcontrol	scheme='esquema del código de control local', value='código de control local'	Código de control nacional	Código de control para referirse al elemento sujeto a descripción, dentro de las fronteras y jurisdicción de un país

Propuesta para el desarrollo de un sistema de metadatos web...

Tabla 3 (cont.)

Metadato	Atributos	Denominación	Descripción
intercontrol	scheme='esquema del código de control internacional', value='código de control internacional'	Código de control internacional	Código de control para identificar el elemento sujeto a descripción fuera de la jurisdicción nacional
locationphis	value='localización física'	Localización física	Datos que identifican la localización física del elemento sujeto a descripción
locationweb	value='url'	Localización electrónica	Localización electrónica por medio de URI o URL permalink del elemento u objeto sometido a descripción
title	value='título'	Título	Título o denominación propiamente dicho que corresponda con la entrada principal del documento, agente, marco legal, actividad, trazo o prueba
othertitle	value='otro título'	Otras formas del título	Otros títulos o denominaciones, subtítulo, título paralelo
fullname	value='apellidos, nombre'	Nombre completo	Nombre completo de una persona, manteniendo el orden alfanumérico según (apellidos, nombre)
relcreator	code='código de agente' value='url del agente'	Sujeto productor o autor	Entidad o persona responsable de la creación del documento, también denominado sujeto productor. Diseñado para establecer una relación vía linked data con el código de agente correspondiente.
relcontributor	code='código de agente' value='url del agente'	Colaboradores	Entidades o personas colaboradoras o coautoras del documento o agrupación documental, vinculados estrechamente al sujeto productor. Diseñado para establecer una relación vía linked data con el código de agente correspondiente.
relauthority	code='código de agente' value='url del agente'	Autoridades	Autoridades que están reseñadas o citadas en el contenido del documento administrativo y que sin embargo no pueden ser consideradas autoras, productoras o colaboradoras. Estas normalmente pueden ser personas, entidades, familias o lugares. Permite generar puntos de acceso añadidos. Diseñado para establecer una relación vía linked data con el código de agente correspondiente.

El acceso a la información gubernamental...

Tabla 3 (cont.)

Metadato	Atributos	Denominación	Descripción
edition	value='datos de la edición'	Edición	Información de la edición, datos de mención de edición, características especiales, etcétera.
datecreated	value='fecha'	Fecha de creación	Fecha de creación o publicación. Puede ser utilizado como la fecha extrema de inicio.
datefinished	value='fecha'	Fecha de finalización	Fecha de finalización o cese. Puede ser utilizado como la fecha extrema final.
dateupdated	value='fecha'	Fecha de actualización	Última fecha de modificación del objeto o sujeto sometido a descripción.
dateapply	value='fecha'	Fecha de los derechos	Fecha de aplicación de una norma, de los derechos de un documento, de funcionamiento de una institución o agente. Depende del objeto o elemento identificado en la descripción.
summary	value='resumen'	Resumen/sumario	Resumen, índice o sumario breve del contenido de la agrupación documental, agente, marco legal, actividad.
keywords	value='palabras clave separadas por comas'	Palabras clave	Descriptores o palabras clave preeminentes para la recuperación del objeto o elemento descrito.
description	value='descripción del contenido'	Descripción	Descripción del contenido y estructura del objeto, elemento, sujeto identificado.
conhistoric	value='definición del contexto histórico'	Contexto histórico	Contexto histórico del documento, cuando es necesario realizar algún tipo de valoración.
constatus	value='definición del contexto jurídico y situación o estatus'	Contexto de situación	Contexto de la situación jurídica del documento.
consocial	value='definición del contexto social'	Contexto social	Contexto social en el que se publica el documento, declaración de intenciones y motivaciones sociales del mismo.
conspatial	value='definición del contexto espacial'	Contexto espacial	Descripción del contexto espacial, cuando sea necesario declarar detalles específicos relativos a direcciones, calles, edificios singulares o ubicaciones topográficas.

Propuesta para el desarrollo de un sistema de metadatos web...

Tabla 3 (cont.)

