

Tutorial: Marcando su conjunto de datos con DCAT

El [vocabulario del catálogo de datos \(DCAT\)](#) define una forma estándar de publicar metadatos legibles por máquina sobre un conjunto de datos.

La forma más sencilla de publicar una descripción de su conjunto de datos es publicar metadatos DCAT mediante RDFa. RDFa permite incrustar metadatos legibles por máquina en una página web. Esto significa que la publicación de los metadatos de su conjunto de datos se puede lograr fácilmente actualizando el HTML de la página de inicio de su conjunto de datos.

Esta guía proporciona una breve introducción a la publicación de metadatos DCAT mediante RDFa. Para casos de uso más avanzados, incluida la publicación de datos en otros formatos, consulte la [documentación oficial del W3C para DCAT](#). El [manual RDFa](#) también puede ser una lectura útil.

La aplicación [de Certificados de Datos Abiertos](#) admite la lectura de DCAT publicados como RDFa. Entonces, además de proporcionar metadatos legibles por máquina para los consumidores de datos, el uso de DCAT simplificará el proceso de certificación de su conjunto de datos, ya que la aplicación podrá completar automáticamente algunas de las respuestas por usted.

Iniciando

Lo primero que debe hacer es informar a las aplicaciones que su página web describe un conjunto de datos. Para hacer esto, debemos declarar los esquemas de metadatos que usaremos para describir el conjunto de datos y luego indicar el tipo de cosa que se describe.

Aquí hay un fragmento de HTML que proporciona un punto de partida. Reemplace {url} con la URL de su página de conjunto de datos.

```
<html prefix="dct: http://purl.org/dc/terms/
          rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
          dcat: http://www.w3.org/ns/dcat#
          foaf: http://xmlns.com/foaf/0.1/">
  <body>
    <div typeof="dcat:Dataset" resource="{url}">
      ...
    </div>
  </body>
</html>
```

El elemento `html` tiene un atributo `prefix` que declara los esquemas. El elemento `div` declara qué recurso está describiendo y el tipo de recurso usando los atributos `resource` y `typeof`.

El resto de los metadatos sobre el conjunto de datos se agregarán a los elementos HTML anidados dentro de este contenedor `<div>`. No tiene que usar un elemento `div`, podría ser cualquier elemento HTML, así que adapte a la estructura de la página de su conjunto de datos.

Cada una de las siguientes secciones ilustra cómo agregar elementos de metadatos adicionales que completan la descripción de su conjunto de datos. Intente proporcionar una descripción lo más completa posible de su conjunto de datos.

De nuevo, los elementos HTML usados en los ejemplos son solo una sugerencia, pueden ser lo que quieras, así que adapta los ejemplos a tu página existente. Es posible que deba agregar algunos elementos adicionales, por ejemplo, para especificar fechas, títulos, etc. que ya están en la página como texto sin formato.

La parte importante de los ejemplos son los atributos RDFa: about, property, content, datatype, etc. Estos atributos definen qué propiedad del conjunto de datos se describe y proporcionan los metadatos legibles por máquina.

Título

Especifique el título de su conjunto de datos usando la propiedad `dct:title`:

```
<h1 property="dct:title">Example Dataset</h1>
```

Fecha de creación

Especifique la fecha en que se creó su conjunto de datos utilizando la propiedad `dct:created`:

```
<p property="dct:created" content='2010-10-25T09:00:00+00:00'
datatype='xsd:dateTime'>25th October 2010</p>
```

En este caso, el texto legible por humanos para la propiedad está contenido dentro de la etiqueta de párrafo. Su valor puede ser cualquier cosa, pero la fecha legible por máquina (especificada en el atributo `content`), debe usar un tipo de datos definido para que pueda analizarse fácilmente.

Se recomienda que utilice el formato [de fecha de esquema XML](#) o formato [de fecha y hora de esquema XML](#).

