

Informe de conformidad con la NTI de Reutilización de recursos de Información.

1	Antecedentes y contexto.....	2
2	Objetivos.....	2
3	Análisis	3
3.1	Consideraciones generales.....	3
3.1.1	Selección de información reutilizable	3
3.1.2	Identificación de información reutilizable	4
3.1.3	Descripción de la información reutilizable	4
3.1.3.1	Descripción de Vocabularios y Ontologías.....	5
3.1.3.2	Vocabularios de base.....	5
3.1.3.3	Vocabularios adicionales.....	6
3.1.3.4	Repositorios de vocabularios	6
3.1.4	<i>Formato de los documentos y recursos de información reutilizables</i>	7
3.1.5	Términos y condiciones de uso aplicables	8
3.1.6	<i>Puesta a disposición de los documentos y recursos de información</i>	9
3.1.7	<i>Catálogo de información pública reutilizable</i>	10
3.2	Esquema de URIs	10
3.2.1	Objetivos	10
3.2.2	Principios básicos del Esquema.....	10
3.2.3	Normalización de los componentes de los URIs.....	11
3.2.4	Estructura de los URIs.....	12
3.2.4.1	Elementos de Ruta.....	13
3.2.4.2	Carácter	13
3.2.4.3	Sector.....	15
3.2.4.4	Dominio	15
3.2.4.5	Concepto:.....	15
3.2.5	Ejemplos.....	15
3.2.5.1	URIs generales.....	15
3.2.5.2	URIs para Linked Open Data	16
3.2.6	Versionado	18
3.2.7	Implementación del esquema de URIs.....	19
3.2.7.1	Implementación mínima.....	19
3.2.7.2	Negociación de contenido	19
3.2.7.3	Aproximación general a la implementación del esquema de URIs	20
3.2.7.4	Ejemplo de implementación basado en Apache 2	21
3.3	Metadatos.....	22
3.3.1	Objetivos	22
3.3.2	Metodología	23
3.3.2.1	Catálogo (Catalog)	24
3.3.2.2	Conjunto de Datos (Dataset)	25
3.3.2.3	Distribución (Distribution).....	27
3.3.3	Comentarios y cuestiones adicionales	28
3.3.4	Ejemplos.....	31
3.3.4.1	Ejemplo de metadatos de Catálogo	31
3.3.4.2	Ejemplo de metadatos de Conjunto de Datos	32
3.3.4.3	Ejemplo de metadatos de Distribución	32
3.3.4.4	Ejemplo de metadatos de Distribución asociado a un Conjunto de Datos.....	33
3.3.5	Recomendaciones	34
3.4	Taxonomías.....	34
3.4.1	Sectores	35
3.4.1.1	Recomendaciones.....	37
3.4.2	Recursos geográficos	39

1 Antecedentes y contexto

El Gobierno de Aragón, mediante acuerdo de Gobierno de 17 de Julio de 2012 se compromete a la efectiva apertura de los datos públicos que obran en su poder siguiendo las pautas de la Directiva 2003/98/CE sobre Reutilización de la Información del Sector Público y su trasposición a nivel nacional en España mediante la Ley 37/2007

Por otra parte se publica en el BOE la **Norma Técnica de Interoperabilidad¹ de reutilización de recursos de la información** en España (publicada en el BOE núm. 54, de 4 de marzo de 2013 y corregida en el BOE núm. 132, de 3 de junio de 2013²) – en adelante la NTI – y que forma parte del Esquema Nacional de Interoperabilidad³.

El objetivo de esta norma es facilitar y garantizar el proceso de reutilización de la información de carácter público procedente de las Administraciones públicas en todos los niveles, asegurando la persistencia de la información, el uso de formatos así como los términos y condiciones de uso adecuados.

La publicación de esta norma supone un hito importante, ya que en ella se establecen las condiciones comunes sobre selección, identificación, descripción, formato, condiciones de uso y puesta a disposición de los documentos y recursos de información elaborados o custodiados por el sector público.

La norma ha sido diseñada teniendo en cuenta en todo momento los **principios del Open Government Data⁴** e incluye instrucciones detalladas sobre:

- El esquema de identificadores únicos para los datos que exponen públicamente.
- Metadatos de los documentos y recursos de información catalogados.
- Taxonomía de sectores primarios donde se especifican los temas relacionados a cada uno de ellos.
- Identificadores correspondientes a los recursos geográficos del territorio español.

2 Objetivos

¹ NTI Reutilización de recursos de la información, BOE núm. 54, de 4 de marzo de 2013: http://www.boe.es/diario_boe/txt.php?id=BOE-A-2013-2380

² Corrección de errores de la NTI de Reutilización de recursos de la información, BOE núm. 132, de 3 de Junio de 2013: http://administracionelectronica.gob.es/pae_Home?nfpb=true&pageLabel=PAE_PG_CTT_General&langPae=es&iniciativa=eni

³ Esquema nacional de interoperabilidad: http://administracionelectronica.gob.es/pae_Home?nfpb=true&pageLabel=PAE_PG_CTT_General&langPae=es&iniciativa=eni

⁴ 8 Principles of Open Government Data: <http://www.opengovdata.org/home/8principles>

Una vez ya desarrollada la infraestructura básica mediante la primera fase del proyecto Aragón Open Data, se concluye la necesidad de llevar a cabo una segunda fase de evolución y mejora entre cuyos principales objetivos se encuentra la adecuación a la reciente normativa de regulación para las plataformas de datos abiertos en España.

Para ello se pretende realizar un estudio en profundidad sobre la **adecuación a la NTI de la plataforma Aragón Open Data**⁵, implementado sobre la plataforma para la gestión de catálogos de datos CKAN⁶ en su versión 1.8, para determinar el grado de conformidad actual, incluyendo una descripción de las medidas a tomar para conseguir la total adecuación a dicha norma.

3 **Análisis**

La principal tarea de este trabajo será el análisis en profundidad del portal actual Aragón Open Data para verificar su grado de conformidad con respecto a la Norma Técnica de Interoperabilidad vigente así como para definir en caso de que sea necesario las medidas adicionales (incluyendo recomendaciones y ejemplos) que se deberían tomar para conseguir una conformidad total.

3.1 *Consideraciones generales*

En esta sección se realiza una revisión de las medidas recogidas por las recomendaciones generales de la NTI en distintos aspectos como la selección de la información reutilizable, su identificación y su descripción, así como los formatos utilizados, los términos y condiciones de uso aplicables o los mecanismos utilizados para su puesta a disposición del público, incluyendo las condiciones básicas para el Catálogo de Información Pública Reutilizable.

3.1.1 Selección de información reutilizable

Según la NTI, a la hora de seleccionar los recursos de información que se van a publicar se deberán priorizar siempre aquellos de mayor relevancia y potencial social y económico. No obstante, no se ofrece ninguna guía específica sobre cuáles son dichos recursos de mayor valor, por tanto, la mejor opción es seguir las distintas recomendaciones internacionales existentes al respecto:

- Los resultados del estudio MEPSIR⁷ de la Unión Europea.
- Las materias que generan negocio por parte del sector infomediario en España, según el análisis del sector realizado por ONTSI⁸.
- Los conjuntos de datos identificados como de alto valor por la Asociación Multisectorial de la Información – ASEDIE⁹.

⁵ Aragón Open Data: <http://opendata.aragon.es>

⁶ CKAN, the open source data portal software: <http://ckan.org/>

⁷ http://ec.europa.eu/information_society/policy/psi/docs/pdfs/mepsir/final_report.pdf

⁸ <http://www.aporta.es/web/guest/estudioRISP2011>

- Los conjuntos de datos para fomentar la reutilización con fines comerciales¹⁰ y la transparencia ¹¹ identificados por el grupo de trabajo de empresas y emprendedores en las jornadas Open Data de Euskadi¹².
- Los conjuntos de datos comunes a las distintas iniciativas Open Data ya existentes en España, persiguiendo la convergencia y armonización de las mismas.
- Los datos de alto valor identificados durante el plan de mejora del proyecto Open Data del Reino Unido¹³, referente en la materia a nivel mundial.
- Los conjuntos de datos prioritario definidos por el G8¹⁴.
- La disponibilidad de datos analizada por el estudio global del Open Data Barometer¹⁵ y el Open Data Census¹⁶.

3.1.2 Identificación de información reutilizable

Para una adecuada identificación de los recursos de información mediante referencias únicas y unívocas se utilizará un Esquema de URIs según lo definido en el apartado 3.2 de este mismo documento.

3.1.3 Descripción de la información reutilizable

Para una adecuada descripción de los recursos de información se asociarán a estos los metadatos correspondientes según lo definido en el apartado 3.3 de este mismo documento y utilizando taxonomías de descripción de acuerdo a lo establecido en el apartado 3.4.

⁹ ASEIE: Asociación multisectorial de la información: <http://www.asedie.es/>

¹⁰ Día Open Data de Euskadi – Grupo de Trabajo de Empresas y Emprendedores: <https://sites.google.com/site/workgroupscid/home/empresas-y-emprendedores-del-sector-infomediario/lista-de-conjuntos-de-datos>

¹¹ Día Open Data de Euskadi – Grupo de Trabajo de Activistas de la Transparencia y Periodistas: <https://sites.google.com/site/workgroupscid/home/activistas-de-la-transparencia-y-periodistas/conjuntos-de-datos-para-fomentar-la-transparencia-de-la-gestion-publica>

¹² Día Open Data en Euskadi – Grupos de Trabajo: <https://sites.google.com/site/workgroupscid/main>

¹³ Gov.uk - Letter to Cabinet Ministers on transparency and open data: <http://www.number10.gov.uk/news/letter-to-cabinet-ministers-on-transparency-and-open-data/>

¹⁴ G8 Open Data Charter and Technical Annex: <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>

¹⁵ Open Data Research Network – Open Data Barometer: <http://www.opendataresearch.org/project/2013/odb>

¹⁶ OKFN – Open Data Census: <http://census.okfn.org/country>

Según lo establecido en la NTI, para la descripción del catálogo de datos y los conjuntos de datos asociados se utilizará el estándar **DCAT**¹⁷, mientras que para facilitar la reutilización de otros vocabularios se utilizará la Relación de modelos de datos¹⁸ del Centro de Interoperabilidad Semántica (CISE) de la Administración una vez esté disponible. Otros vocabularios de referencia internacionalmente reconocidos y que pueden ser utilizados con seguridad son:

3.1.3.1 Descripción de Vocabularios y Ontologías

Existen dos estándares complementarios para la descripción de vocabularios y ontologías:

- **RDF Vocabulary Description Language 1.0: RDF Schema**¹⁹. Estándar del W3C para describir vocabularios RDF.
- **OWL Web Ontology Language**²⁰. Estándar del W3C para descripción de ontologías que permite descripción lógica.

3.1.3.2 Vocabularios de base

Hay una serie de vocabularios básicos que proporcionan una herramienta muy útil a la hora de reutilizar metadatos previamente definidos por su carácter genérico y el consiguiente alto potencial de reutilización:

- **SKOS**²¹. Vocabulario para describir sistemas de organización del conocimiento como diccionarios, taxonomías, tesauros, etc.
- **DCMI Metadata Terms**²². Dublin Core es una iniciativa para la especificación de metadatos básicos para recursos en general.
- **FOAF**²³. Friend Of A Friend es un vocabulario que permite describir personas y relaciones entre ellas.
- **vCard**²⁴. Estándar para el modelado en RDF del perfil de directorio vCard definido por el RFC 2426²⁵. Se puede utilizar para la descripción de información de contacto de personas u organizaciones.