Metadato	Atributos	Denominación	Descripción
contemporal	value='definición del contexto temporal'	Contexto temporal	Descripción exhaustiva de periodos, fechas señaladas en la entidad.
phisholder	value='denominación del soporte del documento'	Soporte del documento	En el caso de entidades documentales se refiere al soporte del documento y su material.
phisextent	value='extensión del documento'	Extensión	Extensión de la entidad documental.
phisdetails	value='detalles de las características físicas del documento'	Detalles físicos	Detalles físicos de la entidad documental. Por ejemplo, ilustraciones, gráficos, tablas estadísticas, grabados, etc.
phisdimensions	value='dimensiones del documento'	Dimensiones	Dimensiones de las entidades documentales descritas.
phisenclosed	value='materiales anejos o adjuntos, empaquetados o embebidos'	Material adjunto	Identificación, volumen y descripción de los materiales adjuntos o anejos a las entidades documentales descritas.
jurisdiction	scheme='esquema de jurisdicciones', value='denominación de la jurisdicción'	Jurisdicción	Jurisdicción o ámbito de aplicación de un marco normativo que se está describiendo. También aplicable a la jurisdicción de agentes, actividades y otros elementos descritos.
valuables	scheme='esquema de valores del documento', value='definición de los valores del documento'	Valores del documento	Especificación de los valores del documento: administrativo, jurídico-legal, probatorio, científico-tecnológico, histórico-testimonial, informativo.
relframework	code='código del marco normativo relacionado', value='url del marco normativo relacionado'	Marco normativo relacionado	Marco jurídico relacionado del que deriva, procede o forma parte el sujeto o elemento que se está describiendo. Diseñado para establecer una relación vía linked data con el código de agente correspondiente.
accessconditions	value='descripción de las condiciones de acceso'	Condiciones de acceso	Todos los aspectos relativos a las condiciones de acceso, legislación y protección legal de la información.
rights	value='descripción de los derechos aplicados'	Derechos	Derechos de propiedad intelectual, explotación y casuística aplicada al documento.
language	scheme='esquema de idioma', value='idioma o lengua de la entidad'	Idioma	Idioma o lengua de la entidad.

El acceso a la información gubernamental...

Tabla 3 (cont.)

Metadato	Atributos	Denominación	Descripción
signature	scheme='tipo de firma digital', value='firma digital'	Firma digital	Firma digital empleada para dar fe pública de la veracidad de una entidad documental.
security	value='descripción de seguridad'	Seguridad	Descripción de la política de seguridad de la entidad
class thematic	scheme='thesaurus tags udc dewey natlang', code='código de tesauro, clasificación decimal', value='denominación temática'	Clasificación temática	Clasificación temática del elemento u objeto que se está describiendo de acuerdo a un esquema normalizado.
class access	scheme='esquema de clasificaciones de acceso aplicado al caso', value='valor de la clasificación de acceso'	Clasificación de acceso	Clasificación de acceso o clasificación de la restricción del elemento u objeto descrito. Por ejemplo, libre acceso, público, oficial, secreto, máximo secreto, sensible, etcétera.
notes	value='anotaciones y textos aclarativos'	Notas	Notas y reseñas añadidas a la descripción de la entidad que está siendo descrita.
rel version prev	code='código del elemento relacionado', value='direccion url del elemento relacionado'	Relación con la versión anterior	Relación con una versión anterior del documento. Diseñado para establecer una relación vía linked data con el código de agente correspondiente.
rel version next	code='código del elemento relacionado', value='direccion url del elemento relacionado'	Relación con la nueva versión	Relación con una versión más moderna del documento. Diseñado para establecer una relación vía linked data con el código de agente correspondiente.
rel hier asc	type='functional organizational documentary', code='código del elemento relacionado', value='direccion url del elemento relacionado'	Relación jerárquica superior	Relación con una entidad, elemento, objeto superior dentro de una escala jerárquica. Pensado para la generación de cuadros de clasificación orgánica y funcional. Diseñado para establecer una relación vía linked data con el código de agente correspondiente.
rel hier desc	type='functional organizational documentary', code='código del elemento relacionado', value='direccion url del elemento relacionado'	Relación jerárquica inferior	Relación con una entidad, elemento, objeto inferior dentro de una escala jerárquica. Pensado para la generación de cuadros de clasificación orgánica y funcional. Diseñado para establecer una relación vía linked data con el código de agente correspondiente.

