



– Opendata Euskadi –
Portal de Datos Abiertos del Gobierno Vasco

Denominación:

Modelo de Contenido y MetaDatos

Datos reutilizables

Autor

Departamento de Justicia y Administración Pública del Gobierno Vasco
Dirección de Innovación y Administración Electrónica

Contenido

Capítulo/sección	Página
1 Introducción	3
2 Modelo de Contenido	4
2.1 Objetos del Modelo y sus MetaDatos	5
2.1.1 Contenido	5
2.2 Documento / Versión	7
2.3 DataFile	8
3 Persistencia del Modelo de Contenido	9
3.1 Detalle de la trasposición física de un contenido	10
3.1.1 XMLs reutilizables	11
3.1.2 Descriptores del Contenido	12
3.1.3 Documento = Versión del contenido	16
3.1.4 Datos comunes del Contenido	20
4 Despliegue de Contenidos (publicación)	21
5 Estructura de URLs	22
5.1 URL directa de un contenido o una página de portal	22
5.2 Combinar un contenido en una página de portal	22



1 Introducción

El presente documento contiene un resumen del **modelo** en el que se basan los contenidos de Euskadi.net. Cualquier contenido en euskadi.net respeta este modelo lo que garantiza:

- Que cualquier contenido de euskadi.net tenga un mínimo de metadatos comunes
- Una estructura física de los contenidos que favorece la organización y mantenimiento

Este documento incluye una descripción de los objetos así como sus propiedades –metadatos– y cómo se traspone este modelo físicamente en los servidores.

Audiencia: Analistas y Desarrolladores

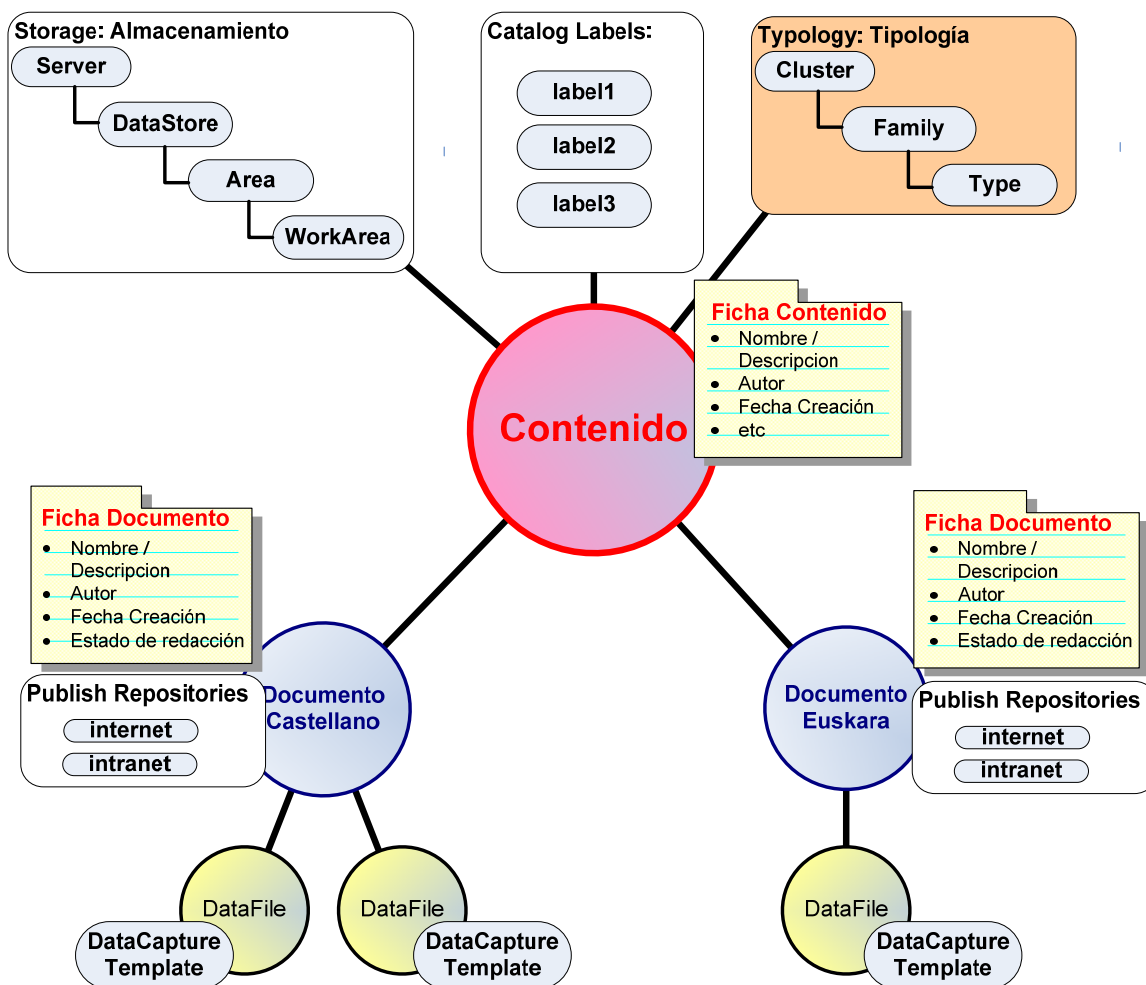
Complejidad: Media / Alta

2 Modelo de Contenido

Todos los contenidos de la iniciativa euskadi.net (y en particular de opendata.euskadi.net) siguen un **modelo de contenido único** que encapsula:

- El comportamiento de los contenidos
- Los metaDatos (datos) mínimos de un contenido.

Toda la arquitectura de soporte al Modelo de Presencia en Internet del Gobierno Vasco se basa en el **modelo de contenido**, cuyos metaDatos básicos se pueden ver en la siguiente figura:



Todo contenido tiene **dos tipos de metaDatos**:

- **MetaDatos comunes:** aquellos que tienen **todos los contenidos**, independientemente de su naturaleza
Ej: Todas las noticias, eventos, informaciones, ayudas, etc tienen asociado un nombre, una descripción, una tipología y en general una serie de metaDatos comunes.
- **MetaDatos específicos:** aquellos propios del contenido en función de su naturaleza
Ej: Las noticias tienen unos metaDatos particulares como la fecha de la noticia o el titular.
Los procedimientos administrativos a su vez, tienen unos metaDatos particulares como el objeto, el estado de vigencia o el lugar de presentación.

A continuación se detalla cada uno de los metaDatos del modelo ya que su comprensión es básica a la hora de entender el modelo de funcionamiento del buscador.

2.1 Objetos del Modelo y sus MetaDatos

2.1.1 Contenido

El contenido es la **unidad mínima** gestionada en el sistema y es un conjunto de información con significado propio en función de la información que facilita y su pertenencia a una misma materia, área u otro criterio de agrupación.

