Normalización de nombres de cubiertas

v2.0 (02/2008)

Febrero 2008

DILVE es una iniciativa de **libro-es.com** promovida por la *Federación de Gremios de Editores de España* y la *Fundación Germán Sánchez Ruipérez,* con el patrocinio de CEDRO.

Introducción

En DILVE coexisten datos de muchas editoriales. Los archivos de recursos (p.ej. imágenes de cubierta o portada) se almacenan en DILVE en ubicaciones separadas para cada editorial. Así, si dos editoriales cargan archivos distintos con el mismo nombre, ambos pueden coexistir.

Sin embargo, cuando un usuario (un *consumidor de información*) de DILVE realiza una extracción de datos, incluyendo los archivos con las imágenes de las portadas de los libros correspondientes, es fundamental dar solución a dos situaciones conflictivas que se pueden presentar:

Colisiones Dado que DILVE no impone, por el momento, ninguna restricción en los nombres de los archivos con las imágenes de cubierta, es posible que entre los libros extraídos haya dos o más con el mismo nombre aunque su contenido sea, lógicamente, distinto. A esta circunstancia se le denomina *colisión de nombres*.
En los registros CSV (u ONIX, en su caso) de todos los libros correspondientes se incluirá ese nombre como archivo de cubierta. Sin embargo, en la extracción de datos de DILVE solo se puede enviar un archivo con ese nombre, por lo que algunos de los libros estarán haciendo mención a una imagen de una portada que no les corresponde.

Huérfanos Si el nombre de un archivo con una imagen no incluye una referencia a su origen y contenido, es fácil que en la cadena de información aparezcan archivos *huérfanos*. Es decir, archivos de los que se desconoce a qué libro corresponden.
Para averiguarlo habría que saber en qué archivo CSV (u ONIX) se referencian y localizar los registros correspondientes. Esto no siempre es posible, por lo que, ante archivos tales como *don_quijote.jpg* o *poemas_amor_jpg*, habrá situaciones en las que la única alternativa que le queda al usuario es borrarlos de su disco ya que, aunque sabe que corresponden a imágenes de portada de *algunos* libros, desconoce de *cuáles* en concreto.

Para ayudar en la solución de estas situaciones, DILVE propone un convenio de normalización de nomenclatura de los archivos de recursos asociados a los libros cargados. Este convenio de normalización se describe, en detalle, en el documento *Cargas de textos e imágenes*, disponible en www.dilve.es.

En el presente documento se expone una versión reducida de este convenio, para el caso concreto de imágenes de cubierta.

DILVE incluye funcionalidades que permiten paliar los posibles problemas derivados de la carga de archivos con nombres no normalizados. Básicamente, esta funcionalidades permiten renombrar a los ficheros correspondientes, durante el proceso de extracción de datos, adaptando su nombre a los criterios de normalización indicados.

Estas funcionalidades serán imprescindibles para los usuarios de DILVE que quieran evitar colisiones o archivos huérfanos en sus extracciones de datos.

Sin embargo, estas funcionalidades solo son de ayuda para quienes utilicen el formato CSV en las extracciones. En el formato ONIX (formato en el que está basado DILVE), no es posible modificar los nombres de los archivos de recursos adjuntos (p.ej. la imagen de la portada).

Por lo tanto, DILVE recomienda a las editoriales que lo deseen que, al margen de los procesos de modificación de nombres en sus bases de datos internas, procedan a la carga de los archivos de cubiertas con los nombres normalizados siguiendo el procedimiento semiautomático cuyas instrucciones se detallan en el presente documento.

Proceso de normalización de nombres de archivos de portada

Para permitir a los consumidores de información la *extracción* de las imágenes de cubierta con nombres de archivos normalizados, y para facilitar a las editoriales la *carga* de los archivos con las imágenes de las portadas con nombres normalizados, se han creado dos campos CSV de salida que permiten la conversión automática.

Así, un usuario de DILVE que quiera obtener los archivos con los nombre normalizados deberá incluir, entre los campos elegidos para la extracción, los nombres *imagen_cubierta_normalizada* (e *imagen_cubierta2_normalizada*) en lugar de *imagen_cubierta* (e *imagen_cubierta2*).

Los campos descritos están disponibles para todos los usuarios y cualquiera puede obtener los archivos normalizados, utilizando extracciones en formato CSV.