Tabla 3 (cont.)

Metadato	Atributos	Denominación	Descripción
reldocument	scheme='esquema del tipo de relaciones', type='tipo de relación', code='código del elemento relacionado', value='direccion url del elemento relacionado'	Documento relacionado	Documento relacionado, necesario para la completar o extender la significación de la entidad. Diseñado para establecer una relación vía linked data con el código de agente correspondiente.
relcopy	scheme='esquema del tipo de copias', type='tipo de copia', code='código del elemento relacionado', value='direccion url del elemento relacionado'	Copia relacionada	Documento copia del original, borrador, copia compulsada, fotocopia. Diseñado para establecer una relación vía linked data con el código de agente correspondiente.
relsource	code='código del elemento relacionado', value='direccion url del elemento relacionado'	Fuente de información relacionada	Entidad o documento considerada fuente de información relacionada. Diseñado para establecer una relación vía linked data con el código de agente correspondiente.

Finalmente, se aporta una visión distinta en relación al control de versiones y jerarquías organizativas de la información. Con los metadatos *relversionprev* y *relversionnext* se pueden relacionar la siguiente versión o la anterior al documento descrito. De la misma manera, con los metadatos *relhierasc* y *relhierdesc* puede definirse si un recurso, elemento o entidad se encuentra en relación superior o inferior jerárquica, desde el punto de vista del cuadro funcional, orgánico o documental. Esto hace posible la creación de estructuras y cuadros de clasificación necesarios para la organización administrativa y archivística.

Integración del método de edición

Un fallo habitual en el desarrollo de esquemas de metadatos es precisamente no desarrollar editores que permitan la codificación automática de los mismos. Es muy recomendable integrar distintos editores, según los tipos de entidad y

El acceso a la información gubernamental...

elementos disponibles.⁸ Esto significa que el editor deberá contemplar distintos esquemas de descripción para cada tipo de agrupación documental, agente, marco normativo y actividad, con los metadatos predefinidos, considerados óptimos para una descripción exhaustiva. Añadido a esto, los esquemas de descripción del editor deberán variar, dependiendo de la institución o centro catalogador, obteniendo un total de 17 esquemas de meta-descripción posibles (*Tabla 4*).

Tabla 4
Edición de metadatos según centro catalogador y entidad descrita

Institución o centro catalogador	Elemento o entidad descrita	
Unidad administrativa	Fondo / subfondo	Persona física
Archivo de oficina	Serie / subserie	Persona jurídica
Archivo central	Sección / subsección	Familia
Archivo histórico	Expediente simple / compuesto	Marco legal / regulatorio
Biblioteca	Institución pública	Actividad
Centro de documentación	Institución privada	

Integración del sistema de recuperación

Junto con los editores y los esquemas de meta-descripción adaptados para cada caso, se requiere de un sistema de recuperación oficial, que posibilite la recuperación de información de los contenidos codificados con los metadatos. El método de recuperación debe ser fiable y sencillo, tanto de desarrollar como de implementar y utilizar por parte de los destinatarios finales. En este caso, se propone la recuperación mediante web-crawler, ya que las estructuras generadas por los metadatos son lo suficientemente identificativas.

De no existir un sistema de recuperación, de nada sirve el esquema de metadatos, el esfuerzo de descripción de los contenidos, elaborados por el centro catalogador y, por ende, no terminará por hacerse transparente la información de la administración pública, al no existir un buscador o un medio de difusión integrado con el sistema de información de las unidades administrativas y los órganos de decisión gubernamental, ministerios, direcciones generales, secretarías generales, etcétera.

RECUPERACIÓN MEDIANTE WEB-CRAWLER

La recuperación por medio de web-crawler implica el diseño de un sistema de filtrado y extracción de metadatos del código fuente de cada página web. Este supuesto obliga al desarrollo del sistema de recuperación en las siguientes fases.