¹⁷ Data Catalog Vocabulary (DCAT), W3C Candidate Recommendation 5 Nov 2013:

<http://www.w3.org/TR/2013/CR-vocab-dcat-20131105/http://vocab.deri.ie/dcat>

¹⁸ Relación de modelos de datos del CISE:

http://administracionelectronica.gob.es/pae_Home/pae_Estrategias/pae_Interoperabilidad_Inicio/pae_Centro_Interoperabilidad_semantica/pae_CISE_PUNTO_SIETE.html#.UpHssWTC8uo

¹⁹ RDF Vocabulary Description Language 1.0: <http://www.w3.org/TR/rdf-schema/>

²⁰ OWL Web Ontology Language: <http://www.w3.org/TR/owl2-overview/>

²¹ SKOS Simple Knowledge Organization System: <http://www.w3.org/2004/02/skos/>

²² DCMI Metadata Terms: <http://dublincore.org/documents/dcmi-terms/>

²³ FOAF Vocabulary Specification: <http://xmlns.com/foaf/spec/>

²⁴ Representing vCard Objects in RDF: <http://www.w3.org/TR/vcard-rdf/>

²⁵ vCard MIME Directory Profile: <http://www.ietf.org/rfc/rfc2426.txt>

3.1.3.3 Vocabularios adicionales

Adicionalmente, existen otros vocabularios representativos más específicos que también resulta interesante conocer por su aplicación en determinados dominios de conocimiento de uso frecuente:

- **VOID**²⁶. Vocabulario orientado a ofrecer metadatos sobre conjuntos de datos semánticos.
- **SIOC**²⁷. Vocabulario orientado a la descripción de comunidades online, tales como foros de debate, blogs, wikis, etc.
- **eGovernment Core Vocabularies**²⁸. Conjunto de vocabularios básicos en el dominio del eGovernment para describir negocios, personas, localizaciones y servicios públicos.
- **Data Cube Vocabulary**²⁹. Vocabulario que sirve para la creación de cubos de datos para definición de estadísticas.
- **Ontologies for e-Government**³⁰. Ontologías para describir estructuras gubernamentales.
- **Geonames**³¹. Ontología para describir lugares.
- **Creative Commons**³². Lenguaje de definición de licencias.
- **Open Data Rights Statement**³³. Vocabulario definido por el Open Data Institute para poder describir los distintos derechos aplicables sobre una publicación Open Data (licencias, copyrights, atribución, etc.) de forma legible por las máquinas.
- **Wordnet** (3.0)³⁴. Gran base de datos léxica en inglés creada por la universidad de Princeton, cuyos términos tienen una representación en RDF. Es muy útil para utilizar como referencia de alto nivel en la definición de términos y clases en vocabularios.

3.1.3.4 Repositorios de vocabularios

A la hora de localizar vocabularios, no existe un repositorio único de vocabularios de referencia, aunque hay varios catálogos y buscadores que permiten descubrir diferentes vocabularios que son compartidos para su reutilizaciones. Algunos de estos sitios son:

²⁶ Vocabulary of Interlinked Datasets: <http://www.w3.org/2001/sw/interest/void/>

²⁷ Semantically-Interlinked Online Communities: <http://rdfs.org/sioc/spec/>

²⁸ EC Joinup e-Government Core Vocabularies:

https://joinup.ec.europa.eu/community/core_vocabularies/description

²⁹ RDF Data Cube Vocabulary: <http://www.w3.org/TR/vocab-data-cube/>

³⁰ Ontologies for e-Government: <http://oegov.org>

³¹ Geonames Ontology: <http://www.geonames.org/ontology/documentation.html>

³² Creative Commons Rights Expression Language: <http://creativecommons.org/ns>

³³ The UK ODI. Open Data Rights Statement Vocabulary: <http://schema.theodi.org/odrs/>

³⁴ Wordnet 3.0 en RDF: <http://semanticweb.cs.vu.nl/lod/wn30/>

- **Joinup Semantic Assets**³⁵. Sitio web de la Comisión Europea que recopila recursos para la interoperabilidad semántica.
- **Swoogle**³⁶. Buscador de ontologías y términos semánticos.
- **Schema.org**³⁷. Colección de esquemas para representar entidades de uso común de forma que puedan ser reconocidas por los buscadores. También disponible documentación adicional³⁸ sobre cómo integrar estos esquemas en proyectos Linked Data.
- **Linked Open Vocabularies**³⁹ (**LOV**). Recopilador de vocabularios, mantenido por la Open Knowledge Foundation.
- **Taxonomy Warehouse**⁴⁰. Catálogo de taxonomías y vocabularios controlados procedentes de varias fuentes.
- **Watson**⁴¹. Buscador de ontologías y documentos semánticos desarrollado por The Open University.
- **Sindice**⁴². Motor semántico que ofrece un buscador de ontologías y recursos semánticos.
- **Vocab.org**⁴³. Repositorio de esquemas RDF.
- **TONES**⁴⁴. Repositorio de ontologías de la Universidad de Manchester

3.1.4 Formato de los documentos y recursos de información reutilizables

Respecto a los formatos utilizados en la actualidad, como se observa en la tabla a continuación, todos ellos son abiertos tal y como aconseja la NTI (**aunque no todos son libres**), y compatibles también con lo establecido en la Norma Técnica de Interoperabilidad de catálogo de estándares⁴⁵.

Formato	Datasets	Abierto	Libre	Limitaciones
CSV	166	S	S	
DXF	2	S	N	Más restringido que el DWG original.
GML	15	S	S	
JPG	2	S	N	
JSON	197	S	S	
KMZ	15	S	S	

³⁵ Joinup: <https://joinup.ec.europa.eu/catalogue/all>

³⁶ Swoogle: <http://swoogle.umbc.edu>

³⁷ Schema.org: <http://schema.org/docs/schemas.html>

³⁸ Schema.RDFS.org: <http://schema.rdfs.org/>

³⁹ Linked Open Vocabularies: <http://lov.okfn.org/dataset/lov/>

⁴⁰ Taxonomy Warehouse: <http://www.taxonomywarehouse.com>

⁴¹ Watson: <http://watson.kmi.open.ac.uk>

⁴² Sindice: <http://sindice.com>

⁴³ Vocab.org: <http://vocab.org>

⁴⁴ TONES Ontology Repository: <http://owl.cs.manchester.ac.uk/repository/>

⁴⁵ BOE num. 262, Norma Técnica de Interoperabilidad de Catálogo de estándares: http://www.boe.es/diario_boe/txt.php?id=BOE-A-2012-13501

RSS	12	S	S	
SHP	31	S	N	No completamente abierto.
SIG	1	S	S	No es realmente un formato de datos, sino de firma.
XLS	155	S	N	Abierto desde 2007.
XML	225	S	S	
ZIP	194	S	N	No todas las versiones del formato son abiertas y no es realmente un formato de datos, sino de compresión.

Sin embargo, algunos aspectos a mejorar son:

- Algunos formatos (SIG, ZIP) no son realmente de representación, por lo que en estos casos deberían ser sustituidos o complementados con la información correspondiente a los formatos finales.
- No se ofrece información sobre los esquemas o vocabularios utilizados para representar la información (por ejemplo en XML), lo que limita significativamente su margen de reutilización.

3.1.5 Términos y condiciones de uso aplicables

Para cumplir con lo establecido en el artículo 8 del Real Decreto 1495/2011, de 24 de octubre sobre las condiciones de reutilización es recomendable añadir un punto adicional a los actuales términos de uso⁴⁶ mediante el cual se especifique que será necesario *“Conservar, no alterar ni suprimir los metadatos sobre la fecha de actualización y las condiciones de reutilización aplicables”*.

Por otro lado, tanto las condiciones de reutilización globales como cualquier otras específicas aplicables a ciertos recursos de información deben estar disponibles en un formato digital que permita su procesamiento electrónico. Para ello se recomienda el uso de los vocabularios RDF propuestos por *Creative Commons* o el *Open Data Institute* y vistos en el apartado 3.1.3.3. La descripción de las condiciones en formato RDF puede asociarse al documento HTML correspondiente a su representación legible por humanos mediante el mecanismo de negociación de contenidos descrito en la sección 3.2.7.2; y después referenciarse directamente desde cualquiera de las fichas RDF de metadatos de los conjuntos de datos que las apliquen de la siguiente manera:

```
<dct:license rdf:resource="http://opendata.aragon.es/portal/terminos"/>
```

En cuanto a la definición de otras condiciones más específicas que las mínimas generales requeridas por la NTI, actualmente se está aplicando una licencia general *Creative Commons-Reconocimiento (CC-by 3.0)*⁴⁷. Si bien **no es imprescindible**, ya que la NTI fija objetivos sobre Reutilización de la Información del Sector Público (RISP) y no necesariamente sobre Open Data, se debe tener en cuenta que aunque la licencia

⁴⁶ Aragón Open Data. Términos de uso: <http://opendata.aragon.es/terminos.html>

⁴⁷ CC-BY 3.0: <http://creativecommons.org/licenses/by/3.0/es/>

seleccionada es compatible⁴⁸ con la *Open Definition*⁴⁹, sólo es de aplicación legal a contenidos y no contempla los derechos “sui generis” aplicables a bases de datos en algunos países del mundo (principalmente en Europa).

Por tanto, si se quiere garantizar una compatibilidad total con los conceptos que subyacen al Open Data, sería recomendable o bien actualizarse a la recientemente publicada **CC-by 4.0**⁵⁰ en la que se da solución a este problema, o bien plantearse el uso de una licencia *Open Data Commons* aplicable tanto a datos como a bases de datos⁵¹:

- Public Domain Dedication and License (PDDL).
- **Attribution License (ODC-By)** – muy similar a la *CC-by* actual.
- Open Database License (ODC-ODbL).

3.1.6 Puesta a disposición de los documentos y recursos de información

Para garantizar el principio de accesibilidad a la información y a los servicios por medios electrónicos establecido en el artículo 4.c) de la Ley 11/2007, de 22 de junio⁵², así como la legislación general en materia de Accesibilidad Web⁵³ que rige actualmente en España, se recomienda revisar la aplicación de las Pautas de Accesibilidad al Contenido Web⁵⁴ en el conjunto del portal y especialmente el catálogo de datos.

Los recursos de información se publicarán según lo previsto por el Esquema de URIs descrito en la sección 3.2 de este documento y contando siempre con una versión legible por los humanos y otra para las máquinas, según lo descrito en la sección 3.2.7 de este documento.

También debe tenerse en cuenta que, si bien actualmente no se están proporcionando Servicios Web o APIs específicos de acceso a los datos (más allá del API que proporciona CKAN para la gestión de los metadatos), si en un futuro así se hiciese sería obligatorio documentar el uso y configuración de estos puntos de acceso dinámico a los recursos de información.

⁴⁸ Open Definition Conformant Licenses: <http://opendefinition.org/licenses/>

⁴⁹ The Open Definition: <http://opendefinition.org/>

⁵⁰ CC-BY 4.0: <http://creativecommons.org/licenses/by/4.0/>

⁵¹ Open Data Commons. data/databases license:

<http://opendatacommons.org/licenses/>

⁵² Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos: <https://www.boe.es/buscar/doc.php?id=BOE-A-2007-12352>

⁵³ INTECO. Legislación en Accesibilidad Web en España:

http://www.inteco.es/Accesibilidad/difusion/Publicaciones/Legislacion_en_Accesibilidad_Web_en_Espana_y

http://www.inteco.es/Accesibilidad/difusion/Publicaciones/Legislacion_en_Accesibilidad_Web_en_Espana_1

⁵⁴ Web Content Accessibility Guidelines (WCAG) 2.0: <http://www.w3.org/TR/WCAG20/>

3.1.7 Catálogo de información pública reutilizable

Una vez implementadas las mejoras indicadas en este documento se cumplirían con los requisitos previstos por la NTI para los catálogos de información pública reutilizable.