Fecha de modificación

Especifique la fecha en que su conjunto de datos se actualizó por última vez usando la propiedad `dct:modified`:

```
<p property="dct:modified" content='2010-10-25T09:00:00+00:00'
datatype='xsd:dateTime'>25th October 2010</p>
```

Consulte más arriba las notas sobre los formatos de fecha.

Descripción

Marque la descripción de su conjunto de datos usando la propiedad `dct:description`:

```
<p property="dct:description">This is the description.</p>
```

Licencia

El marcado para declarar la licencia del conjunto de datos es un poco más complejo. Debe declarar la propiedad de la licencia (`dct:license`), así como el nombre y la URL de la licencia:

Sustituya los marcadores de posición `{license URL}` y `{license name}` por los valores que se aplican a su conjunto de datos.

```
<div property="dct:license"
  resource="{license URL}">
  <a href="{license URL}">
    <span property="dct:title">{license name}</span>
  </a>
</div>
```

Para obtener una guía más detallada sobre la publicación de una declaración de derechos integral para su conjunto de datos, incluida la licencia, las declaraciones de derechos de autor y la forma preferida de atribución, lea la [Guía para editores sobre el vocabulario de declaración de derechos de datos abiertos](#) .

Editor

Declare el editor de su conjunto de datos usando la propiedad `dct:publisher` . Nuevamente, hay varios elementos para declarar aquí, incluidos el nombre y la URL de la página de inicio del editor.

Reemplace las propiedades `publisher URL` y `publisher name` con los valores apropiados.

```
<div property="dct:publisher"
  resource="{publisher URL}">
  <a href="{publisher URL}" about="{publisher URL}" property="foaf:homepage">
    <span property="foaf:name">{publisher name}</span>
  </a>
</div>
```

Palabras clave

Las palabras clave se pueden adjuntar a un conjunto de datos usando la propiedad `dcat:keyword`. Los valores de propiedad son etiquetas o etiquetas simples. Puede tener tantos o tan pocos (¡o ninguno!) como desee.

```
<span property="dcat:keyword">Examples</span>, <span property="dcat:keyword">DCAT</span>
```

Frecuencia de actualización

La propiedad `dcat:accrualPeriodicity` se utiliza para definir la frecuencia con la que se actualiza un conjunto de datos. Los valores de la propiedad son URI que se toman como un vocabulario controlado simple.

```
<a href="{frequency}" property="dcat:accrualPeriodicity">{frequency} (human readable)</a>
```

Sustituya el marcador de posición `{frequency}` por uno de los siguientes URI:

- <http://purl.org/linked-data/sdmx/2009/code#freq-A> - Anual
- <http://purl.org/linked-data/sdmx/2009/code#freq-B> – Todos los días laborables (Lun – Vie)
- <http://purl.org/linked-data/sdmx/2009/code#freq-D> – Diario (7 días a la semana)
- <http://purl.org/linked-data/sdmx/2009/code#freq-M> - Mensual
- <http://purl.org/linked-data/sdmx/2009/code#freq-N> - Cada minuto
- <http://purl.org/linked-data/sdmx/2009/code#freq-Q> - Cada cuarto
- <http://purl.org/linked-data/sdmx/2009/code#freq-S> - Medio año
- <http://purl.org/linked-data/sdmx/2009/code#freq-W> - Semanalmente

Distribuciones

Un conjunto de datos puede tener varias distribuciones. [Las distribuciones](#) describen cómo se empaqueta y publica un conjunto de datos. Su conjunto de datos puede tener varias distribuciones, por ejemplo, si publica una serie de datos durante un período de tiempo como paquetes separados, o si está disponible en diferentes formatos.