Registro de las fuentes de información de la administración pública

El registro de fuentes de información de la administración pública consiste en una lista o semilla de enlaces de direcciones URL, en las que se publican los contenidos con los metadatos propuestos. Conocer el origen de la publicación de la información permite mejorar la efectividad del proceso de análisis del web-crawler que se dirige exclusivamente a las fuentes indicadas y, a la vez, establecer un control de las mismas.

Diseño del código de extracción y filtrado de metadatos

Un web-crawler puede analizar las páginas web de las direcciones URL indicadas en la lista de fuentes de información de la administración pública, descargando su código fuente e identificando los metadatos mediante la técnica de filtrado XPath (XML Path Language) (*Tabla 5*).

Tabla 5
Método XPath de recuperación de metadatos

```
$metadata01 = $xpath1->query("meta[@name='DOCGROUP.identifier']");  
for($i=0; $i<$metadata01->length; ++$i) {  
  $item_metadata01 = $metadata01->item($i);  
  $array_metadata01type[] = utf8_decode($item_metadata01->getAttribute('type'));  
  $array_metadata01code[] = utf8_decode($item_metadata01->getAttribute('code'));  
  $array_metadata01title[] = utf8_decode($item_metadata01->getAttribute('title'));  
  $array_metadata01value[] = utf8_decode($item_metadata01->getAttribute('value'));  
}
```

Se observa que los metadatos pueden ser identificados fácilmente mediante la consulta XPath `meta[@name='DOCGROUP.identifier']` indicando que el nombre del metadato se encuentra en el atributo `name`. El resto de atributos del metadato se obtienen mediante la función `getAttribute` de DOM, extrayendo en tal caso el tipo, código, título y valor del metadato. Este método permite revisar las 240 posibilidades de combinación de etiquetas y construir un registro completo de la descripción de la entidad descrita.

Almacenamiento en base de datos

El almacenamiento del registro completo de la descripción de la entidad debería realizarse en base de datos Oracle, MySQL, PostgreSQL o similar, para garantizar un acceso a datos rápido y efectivo. También tiene la ventaja de favore-

cer el proceso de indexación y recuperación a texto completo, facilitando el desarrollo de servicios de consulta.

FUTUROS DESARROLLOS

Planteadas las bases para el desarrollo de un esquema de metadatos para la administración pública, se propone su desarrollo completo. Ello conlleva la publicación de la normativa, sintaxis, vocabulario y metodología de aplicación, basadas en las recomendaciones de la presente investigación. Por otra parte, se propone el diseño y desarrollo de los editores y sistema de recuperación integral para los metadatos propuestos, de forma tal que pueda ser aprovechado por las administraciones públicas en diversos países, fomentando la reutilización de la información pública y la transparencia informativa de los gobiernos.

CONCLUSIONES

El esquema de metadatos e-EMGDE permite diferenciar mejor que el esquema AGLS, los distintos agentes, agrupaciones documentales, relaciones y actividades de la administración pública, al contemplarlos en la tipificación de entidades sujetas a descripción. El esquema de interoperabilidad de e-EMGDE está adaptado a la casuística administrativa española, y no lo está para su compatibilidad europea e internacional. Ello es debido a que los metadatos están escritos en español y no en inglés, lo que dificulta su internacionalización. Los metadatos AGLS resultan más sencillos de identificar y codificar al emplear prefijos que delimitan las áreas de aplicación de los mismos. En el esquema de metadatos e-EMGDE,