3.2 Esquema de URIs

En este apartado se definirá un **esquema de identificadores mediante URIs** adecuado para el catálogo, incluyendo estructura básica, normalización y pautas para su gestión, de forma que se cumplan con las condiciones expresadas en el **Anexo II de la Norma Técnica de Interoperabilidad de Reutilización de recursos de la información** – en adelante la NTI (publicada en el BOE núm. 54, de 4 de marzo de 2013 y corregida en el BOE núm. 132, de 3 de junio de 2013).

3.2.1 Objetivos

Los **Identificadores de Recursos Uniformes** –del inglés, *Uniform Resource Identifiers* o **URIs**⁵⁵– ofrecen la capacidad de identificación de forma unívoca de cualquier recurso a través de la Web, lo que permite establecer identificadores universales para cualquier pieza de información que se desee exponer. Esencialmente, son direcciones web que identifican cualquier elemento que representemos en la Web.

Con la implantación de un esquema de URIs se pretende proporcionar un mecanismo de identificación común para los datos que se exponen públicamente, de forma que se pueda hacer referencia a estos de forma única, fiable y persistente en el tiempo, requisito clave para facilitar su posterior reutilización.

Mediante el esquema se indican las partes que conforman los URIs de una plataforma de reutilización de información del sector público, las cuales siguen patrones claramente definidos, lo que permite ofrecer identificadores semánticos e intuitivos, y coherencia en la ampliación del sistema.

El objetivo principal será obtener un esquema de URIs que funcione como una guía de estilo de referencia a la hora de diseñar nuevos identificadores para los elementos de información reutilizables del catálogo y sus servicios relacionados.

3.2.2 Principios básicos del Esquema

Además de la propia definición de las reglas de construcción de los URIs del esquema deberán también tenerse en cuenta una serie de requisitos básicos en el diseño del esquema, necesarios para cumplir con la NTI respetando los principios básicos de la arquitectura de la *World Wide Web*⁵⁶ y siguiendo las buenas prácticas para la utilización

⁵⁵ Naming and Addressing: URI, URLs: <http://www.w3.org/Addressing/>

⁵⁶ Architecture of the World Wide Web, Volume One:
<http://www.w3.org/TR/2004/REC-webarch-20041215/>

de URIs en la Web Semántica⁵⁷ y las buenas prácticas para la publicación de Linked Data⁵⁸:

1. Se debe utilizar siempre el **protocolo HTTP**⁵⁹ –en inglés, HyperText Transfer Protocol– y no otros. De esta forma se garantiza que la resolución de cualquier URI en la Web es posible.
2. Utilizar siempre una estructura de composición de URIs **consistente, extensible y persistente**. Las normas de construcción de los URIs deberán seguir unos patrones determinados que ofrezcan coherencia en la uniformidad y sean *predecibles*, pero a la vez deben ofrecer flexibilidad, pudiendo dichos patrones ser ampliados o adaptados en caso de necesidad.

Todo identificador que ya ha sido creado y publicado deberán mantenerse siempre en el tiempo, independientemente del estado del recurso al que identifica. En el caso de que dicho recurso haya sido modificado o eliminado, esos cambios se comunicarán usando los códigos de estado de HTTP. Si el recurso ha cambiado de ubicación se utilizarán los códigos HTTP 3XX, mientras que si el recurso ha desaparecido permanentemente se utilizará el código HTTP 410.

3. Los URI deben seguir una estructura de composición **comprensible y significativa**. Esto significa que el propio identificador debe ofrecer información semántica autocontenida, es decir, que las distintas partes que componen el identificador final sean términos con significado por sí mismos, lo que permitirá a cualquier agente reutilizador disponer de valiosa información adicional sobre el propio recurso, así como su procedencia.
4. No se debe exponer información sobre la implementación técnica de los recursos que representan los URI. En la medida de lo posible se debería omitir información específica sobre la tecnología subyacente del recurso representado –por ejemplo, no se debería incluir las extensiones correspondientes a tecnologías con las que se generan los recursos Web como .php, .jsp, .asp, etc.– presentándose las URIs de referencia sin ninguna extensión asociada.

3.2.3 Normalización de los componentes de los URIs

Para mantener la coherencia y facilitar el mantenimiento posterior del esquema de URIs, y siguiendo las recomendaciones de la NTI, se tendrán en cuenta unas sencillas reglas para normalizar las partes que componen los URIs y que consistirán en:

- Seleccionar identificadores alfanuméricos cortos únicos, que sean representativos, intuitivos y semánticos (con significado).

⁵⁷ Cool URIs for the Semantic Web: <http://www.w3.org/TR/coolURI/>

⁵⁸ (Draft) Best Practices for Publishing Linked Data: <https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html>

⁵⁹ Hypertext Transfer Protocol – HTTP/1.1:
<http://www.w3.org/Protocols/rfc2616/rfc2616.html>

- Usar siempre minúsculas, salvo en los casos en los que se recomienda utilizar el nombre de la clase o concepto. Habitualmente, los nombres de las clases se representan con el primer carácter de cada palabra en mayúsculas (técnica también conocida como *upper camel case*).
- Eliminar todos los acentos, diéresis y símbolos de puntuación. Únicamente podrá utilizarse el guión medio (-) como excepción, otros caracteres no ASCII deben también evitarse (por ejemplo cambiando las *eñes* por *enes*).
- Eliminar conjunciones y artículos en los casos de que el concepto a representar contenga más de una palabra, pudiendo usarse el guión medio (-) como separador entre palabras o la técnica de *lower camel case*.
- Evitar en la medida de lo posible la abreviatura de palabras, salvo que la abreviatura sea ampliamente utilizada y claramente reconocible.

Aunque la NTI recomienda usar términos en castellano o cualquiera de las lenguas oficiales para facilitar su legibilidad, los URIs podrían tener también distintas representaciones adicionales compuestas por términos equivalentes en otros idiomas. Después, desde el servidor Web, se podría realizar la negociación de idioma y hacer las redirecciones oportunas para servir los recursos equivalentes.

3.2.4 Estructura de los URIs

Todos los URIs que formen parte del Esquema diseñado deberán mantener una estructura uniforme, ofreciendo la coherencia necesaria para el sistema de representación de los recursos, cubriendo los principios básicos anteriormente vistos.

En el caso que nos concierne podríamos decir que la estructura básica general de un URI⁶⁰ es la siguiente:

protocolo:://{base}/{ruta}[?parámetros][#fragmento]

Cuyos componentes son:

- El esquema del protocolo utilizado.
- La base de los URIs, que puede incluir el dominio de la máquina que sirve los datos –incluyendo elementos opcionales como el *puerto*– o incluso una parte fija de la ruta de los URIs.
- Los *parámetros* variables de un recurso se podrán incluir libremente y no se tendrán en cuenta en esta especificación, ya que son elementos técnicos propios de las soluciones utilizadas.
- Una ruta, que a su vez está compuesta por distintas variables, que se detallarán en las siguientes secciones.

En este caso concreto, y según lo definido en los principios básicos el protocolo a utilizar será **http**, la base será **opendata.aragon.es**. Para simplificar se servirá siempre por el

⁶⁰ Sintaxis genérica de URI: <http://tools.ietf.org/html/rfc3986>

puerto por defecto (80) y no se hará uso de los parámetros variables por ser elementos técnicos propios de otras soluciones que en este caso no se van a utilizar.

Por tanto, nuestro Esquema básico será:

`http://opendata.aragon.es/{ruta}[/#fragmento]`

Es importante recordar que el dominio elegido para la base debe mantenerse indefinidamente en el futuro y, en caso de ser obligado el cambio, deberían realizarse siempre las redirecciones necesarias para mantener todas los identificadores que se hubiesen construido mediante el esquema de URIs.

3.2.4.1 Elementos de Ruta

Según lo previsto en los principios básicos, la ruta incluirá información descriptiva sobre el tipo de recurso que se representa. Los elementos que componen la ruta de una URI son:

- Carácter de la información.
- Sector.
- Dominio o temática.
- Conceptos específicos.

Por lo tanto la estructura completa del Esquema de URIs será:

`http://opendata.aragon.es/{carácter}/{sector}/{dominio}/{concepto}[/#fragmento]`

O, alternativamente, si se opta por utilizar la sintaxis de identificadores de fragmento:

`http://opendata.aragon.es/{carácter}/{sector}/{dominio}[/#fragmento]`

Si bien la NTI permite cierta flexibilidad en cuanto a la estructura general, siempre será **obligatorio mantener al menos los elementos base y carácter** en todos URIs del esquema.

3.2.4.2 Carácter

Este elemento de composición de los URIs indica el carácter general de los datos que representa. A continuación se describen los distintos tipos de posibles según la NTI:

Valor	Tipo de recurso que representa
catalogo	Fichas de cada uno de los conjuntos de datos – <i>datasets</i> – incluido en el catálogo. Estos recursos de información contendrán los distintos metadatos contemplados en la sección 3.3 <i>Metadatos</i> . También identifica al catálogo en sí cuando no está asociado a ningún otro elemento.

def	Vocabulario u ontología de desarrollo propio en la iniciativa de Aragón y utilizadas para modelar aquellos datos que estén disponibles en formato semántico. Se trata habitualmente de esquemas en formato RDF-S ⁶¹ u ontologías representadas mediante OWL ⁶² .
kos	Sistemas de organización del conocimiento de desarrollo propio en la iniciativa de Aragón sobre un dominio concreto. Se trata habitualmente taxonomías, diccionarios o tesauros, representados normalmente mediante SKOS ⁶³ .
recurso	Identificación abstracta única y unívoca de un recurso físico o conceptual. Estos recursos son cada una de las instancias de los distintos conceptos que han sido modelados semánticamente mediante vocabularios propios o de terceros.

Se recomienda también que, aparte de los tipos de carácter definidos ya por la NTI, se utilice uno adicional que sirva para el resto de contenidos del portal. Esta aproximación contribuiría a mejorar también la organización del resto de contenidos del portal, al mismo tiempo que se mantiene compatible con todos los requisitos definidos por la NTI para el Esquema de URIs:

portal *Resto de documentos pertenecientes a otras secciones que, además del propio catálogo, se encuentra también con frecuencia en cualquier portal Open Data, como por ejemplo: información legal, ayuda, noticias, tutoriales, aplicaciones, participación, etc.*

Finalmente, una opción también a valorar será la posible integración completa del desarrollo previsto para la AragoPedia en el catálogo principal, tanto a nivel de contenidos como en el propio esquema de identificación. En este caso se podría incorporar un elemento nuevo de carácter de la forma:

aragopedia *Contenidos con un carácter principal de difusión hacia el ciudadano, organizados y desglosados según los distintos municipios aragoneses y estructurados también semánticamente.*

Si se opta no obstante por una implementación completamente independiente de la AragoPedia, éste debería seguir exactamente el mismo esquema de identificación aquí descrito, cambiando únicamente la base por una nueva identificativa del proyecto, por ejemplo: <http://aragopedia.aragon.es/>

⁶¹ RDF Vocabulary Description Language 1.0. RDF Schema: <http://www.w3.org/TR/rdf-schema/>

⁶² OWL Web Ontology Language: <http://www.w3.org/TR/owl-features/>

⁶³ SKOS Simple Knowledge Organization System Primer: <http://www.w3.org/TR/skos-primer/>

3.2.4.3 Sector

La selección de un sector adecuado, acompañado del dominio específico del origen, le dará a cualquier usuario la confianza de conocer el tipo de información que está manejando y la fuente de la misma. Se seleccionará un identificador del sector (primario), según lo especificado en el anexo IV de la NTI. Cada documento o recurso de información, vocabulario o esquema de conceptos debe pertenecer a un único sector. Si pertenece a más de uno, se utilizará el más representativo o alguno que se pueda considerar común.