Sin embargo, como se ha mencionado en la introducción, es muy conveniente que las editoriales vayan abordando el proceso de carga en DILVE de archivos de imágenes con nombres normalizados, puesto que en las extracciones en formato ONIX las imágenes se recuperan con el nombre utilizado por la editorial en la carga en DILVE.

En el resto del presente documento se describe en detalle el proceso a seguir por las editoriales que deseen normalizar los nombres de los archivos de imágenes sus portadas, cargados en DILVE. Para cualquier consulta tienen a su disposición el servicio de asistencia:

asistencia@dilve.es (914 135 127)

Campos CSV

Imagen de la cubierta normalizada

Tipo	Este campo solo se puede utilizar en extracciones CSV.
Nombre	`imagen_cubierta_normalizada`
Función	Permite renombrar los archivos de imágenes de cubierta (que deben estar cargados anteriormente) según el convenio para los nombres de imágenes, descrito a continuación y explicado con más detalle en el documento *Cargas de textos e imágenes*, disponible en www.dilve.es.
Comentarios	Este campo facilitará a las editoriales la actualización de los datos cargados en DILVE que no cumplan el convenio para los nombres de las imágenes. Asegurará a los usuarios de DILVE la obtención de los archivos de las imágenes con los nombres normalizados independientemente de cómo haya sido realizada la carga por parte de las editoriales.

Imagen de la cubierta normalizada (2)

Tipo	Este campo solo se puede utilizar en extracciones CSV.
Nombre	`imagen_cubierta2_normalizada`
Función	Véase el campo *Imagen de cubierta normalizada*.
Comentarios	Véase el campo *Imagen de cubierta normalizada*.

Convenio para los nombres de archivos de imágenes en DILVE

Los nombres de archivos deberán tener la siguiente estructura:

*nnnnnnnnnnnnn*_04_*X*.jpg

donde:

nnnnnnnnnnnnn	será el ISBN10 o el ISBN13 de la publicación. En todo caso, sin guiones ni otros separadores.
04	Indica el tipo de archivo. Debe ser 04 para la cubierta.
X	Es un código propio, de uso interno, que facilita más detalles sobre el tipo de archivo (ver más adelante).
jpg	Es la extensión del archivo. Este debe estar en formato JPEG (también se admite el formato png)

Para imágenes de cubierta e ilustraciones en general se utilizarán los siguientes códigos para la letra ***X***, en función del tamaño aproximado de la imagen contenida en el archivo (en píxeles):

h entre 675 x 675 y 450 x 450
l entre 450 x 450 y 300 x 300 (es una letra "L" minúscula)
m entre 300 x 300 y 200 x 200
n entre 200 x 200 y 135 x 135
s entre 135 x 135 y 90 x 90
t 90 x 90 o menos

Ejemplo:

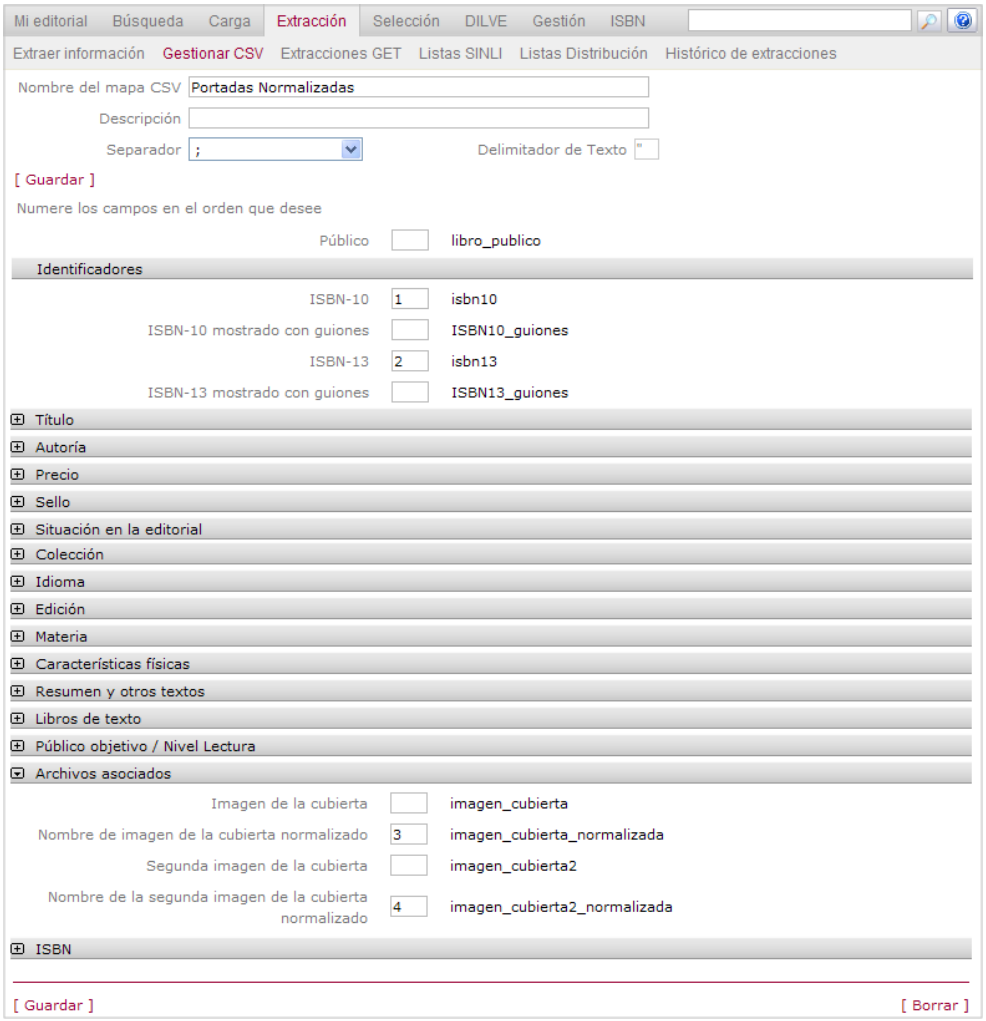
9781572316218_04_n.jpg es una imagen de la cubierta ("04") del libro con ISBN13 = 9781572316218, de tamaño aproximado 135x200 píxeles ("*n*") y en formato JPEG.

IMPORTANTE:	Para más detalles respecto al tratamiento de imágenes en DILVE, consulte el documento *Cargas de textos e imágenes*, disponible en www.dilve.es.

¿Cómo actualizar los datos?

1. *Extracción de los archivos cargados*

Crear un *mapa CSV* para incluir, en los archivos extraídos de DILVE, los campos con las imágenes de la cubierta con los nombres normalizados:



En el ejemplo, este mapa CSV se ha creado con el nombre *Portadas Normalizadas*.

Para el proceso descrito, únicamente son necesarios los campos de imagen de cubierta normalizada y un identificador del libro. Si se desea se puede incluir algún campo adicional, como el título, autor, etc.

Si se ha realizado la carga de una única imagen de la cubierta, no es necesario incluir el campo *imagen_cubierta2_normalizada*.

Para extraer datos de DILVE es necesaria la **creación de una *programación***:

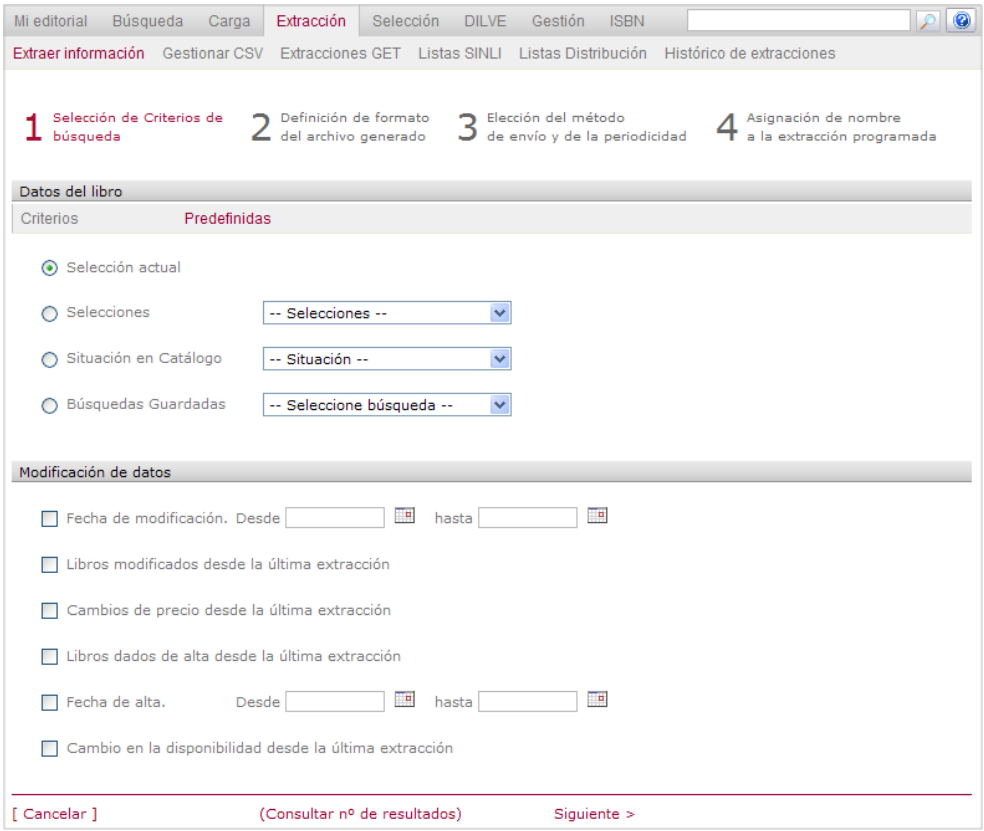
Posteriormente, *cada vez que se ejecute* la programación, DILVE *generará una extracción* de datos incluyendo los registros (fichas) y campos que, en el momento de la ejecución, cumplan los criterios inicialmente determinados en la programación

La creación de una programación se realiza en los siguientes pasos:

PASO 1: Criterios para selección de registros (fichas)

Existen muchas posibilidades para determinar los criterios de selección de los registros (fichas de libros) que se incluirán en cada extracción. Es importante recordar que estos criterios se aplican sobre los registros y las circunstancias existentes en DILVE en el momento de ejecutar la programación.

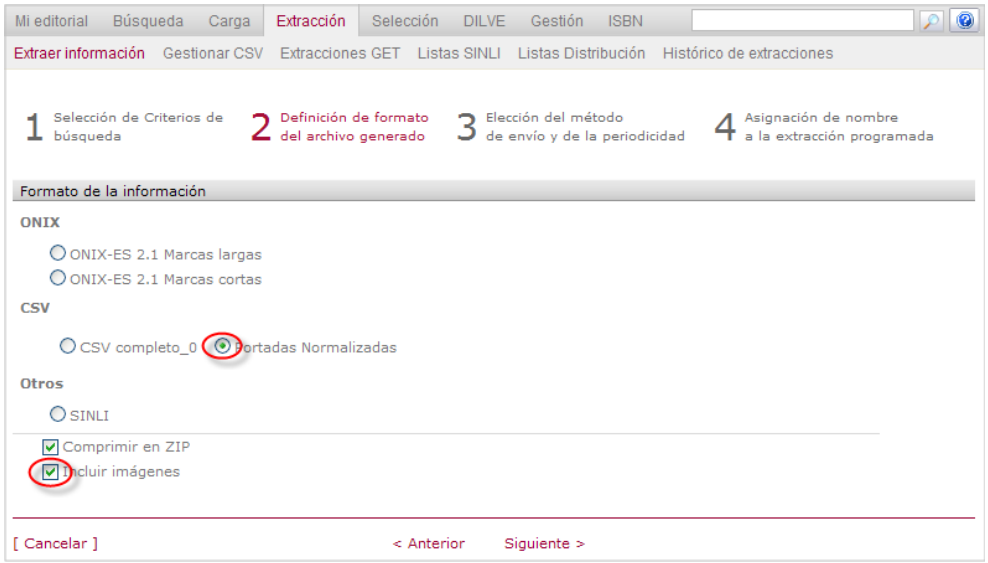
En este ejemplo, se supone que el usuario habrá seleccionado manualmente un conjunto de fichas antes de llevar a cabo la extracción.



En la pestaña *Predeterminadas* de debe marcar la opción *Selección actual*, para indicar que se desean incluir en la extracción aquellos libros que estén marcados como seleccionados en el momento de la ejecución de la programación.