El contenido es una entidad lógica que tiene los siguientes metaDatos:

MetaDato	Utilidad								
oid	Identificador único del contenido Se trata de una serie de caracteres sin significado que identifican de forma única el contenido								
storeLocation	Datos de almacenamiento del contenido: todo contenido está localizado para su gestión en un lugar identificado por: <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 15%;">server</td> <td>Servidor de Gestión de contenidos</td> </tr> <tr> <td>dataStore</td> <td>Almacén de datos dentro del servidor: Un servidor se puede dividir en varios dataStores donde efectivamente se almacenan los contenidos</td> </tr> <tr> <td>area</td> <td>División de un dataStore en función de criterios organizativos</td> </tr> <tr> <td>workArea</td> <td>División de un area en función de los grupos de personas que trabajan con contenidos. Una personas pertenece a uno o varios grupos de trabajo, cada uno de los cuales tiene una zona de trabajo donde se encuentran sus contenidos.</td> </tr> </table>	server	Servidor de Gestión de contenidos	dataStore	Almacén de datos dentro del servidor: Un servidor se puede dividir en varios dataStores donde efectivamente se almacenan los contenidos	area	División de un dataStore en función de criterios organizativos	workArea	División de un area en función de los grupos de personas que trabajan con contenidos. Una personas pertenece a uno o varios grupos de trabajo, cada uno de los cuales tiene una zona de trabajo donde se encuentran sus contenidos.
server	Servidor de Gestión de contenidos								
dataStore	Almacén de datos dentro del servidor: Un servidor se puede dividir en varios dataStores donde efectivamente se almacenan los contenidos								
area	División de un dataStore en función de criterios organizativos								
workArea	División de un area en función de los grupos de personas que trabajan con contenidos. Una personas pertenece a uno o varios grupos de trabajo, cada uno de los cuales tiene una zona de trabajo donde se encuentran sus contenidos.								
typoInfo	Información de tipificación del contenido: todo contenido está tipificado según su naturaleza (noticia, información, evento, procedimiento administrativo, etc) en una jerarquía de tipificación de tres niveles: <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 15%;">cluster</td> <td>Conjunto de familias de contenidos</td> </tr> <tr> <td>family</td> <td>Conjunto de contenidos con naturaleza similar Ej: Familia Comunicación que contiene los tipos: noticia y nota de prensa Familia Procedimientos Administrativos que contiene los tipos ayuda, contratación, autorización, etc</td> </tr> <tr> <td>Type</td> <td>Tipo concreto de contenido Ej: Noticia, Evento, Información, Nota de Prensa, Ayuda, Procedimiento de Contratación, etc</td> </tr> </table> <p>La tipificación de contenidos es importante ya que el nombre del tipo forma parte de la URL final del contenido desplegado.</p>	cluster	Conjunto de familias de contenidos	family	Conjunto de contenidos con naturaleza similar Ej: Familia Comunicación que contiene los tipos: noticia y nota de prensa Familia Procedimientos Administrativos que contiene los tipos ayuda, contratación, autorización, etc	Type	Tipo concreto de contenido Ej: Noticia, Evento, Información, Nota de Prensa, Ayuda, Procedimiento de Contratación, etc		
cluster	Conjunto de familias de contenidos								
family	Conjunto de contenidos con naturaleza similar Ej: Familia Comunicación que contiene los tipos: noticia y nota de prensa Familia Procedimientos Administrativos que contiene los tipos ayuda, contratación, autorización, etc								
Type	Tipo concreto de contenido Ej: Noticia, Evento, Información, Nota de Prensa, Ayuda, Procedimiento de Contratación, etc								
name	Nombre del contenido El nombre es importante ya que forma parte de la URL final del contenido desplegado en internet								
description	Descripción interna del contenido. De uso (utilidad) únicamente para el usuario que gestiona el contenido								
createDate	Fecha de creación del contenido								
newnessUntilDate	Un contenido puede ser considerado como "novedad" hasta la fecha que se indique en este campo. Esto permite hacer búsquedas de todos los contenidos que son "novedad"								
author	Autor del contenido								
chargePerson	Responsable del contenido: persona que se hace responsable de la información que se suministra en el contenido								

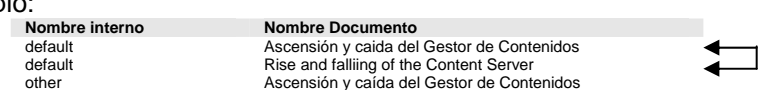
source	Fuente del contenido: lugar de donde se ha tomado la información expresada (agencia de prensa, fuente interna, etc)
labels	<p>Etiquetas en las que está catalogado el contenido</p> <p>Lista de identificadores de las etiquetas de catalogación que el gestor ha “asociado” al contenido.</p> <p>Ej: Si el gestor cataloga un contenido con la etiqueta “empresas” (el oid), a partir de este momento, el contenido se podrá encontrar cuando se busquen todos los contenidos que tienen que ver con “empresas”.</p>
geoLabels	<p>Etiquetas de catalogación geográfica</p> <p>Lista de identificadores de etiquetas de catalogación geográfica que el gestor ha “asociado” al contenido y que permiten geo-localizarlo</p> <p>Ej: Si el gestor cataloga un contenido con la etiqueta “Bilbao” (el oid), a partir de este momento, el contenido se podrá encontrar cuando se busquen todos los contenidos sobre Bilbao por ejemplo para “pintar” un mapa.</p>

2.2 Documento / Versión

Un documento es una **versión** de un contenido, bien sea por idioma, destinatario, etc: El documento contiene toda la información del contenido expresada en un lenguaje concreto y con una narrativa específica para un destinatario (jóvenes, empresarios, etc)

Ej: Un contenido puede expresarse en castellano y con una narrativa simple o directa para que pueda ser entendido por jóvenes o bien expresarse en euskera y con una narrativa técnica para empresarios.

El documento tiene a su vez una serie de metaDatos propios:

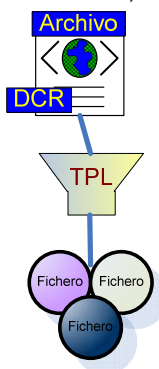
MetaDato	Utilidad								
oid	Identificador único del documento Se trata de una serie de caracteres sin significado que identifican de forma única el documento								
language	Lenguaje del documento								
depthLevel	Nivel de profundidad: complejidad de la narrativa en función del destinatario								
internalName	Nombre interno de un documento: Permite asociar entre sí un documento con sus traducciones Un documento se puede crear “de cero” o partir de otro simplemente traduciéndolo, lo cual crea una copia exacta del contenido original, que posteriormente se traduce. Sin embargo, el documento original y la copia mantienen el mismo nombre interno. Ejemplo: <table border="1" style="margin-left: 20px;"> <thead> <tr> <th>Nombre interno</th> <th>Nombre Documento</th> </tr> </thead> <tbody> <tr> <td>default</td> <td>Ascensión y caída del Gestor de Contenidos</td> </tr> <tr> <td>default</td> <td>Rise and falling of the Content Server</td> </tr> <tr> <td>other</td> <td>Ascensión y caída del Gestor de Contenidos</td> </tr> </tbody> </table> 	Nombre interno	Nombre Documento	default	Ascensión y caída del Gestor de Contenidos	default	Rise and falling of the Content Server	other	Ascensión y caída del Gestor de Contenidos
Nombre interno	Nombre Documento								
default	Ascensión y caída del Gestor de Contenidos								
default	Rise and falling of the Content Server								
other	Ascensión y caída del Gestor de Contenidos								
name	Nombre del documento expresado en el lenguaje del mismo El nombre del documento es muy importante ya que es lo que aparece en los resultados de búsqueda								
description	Descripción del documento expresada en el lenguaje del mismo El nombre del documento es muy importante ya que es lo que aparece en los resultados de búsqueda								
createDate	Fecha de creación del documento								
publicationInfo	Información del publicación: en qué repositorios está publicado el documento Los documentos de un contenido pueden ser publicados independientemente , en distintos repositorios de publicación: Por ejemplo, un contenido puede tener dos documentos A y B, pudiéndose publicar el documento A en Internet y el documento B en Internet e Intranet								

2.3 DataFile

Un documento (versión) en sí mismo es una entidad lógica (no contiene información), sin embargo, los documentos contienen archivos de datos, adjuntos, imágenes, etc que es donde realmente está la información a ofrecer.