3.2.4.4 Dominio

Para identificar las temáticas específicas dentro de un sector se creará una referencia adecuada que represente al dominio (o sector secundario) de la información tratada. Este identificador, acompañado del sector, ofrecerá a cualquier usuario la oportunidad de intuir y reconocer fácilmente el tipo de datos concreto que está manejando.

3.2.4.5 Concepto:

Los últimos elementos de ciertos URI –tras el carácter, sector y dominio de la información– hacen referencia a los conceptos e instancias específicas de recursos. Los conceptos son representaciones abstractas que se corresponden con las clases o propiedades de los vocabularios u ontologías utilizados para representar semánticamente los recursos. Además de los conceptos, se podrán representar referencias unívocas a instancias concretas u otros esquemas de conceptos abstractos, dentro de sistemas de gestión del conocimiento (taxonomías, tesauros, etc.).

3.2.5 Ejemplos

A continuación se presentan algunos ejemplos del esquema de URIs a aplicar en Aragón Open Data según los patrones anteriormente descritos y de acuerdo a la NTI. Se ha optado por utilizar la nomenclatura *slash* (/) para los recursos [{dominio}] y la nomenclatura *hash* (#) para los conceptos y propiedades [{Concepto|propiedad}].

3.2.5.1 URIs generales

Este tipo de URIs podrán ser utilizadas tanto en iniciativas Open Data convencionales como en iniciativas que utilicen tecnologías Linked Open Data y de Web Semántica.

URI para identificar el catálogo: `http://{base}/catalogo`

`http://opendata.aragon.es/catalogo`

NOTA: podría considerarse que el equivalente actual en el portal sería

`http://opendata.aragon.es/dataset`

URIs para otros contenidos del portal: `http://{base}/portal/{sector}/{dominio}`

<http://opendata.aragon.es/portal/participacion>

<http://opendata.aragon.es/portal/aplicaciones>

<http://opendata.aragon.es/portal/noticias>

NOTA: podría considerarse que el equivalente actual en el portal sería

<http://opendata.aragon.es/{sector.ext}>

URIs para identificar los conjuntos de datos: <http://{base}/catalogo/{dataset}>

<http://opendata.aragon.es/catalogo/nombre-dataset1>

<http://opendata.aragon.es/catalogo/nombre-dataset2>

<http://opendata.aragon.es/catalogo/nombre-dataset3>

NOTA: podría considerarse que el equivalente actual en el portal sería

<http://opendata.aragon.es/dataset/nombre-dataset1>

O en su lugar se podría utilizar también de forma alternativa:

<http://{base}/catalogo/{sector}/{dataset}>

<http://opendata.aragon.es/catalogo/turismo/nombre-dataset1>

<http://opendata.aragon.es/catalogo/salud/nombre-dataset2>

<http://opendata.aragon.es/catalogo/vivienda/nombre-dataset3>

NOTA: Si bien esta aproximación **alternativa** a la anterior da lugar a URIs con una estructura más comprensible y significativas - y en principio sería también totalmente compatible con la definición genérica del Esquema de URIs que se propone en la NTI

<http://{base}/{carácter}/{sector}/{dominio}/{concepto}/{ext}>; indicando

además la propia norma que dicha estructura general puede variar dependiendo de las necesidades o preferencias de la organización que la aplique siempre y cuando se mantengan invariables los elementos base y carácter - sería igualmente conveniente verificar su validez formal en cuanto a conformidad con la NTI antes de adoptarla definitivamente, ya que la norma cita actualmente <http://{base}/catalogo/{dataset}> como el tipo a seguir en el caso de los conjuntos de datos.

Y de forma análoga podemos utilizar: <http://{base}/datos/{sector}/{dataset}.{ext}> para identificar las distintas distribuciones de un conjunto de datos

<http://opendata.aragon.es/datos/turismo/nombre-dataset1.xml>

<http://opendata.aragon.es/datos/salud/nombre-dataset2.csv>

<http://opendata.aragon.es/datos/vivienda/nombre-dataset3.xls>

NOTA: Si bien esta forma de representación de las distribuciones no es una exigencia de la NTI y tampoco aporta ninguna ventaja desde el punto de vista del contenido *auto-descubrible*, si la aplicamos conseguiríamos mantener un esquema más predecible e intuitivo sin interferir en el correcto cumplimiento de la norma.

3.2.5.2 URIs para Linked Open Data

Este tipo de URIs están específicamente pensadas para la implementaciones de Linked Open Data que se basen en tecnologías para la Web Semántica, si bien el uso de

esquemas de conceptos está cada vez más extendido también en iniciativas Open Data convencionales.

URIs para identificar esquemas de conceptos:**`http://{base}/kos/{sector}/{dominio}`**

`http://opendata.aragon.es/kos/seguridad/delitos`
`http://opendata.aragon.es/kos/sociedad-bienestar/ayudas`
`http://opendata.aragon.es/kos/deporte/deportes`

NOTA: Actualmente no hay ningún esquema de conceptos definido en el portal, por lo que no existen equivalentes. Los ejemplos anteriores representarían taxonomías que permitiesen clasificar los tipos de delito, de ayudas y de deportes respectivamente.

URIs para identificar conceptos:**`http://{base}/kos/{sector}/{dominio}#{Concepto}`**

`http://opendata.aragon.es/kos/seguridad/delitos#Hurto`
`http://opendata.aragon.es/kos/seguridad/delitos#Agresion`
`http://opendata.aragon.es/kos/seguridad/delitos#Homicidio`

NOTA: Actualmente no hay ningún esquema de conceptos definido en el portal, por lo que no existen equivalentes. Los ejemplos anteriores representarían distintos conceptos dentro de la taxonomía de delitos.

URIs para identificar vocabularios: `http://{base}/def/{sector}/{dominio}`

`http://opendata.aragon.es/def/turismo/hosteleria`
`http://opendata.aragon.es/def/sector-publico/contratacion`
`http://opendata.aragon.es/def/transporte/horarios`

NOTA: actualmente no hay ningún vocabulario definido en el portal, por lo que no existen equivalentes. Los ejemplos anteriores representarían vocabularios que permitiesen definir la oferta turística de hostelería, la contratación pública y los horarios de transportes respectivamente.

URIs para identificar clases y propiedades:**`http://{base}/def/{sector}/{dominio}#{propiedad|Clase}`**

`http://opendata.aragon.es/def/sector-publico/contratacion#Concurso`
`http://opendata.aragon.es/def/sector-publico/contratacion#importe`
`http://opendata.aragon.es/def/sector-publico/contratacion#fecha`

NOTA: actualmente no hay ningún vocabulario definido en el portal, por lo que no existen equivalentes. Los ejemplos anteriores representarían distintas clases y propiedades dentro del vocabulario de contratación.

URIs para identificar instancias físicas o conceptuales:**`http://{base}/recurso/{sector}/{dominio}/{clase}/{ID}`**

<http://opendata.aragon.es/recurso/turismo/alojamiento/Hotel/Hotel1>
<http://opendata.aragon.es/recurso/sector-publico/tramites/Subvencion/Subvencion2>
<http://opendata.aragon.es/recurso/salud/equipamientos/Hospital/Hospital3>

NOTA: Actualmente no hay instancias de ningún esquema de conceptos definido en el portal, por lo que no existen equivalentes. Los ejemplos anteriores representarían un hotel específico, un tipo de subvención concreta y un determinado hospital respectivamente, cada uno con todas sus propiedades modeladas según los correspondientes vocabularios.

Una buena práctica sería que la clase se correspondiese siempre con el tipo de concepto que se está representado y el dominio con el vocabulario utilizado. Así, según los ejemplos anteriores:

- `Hotel1` sería una instancia específica de la clase `Hotel`, que estaría definida en el vocabulario de alojamientos (dominio).
- `Subvencion2` sería una instancia específica de la clase `Subvención`, que estaría definida en el vocabulario de trámites (dominio).
- `Hospital3` sería una instancia específica de la clase `Hospital`, que estaría definida en el vocabulario de equipamientos (dominio).

3.2.6 Versionado

Con el tiempo, y según la iniciativa vaya creciendo y madurando, es frecuente que sea necesario evolucionar y adaptar también los vocabularios de creación propia que hayan sido elaborados para modelar los datos. Para mantener la compatibilidad y coherencia entre todas las versiones de los vocabularios y esquemas de conceptos será entonces necesario establecer un mecanismo de control de versiones sencillo, fácil de mantener y compatible con el esquema de URIs aplicado.

La estrategia general a seguir será utilizar para la versión más reciente el espacio de nombres por defecto los URI genéricos especificados según el esquema definido anteriormente, aunque de forma simultánea mantendremos también otros documentos históricos con las versiones anteriores.

Así por ejemplo, si los vocabularios se representan generalmente de la forma **`http://{base}/def/{sector}/{dominio}`**

La última versión de un vocabulario utilizado para representar horarios de transportes estaría localizada siempre en `http://opendata.aragon.es/def/transporte/horarios`

Para las versiones anteriores usaríamos un sufijo adicional de la forma `_YYYYMMDD` que indicaría la fecha de revisión correspondiente (año, mes y día), dando lugar por ejemplo a versiones del tipo:

`http://opendata.aragon.es/def/transporte/horarios_20110505`

para la versión del 5 de Mayo del 2011

http://opendata.aragon.es/def/transporte/horarios_20120208

para la versión del 8 de Febrero de 2012

3.2.7 Implementación del esquema de URIs

Cualquier información semántica, previamente identificada mediante un URI unívoco, debería contar con dos formas básicas de representación simultánea que, a su vez, tendrán asociados los URI descritos en el presente documento:

- RDF (en cualquiera de sus formatos: xml/rdf, ttl, n3) para facilitar la explotación automatizada de los datos y los vínculos con otros conjuntos externos.
- HTML, para ofrecer una versión legible del contenido destinada a los humanos.

El catálogo actual cuenta con un mecanismo básico de negociación de contenidos para servir las distintas representaciones (HTML y RDF) de los conjuntos de datos.

A continuación se describen las posibles estrategias para una implementación completa del esquema de URIs previsto, así como un ejemplo con negociación de contenidos basada en una infraestructura que cuente con el servidor web *Apache HTTP Server*⁶⁴ (versión 2.X).

3.2.7.1 Implementación mínima

En el caso de que no se realice una negociación del contenido desde el servidor, y con el objetivo de favorecer el descubrimiento automático de contenido RDF desde los documentos HTML relacionados con las descripciones de los recursos, se incluirán como mínimo enlaces a la representación alternativa en cualquiera de las representaciones en RDF desde los propios documentos HTML incrustando el siguiente código en la sección HEAD de los documentos:

```
<link rel="alternate" type="application/rdf+xml" href="../../../documento.rdf">
```

La URL del documento y el tipo de formato MIME serán los adecuados en cada caso (application/rdf+xml, text/n3, text/turtle, etc.)

3.2.7.2 Negociación de contenido

El mecanismo de negociación de contenido permite que, al solicitar una URI que identifica los datos que se quieren consultar y que ha sido definida según el esquema anteriormente descrito, la información se muestre según la representación más apropiada para el usuario que realiza la petición, después de analizar sus características y preferencias.

⁶⁴ Apache HTTP Server: <http://httpd.apache.org/>

Para llevar a cabo la negociación de contenido, se utilizará un mecanismo conocido como **redirecciones 303**⁶⁵ y que consiste en utilizar el código de respuesta HTTP 303 (*See other*) para redirigir a los clientes hacia las distintas representaciones alternativas de los recursos.