PASO 2: Formato del archivo e inclusión de imágenes

En este paso de la programación se indica el formato en el que se desea recibir el archivo resultante de la extracción. También se indica si se desea que se incluyan las imágenes de la cubierta.

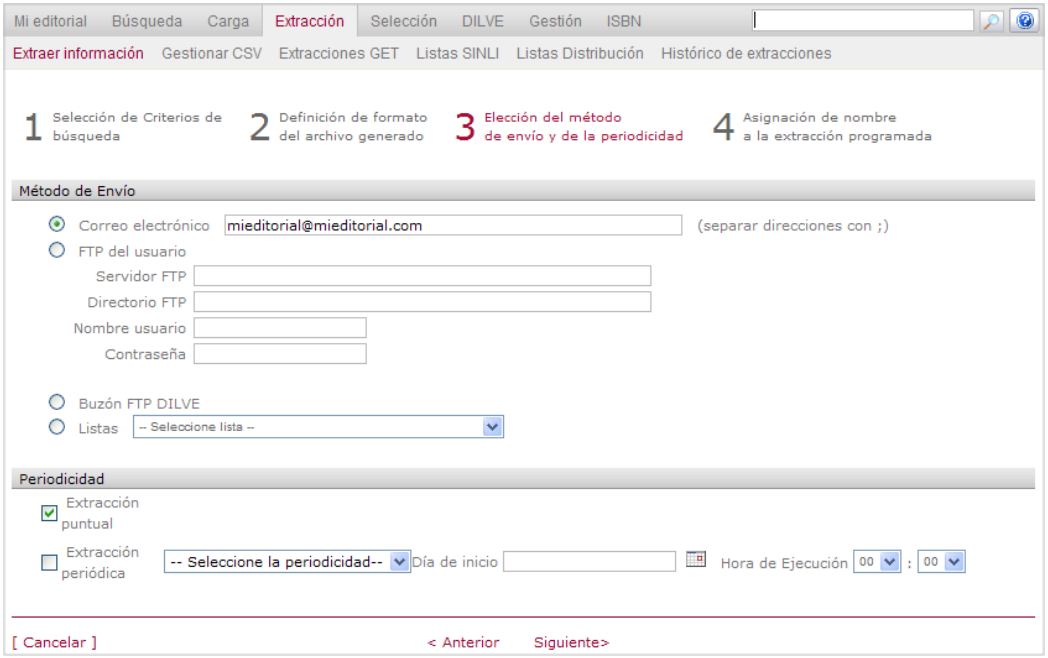


Para el formato del archivo, habrá que seleccionar el Mapa CSV creado anteriormente (*Portadas normalizadas*). También habrá que seleccionar la opción *Incluir imágenes* para que se incluyan los archivos con las imágenes de la portada en la extracción.

El sistema marcará automáticamente la opción *Comprimir en ZIP*, ya que en las extracciones que incluyan archivos de imágenes, además del propio CSV, todos estos archivos se *empaquetan* en un único archivo comprimido.

PASO 3: Método de envío y periodicidad

En este paso de la programación se indica el método de envío que debe aplicarse al archivo resultante de la extracción, y la periodicidad con la que debe ejecutarse.