Un archivo de datos es por lo tanto un fichero capturado utilizando una **plantilla de introducción de datos** y que contiene los **metaDatos específicos** (aquellos que no son los comunes expuestos con anterioridad).

Ejemplo: Las noticias se “capturan” utilizando una plantilla de introducción de datos específica que permite al usuario introducir metaDatos específicos de las noticias: fecha de la noticia, titular, entrada, desarrollo, conclusión, etc.



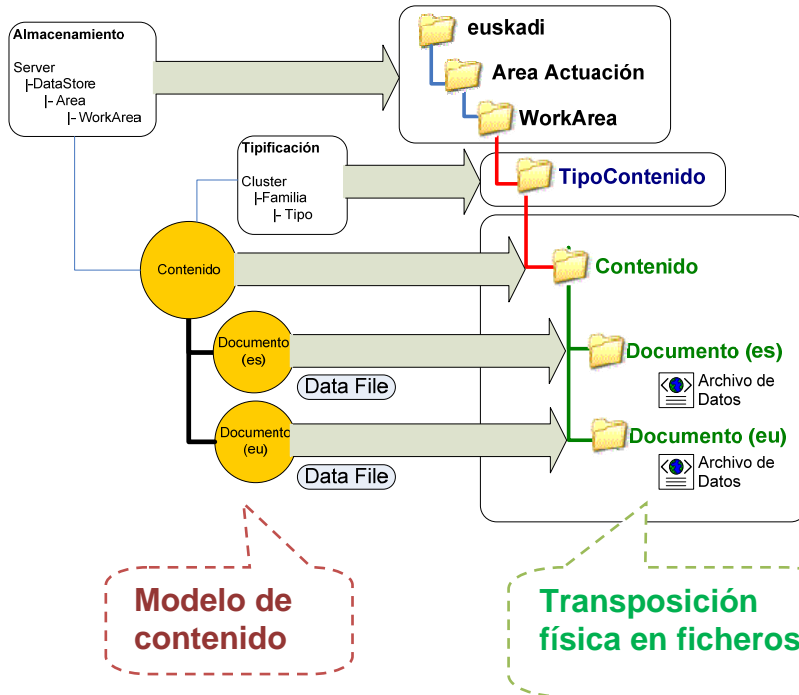
Los ficheros de datos se guardan como XML, sin embargo, para su publicación en Internet hay que “generarlos” en HTML, para lo cual se utilizan **plantillas de presentación**.

Los dataFiles, a su vez tienen una serie de metaDatos:

MetaDato	Utilidad
oid	Identificador único del dataFile Se trata de una serie de caracteres sin significado que identifican de forma única el dataFile
name	Nombre del dataFile El nombre del dataFile es normalmente el nombre del fichero html generado.
description	Una descripción interna (de uso únicamente para el gestor)
templateOid	Identificación de la plantilla de captura utilizada
author	Autor del dataFile
editURL	URL de edición del dataFile
previewURL	URL de previsualización del dataFile en el Gestor de Contenidos
generateURL	URL de generación en HTML (o cualquier otro formato)

3 Persistencia del Modelo de Contenido

La versión que se publica de un contenido está **compuesto únicamente por ficheros en una estructura de carpetas** – recordar que una de las premisas de euskadi.net es que los contenidos se generan en “estático” para ser servidos únicamente por un servidor web –



En la figura se observa la transposición del modelo de contenido en el sistema de ficheros del Gestor de Contenidos.

Lo principal es que **cada contenido tiene su propia carpeta que contiene una carpeta por cada uno de los documentos** (versiones idiomáticas).

En la carpeta del documento es donde se encuentran los dataFiles, ficheros adjuntos, etc.

3.1 Detalle de la trasposición física de un contenido

Físicamente un contenido tiene la siguiente estructura:



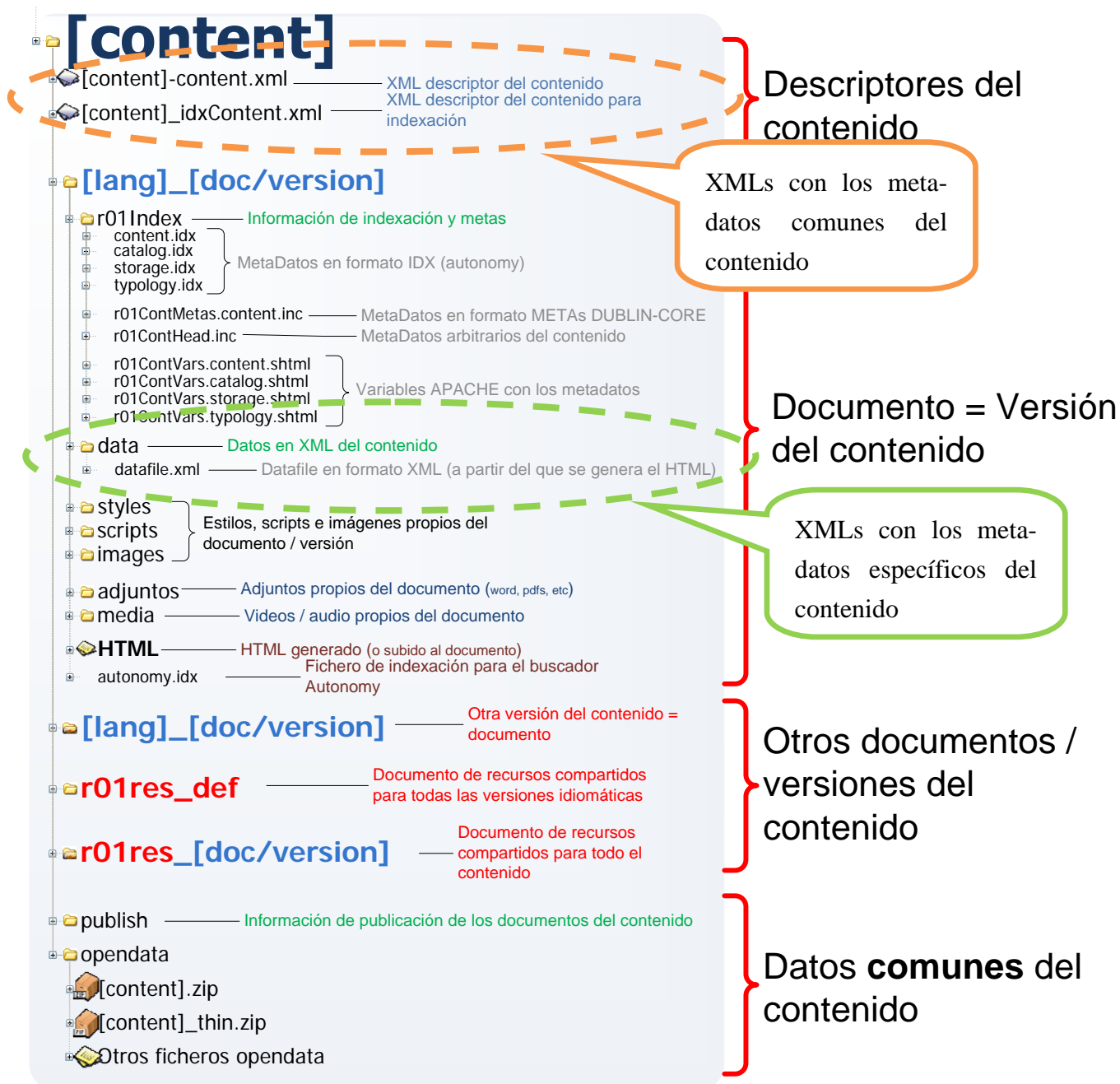
A continuación se describen cada uno de los elementos; muchos de estos ficheros son re-utilizables en aras a obtener productos derivados.