Se tendrán en cuenta los dos casos de uso que se pueden presentar, donde dos agentes distintos hacen una petición sobre un mismo recurso que se supone identificado mediante el siguiente URI de ejemplo:

`http://opendata.aragon.es/catalogo/nombre-dataset1`

CASO1: Cuando el agente que realiza la petición es un usuario desde un navegador web entonces el servidor procesa el identificador y ofrece el recurso alternativo más adecuado, en este caso la representación legible para humanos en versión `html`:

`http://opendata.aragon.es/catalogo/nombre-dataset1.html`

CASO 2: Cuando el agente que realiza la representación tiene soporte para contenidos RDF y realiza una petición usando el mismo identificador, el servidor Web ofrecería la representación alternativa apropiada para este agente, es decir la versión `rdf` en cualquiera de sus formas aceptadas:

`http://opendata.aragon.es/catalogo/nombre-dataset1.rdf`

`http://opendata.aragon.es/catalogo/nombre-dataset1.n3`

`http://opendata.aragon.es/catalogo/nombre-dataset1.ttl`

La diferenciación entre ambos casos de uso se realizará gracias al campo `Accept`⁶⁶ de los mensajes que se transmiten en las cabeceras `http` durante el intercambio de mensajes que forma parte de las peticiones de recursos.

3.2.7.3 *Aproximación general a la implementación del esquema de URIs*

Una vez habilitado el mecanismo de negociación de contenido, el siguiente paso consistirá en la implantación del esquema de URIs mediante el establecimiento de relaciones entre la localización de los recursos y los URIs que les corresponde según el esquema.

Existen dos estrategias diferenciadas para implementar el esquema de URI con Apache:

1. **alias.** Utilizando la directiva `alias` para asignar una URL real a cada una de los URI del esquema.

⁶⁵ Best Practice Recipes for Publishing RDF Vocabularies:

<http://www.w3.org/TR/swbp-vocab-pub/>

⁶⁶ HTTP 1.1 Header field definitions: <http://www.w3.org/Protocols/rfc2616/rfc2616-sec14.html>

2. **mod_rewrite**. Utilizando `mod_rewrite` y estableciendo relaciones para cada una de los URI a la ubicación real donde se encuentra el fichero a servir. `mod_rewrite` permite utilizar expresiones regulares, de forma que si se detecta un patrón en los URI no será necesario definir tantas reglas como recursos.

En ambos casos el número de reglas a definir puede llegar a ser extenso, ya que si no se pueden distinguir patrones comunes habrá que crear una regla para cada uno de los elementos a publicar. Además, Apache evaluará dichas reglas de forma secuencial, lo que a la larga podría traducirse en una notoria pérdida de eficiencia a medida que el número de reglas vaya aumentando.

En conclusión, si bien estos mecanismos podrían utilizarse como primera opción provisional para llevar a cabo la implantación del esquema de URI, es necesario plantearse una solución destinada a su funcionamiento a largo plazo y que mejoraría considerablemente el rendimiento del mecanismo de traducción de URI.

En líneas generales, dicha solución consistiría en utilizar un lenguaje de scripting interpretado (por ejemplo, *php*, *python* o *perl*) para definir estructuras en el que se almacenen las reglas, de manera que Apache simplemente tuviera que consultar dicha estructura para saber cuál es la ubicación real del recurso solicitado.

3.2.7.4 Ejemplo de implementación basado en Apache 2

Para poder llevar a cabo las operaciones que se describen en este apartado será necesario tener instalado y habilitado el módulo `mod_rewrite` para Apache 2. Este módulo permite evaluar las cabeceras de una petición HTTP y, en base a su valor, redirigir a uno u otro contenido. Además, se suponen definidas y localizadas ambas representaciones de la información en dos archivos situados en el servidor HTTP:

- El documento HTML en `/catalogo/dataset.html`
- La descripción RDF(XML) en `/catalogo/dataset.rdf`

Nótese que no existe ningún directorio directamente asociado con el identificador del recurso (id).

Una de las configuraciones que se podría aplicar, dentro del fichero de configuración `.htaccess` de Apache para resolver la redirección de estos casos de uso planteados es la siguiente:

```
# Desactiva las vistas múltiples si no está hecho por defecto
# Options +FollowSymlinks -MultiViews
Options +FollowSymlinks

# Directiva para asegurar que los archivos *.rdf se sirven con el tipo
# de contenido apropiado, si no está presente en la configuración
# por defecto de Apache
AddType application/rdf+xml .rdf
```

```
# Configuración del motor de reescritura
RewriteEngine On
RewriteBase /

# Regla de reescritura para servir contenido HTML desde la URI de
identificación
RewriteCond %{HTTP_ACCEPT}
!application/rdf\+xml.*(text/html|application/xhtml\+xml)
RewriteCond %{HTTP_ACCEPT} text/html [OR]
RewriteCond %{HTTP_ACCEPT} application/xhtml\+xml [OR]
RewriteCond %{HTTP_USER_AGENT} ^Mozilla/.*
RewriteRule ^catalogo/dataset$ catalogo/dataset.html [R=303]

# Regla de reescritura para servir contenido RDF/XML si se pide
RewriteCond %{HTTP_ACCEPT} application/rdf\+xml
RewriteRule ^catalogo/dataset$ catalogo/dataset.rdf [R=303]

# Selección de la respuesta por defecto
# -----

# Reglas de reescritura para servir contenido HTML por defecto
RewriteRule ^catalogo/dataset$ catalogo/dataset.html [R=303]

# Regla de reescritura para servir contenido RDF/XML por defecto
(deshabilitado)
# (Para permitir esta opción, descomentar la regla siguiente y comentar la
# regla de reescritura anterior)
#RewriteRule ^catalogo/dataset$ catalogo/dataset.rdf [R=303]
```

3.3 Metadatos

En este apartado se verificará la adecuación de los **metadatos** del **Catálogo** y los **Conjuntos de Datos** de acuerdo al **Anexo III de la Norma Técnica de Interoperabilidad de Reutilización de recursos de la**.

3.3.1 Objetivos

El objetivo principal será cumplir completamente con los requisitos mínimos marcados por dicha Norma. Sin ser un requisito prioritario, se analizarán también posibles problemas de compatibilidad con la evolución actual del **vocabulario estándar para Catálogo de Datos (DCAT)**⁶⁷, según la versión Candidata a Recomendación del 5 de Noviembre de 2013⁶⁸ (revisado posteriormente con la publicación de la versión definitiva de 16 de Enero de 2014⁶⁹), y especialmente con el **perfil de aplicación**

⁶⁷ Si bien el estándar DCAT está todavía en desarrollo, actualmente se encuentra en su segunda iteración de Candidato a Recomendación, por lo que cabría esperar que su grado de estabilidad actual sea ya muy alto y que de producirse cambios posteriores estos fuesen mínimos.

⁶⁸ Data Catalog Vocabulary (DCAT), W3C Candidate Recommendation 5 Nov 2013
<http://www.w3.org/TR/2013/CR-vocab-dcat-20131105/>

⁶⁹ Data Catalog Vocabulary (DCAT), W3C Recommendation 16 January 2014:
<http://www.w3.org/TR/vocab-dcat/>

específico (DCAT-AP)⁷⁰ en su versión 1.0 del 2 de Septiembre del 2013 - desarrollado para garantizar la interoperabilidad de los portales de datos Europeos - tratando en la medida posible de establecer pautas para el mayor grado de cumplimiento de ambos estándares.

Teniendo en cuenta que:

- Dentro de la NTI se especifican **metadatos obligatorios y opcionales**; siendo los primeros siempre de obligado uso para cumplir la norma.
- El DCAT-AP Europeo distingue también entre **propiedades obligatorias, recomendadas y opcionales**; siendo las obligatorias siempre de obligado uso y las recomendadas también siempre y cuando la información esté disponible en el sistema.
- El Vocabulario estándar de Catálogos de Datos (DCAT) **no define ninguna restricción** respecto a conjuntos de metadatos mínimos, dejando esa responsabilidad a los distintos perfiles de aplicación.

Entonces los objetivos específicos por orden de prioridad serán:

1. Cumplir con los metadatos obligatorios de la NTI
2. Cumplir con los metadatos obligatorios y recomendados del perfil de aplicación de DCAT-AP.
3. Asegurar la compatibilidad con los requisitos de conformidad del vocabulario DCAT⁷¹.
4. Tener en cuenta cualquier otra característica adicional de compatibilidad que pudiera ser necesaria para cumplir con los metadatos opcionales.

3.3.2 Metodología

El análisis se realizará desde el punto de vista del **productor de datos**, y no desde el del receptor, por ser este el rol principal de un catálogo de datos. Sin embargo, dado que cualquier catálogo podría en un momento dado actuar también como receptor de datos (por ejemplo si ofrece un servicio de *federación* de conjuntos provenientes de otros catálogos), se realizarán también las observaciones que se consideren pertinentes para tener en cuenta también dicho escenario.

El análisis se ha efectuado mediante el estudio de las características del portal actual, la documentación oficial de CKAN tanto en su versión 1.8⁷² (implementación inicial de Aragón Open Data) como en la versión más reciente⁷³ (v2.2a en el momento de

⁷⁰ DCAT Application Profile for Data portal in Europe – Final versión:

https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/dcat-application-profile-data-portals-europe-final

⁷¹ DCAT Conformance: <http://www.w3.org/TR/vocab-dcat/#conformance>

⁷² CKAN's 1.8 administrator guide: <http://docs.ckan.org/en/ckan-1.8/>

⁷³ CKAN 2.2a documentation: <http://docs.ckan.org/en/latest/>

realización del informe) y una muestra aleatoria de conjuntos de datos del mismo, incluyendo:

<http://opendata.aragon.es/dataset/cifras-poblacion-revision-padron-municipal-1998-comarcas>
http://opendata.aragon.es/dataset/ejes-de-vias-de-comunicacion_06_carreteras_aragon
http://opendata.aragon.es/dataset/ambitos-de-proteccion-de-especies-protegidas_rednat_appe
http://opendata.aragon.es/dataset/mapa-de-aragon-1-400000-geopdf_mapaaragon_400k
<http://opendata.aragon.es/dataset/anuncios-informacion-publica-concesiones-uso-privativo>
http://opendata.aragon.es/dataset/terrenos-cinegeticos_rtc_cotos
<http://opendata.aragon.es/dataset/anuncio-publico-expedientes-determinacion-procedencia-eia>
http://opendata.aragon.es/dataset/planes-de-ordenacion-de-los-recursos-naturales_rednat_porn
<http://opendata.aragon.es/dataset/boletin-oficial-aragon-diario>
<http://opendata.aragon.es/dataset/accidentes-circulacion-y-victimas>
<http://opendata.aragon.es/dataset/registro-de-convenios-de-aragon-1986>
http://opendata.aragon.es/dataset/anejo-fotografico-alto-gallego-c02_album_c02_album
http://opendata.aragon.es/dataset/ortofoto-color-05m-pnoa-2012_2012_pnoa
http://opendata.aragon.es/dataset/limites-de-municipios-comarcas-y-provincias_02_municipios_aragon

Así mismo, el análisis se centrará exclusivamente en los metadatos pertenecientes a los dominios del Catálogo (*Catalog*), los Conjuntos de Datos (*Dataset*) y las Distribuciones (*Distribution*), quedando explícitamente excluido el dominio de los Registros del Catálogo (*Catalog Record*), por tratarse según su propia definición en el estándar DCAT⁷⁴ de un elemento opcional cuya utilidad se limita a aquellos casos en los que el catálogo realiza distinción entre los metadatos de un conjunto de datos cualquiera y los metadatos del registro que se crea en el catálogo para ese conjunto. Esta cualidad no es aplicable a todos los catálogos y dependerá de su implementación final, quedando además excluida de la NTI y siendo también opcional en el DCAT-AP Europeo.