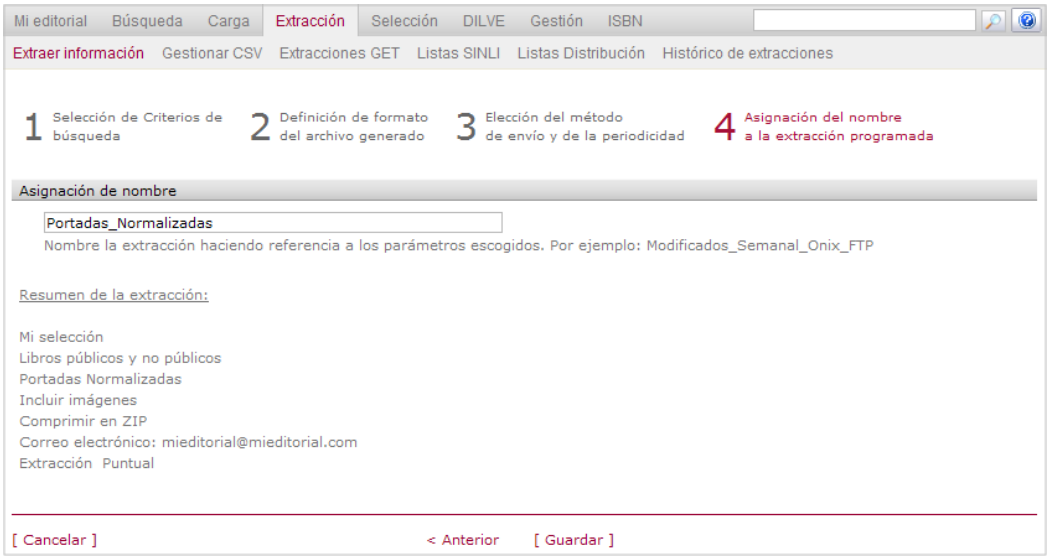


En este ejemplo suponemos que se desea recibir por correo el archivo resultante de la extracción. Se debe comprobar que la dirección de correo electrónico indicada es correcta.

También, dado que la finalidad de esta programación es la obtención de archivos de imágenes de cubierta con nombres normalizados para su posterior tratamiento, y dado, también, el criterio señalado para la selección de registros en el Paso 1, se seleccionará a opción *Extracción puntual*. Esto significa que la programación se ejecutará solo cuando el usuario lo indique.

PASO 4: Asignación de un nombre a la programación

Es el último paso de la programación, en el que se le debe asignar un nombre.
En el ejemplo es *Portadas_Normalizadas*.



Seleccionar los libros a extraer. Deberán seleccionarse los libros que tenga asignado un archivo de portada cuyo nombre no siga el estándar de normalización descrito anteriormente.

En caso de que se trate de un número muy elevado de libros es recomendable realizar la actualización por grupos, ya que se va a realizar una extracción que enviará por correo electrónico todos los archivos de portada de los libros seleccionados y dicho correo electrónico puede llegar a ser muy grande.

Para la realización de varias actualizaciones será necesario llevar a cabo este paso y los siguientes tantas veces como sea necesario.

La selección de los libros se lleva a cabo marcando el *check-box* de la primera columna en los listados de libros.

		ISBN-13	Título Todos	Autor	Editorial	Obra
☑		978-11-11-11831-0	Nuevo Título 1111118310	Autor, 1111118310	Mi editorial	M
☑		978-11-11-11832-7	Nuevo Título 1111118329	Autor, 1111118329	Mi editorial	M
☑		978-11-11-11833-4	Nuevo Título 1111118337	Autor, 1111118337	Mi editorial	M

Finalmente, tras comprobar que están seleccionados las fichas deseadas (esto se puede hacer accediendo al menú *Selección*), habrá que ejecutar la programación (en el ejemplo será la programación *Portadas_Normalizadas*) para obtener la extracción de los datos:

Mi editorial | Búsqueda | Carga | Extracción | Selección | DILVE | Gestión | ISBN

Extraer información | Gestionar CSV | Extracciones GET | Listas SINLI | Listas Distribución | Histórico de extracciones

Programaciones	Periodicidad	Formato	
Portadas_Normalizadas	Puntual	CSV Portadas Normalizadas	Editar / Ejecutar / Clonar / Borrar / Descargar
Mi_edi_ONIX	Puntual	ONIX-ES 2.1 Marcas Largas	Editar / Ejecutar / Clonar / Borrar / Descargar

DILVE realizará una extracción de los registros seleccionados, renombrará los correspondientes archivos con las imágenes de las cubiertas asignándoles nombres normalizados de acuerdo con el convenio DILVE, preparará un archivo CSV con el formato indicado y *empaquetará* todo el conjunto (archivo CSV y archivos de imágenes) mediante compresión ZIP.

Luego enviará el archivo resultante, como adjunto, en un mensaje de correo a la dirección indicada en la programación.