3.1.1 XMLs reutilizables

Todos los contenidos capturados en euskadi.net tienen una representación XML que recoge TODOS sus metadatos que pueden ser de dos tipos:

MetaDatos Comunes	Aquellos que tienen cualquier tipo de contenido independientemente de su tipo
MetaDatos Específicos	Aquellos que tienen los contenidos en función de su tipo

Desde el punto de vista de la reutilización estos XMLs se encuentran en dos localizaciones:

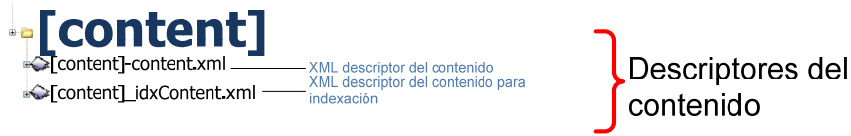


Por lo tanto, los XMLs reutilizables se encuentran en:

MetaDatos Comunes	Directorio raíz del contenido con el nombre: [contentName]-content.xml [contentName]-idxContent.xml
MetaDatos Específicos	En el directorio data de cada uno de los documentos (versiones) del contenido

3.1.2 Descriptores del Contenido

Todos los contenidos de euskadi.net tienen dos descriptores



Todos los datos del contenido se guardan en un **XML descriptor del contenido**; este XML descriptor referencia todos los datos del modelo de contenido: metaDatos en definitiva.

Para facilitar el trabajo a los indexadores utilizados por los buscadores se dispone de un **XML indexable** que contiene todos los datos del XML descriptor del contenido pero “**expandidos**”: el XML descriptor del contenido únicamente referencia oids (identificadores) de elementos auxiliares, pero **no** tiene sus descripciones.

Por ejemplo, el XML descriptor contiene los oids de los objetos de almacenamiento (servidor, dataStore, area y workArea), de los objetos de tipificación (cluster, familia y tipo) y de las etiquetas de catalogación pero **no** contiene sus descripciones que **si** están incluidos en el XML indexable.

- En el XML descriptor figura únicamente la identificación (workAreaOid='wr01_comun'), sin embargo en el XML indexable además del oid figuran las descripciones de la workarea en todos los idiomas
- Esto es especialmente significativo en el caso de las **catalogaciones**: en el XML descriptor únicamente aparece el identificador de la etiqueta (labelOid='r01ads1lsf'), sin embargo en el XML indexable además del oid figuran las descripciones en todos los idiomas, así como cada uno de los términos sinónimos, también en todos los idiomas.

A continuación se incluyen dos ejemplos de los XML descriptores:

Ejemplo de XML descriptor del contenido

<pre><?xml version="1.0" encoding="UTF-8"?> <content oid='r01dpd01idf69d909c17f6f3298aad93ea43811b0'> <name><![CDATA[ejje_test]]></name> <author>C001601T</author> <description><![CDATA[Contenido de testeo de funcionamiento general]]></description> <chargePerson><![CDATA[ejje - adactor]]></chargePerson> <createDate>02/12/2008 [08:33:20:000]</createDate> <source><![CDATA[]]></source> <newnessUntilDate>02/12/2008 [00:00:00]</newnessUntilDate> <typo> <familyOid>informacion</familyOid> <typeOid>informacion</typeOid> <clusterOid>euskadi</clusterOid> </typo> <storeLocation> <workAreaOid>whoedpl</workAreaOid> <serverOid>ejlp013</serverOid> <areaOid>horizontal_educacion</areaOid> <dataRepositoryOid>euskadi</dataRepositoryOid> </storeLocation> <labels> <label oid='r01e0000ff26d461a5a470b832d9102c1a18bea0'> <role>l</role> <structureOid>r01e00000fe4e6676dda470b8384c1e45db239b9c</structureOid> </label> <label oid='r01e0000ff26d461a5a470b8618f7e296c596449'> <role>l</role> <structureOid>r01e00000fe4e6676dda470b8384c1e45db239b9c</structureOid> </label> </labels></pre>	<p>Metadatos comunes del contenido: nombre, descripción, autor, etc</p> <hr/> <p>Información de tipología: Cluster > Familia > tipo (tres niveles)</p> <hr/> <p>Información de almacenamiento en el gestor de contenidos</p> <hr/> <p>Información de catalogación del contenido en las taxonomías normalizadas del GV</p>
--	--

<pre><documents> <document oid='r01dpd011df69eb6cb17f6f3287794717250df4a1'> <language>es</language> <name>![CDATA[::: Documento para testeos :::]]</name> <description>![CDATA[Documento de pruebas]]</description> <internalName>![CDATA[doctest]]</internalName> <createDate>02/12/2008 [08:34:35:000]</createDate> <author>CO01601T</author> <publicationInfo> <publicationTarget oid='Inter'> <publisherConfirmation>0</publisherConfirmation> <publishDate>13/10/2009 [22:12:06:000]</publishDate> <publicationOid>r01jpd01244f8747eb892e3a958f748b10142ff66</publicationOid> <state>1</state> <url>/contenidos/informacion/ejie_test/es_doctest/file1.html</url> </publicationTarget> </publicationInfo> <revisionState> <remainingRevisions>0</remainingRevisions> <state>1</state> </revisionState> <dataFiles> <dataFile oid='r01dpd011df69edfe117f6f325fc5955591d2aa72'> <metaDataContainer>true</metaDataContainer> <name>![CDATA[file1]]</name> <author>CO01601T</author> <createDate>02/12/2008 [08:34:46:476]</createDate> <templateOid>informacion</templateOid> <generateUrl>![CDATA[URL de generación]]</generateUrl> <editUrl>![CDATA[URL de edición]]</editUrl> <previewUrl>![CDATA[URL de previsualización]]</previewUrl> <initialPage>1</initialPage> <description>![CDATA[]]</description> </dataFile> </dataFiles> </document> </documents> </content></pre>	<p>Metadatos comunes del documento (versión idiomática)</p> <hr/> <p>Información de publicación del documento Recordar que lo que se publica es el documento</p> <hr/> <p>Información de revisión</p> <hr/> <p>Datafiles de datos a partir de los que se genera el HTML</p>
---	---

Ejemplo de **XML indexable**; observar como los OIDs del XML descriptor se “expanden” para incluir textos descriptivos en diferentes idiomas por lo que desde el punto de vista de la reutilización de datos parece más “útil”.