En las siguientes secciones se detalla el análisis de cada uno de los dominios mencionados (**Si**, **No** o **Parcial**) tanto a nivel de metadato como para su implementación en RDF con DCAT, si bien es ya conocido de antemano que el soporte de DCAT en la herramienta CKAN es de carácter bastante básico, incluyendo las versiones más recientes.

Una tabla completa con el análisis comparativo de los metadatos disponibles en las distintas especificaciones objeto de este estudio y su estado de implementación en CKAN 1.8 puede consultarse también en el Anexo I (*Análisis de Metadatos*).

3.3.2.1 Catálogo (*Catalog*)

CKAN no proporciona actualmente exportación de ningún tipo de metadato a nivel de Catálogo a través de DCAT ni tampoco acceso directo programáticamente a través de las entidades de dominio implementadas en su modelo de datos⁷⁵, por lo que la implementación de la NTI deberá a priori hacerse *ad-hoc* en este caso.

⁷⁴ DCAT Catalog record class: <http://www.w3.org/TR/vocab-dcat/#class-catalog-record>

⁷⁵ CKAN Domain model: <http://docs.ckan.org/en/ckan-1.8/domain-model.html>

Metadato NTI	Propiedad RDF	Estado		Comentarios
REQUERIDOS		Meta.	RDF	
Nombre	dct:title	N	N	
Descripción	dct:description	N	N	
Órgano publicador	dct:publisher	N	N	
Fecha de creación	dct:issued	N	N	Ver Nota 4 Caso 1
Fecha de actualización	dct:modified	N	N	Ver Nota 4 Caso 1
Idioma(s)	dc:language dct:language	N	N	Ver Nota 1, Nota 4 C1, Nota 8
Temáticas	dcat:themeTaxonomy	N	N	Ver Nota 4 Caso 2
Página Web	foaf:homepage	N	N	Ver Nota 4 Caso 1
Términos de uso	dct:license	N	N	Ver Nota 4 Caso 1
Documento(s) y recurso(s)	dcat:dataset	N	N	
OPCIONALES				
Tamaño del catálogo	dct:extent	N	N	
Identificador	dct:identifier	N	N	
Cobertura geográfica	dct:spatial	N	N	Ver Nota 4 Caso 2
Derechos	dct:rights	N	N	
Registro	dcat:record	N	N	

Una vez hayamos cumplido con los metadatos requeridos por la NTI estaríamos también cumpliendo con los metadatos obligatorios y recomendados por el DCAT-AP Europeo, por lo que el resto de metadatos serían completamente opcionales.

La opción más recomendable es que los metadatos de Catálogo estén directamente ubicados en la URI prevista para su identificación según el diseño del Esquema:

URI para identificar el catálogo: `http://{base}/catalogo`

`http://opendata.aragon.es/catalogo`

3.3.2.2 Conjunto de Datos (Dataset)

CKAN proporciona cierta implementación básica de los metadatos para los Conjuntos de Datos (*Datasets*) en formato RDF legible por las máquinas, aunque con varios problemas que será necesario corregir.

En la siguiente tabla se hace un resumen del estado de implementación de cada una de las propiedades en las distintas normas que son objeto de este análisis. En cada caso se indica tanto si el metadato es recogido o no en el portal como si se proporciona además en formato RDF para su reutilización.

Metadato NTI	Propiedad RDF	Estado		Comentarios
REQUERIDOS		Meta.	RDF	
Nombre	dct:title	S	S	Ver Nota 3
Descripción	dct:description	S	S	Ver Nota 3
Temática(s)	dcat:theme	S	N	Ver Nota 4 C2
Organismo publicador	dct:publisher	S	N	
Distribución(es)	dcat:distribution	S	N	Ver Nota 11
RECOMENDADOS – DCAT-AP				
Etiqueta(s)	dcat:keyword	S	S	Ver Nota 3
Contacto	dcat:contactPoint adms:contactPoint	S	N	Ver Nota 8
OPCIONALES				
Identificador	dct:identifier	S	P	Ver Nota 2, Nota 4 C3
Fecha de creación	dct:issued	S	N	Ver Nota 4 C1
Fecha de actualización	dct:modified	S	N	Ver Nota 2, Nota 4 C1, Nota 9
Frecuencia de actualización	dct:accrualPeriodicity	S	N	Ver Nota 9
Idioma	dc:language dct:language	S	P	Ver Nota 4 C1, Nota 5, Nota 8
Condiciones de uso	dct:publisher	S	N	
Cobertura geográfica	dct:spatial	S	P	Ver Nota 4 C2, Nota 5
Cobertura temporal	dct:temporal	S	P	Ver Nota 5
Vigencia	dct:valid	N	N	
Recursos relacionado	dct:references	P	N	Ver Nota 10
Normativa	dct:conformsTo	N	N	Ver Nota 7
Página de acceso	dcat:landingPage	S	N	
Identificador secundario	adms:identifier	N	N	Ver Nota 7
Versión	adms:version	S	N	Ver Nota 7
Notas de la versión	adms:versionNotes	N	N	Ver Nota 7

En el caso de los conjuntos de datos, los metadatos deben estar directamente asociados a los URIs correspondientes que identifican cada conjunto según el diseño del esquema establecido.

URIs para identificar los conjuntos de datos: `http://{base}/catalogo/{dataset}`

Ej. `http://opendata.aragon.es/catalogo/nombre-dataset1`

3.3.2.3 Distribución (Distribution)

En el caso de las distribuciones la implementación RDF actual ofrecida por CKAN es nuevamente insuficiente y limitada tan sólo a ciertos elementos opcionales. Nuevamente se distingue entre los casos en los que el metadato está siendo gestionado por el portal y aquellos en los que además se proporciona en formato RDF que facilita su reutilización.

Metadato NTI	Propiedad	Estado		Comentarios
REQUERIDOS		Meta.	RDF	
URL de acceso	dc:accessURL	S	N	
Formato	dc:mediaType	S	N	
RECOMENDADOS – DCAT-AP				
Descripción	dct:description	S	N	
Condiciones de uso	dct:license	S	N	
Formato	dct:format	S	N	
OPCIONAL				
Identificador	dct:identifier	S	N	Ver Nota 2, Nota 11
Nombre	dct:title	S	P	Ver Nota 3, Nota 6
Tamaño	dc:byteSize	S	N	Ver Nota 4 Caso 1
Información adicional	dct:relation	N	P	Ver Nota 5
Fecha de creación	dct:issued	S	N	Ver Nota 11
Fecha de actualización	dct:modified	S	N	Ver Nota 11
Derechos	dct:rights	N	N	
URL de descarga	dc:downloadURL	S	N	
Estado	adms:status	N	N	Ver Nota 7

En este caso los metadatos generalmente no aparecerán de forma independiente, sino asociados al conjunto de datos correspondiente (ver ejemplo 3.3.4.4). Por lo tanto se accederá a ellos a través de los URIs que identifican al conjunto de datos asociado.

URIs para identificar los conjuntos de datos: `http://{base}/catalogo/{dataset}`

Ej. `http://opendata.aragon.es/catalogo/nombre-dataset1`

3.3.3 Comentarios y cuestiones adicionales

Nota 1: Si bien en la NTI la propiedad de representación indicada es `dc:language`, perteneciente al espacio de nombres de elementos⁷⁶ de Dublin Core y no al de términos⁷⁷, dado que:

- este último espacio de nombres incluye también los elementos del anterior;
- además dicho espacio de nombres ha sido creado como soporte a las buenas prácticas de la Web Semántica en cuanto a la definición de rangos y dominios;
- la compatibilidad hacia atrás se mantiene por estar las propiedades definidas en el espacio de nombres de términos declaradas como subpropiedades del espacio de nombres de elementos;
- tanto la especificación de DCAT del W3C como el DCAT-AP Europeo se utiliza este último espacio de nombres;

Se recomienda entonces utilizar la propiedad definida en el espacio de nombres de términos (`dct`) en su lugar. No obstante será necesario antes consultar la validez oficial de dicha interpretación para evitar problemas futuros con el cumplimiento de la NTI.

Nota 2: Debido a lo que parecen ser simples errores tipográficos algunas de las propiedades de representación utilizadas en la NTI son incorrectas y en este documento aparecen directamente corregidas. Dichas propiedades incluyen:

Propiedad	Original en la NTI	Dominio
<code>dct:identifier</code>	<code>det:identifie</code>	Dataset
<code>dct:modified</code>	<code>det:modified</code>	Dataset
<code>dct:identifier</code>	<code>la dct:identifie</code>	Distribution

Nota 3: Se han detectado problemas con la **codificación de caracteres** en los valores literales de algunas de las propiedades, lo que da lugar a caracteres extraños que dificultan o imposibilitan la legibilidad de dichos valores. Dado que el análisis se ha realizado sobre un muestreo aleatorio de conjuntos de datos del portal, y que no todos los literales analizados presentaban caracteres potencialmente conflictivos, es posible

⁷⁶ Dublin Core metadata element set: <http://dublincore.org/documents/dces/>

⁷⁷ DCMI Metadata terms: <http://purl.org/dc/terms/>

que estos problemas de codificación existan también en otras propiedades literales, por lo que se deberán revisar en general.

Nota 4: Pueden existir algunos **conflictos entre los tipos y esquemas de valores** definidos por la NTI y los rangos establecidos para las mismas propiedades en las otras especificaciones, especialmente el DCAT-AP Europeo. A continuación se describe como poder abordar dichos conflictos según su tipo:

CASO 1: Cuando la NTI establece un **tipo de datos menos estricto pero compatible** con los definidos para las propiedades equivalentes en DCAT ó DCAT-AP se recomienda utilizar el tipo más estricto para maximizar la compatibilidad.

Por ejemplo, la NTI establece que las fechas deben usar un formato Literal (texto) siguiendo el formato ISO-8601: `YYYY-MM-DDThh:mm:ssTZD`; sin embargo tanto DCAT como DCAT-AP añaden el requisito adicional de utilizar los tipos de XML Schema `date` or `dateTime`. Por lo tanto, la recomendación en este caso sería utilizar un Literal de texto con formato ISO-8601 y aplicando uno de los tipos indicados para maximizar la compatibilidad, como en el siguiente ejemplo (ver también sección de Ejemplos):

```
...  
< dct:issued rdf:datatype="xsd:dateTime">2013-11-11T12:00:00CET</ dct:issued>  
...
```

CASO 2: Cuando la NTI recomienda utilizar un **Recurso específico** de una de las taxonomías establecidas por el Catálogo Nacional y en DCAT ó DCAT-AP el ámbito es más abierto a cualquier Recurso previamente definido se recomienda reutilizar los Recursos de la taxonomías ya definidas cuando sean de aplicación, pero pudiendo utilizar también otros recursos de taxonomías propias o de terceros cuando lo necesitemos, ya que no se originaría ningún problema de compatibilidad.

Por ejemplo, la NTI establece taxonomías e identificadores correspondientes a distintos recursos geográficos del territorio español –País, Autonomías y Provincias; sin embargo no se incluyen otros recursos territoriales de menor nivel como los Municipios. Por lo tanto, la recomendación en este caso sería utilizar las taxonomías ya definidas por la NTI cuando necesitemos hacer referencia al País, las Autonomías o las Provincias, pero podríamos también utilizar otra taxonomía de referencia para los Municipios o incluso crear la nuestra propia en caso de no existir ninguna previa.