El marcado aquí es un poco más complejo. Define un nuevo recurso (a `dcat:Distribution`) y lo asocia con su conjunto de datos. El marcado anidado luego proporciona metadatos sobre el nuevo recurso, por ejemplo, su formato, tamaño, fecha de publicación, etc.

```
<div property='dcat:distribution' typeof='dcat:Distribution'>
  <span property="dct:title">{Distribution title}</span>
  <ul>
    <li><strong>Format</strong> <span content='{format}'
property='dcat:mediaType'>{format (human readable)}</span></li>
```

```

        <li><strong>Size</strong> <span content='{size in bytes}'
datatype='xsd:decimal' property='dcat:byteSize'>{size (human
readable)}</span></li>
        <li><strong>Issued</strong> <span property='dct:issued' content='{date
issued}' datatype='xsd:date'>{date issued (human readable)}</span></li>
    </ul>
    <p><a href='{link to data}' property='dcat:accessURL'>Download the full
dataset</a></p>
</div>

```

El marcador de posición {format} debe ser un [tipo MIME reconocido](#) (por ejemplo, text/csv o application/json)

La propiedad dct:issued especifica la fecha en que se publicó la distribución. La propiedad debe seguir las mismas pautas que para la propiedad dct:created descrita anteriormente.

Poniéndolo todo junto

Aquí hay un ejemplo completo de una página HTML marcada con DCAT. Proporciona todos los metadatos básicos para el conjunto de datos, incluida una descripción de una sola distribución:

```

<!DOCTYPE html>
<html prefix="dct: http://purl.org/dc/terms/
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
dcat: http://www.w3.org/ns/dcat#
foaf: http://xmlns.com/foaf/0.1/">
    <head>
        <title>DCAT in RDFa</title>
    </head>
    <body>
        <div typeof="dcat:Dataset" resource="http://gov.example.org/dataset/finances">
            <h1 property="dct:title">Example DCAT Dataset</h1>

            <p property="dct:created" content='2010-10-25T09:00:00+00:00'
datatype='xsd:dateTime'>25th October 2010</p>
            <p property="dct:modified" content='2013-05-10T13:39:36+00:00'
datatype='xsd:dateTime'>10th March 2013</p>

            <p property="dct:description">This is the description.</p>

            <div property="dct:license"
resource="http://reference.data.gov.uk/id/open-government-
licence">
                <a href="http://reference.data.gov.uk/id/open-government-licence">
                    <span property="dct:title">UK Open Government Licence
(OGI)</span>
                </a>
            </div>

            <div property="dct:publisher"

```

```

        resource="http://example.org/publisher">
        <a href="http://example.org/publisher"
about="http://example.org/publisher" property="foaf:homepage">
        <span property="foaf:name">Example Publisher</span>
        </a>
    </div>

    <div>
        <span property="dcat:keyword">Examples</span>, <span
property="dcat:keyword">DCAT</span>
    </div>

    <div>
        <a href="http://purl.org/linked-data/sdmx/2009/code#freq-W"
property="dcat:accrualPeriodicity">Weekly</a>
    </div>

    <div property='dcat:distribution' typeof='dcat:Distribution'>
        <span property="dct:title">CSV download</span>
        <ul>
            <li><strong>Format</strong> <span content='text/csv'
property='dcat:mediaType'>CSV</span></li>
            <li><strong>Size</strong> <span content='240585277'
datatype='xsd:decimal' property='dcat:byteSize'>1024MB</span></li>
            <li><strong>Issues</strong> <span property='dct:issued'>2012-01-
01</span></li>
        </ul>
        <p><a class='btn btn-primary'
href='http://example.org/distribution.csv.zip'
property='dcat:accessURL'>Download the full dataset</a></p>
    </div>

</body>
</html>

```

Usando la información de este tutorial cree una página HTML que permite anotar los siete conjuntos de datos relacionados a cultura y patrimonio de la ciudad de Cuenca. La información sobre estos recursos está en el siguiente enlace

<https://cuencaendatos.cuenca.gob.ec/group/cultura-y-patrimonio>

Puede usar el servidor utilizado en las primeras prácticas de la asignatura para publicar el enlace a la página anotada con DCAT