<pre><?xml version="1.0" encoding="UTF-8"?> <content oid='r01dpd011df69d909c17f6f3298aad93ea43811b0' objectType='0'> <boosting>1</boosting></pre>	<p>Potenciador del contenido en el buscador (preferencia frente a otros)</p>
<pre><name>![CDATA[ejie_test]]</name> <internalName>ejie_test</internalName> <createDate>02/12/2008 [08:33:20:000]</createDate> <responsible>ejie - adocctor</responsible> <description>![CDATA[Contenido de testeo de funcionamiento general.]]</description> <workAreaRelativePath>informacion/ejie_test</workAreaRelativePath> <newnessUntilDate>02/12/2008 [00:00:00:000]</newnessUntilDate> <author>CO01601T</author> <storage> <server oid='ejlpl013'> <path>/iwmnt/</path> <name> <es>Gestor Contenidos PRODUCCIÓN (ejslp013)</es> <eu>e_Gestor Contenidos PRODUCCIÓN (ejslp013)</eu> </name> </server> <workArea oid='whoedpl1'> <path>/iwmnt/euskadi/main/horizontal_educacion/WORKAREA/whoedpl1/</path> <name> <es>Comunicación y Divulgación</es> <eu>Comunicación y Divulgación</eu> </name> </workArea> <area oid='horizontal_educacion'> <path>/iwmnt/euskadi/main/horizontal_educacion/</path> <name> <es>horizontal_educacion</es> <eu>e_horizontal_educacion</eu> </name> </area> <dataStore oid='euskadi'> <path>/iwmnt/euskadi/</path> <name> <es>Euskadi.net</es> <eu>Euskadi.net</eu> </name> </dataStore> </storage></pre>	<p>Metadatos comunes del contenido: nombre, descripción, autor, etc</p> <hr/> <p>Información de almacenamiento en el gestor de contenidos</p> <p>* Las descripciones de los objetos de almacenamiento se “expanden” en los idiomas en los que están configurados</p>

<pre><typo numericId='1040040'> <cluster oid='euskadi' numericId='1'> <name> <es>Contenidos de Euskadi.net</es> <eu>Euskadi.neteko edukiak</eu> </name> </cluster> <family oid='informacion' numericId='4'> <name> <es>Información General</es> <eu>Informazio orokorra</eu> </name> </family> <type oid='informacion' numericId='40'> <name> <es>Información</es> <eu>Informazioa</eu> </name> </type> </typo> <editURLs> <es><![CDATA[/r01dContentManagerWar/r01dLoadAdminContent.jsp? oid=r01dpd011df69d909c17f6f3298aad93ea43811b0]]></es> <eu><![CDATA[/r01dContentManagerWar/r01dLoadAdminContent.jsp? oid=r01dpd011df69d909c17f6f3298aad93ea43811b0]]></eu> </editURLs> <structureCatalogs> <structureCatalog labelOid='r01e0000ff26d461a5a470b832d9102c1a18bea0' labelRoleInStructure='1' structureOid='r01e0000fe4e6676dda470b8384c1e45db239b9c' labelNumericId='1'> <terms> <term lang='de'><termText><![CDATA[Personen]]></termText></term> <term lang='es'><termText><![CDATA[Personas]]></termText></term> <term lang='en'><termText><![CDATA[People]]></termText></term> <term lang='eu'><termText><![CDATA[Pertsonak]]></termText></term> <term lang='fr'><termText><![CDATA[Personnes]]></termText></term> </terms> </structureCatalog> <structureCatalog labelOid='r01e0000ff26d461a5a470b8618f7e296c596449' labelRoleInStructure='1' structureOid='r01e0000fe4e6676dda470b8384c1e45db239b9c' labelNumericId='3'> <terms> <term lang='de'><termText><![CDATA[Verwaltungen]]></termText></term> <term lang='es'><termText><![CDATA[Administraciones]]></termText></term> <term lang='eu'><termText><![CDATA[Administrazioak]]></termText></term> <term lang='en'><termText><![CDATA[Administrations]]></termText></term> <term lang='fr'><termText><![CDATA[Administrations]]></termText></term> </terms> </structureCatalog> </structureCatalogs> <documents> <document oid='r01dpd011df69eb6cb17f6f3287794717250df4a1' objectType='1'> <language>es</language> <internalName>doctest</internalName> <name><![CDATA[::: Documento para testeos :::]]></name> <description><![CDATA[Documento de pruebas]]></description> <createDate>02/12/2008 [08:34:35:000]</createDate> <author>CO01601T</author><depthLevel></depthLevel> <approvalStatus>0</approvalStatus> <contentRelativePath>es_doctest</contentRelativePath> <editURLs> <es><![CDATA[/r01dContentManagerWar/r01dLoadAdminContent.jsp? oid=r01dpd011df69d909c17f6f3298aad93ea43811b0]]></es> <eu><![CDATA[/r01dContentManagerWar/r01dLoadAdminContent.jsp? oid=r01dpd011df69d909c17f6f3298aad93ea43811b0]]></eu> </editURLs> <workAreaRelativePath>informacion/ejie_test/es_doctest</workAreaRelativePath> </document> </documents> <publicationInfo> <publishContext repositoryNumericId='1' repositoryAlias='contenidos' state='1' repositoryOid='Inter'> <publishDate>13/10/2009 [22:12:06:000]</publishDate> </publishContext> </publicationInfo></pre>	<p>Información de tipología Cluster > Familia > tipo (tres niveles)</p> <p>* Las descripciones de los objetos de typo se "expanden" en los idiomas en los que están configurados</p> <hr/> <p>URL de edición del contenido</p> <hr/> <p>Información de catalogación del contenido en las taxonomías normalizadas del GV</p> <p>* Las descripciones de las etiquetas de catalogación se "expanden" en los idiomas en los que están configuradas en el sistema de ejes</p> <hr/> <p>Documento (versión idiomática)</p> <hr/> <p>Metadatos comunes del documento (versión idiomática)</p> <hr/> <p>Información de publicación del documento Recordar que lo que se publica es el documento</p>
--	--

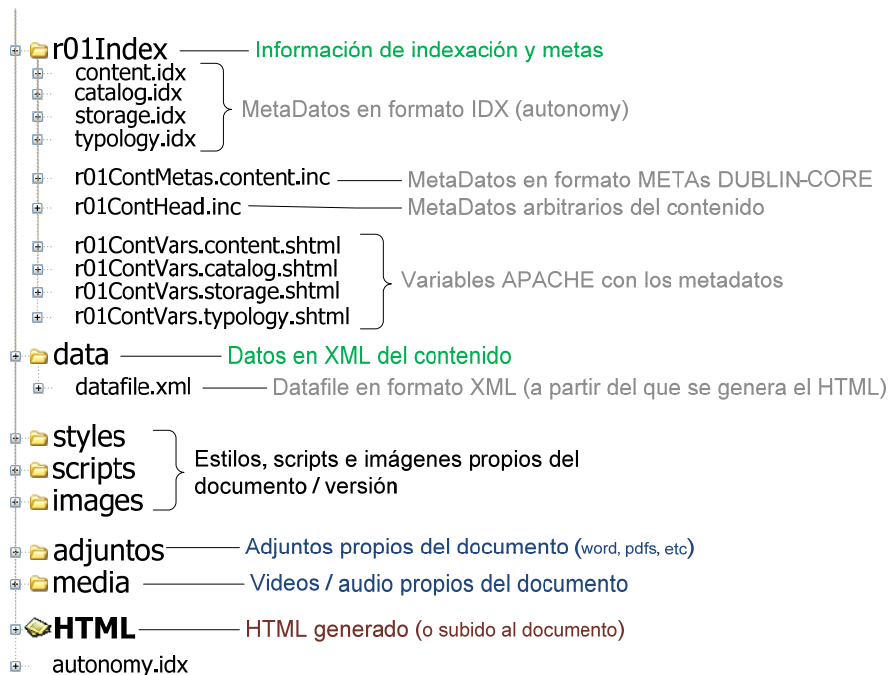
```

<dataFiles>
  <dataFile metaDataContainer='true'
    mainDataFile='true'
    oid='r01dpd011df69edfe117f6f325fc595591d2aa72'
    objectType='2' >
    <defaultTemplateGeneratedFileDocumentRelativePath>
      file1.html
    </defaultTemplateGeneratedFileDocumentRelativePath>
    <editURLs>
      <es><![CDATA[URL edición]]></es>
      <eu><![CDATA[URL edición]]></eu>
    </editURLs>
    <generatedFilesDocumentRelativePaths>
      <informacion>main:file1.html</informacion>
    </generatedFilesDocumentRelativePaths>
    <internalName>file1</internalName>
    <dataCaptureTemplate>informacion</dataCaptureTemplate>
    <description><![CDATA[]]></description>
    <workAreaRelativePath>
      templatedata/editorial/informacion/
      data/es_r01dpd011df69edfe117f6f325fc595591d2aa72
    </workAreaRelativePath>
    <createDate>02/12/2008 [08:34:46:476]</createDate>
    <name><![CDATA[file1]]></name>
    <author>C001601T</author>
  </dataFile>
</dataFiles>
</document>
</documents>
</content>