CASO 3: Cuando la NTI establece un **tipo de datos más estricto pero compatible** con los definidos para las propiedades equivalentes en DCAT ó DCAT-AP se recomienda nuevamente utilizar el tipo más estricto para maximizar la compatibilidad.

Por ejemplo, establece que los identificadores de los conjuntos de datos deben ser del tipo `anyURI` definido por XML Schema cuando tanto DCAT como DCAT-AP permiten cualquier tipo de Literal (texto). Por lo tanto la recomendación en este caso sería utilizar un Literal de texto aplicando el tipo `anyURI` para maximizar la compatibilidad, como en el siguiente ejemplo (ver también sección de Ejemplos):

```
...
< dct:identifier
rdf:datatype="xsd:anyURI">http://opendata.aragon.es/id101</ dct:identifier>
...
```

Nota 5: La implementación que se hace de DCAT en CKAN 1.8 utiliza con cierta frecuencia **atributos genéricos** `dct:relation` para hacer referencia a entidades descritas mediante pares etiquetas-valor, dando lugar a una semántica genérica *pobre* que hace que en gran parte se pierda el potencial valor añadido al utilizar tecnologías de Web Semántica.

Por ejemplo, la descripción de los idiomas se realiza de la siguiente forma:

```
...
< dct:relation>
  < rdf:Description>
    < rdfs:label>Language</ rdfs:label>
    < rdf:value>es</ rdf:value>
  </ rdf:Description>
</ dct:relation>
...
```

Que, si bien es sintácticamente correcta, semánticamente sería mucho más adecuado y simple representarlo de la siguiente forma:

```
...
< dct:language>es</ dct:language>
...
```

Nota 6: Es importante no sólo utilizar los metadatos adecuados, sino que los **valores** que se aplican a dichos metadatos también lo sean. En el caso de los títulos y descripciones deben siempre ser representativos para el caso concreto y evitarse la utilización de textos genéricos.

Por ejemplo, la implementación analizada de DCAT en CKAN reutiliza con cierta frecuencia los valores de algunas propiedades para entidades distintas, como igual título para el *Dataset* y la *Distribución* o igual Organismo o Entidad como autor y al mismo tiempo colaborador de un Dataset específico.

Nota 7: Algunas propiedades no están incluidas dentro del estándar DCAT, pero al ser parte de las propiedades que el NTI, el DCAT-AP Europeo o ambas normas toman de otros vocabularios de referencia, se incluyen igualmente en el estudio.

Nota 8: En algún caso excepcional las distintas normas y estándares analizados (NTI, DCAT, DCAT-AP) pueden haber optado por utilizar distintas propiedades para representar metadatos iguales. Esto es debido generalmente a algunos problemas de sincronización entre las distintas normativas durante su desarrollo (hay que tener en cuenta tanto DCAT como DCAT-AP estuvieron en estado de borrador hasta hace relativamente poco tiempo).

En esos casos, si bien no existe realmente un problema semántico, al ser generalmente las distintas propiedades utilizadas equivalentes, sí puede existir un problema sintáctico

que afecte a las distintas herramientas que vayan a tratar esos datos. Por lo tanto, se recomienda la utilización redundante de ambas propiedades para garantizar la compatibilidad. En el siguiente ejemplo en el que se utilizan propiedades redundantes de contacto para garantizar la compatibilidad tanto con DCAT como con DCAT-AP (ver también la sección de Ejemplos):

```
...  
<dc:contactPoint rdf:resource="http://opendata.aragon.es/contacto"/>  
<adms:contactPoint rdf:resource="http://opendata.aragon.es/contacto"/>  
...
```

Nota 9: Si bien en el Anexo III de la NTI donde se recogen los metadatos asociados a los documentos y recursos de información se indica que tanto la fecha de actualización como la frecuencia de la misma son opcionales en el caso de los recursos de información (*Dataset*), en el apartado III de la misma – Selección de información reutilizable – se indica también en el punto 5 que los documentos y recursos de información deberán indicar la fecha de última actualización y el periodo de la misma. Por tanto, es recomendable incluir estos metadatos también como obligatorios para evitar posibles problemas de cumplimiento de la NTI.

Nota 10: En este caso el metadato se limita a un tipo específico de recurso relacionado (diccionario de datos) y además no se adecúa al tipo de datos esperado (en la actualidad texto *embebido* en lugar de referencia al recurso en cuestión)

Nota 11: Los valores en estos casos se generan de forma automática por parte del sistema.

3.3.4 Ejemplos

Los siguientes ejemplos muestran el conjunto mínimo de metadatos que se debería crear para cumplir con la NTI y los metadatos obligatorios y recomendados por el DCAT-AP Europeo (indicados estos últimos en gris y cursiva en el texto) por el DCAT-AP. Un ejemplo con todos los metadatos puede consultarse en el Anexo VI de la propia Norma. Los vocabularios mínimos que será necesario utilizar para ello son los siguientes:

Vocabulario	URI
RDF Syntax	rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
XML Schema	xsd: http://www.w3.org/2001/XMLSchema#
Dataset Catalog (DCAT)	dc: http://www.w3.org/ns/dcat#
Dublin Core Terms (DC)	dct: http://purl.org/dc/terms/
Friend Of A Friend (FOAF)	foaf: http://xmlns.com/foaf/0.1/

3.3.4.1 Ejemplo de metadatos de Catálogo

```
<rdf:RDF xmlns:dct="http://purl.org/dc/terms/"  
  xmlns:dcat="http://www.w3.org/ns/dcat#"
```

```
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

<dcatalog:Catalog rdf:about="http://opendata.aragon.es/catalogo/">
  <dc:title xml:lang="es">Título del Catálogo</dc:title>
  <dc:description xml:lang="es">Descripción del Catálogo</dc:description>
  <dc:publisher rdf:resource="http://aragon.es/" />
  <dc:issued rdf:datatype="xsd:dateTime">2013-11-
11T12:00:00CET</dc:issued>
  <dc:modified rdf:datatype="xsd:dateTime">2013-11-
11T12:00:00CET</dc:modified>
  <dc:language>es</dc:language>
  <dc:themeTaxonomy rdf:resource="http://opendata.aragon.es/temas" />
  <foaf:homepage rdf:resource="http://opendata.aragon.es/" />
  <dc:license rdf:resource="http://opendata.aragon.es/portal/terminos" />
  <dc:dataset rdf:resource="http://opendata.aragon.es/data/dataset1" />
  <dc:dataset rdf:resource="http://opendata.aragon.es/data/dataset2" />
  <dc:dataset rdf:resource="http://opendata.aragon.es/data/dataset3" />
</dcatalog:Catalog>
</rdf:RDF>
```

3.3.4.2 Ejemplo de metadatos de Conjunto de Datos

```
<rdf:RDF xmlns:dct="http://purl.org/dc/terms/"
  xmlns:dcatalog="http://www.w3.org/ns/dcat#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

  <dcatalog:Dataset rdf:about="http://opendata.aragon.es/data/dataset1/">
    <dc:title xml:lang="es">Título del Conjunto de Datos</dc:title>
    <dc:description xml:lang="es">Descripción del Conjunto</dc:description>
    <dc:theme rdf:resource="http://opendata.aragon.es/temas/tema1" />
    <dc:theme rdf:resource="http://opendata.aragon.es/temas/tema2" />
    <dc:theme rdf:resource="http://opendata.aragon.es/temas/tema3" />
    <dc:keyword>palabraClave1</dc:keyword>
    <dc:keyword>palabraClave2</dc:keyword>
    <dc:keyword>palabraClave3</dc:keyword>
    <dc:publisher rdf:resource="http://opendata.aragon.es/entidades/sarga" />
    <dc:distribution
rdf:resource="http://opendata.aragon.es/data/dataset1/distribution1" />
    <dc:distribution
rdf:resource="http://opendata.aragon.es/data/dataset1/distribution2" />
    <dc:contactPoint
rdf:resource="http://opendata.aragon.es/entidades/sarga/contacto" />
    <adms:contactPoint
rdf:resource="http://opendata.aragon.es/entidades/sarga/contacto" />
  </dcatalog:Dataset>
</rdf:RDF>
```

3.3.4.3 Ejemplo de metadatos de Distribución

```
<rdf:RDF xmlns:dct="http://purl.org/dc/terms/"
  xmlns:dcatalog="http://www.w3.org/ns/dcat#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
```



```
<dcat:Distribution
rdf:about="http://opendata.aragon.es/data/dataset1/distribution1">
  <dcat:accessURL rdf:datatype="xsd:anyURI">
http://opendata.aragon.es/data/dataset1/dataset1.csv</dcat:accessURL>
  <dcat:mediaType>
    < dct:IMT>
      <rdf:value>text/csv</rdf:value>
      <rdfs:label>CSV</rdfs:label>
    </dct:IMT>
  </dcat:mediaType>
  <dct:description xml:lang="es">Descripción de la
Distribución</dct:description>
  <dct:license rdf:resource="http://sarga.es/terminos"/>
</dcat:Distribution>
</rdf:RDF>
```

3.3.4.4 Ejemplo de metadatos de Distribución asociado a un Conjunto de Datos

```
<rdf:RDF xmlns:dct="http://purl.org/dc/terms/"
xmlns:dcat="http://www.w3.org/ns/dcat#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

<dcat:Dataset rdf:about="http://opendata.aragon.es/data/dataset1/">
  <dct:title xml:lang="es">Título del Conjunto de Datos</dct:title>
  <dct:description xml:lang="es">Descripción del Conjunto</dct:description>
  <dcat:theme rdf:resource="http://opendata.aragon.es/temas/tema1"/>
  <dcat:theme rdf:resource="http://opendata.aragon.es/temas/tema2"/>
  <dcat:theme rdf:resource="http://opendata.aragon.es/temas/tema3"/>
  <dcat:keyword>palabraClave1</dcat:keyword>
  <dcat:keyword>palabraClave2</dcat:keyword>
  <dcat:keyword>palabraClave3</dcat:keyword>
  <dct:publisher rdf:resource="http://opendata.aragon.es/entidades/sarga"/>
  <dcat:distribution>
    <dcat:Distribution
rdf:about="http://opendata.aragon.es/data/dataset1/distribution1">
      <dcat:accessURL rdf:datatype="xsd:anyURI">
http://opendata.aragon.es/data/dataset1/dataset1.csv</dcat:accessURL>
      <dcat:mediaType>
        < dct:IMT>
          <rdf:value>text/csv</rdf:value>
          <rdfs:label>CSV</rdfs:label>
        </dct:IMT>
      </dcat:mediaType>
      <dct:description xml:lang="es">Descripción de la
Distribución</dct:description>
      <dct:license rdf:resource="http://sarga.es/terminos"/>
    </dcat:Distribution>
  </dcat:distribution>
  <dcat:contactPoint
rdf:resource="http://opendata.aragon.es/entidades/sarga/contacto"/>
    <adms:contactPoint
rdf:resource="http://opendata.aragon.es/entidades/sarga/contacto"/>
  </dcat:Dataset>
</rdf:RDF>
```

3.3.5 Recomendaciones

Si bien es evidente que en general en el portal actual se están recopilando los metadatos de una forma bastante completa, a la hora de presentarlos en el formato estándar de intercambio de datos DCAT, en el que se basan tanto la NTI como el DCAT-AP Europeo, se delega en la implementación que realiza CKAN de DCAT⁷⁸, que en la actualidad es bastante limitada - e incluso inexistente para los metadatos del propio Catálogo - dando lugar a documentos de intercambio de metadatos incompletos y con problemas de semántica.