2. Actualización de los datos en DILVE

Una vez recibido el mensaje de correo con el archivo comprimido que contiene tanto el CSV que relaciona los libros con las imágenes como los propios archivos de las imágenes, se deberá abrir el archivo CSV para cambiar los nombres de los campos con las imágenes:

	A	B	C	D
1	isbn10	isbn13	imagen_cubierta_normalizada	imagen_cubierta2_normalizada
2	1111115095	9781111115098	9781111115098_04_t.jpg	9781111115098_04_s.jpg
3	1111115109	9781111115104	9781111115104_04_n.jpg	9781111115104_04_m.jpg
4	1111115117	9781111115111	9781111115111_04_l.jpg	9781111115111_04_l.jpg
5	1111115125	9781111115128	9781111115128_04_h.jpg	9781111115128_04_h.jpg

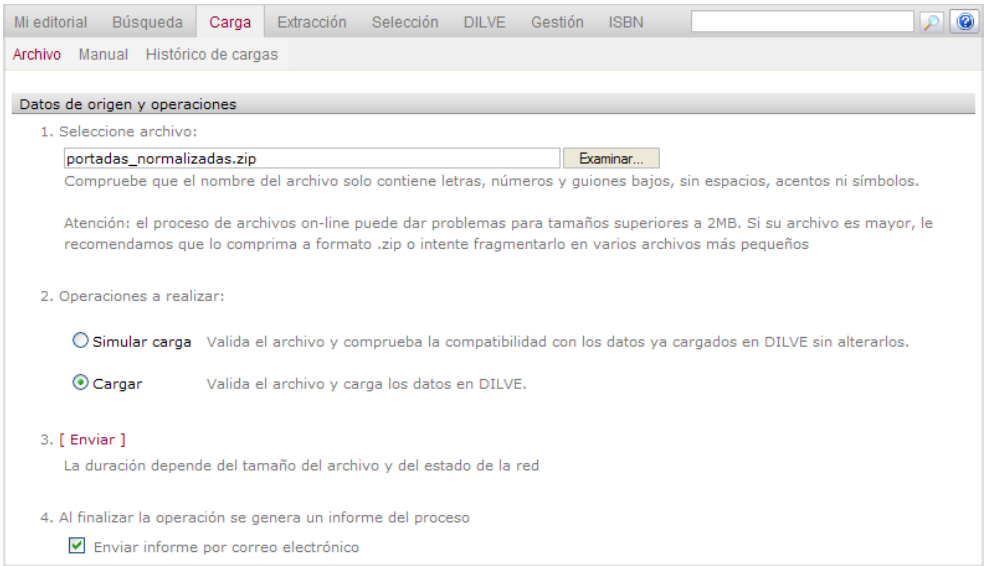
Cambiar:

- **imagen_cubierta_normalizada** por **imagen_cubierta**
- **imagen_cubierta2_normalizada** por **imagen_cubierta2**

En caso de que sólo se haya cargado una imagen de cubierta para cada libro y, por tanto, la última columna del archivo esté vacía, es conveniente eliminarla.

Si algunos libros tienen datos en dicha columna, pero quedan algunas celdas vacías, se recomienda situar en último lugar una de las columnas de ISBN (que no puede contener ningún valor vacío). Esto es porque si, en la última columna, existen celdas vacías la transformación a formato CSV por parte de Excel puede dar resultados incorrectos, que generarán problemas en la carga en DILVE.

Tras la modificación del CSV y su inclusión de nuevo en el archivo comprimido, se debe llevar a cabo la actualización de los datos en DILVE mediante la carga de este archivo comprimido:



Es recomendable leer el Informe de resultados para comprobar que la carga de todos los registros se ha realizado correctamente. También conviene comprobar los datos correspondientes a alguno de los libros actualizados, accediendo a su ficha.

Con este proceso se consiguen cargar en DILVE los archivos con las imágenes de cubierta, con los nombres normalizados.

A partir de este momento tanto las extracciones en formato nativo ONIX, como en formatos CSV que incluyan el campo *imagen_cubierta* (o *imagen_cubierta2*) obtendrán las imágenes con los nombres normalizados, dado que así es como figuran cargadas en DILVE.

Además, la editorial dispondrá, en su sistema local, de copias de estos archivos con los nombres normalizados.

Esto le permitiría a la editorial ir sustituyendo, en sus propios sistemas, los archivos de imágenes con nombres "internos" por los archivos con los nombres normalizados (p.ej. *9781572316218_04_n.jpg* en lugar de *don_quijote.jpg)*.

Lógicamente, la viabilidad y pasos a seguir en este proceso dependerán de las circunstancias de cada entidad.