```

Datafiles de datos a partir de los que se genera el HTML

3.1.3 Documento = Versión del contenido



Los contenidos tienen una o varias “versiones”, lo que se denominan **documentos** que normalmente son **idiomáticas** pero que pueden seguir otro criterio como nivel de profundidad (para expertos / no expertos), pero que todas ellas tienen en común que tienen como objeto el “tema” del contenido.

Los documentos tienen varios componentes:

3.1.3.1 DataFiles (XML de datos) y HTML Generado



Normalmente (no siempre) los HTML se generan a partir de un XML que es “capturado” utilizando una plantilla de captura (formulario) y a partir del cual se genera normalmente HTML (aunque se pueden generar otros formatos)

El **XML capturado** se almacena y publica en la carpeta **data**

El **HTML generado** a partir del XML se publica en la carpeta del documento

NOTA: El HTML habitualmente se genera a partir de los datos capturados en la plantilla aunque en ocasiones el HTML se maqueta externamente al gestor de contenidos y se “sube” como un fichero adjunto por lo que no existen datos en formato XML asociados al XML.

Los datafiles (localizados en la carpeta **data**) normalmente son **ficheros XML cuya estructura depende del tipo de contenido**, por ejemplo, si el contenido es una ayuda, este XML contiene los datos de la ayuda, sin embargo, si el contenido es una predicción meteorológica, este XML contiene los datos de la predicción.

3.1.3.2 Estilos, imágenes y scripts

El HTML del documento en ocasiones (aunque raramente) puede necesitar algún recurso propio. En este caso estos recursos se colocan en el directorio adecuado

3.1.3.3 Adjuntos y media

Es habitual que desde los contenidos se haga referencia a **adjuntos** (PDFs, documentos de Word, Excel, etc); normalmente (aunque no es obligatorio) estos adjuntos se colocan en la carpeta **adjuntos**.

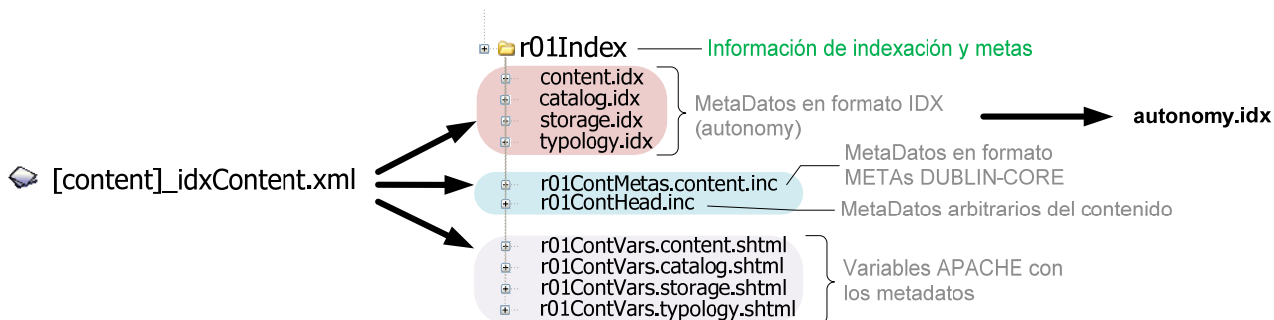
Algo similar ocurre con los recursos multimedia asociados al contenido (imágenes, video y sonidos) que normalmente (aunque no es obligatorio) se colocan en la carpeta **media**.

3.1.3.4 Datos de Indexación

Anteriormente se ha descrito como todos los contenidos incorporaban un fichero XML “indexable” que era un “descriptor mejorado” del contenido; este XML indexable, normalmente no es útil para nada salvo para generar otros formatos con un fin específico, como:

- Formato fuente para el indexador del buscador de euskadi.net (autonomy)
- Formatos META – *Dublin Core* para incluir en los contenidos entregados en internet
- Formatos VAR de apache para operaciones avanzadas en el servidor web

Por lo tanto, a partir del XML indexable del contenido se generan otros formatos más útiles:



Desde el punto de vista de la reutilización, cualquiera de los ficheros anteriores puede ser utilizado.

Ficheros idx de Autonomy

El indexador del buscador Autonomy utiliza ficheros en un formato en texto plano propietario (IDX)

El gestor de contenidos y portales generan ficheros IDX a partir del XML indexable para cuatro “grupos” de metadatos:

- Metadatos generales del contenido / documentos
- Metadatos de catalogación
- Metadatos de almacenamiento
- Metadatos de tipología

Los anteriores ficheros de metadatos se unen en el destino de publicación en un solo fichero (**autonomy.idx**) que es el que utiliza autonomy para su indexación.

Esto se ha hecho así para permitir cambios individuales en la indexación (metadatos, catalogaciones, tipología) sin necesidad de publicar todo el contenido y en especial el HTML o ficheros adjuntos.

El formato IDX de autonomy es muy sencillo y está formado por líneas para cada metadato:

#DREFIELD [nombreMetaDato]="[Valor del metadato]"