La opción disponible para corregir estos problemas sería actuar directamente sobre las plantillas de generación de los ficheros RDF en CKAN, y concretamente el fichero "read.rdf"⁷⁹, localizado en:

```
(ckan_path)/default/src/ckan/ckan/templates_legacy/package/
```

La sintaxis utilizada para la configuración de dichos ficheros no es especialmente intuitiva, por lo que será necesario también consultar la documentación de las funciones auxiliares⁸⁰ para las plantillas que se utilizan.

No obstante, también hay que tener en cuenta que no todos los datos que necesitamos para construir la plantilla de forma correcta estarán disponibles a través de las entidades de dominio del modelo de datos de CKAN⁸¹ (ver Anexo I), por lo que en algún caso podría ser necesario llegar a implementar dicha propiedad de forma "manual", estrategia que también deberemos seguir a la hora de generar el fichero DCAT para el Catálogo (no confundir con los metadatos actuales que genera CKAN para los conjuntos de datos), hasta ahora inexistente.

Una posible alternativa y solución de futuro podría pasar por el uso de la extensión de CKAN para DCAT⁸². No obstante, hay que tener en cuenta que ésta se encuentra todavía en fase de desarrollo y no ofrece aún una implementación completa del estándar.

3.4 Taxonomías

En la actualidad el portal Aragón Open Data utiliza únicamente una taxonomía propia para la categorización de contenidos, aún cuando esta no ha sido formalmente descrita. Tanto en este caso como para posibles taxonomías nuevas (organismos, formatos,

⁷⁸ CKAN Linked Data and RDF: <http://docs.ckan.org/en/latest/linked-data-and-rdf.html>

⁷⁹ CKAN RDF template:

https://github.com/okfn/ckan/blob/master/ckan/templates_legacy/package/read.rdf

⁸⁰ CKAN Template helper functions: <http://docs.ckan.org/en/847-new-theming-docs/template-helper-functions.html>

⁸¹ CKAN Domain model: <http://docs.ckan.org/en/ckan-1.8/domain-model.html>

⁸² CKAN DCAT extensión: <https://github.com/okfn/ckanext-dcat>

licencias, etc.) la recomendación inicial sería utilizar siempre SKOS⁸³ como vocabulario de representación formal de dichas taxonomías y asociarlas a los identificadores correspondientes según lo definido en el esquema de URIs:

URIs para identificar esquemas de conceptos: `http://{base}/kos/{sector}/{dominio}`

Ej. `http://opendata.aragon.es/kos/seguridad/delitos`

URIs para identificar conceptos: `http://{base}/kos/{sector}/{dominio}#{Concepto}`

Ej. `http://opendata.aragon.es/kos/seguridad/delitos#Homicidio`

3.4.1 Sectores

Si bien la taxonomía de categorización de contenidos utilizada se asemeja bastante a la categorización propuesta en el Anexo IV de la NTI como taxonomía de sectores primarios, ésta no es completamente equivalente.

En la siguiente tabla se realiza un mapeo entre los sectores primarios definidos por Aragón Open Data a la hora de categorizar los conjuntos de datos y sus equivalencias tanto en la taxonomía de sectores de la NTI (teniendo también en cuenta su descripción temática⁸⁴) como en EUROVOC⁸⁵, el tesoro multilingüe de la Unión Europea que adopta el DCAT-AP Europeo como sistema de clasificación primaria de contenidos.

Aragón OD	Taxonomía NTI	EUROVOC	Comentarios
Administración Pública	Sector público	04 Vida Política	
Asuntos Sociales	Sociedad y Bienestar Demografía	28 Asuntos Sociales	Ver Nota 1
Ciencia y Tecnología	Ciencia y Tecnología	36 Ciencia 32 Educación y Comunicación 64 Producción, Tecnología e Investigación	
Cultura	Cultura y Ocio	28 Asuntos Sociales	
Demografía	Demografía	28 Asuntos Sociales	
Deporte	Deporte	28 Asuntos Sociales	

⁸³ SKOS Simple Knowledge Organization System

Reference: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>

⁸⁴ Documento de trabajo de la Taxonomía del Catálogo datos.gob.es:

http://datos.gob.es/datos/sites/default/files/files/12_tax_02.pdf

⁸⁵ EUROVOC, Tesoro multilingüe de la Unión Europea:

<http://eurovoc.europa.eu/drupal/?q=es>

Economía	Economía	16 Vida Económica 24 Asuntos Financieros	
Educación	Educación	32 Educación y Comunicación	
Empleo	Empleo	44 Trabajo y Empleo	
Interior y Justicia	Legislación y Justicia Sector Público Seguridad	12 Derecho	Ver Nota 2
Medio Ambiente	Medio Ambiente	52 Medio Ambiente 72 Geografía	
Movilidad e Infraestructuras	Transporte Urbanismo e Infraestructuras	48 Transportes 28 Asuntos Sociales	Ver Nota 3
Ordenación del Territorio	Medio Ambiente Urbanismo e Infraestructuras	52 Medio Ambiente 72 Geografía 28 Asuntos Sociales	Ver Nota 4
Salud	Salud	28 Asuntos Sociales	
Turismo	Turismo	28 Asuntos Sociales	
Urbanismo y Vivienda	Urbanismo e Infraestructuras Vivienda	28 Asuntos Sociales	
	Comercio	20 Intercambios Económicos y Comerciales	Ver Nota 5
	Energía	66 Energía	Ver Nota 5
	Hacienda	24 Asuntos Financieros	Ver Nota 5

NOTA 1: Varios de los conjuntos de datos actualmente catalogados como *Asuntos Sociales* pasarían a formar parte de *Demografía*.

NOTA 2: Si bien los conjuntos actuales corresponderían probablemente al área de *Sector Público* es muy probable que otros potenciales conjuntos de esta Categoría tengan su correspondencia en *Legislación y Justicia* o *Seguridad* (tópico éste último tradicionalmente considerado como parte de “*Interior*”).

NOTA 3: Si bien actualmente no existen conjuntos en esta categoría su correspondencia sería claramente la indicada.

NOTA 4: Los datos *Cartográficos* y *Geoespaciales* tendrían su correspondencia en la categoría de *Urbanismo e Infraestructuras*.

NOTA 5: Algunos sectores de la NTI no tienen actualmente un equivalente en las categorías de Aragón Open Data, pero hay que tener en cuenta que en el futuro podrían introducirse nuevos conjuntos de datos que correspondiesen a estos sectores.

3.4.1.1 Recomendaciones

Para garantizar la compatibilidad entre el sistema de categorías utilizado por Aragón Open Data y los sectores definidos por la NTI existen dos aproximaciones posibles:

1. Mantener las categorías actuales, haciendo un mapeo semántico entre dichas categorías y los sectores de la NTI a través de las propiedades disponibles en SKOS al respecto (`closeMatch`, `exactMatch`, `broadMatch`, `narrowMatch` y `relatedMatch`). Por ejemplo:

```
<rdf:RDF xmlns:skos="http://www.w3.org/2004/02/skos/core">

  <skos:Concept
rdf:about="http://opendata.aragon.es/kos/categorias/administracion-publica">
    <skos:prefLabel xml:lang="es">Administración Pública</skos:prefLabel>
    <skos:closeMatch rdf:resource="http://datos.gob.es/kos/sector-
publico/sector/sector-publico"/>
  </skos:Concept>

  <skos:Concept rdf:about="http://opendata.aragon.es/kos/categorias/cultura">
    <skos:prefLabel xml:lang="es">Cultura</skos:prefLabel>
    <skos:broadMatch rdf:resource="http://datos.gob.es/kos/sector-
publico/sector/cultura-ocio"/>
  </skos:Concept>

  <skos:Concept rdf:about="http://opendata.aragon.es/kos/categorias/deporte">
    <skos:prefLabel xml:lang="es">Deporte</skos:prefLabel>
    <skos:exactMatch rdf:resource="http://datos.gob.es/kos/sector-
publico/sector/deporte"/>
  </skos:Concept>

  <skos:Concept rdf:about="http://opendata.aragon.es/kos/categorias/urbanismo-
vivienda">
    <skos:prefLabel xml:lang="es">Urbanismo y Vivienda</skos:prefLabel>
    <skos:closeMatch rdf:resource="http://datos.gob.es/kos/sector-
publico/sector/urbanismo-infraestructuras"/>
    <skos:narrowMatch rdf:resource="http://datos.gob.es/kos/sector-
publico/sector/vivienda"/>
  </skos:Concept>

</rdf:RDF>
```

2. Adoptar directamente la taxonomía de sectores de la NTI sustituyendo completamente a la categorización actual mediante la siguiente tabla de equivalencias:

Taxonomía NTI	Aragón OD
Ciencia y tecnología	Ciencia y tecnología
Comercio	
Cultura y ocio	Cultura
Demografía	Demografía Asuntos Sociales

Deporte	Deporte
Economía	Economía
Educación	Educación
Empleo	Empleo
Energía	
Hacienda	
Industria	Economía Medio Ambiente
Legislación y justicia	Interior y Justicia
Medio ambiente	Medio ambiente Ordenación del Territorio
Medio Rural	Economía
Salud	Salud
Sector público	Administración publica Interior y Justicia
Seguridad	Interior y Justicia
Sociedad y bienestar	Asuntos sociales
Transporte	Movilidad e infraestructuras Medio Ambiente
Turismo	Turismo
Urbanismo e infraestructuras	Movilidad e infraestructuras Urbanismo y Vivienda Ordenación del Territorio
Vivienda	Urbanismo y Vivienda

Si bien quizás la solución más rápida sencilla a corto plazo sería optar por el mapeo de categorías propuesto en la opción 1, es posible sin embargo que a largo plazo – y según vaya creciendo la complejidad y tamaño del catálogo – ésta suponga una opción difícil de mantener y que pueda dar lugar a inconsistencia e incompatibilidades.

Por tanto la recomendación en este caso sería optar por la segunda opción propuesta y **adoptar directamente la misma taxonomía que utiliza la NTI** aunque pueda suponer un esfuerzo mayor en cuanto a la transición inicial.

Por otro lado, la compatibilidad con EUROVOC en principio no sería una cuestión prioritaria, ya que es previsible que en un futuro próximo se realice el mapeo correspondiente entre la taxonomía de sectores de la NTI y EUROVOC. Por ello, tanto si se decide adoptar directamente la taxonomía de la NTI como si se opta por hacer un

mapeo entre el sistema de categorías actual y dicha taxonomía, en ese momento tendríamos ya directamente una correspondencia establecida también con EUROVOC.

3.4.2 Recursos geográficos

Para el caso de los ámbitos **nacional**, **autonómico** y **provincial** la recomendación es reutilizar directamente las taxonomías definidas al respecto por la NTI en su anexo V.

Para otros casos más específicos (**municipal** y/o **comarcal**), que quedan fuera del ámbito de la norma nacional, se puede optar por reutilizar algún servicio externo como GeoNames⁸⁶ para simplificar la tarea pero, dado los problemas de autenticidad y gestión de los datos a los que puede dar lugar, la recomendación es modelar una taxonomía propia que sirva como referencia regional a partir de alguna fuente de datos oficial fiable como puede ser el Nomenclátor⁸⁷.

⁸⁶ GeoNames Search Web Service: <http://www.geonames.org/export/geonames-search.html>

⁸⁷ Nomenclátor de Aragón: http://www.aragon.es/DepartamentosOrganismosPublicos/Organismos/InstitutoAragonesEstadistica/AreasTematicas/02_Demografia_Y_Poblacion/01_CifrasPoblacion_Y_Censos/01_Padron/ci.05_Nomenclator.detalleDepartamento#section1