

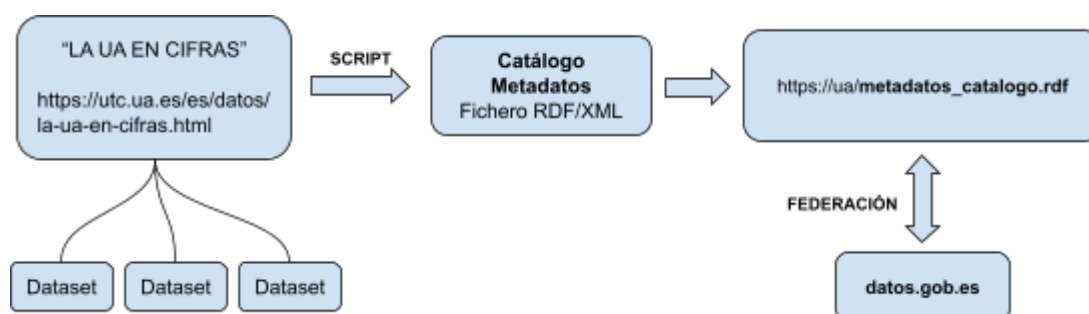
FEDERACIÓN DE DATOS DE LA UTC

1. Introducción

Se plantea la necesidad de recopilar los metadatos asociados a los conjuntos de datos expuestos en el portal [“La UA en cifras”](https://utca.ua.es/es/datos/la-ua-en-cifras.html) de la Universidad de Alicante. La principal motivación es la federación de todos estos conjuntos en el catálogo de datos abiertos del gobierno ([datos.gob](https://datos.gob.es)). Para realizar esta tarea, es necesario generar un fichero que contenga todos los metadatos necesarios en un formato determinado, además es necesario que se realice de manera automática.

Los metadatos se organizan de acuerdo al modelo RDF y se empleará la serialización RDF/XML, es decir, finalmente los metadatos quedarán recogidos en un fichero .xml con una estructura determinada de etiquetas. La estructura y el vocabulario empleado se encuentran reflejados en la [Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información](#). La generación del fichero se realizará por medio de un script de PHP que se ejecutará de manera periódica o cuando se detecte algún cambio en los conjuntos de datos; el código se encuentra adjunto a este informe.

Finalmente, se expondrá este fichero en una dirección para que pueda ser consumido desde el portal de datos del gobierno ([datos.gob](https://datos.gob.es)) y se realice la federación del catálogo completo.



2. Metadatos

Para la obtención de los metadatos se nos proporciona una tabla que recoge toda la estructura de ficheros y direcciones del apartado “La UA en cifras”. Algunos están directamente disponibles en esta tabla, mientras que otros requieren ser extraídos de los propios ficheros asociados a cada conjunto, generalmente en formato .xlsx.

Los metadatos se dividen en aquellos asociados al catálogo completo, a cada conjunto de datos y a cada distribución. Se han considerado los siguientes:

Catálogo:

- | | | |
|---------------|--------------------|-----------------------|
| - Título | - Temática | - Licencia |
| - Descripción | - Idioma(s) | - Fecha publicación |
| - Publicador | - Página principal | - Fecha actualización |

Conjunto:

- | | | |
|---------------|-------------|------------------------|
| - Título | - Idioma(s) | - Cobertura geográfica |
| - Descripción | - Temática | - Fecha publicación |
| - Publicador | - Licencia | - Fecha actualización |

Distribución:

- | | |
|-----------------|-----------|
| - Título | - Formato |
| - Url de acceso | |
| - Tamaño | |

3. Script y federación

Se emplea un script de PHP para recopilar los metadatos de la tabla y generar un fichero .xml que los contenga con la estructura adecuada. Las herramientas empleadas son:

- La clase *DOMDocument* para manipular el fichero .xml.
- Librería *PHPOffice/PhpSpreadsheet* para manipular ficheros .xlsx.

En general, la principal dificultad es la extracción de metadatos contenidos en los propios ficheros, esto se debe a que no existe consistencia en el formato de su contenido, lo que ha requerido considerar diferentes casos. Esto puede observarse en el método *getTitle()*, empleado para extraer el metadato *título* a partir del .xlsx.

Hay que destacar que el script parte de una tabla de metadatos que se nos proporciona en formato .xlsx, en caso de considerar otra fuente distinta habría que adaptar el código a ese caso en particular.

Resumen del script:

- Carga de la tabla con los metadatos y de un .xml que contiene la estructura básica del fichero a generar ("catalogo_base.rdf").
- Se añaden los metadatos generales (catálogo).
- Itera sobre cada fila de la tabla y recoge los metadatos disponibles.
- Para cada metadato se construye una etiqueta que se va añadiendo al .xml base.
- Una vez recorridos todos los conjuntos, se guarda el fichero con los metadatos ("catalog.rdf").

El objetivo es que el script se ejecute cada vez que se detecte algún cambio en los conjuntos de datos, ya sea por modificación o adición de uno nuevo, sin embargo, esto aún no se ha implementado.

Finalmente, para realizar la federación del catálogo es necesario que el fichero generado se exponga en una dirección pública de la UA. La configuración de este proceso se realiza desde el portal de datos.gob, indicando la dirección así como la frecuencia con la que se desea federar.