Por ejemplo:

```
#DRETITLE ::: Documento para testeos :::
#DREFIELD documentAuthor="CO01601T"
#DREFIELD contentName="ejie_test"
#DREFIELD documentLanguage="es"
#DREFIELD typologyTypeOid="informacion"
#DREFIELD contentOid="r01d9d011df69d909c17f6f3298aad93ea43811b0"
#DREFIELD documentCreateDate="1228203275"
#DREFIELD workAreaPath="//iwmnt/euskadi/main/horizontal_educacion/WORKAREA/whoedpl/"
#DREFIELD documentDepthLevel=""
#DREFIELD contentWorkAreaRelativePath="informacion/ejie_test"
#DREFIELD contentNewnessUntilDate="1228172400"
#DREFIELD dataStorePath="//iwmnt/euskadi/"
#DREFIELD contentResponsible="ejie - adocor"
#DREFIELD areaPath="//iwmnt/euskadi/main/horizontal_educacion/"
#DREFIELD contentAuthor="CO01601T"
#DREFIELD documentContentRelativePath="es_doctest"
#DREFIELD typologyFamilyOid="informacion"
#DREFIELD storageDataStoreOid="euskadi"
#DREFIELD documentOid="r01d9d011df69e9eb6cb17f6f3287794717250df4a1"
#DREFIELD storageAreaOid="horizontal_educacion"
#DREFIELD typologyFamilyNumericId="4"
#DREFIELD serverPath="//iwmnt/"
#DREFIELD documentWorkAreaRelativePath="informacion/ejie_test/es_doctest"
#DREFIELD pubTargetsDetails="(Inter;contenidos;l;1247172229),(Inter;contenidos;0;1231738325)"
#DREFIELD pubTargetsOids="Inter"
#DREFIELD pubTargetsNumericIds="1"
#DREFIELD pubTargets_Inter="1247172229"
#DREFIELD documentMainDataFileGeneratedFilesDocumentRelativePaths="(informacion;main:file1.html)"
#DREFIELD documentName::: Documento para testeos :::
#DREFIELD typologyClusterOid="euskadi"
#DREFIELD contentCreateDate="1228203200"
#DREFIELD typologyTypeNumericId="40"
#DREFIELD typologyNumericId="1040040"
#DREFIELD structureCatalogsOids="r01e00000fe4e66771ba470b8101d0b01550f37ed,r01e000000ff26d46212a470b8ae993756a24b76"
#DREFIELD structureCatalogsOidsBool="r01e00000fe4e66771ba470b8101d0b01550f37ed OR
r01e000000ff26d46212a470b8ae993756a24b76"
#DREFIELD structureCatalogsNumeric="185,60,54,51,460,255,1,98,3"
#DREFIELD structureCatalogsNumericBool="185 OR 60"
#DREFIELD structureCatalogsTerms="Cultura,Transportes y Obras Públicas,Seguridad e Interior,Acción exterior,
Consejero/a de Cultura;Consejería de Cultura,Red viaria"
#DREFIELD structureCatalogsCats="
(r01e00000fe4e6676dda470b8384c1e45db239b9c;r01e00000ff26d461a5a470b8618f7e296c596449;1),
(r01e00000fe4e6676dda470b898e584a4a1047312;r01e00000ff26d461e3a470b89602cd52e9b70e15;1)"
#DREFIELD documentApprovalStatus="1"
#DREFIELD contentDescription="Contenido de testeo de funcionamiento general"
#DREFIELD documentDataFilesGeneratedFilesDocumentRelativePaths=
"r01d9d011df69edfe117f6f325fc5955591d2aa72>>(informacion;main:file1.html)"
#DREFIELD documentMainDataFileDefaultGeneratedFileDocumentRelativePath="file1.html"
#DREFIELD documentDescription="Documento de pruebas"
#DREFIELD storageWorkAreaOid="whoedpl"
#DREFIELD typologyClusterNumericId="1"
#DREFIELD contentEditURLs="
(es;/r01dContentManagerWar/r01dLoadAdminContent.jsp?oid=r01d9d011df69d909c17f6f3298aad93ea43811b0),
(eu;/r01dContentManagerWar/r01dLoadAdminContent.jsp?oid=r01d9d011df69d909c17f6f3298aad93ea43811b0)"
#DREFIELD documentInternalName="doctest"
#DREFIELD documentEditURLs="
(es;/r01dContentManagerWar/r01dLoadAdminContent.jsp?oid=r01d9d011df69d909c17f6f3298aad93ea43811b0),
(eu;/r01dContentManagerWar/r01dLoadAdminContent.jsp?oid=r01d9d011df69d909c17f6f3298aad93ea43811b0)"
#DREFIELD storageServerOid="ej1p013"
#DREFIELD guide_typo_cluster="euskadi"
#DREFIELD guide_typo_family="informacion"
#DREFIELD guide_typo_type="informacion"
```

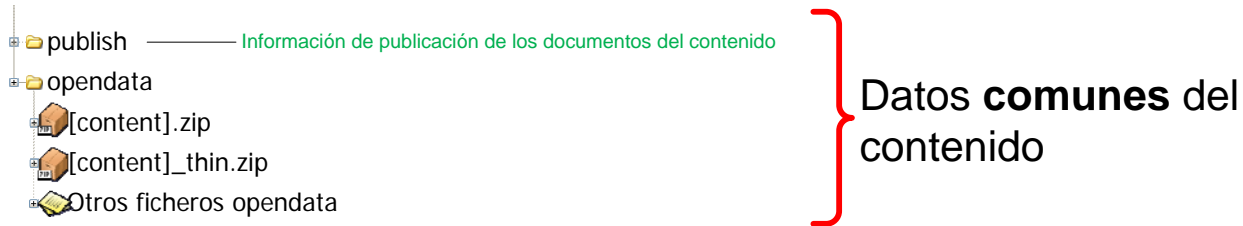
Ficheros de METAS y Variables de Apache

Contenidos	r01ContVars.shtml	Variables APACHE con metadatos del contenido que pueden ser utilizadas en cualquier sitio
	r01ContMetas.inc	METAS (en formato HTML dublin core) del contenido para ser incluidos en el HEAD de la página
	r01ContHead.shtml	Recursos (scripts y estilos) propios del contenido para ser incluidos en el HEAD de la página
Portales	r01[PageName]PortalVars.shtml	Variables APACHE con metadatos de la página de portal que pueden ser utilizadas en cualquier sitio
	r01[PageName]PortalMetas.inc	METAS, scripts y estilos de la página de portal para ser incluidos en el HEAD de la página de portal
	r01[PageName]PortalHead.shtml	Recursos (scripts y estilos) propios del contenido para ser incluidos en el HEAD de la página
Aplicaciones	Un fichero arbitrario que incluye el HTML que hay que insertar en el HEAD de la página.	

Como se ve hay tres tipos de ficheros:

<p>Variables</p>	<p>xxxVars.shtml</p>	<p>Es un fichero shtml (es procesado por el filtro SSI de apache) que contiene definiciones de variables apache en la forma:</p> <pre><!--#set var="doc_author" value="EJ03680M" --> <!--#set var="cont_name" value="prensa_sanidad79" --> <!--#set var="doc_lang" value="es" --> <!--#set var="cont_author" value="EJ03680M" --> <!--#set var="doc_name" value="My doc name" --> <!--#set var="cont_description" value="My doc description" --> <!--#set var="doc_description" value="Nota de Prensa. My doc description" --> <!--#set var="doc_internal_name" value="np79" --></pre> <p>Como se ve, simplemente son metadatos del contenido / página expresados en forma de variables apache que pueden utilizarse en cualquier lugar del HTML de las páginas (y de los contenidos) en la forma:</p> <pre><!--#echo var="[nombre variable]" --></pre> <p>Por ejemplo para incluir el nombre del contenido en el title de la página:</p> <pre><title> Euskadi.net <!--#if expr="\$doc_name" --><!--#echo var="doc_name" --><!--#endif --> </title></pre> <p>Estos ficheros son generados automáticamente por el Gestor de Contenidos y Portales.</p>
<p>MetaDatos</p>	<p>xxxMetas.inc</p>	<p>Es un fichero inc (no es procesado por el filtro SSI de apache) que contiene METAS en formato HTML dublicon-core en el formato:</p> <pre><meta name="DC.creator" content="EJ03680M"> <meta name="DC.language" content="es"> <meta name="DC.date" content="11/09/2006"> <meta name="DC.identifier" content="r01dp010d9c8c004683e047a6a10aedaccd245"> <meta name="DC.title" content="My description"> <meta name="DC.description" content="Nota de Prensa. My Description"></pre> <p>Estos ficheros son generados automáticamente por el Gestor de Contenidos y Portales y se incluyen automáticamente en el HEAD de las páginas que tienen un AV contenedora de contenidos</p>
<p>Recursos del Head</p>	<p>xxxHead.shtml</p>	<p>Es un fichero shtml (es procesado por el filtro SSI de apache) que contiene recursos (scripts, hojas de estilo y metas) que se han de incluir en la sección HEAD de la página.</p> <p>Estos ficheros NO se generan automáticamente sino que el usuario debe dejarlos en la carpeta del contenido / aplicación o el portal (ver más adelante) y su contenido es arbitrario (el que el usuario quiera)</p> <p>Ej:</p> <pre><meta name="WT.ti" content="Home"> <link rel="stylesheet" type="text/css" href="styles/myContentStyles.css"/> <script type="text/javascript" src="scripts/myContentScripts.js"></script></pre>