los metadatos se construyen a partir de palabras y sintagmas completos que aumentan notablemente la extensión de los metadatos, llegando a superar los 50 caracteres y dificultando su codificación y uso manual. Los metadatos AGLS y e-EMGDE no disponen de editores de metadatos públicos y oficiales, lo que dificulta su implementación y tampoco cuentan con sistemas de recuperación normalizados para crear buscadores y directorios automatizados para aprovechar las descripciones efectuadas con ambos modelos. El esquema e-EMGDE destaca por presentar más metadatos dedicados a la descripción física y a las condiciones de acceso y uso. Sin embargo, carece de metadatos bibliográficos y de contexto. Por su parte, el esquema AGLS destaca en metadatos de documentación asociada para establecer relaciones vía linked data y en autoridades. La cantidad de metadatos destinados a la identificación de las entidades y agrupaciones documentales es similar, y no baja de 15 metadatos disponibles. Finalmente, es reseñable que el esquema de AGLS está en gran medida determinado por los metadatos DC Qualified, que ocupan cerca del 50% del esquema. Se propone que los elementos y objetos que deben ser descritos sean las agrupaciones documentales, agentes, actividades, ya propuestas en el esquema e-EMGDE y, además, se inscriban los marcos regulatorios y trazos propios de la tramitación administrativa. Este modelo plantea un nuevo modelo de metadatos para la descripción de la documentación que identifique, en primera instancia, el elemento sujeto a descripción por medio de un prefijo significativo (DOCGROUP, AGENT, ACTIVITY, FRAMEWORK) y un metadato de identificación, denominado (*identifier*). El método sintáctico de combinación de prefijos y metadatos expuesto proporciona 240 metadatos distintos, duplicando la capacidad descriptiva de los esquemas AGLS, e-EMGDE y DC Qualified y mejorando la flexibilidad y

adaptabilidad de la norma de descripción. También se logra simplificar el uso y manejo de los metadatos, facilitando su recuperación y edición posterior. Cualquier modelo de metadatos, con independencia de su objetivo, debe contemplar la creación de editores y esquemas de descripción para el entorno en el que será aplicado. Esto es tener en cuenta qué centros catalogadores lo utilizarán y, por otra parte, qué esquemas de descripción serán necesarios para cada objeto, elemento o entidad descrita con el sistema de metadatos. El método propuesto permite reflejar relaciones jerárquicas y cuadros de clasificación funcional y orgánica, aspectos claves que no son reflejados por ningún otro esquema de metadatos, excepto SKOS-Core y OWL.

REFERENCIAS BIBLIOGRÁFICAS

1. J. Hendler & T. Berners-Lee (2010), "From the Semantic Web to social machines: A research challenge for AI on the World Wide Web", en *Artificial Intelligence*, 174 (2), pp. 156-161 [en línea], <http://www.stanford.edu/class/cs227/Readings/hendler-berners-lee-semantic-web.pdf>
2. A. Cunningham (2010), "The Uses of Metadata in Public Administration", en *Digital Preservation Europe* [en línea], http://www.digitalpreservationeurope.eu/publications/briefs/uses_of_metadata_in_public_administration.pdf
3. F. Schenk; F. Kruse; M. Klenke (2010) "From Simple Data Sources to a Complex Information System: Integrating Heterogeneous Data Models into an Information Infrastructure for the Public Administration", en *Gebalten auf der 24th International Conference on Informatics for Environmental Protection in Cooperation with Intergeo* [en línea], <http://enviroinfo.eu/sites/default/files/pdfs/vol6516/0540.pdf>

El acceso a la información gubernamental...

4. National Archives of Australia: AGLS Metadata Standard (2010) [en línea], <http://www.agls.gov.au/>

5. España. Ministerio de Hacienda y Administraciones Públicas, Metadatos para la gestión del documento electrónico (e-EMG-DE) (2013), en PAE, Portal Administración Electrónica [en línea], <http://administracionelectronica.gob.es/ctt/verPestanaGeneral.htm?idIniciativa=381>

6. J. V. F. Dombeu & M. Huisman (2011), “Combining ontology development methodologies and semantic web platforms for e-government domain ontology development. arXiv preprint arXiv: 1104.4966” [en línea], <http://arxiv.org/ftp/arxiv/papers/1104/1104.4966.pdf>

7. L. Bountouri; C. Papatheodorou; V. Soulikias & M. Stratis (2009), “Metadata interoperability in public sector information”, en *Journal of Information Science*, 35 (2) 2007, pp. 204-231 [en línea], http://eprints.rclis.org/12819/1/JIS-0707_-_v4_-_Metadata_Interoperability_in_Public_Sector_Information.pdf

8. National Archives of Australia: AGLS Metadata Standard: Part1 – Reference Description (2010). [en línea], <http://www.agls.gov.au/pdf/AGLS%20Metadata%20Standard%20Part%201%20Reference%20Description.PDF>