3.1.4 Datos comunes del Contenido



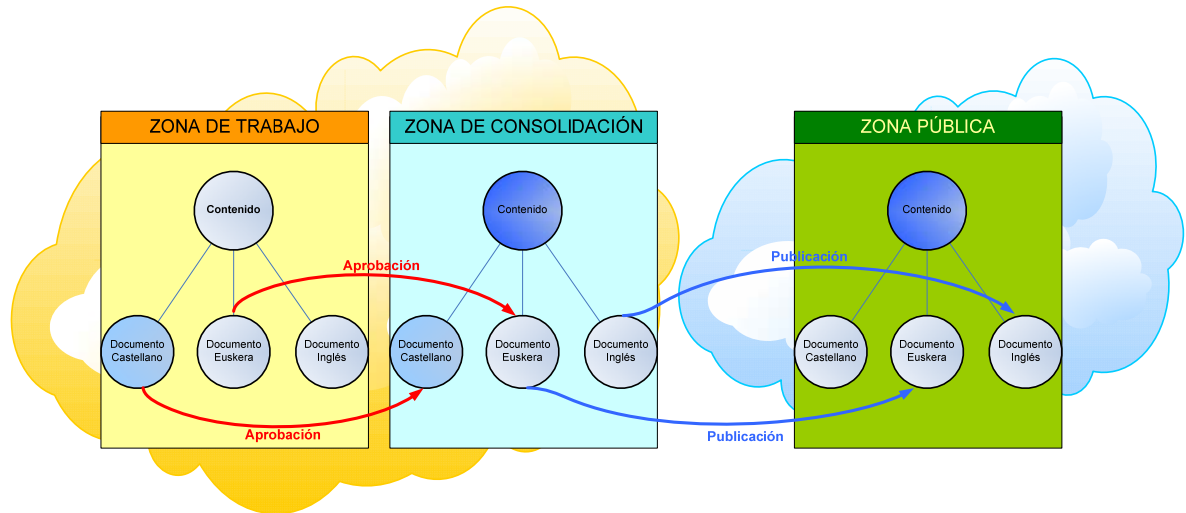
Desde el punto de vista de la reutilización de datos es muy importante la carpeta **opendata** ya que en ella se dejan ficheros reutilizables:

[content]_thin.zip	<p>Se trata de un fichero empaquetado en formato ZIP que contiene todos los ficheros del contenido excepto los ficheros adjuntos y media. Específicamente, este empaquetado contiene:</p> <ul style="list-style-type: none"> • Los descriptores en XML y XML indexable • El directorio r01Index con los ficheros de indexación • El directorio data y los XML de los data files a partir de los que se generan los contenidos • El HTML generado
[Content].zip	<p>Se trata de un fichero empaquetado igual que el anterior pero que además incorpora los ficheros adjuntos y media</p>
Otros ficheros opendata	<p>Se trata de ficheros con datos en formato fuente dependientes del tipo de contenido. Por ejemplo en el caso de los datos abiertos de las estaciones meteorológicas, en este fichero se dejan los datos fuente de las estaciones.</p>

4 Despliegue de Contenidos (publicación)

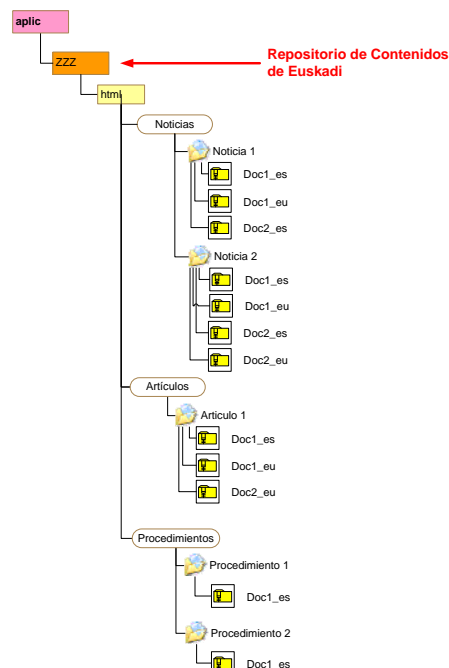
Los contenidos se crean y generan en la zona de trabajo (jakinaplus) y se despliegan como **ficheros estáticos** en los servidores web que se encargan de servirlos en la zona pública de entrega.

El paso de una zona a otra se hace según el siguiente esquema:



- Los contenidos se crean y mantienen en la **zona trabajo** interna (jakinaplus).
- La **aprobación** se hace a nivel de **documento** desde la zona de producción de contenidos: una vez **aprobado** un documento, se pasa a la **zona de consolidación** donde existe una versión completa del contenido.
- La **publicación** se hace también a nivel de **documento** desde la zona de consolidación a la **zona pública** (de entrega) donde son accesibles a los ciudadanos/as.

En la **zona de entrega**, los contenidos se despliegan en un **repositorio** con la siguiente estructura:



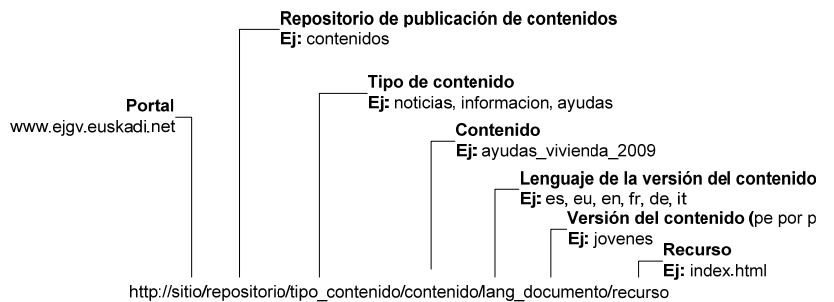
5 Estructura de URLs

Con lo señalado anteriormente, se puede ahora entender la estructura de las URLs de los contenidos.

5.1 URL directa de un contenido o una página de portal

Una característica de la infraestructura es que los contenidos y las páginas de portal están **separados** y se unen en tiempo de visualización, de forma que un mismo contenido puede ser visualizado en en cualquier página de cualquier portal.

Para acceder directamente a un contenido la URL es:



5.2 Combinar un contenido en una página de portal

Dado que los contenidos y las páginas de portal son independientes, se utiliza un mecanismo basado en la URL para “combinar” un contenido y una página de portal:

Las URL de Euskadi.net tienen una **estructura uniforme**:

Sitio web	Portal - Página	Idioma	Recurso
-----------	-----------------	--------	---------

URL de un contenido visualizado en una página de portal (la url depende del portal y la página)

