Distribuidor de información
del libro español en venta

Cómo revisar y mejorar los datos de mi catálogo

v1 07/2008

Contenido

Cómo mejorar mis datos 3
Niveles y plantillas CSV 3
Recomendaciones del Informe 4

1 **Extracción de los datos** 5
Preparar el mapa CSV 5
Extracción de datos 6

2 **Actualizar los datos** 8
Autofiltros en Excel 9
Completar datos 10
Otras plantillas 11

3 **Carga de datos** 11

Anexos
Abrir CSV con Excel 12

DILVE es un proyecto de libro-es.com

Promovido por

FEDERACIÓN DE GREMIOS DE EDITORES DE ESPAÑA

Fundación Germán Sánchez Ruipérez

Con el patrocinio de

CEDRO

Cómo mejorar mis datos

En el Informe DILVE enviado a las editoriales (julio 2008) se recomiendan posibles mejoras de los datos tanto en el número de campos (ausencia de algún dato) como en la normalización de los datos (todo mayúsculas y nombres de archivos de cubierta).

El procedimiento para revisar y mejorar los datos del catalogo por medio de archivos CSV (editables en Excel o similar) será el siguiente:

1 Extraer los datos presentes en DILVE en forma de archivo CSV.

2 Editar en Excel el archivo y localizar los títulos a modificar (aprovechando las funcionalidades del programa para automatizar tareas) y corregir y/o añadir los datos correspondientes.

3 Cargar los nuevos datos en DILVE.

Este proceso puede realizarse bien mediante un único archivo CSV con todos los datos de todos los títulos o bien mediante la utilización de las plantillas que se describen a continuación que permiten manejar subconjuntos de campos más reducidos. Para aquellos usuarios con un conocimiento básico de Excel recomendamos esta segunda opción.

Niveles y plantillas CSV

En DILVE la riqueza de contenidos se estructura a partir de cuatro niveles:

- **Nivel 0 Inicial:** datos mínimos de identificación del libro.
- **Nivel 1 Básico:** datos sobre autorías, idiomas, edición, colecciones, características físicas, otras menciones de responsabilidad, libro de texto y obras completas.
- **Nivel 2 Rico:** cubiertas, resúmenes, títulos, disponibilidad, edición, materias, características físicas y destinatarios.
- **Nivel 3 Avanzado:** índice, primer capítulo, reseña, comentarios, etc.

Para cada uno de estos niveles DILVE dispone de una serie de plantillas de mapas CSV (o listas de campos) que facilitarán la gestión organizada de los datos:

MAPA CSV		CAMPOS QUE INCLUYE
Nivel 0 Inicial	Identificación	isbn13
		titulo
		autor
		fecha_public
		precio_sin_iva
		libro_publico
Nivel 1A Campos Básicos		subtitulo
		idioma
		idioma_orig
		idioma_traduccion
		codigo_isbn_materia
	Colecciones	coleccion
		num_en_colec
		coleccion2
		num_en_colec2
	Autorías	traductor
		ilustrador
	Edición	num_edicion
		tipo_edic_isbn
		isbn13_edicion_anterior
		num_impresion
		fecha_impresion
		situ_catalogo_editorial
		publico_objetivo
	Caract. físicas	ancho
		alto
		peso
		num_pags
		encuad_isbn
		descripcion_fisica
		num_ilustr
		num_ilustr_color
		material_anejo

Nivel 1B Menciones Responsabilidad	Otras menciones responsabilidad	ilustrador_cubierta
		autor_prologo
		autor_notas
		editor_literario
		otras_menciones1
		otras_menciones2
		otras_menciones3
Nivel 1C Libro de Texto	Libro de texto	precio (si precio libre)
		material_educ
		tipo_material_educ
		tipo_enseñanza
		nivel
		ciclo
		curso
		com_autonoma
		etapa_ciclo_curso
		titulo_isbn
		codigo_asignatura
Nivel 1D Obras Completas	Obras Completas	obra_completa
		num_volumenes
		titulo_volumen
		num_volumen
		titulo_obra_completa
		isbn13_obra_completa
Nivel 2A Cubiertas	Cubiertas	imagen_cubierta
		imagen_cubierta2
Nivel 2B Resumenes	Resúmenes (3)	texto_resumen
		idioma_resumen
Nivel 2C Completar Informacion	Títulos	titulo_orig
		titulo_distribuidor
	Disponibilidad	disponibilidad
		fecha_disponibilidad
	Edición	ciudad_public
		pais_public
		coment_edicion
	Materias	codigo_eci_materia
		términos_clave_materia
	Caract. físicas	grueso_cm
	Destinatarios	nivel_lectura
		edad_interes
Nivel 3 Textos	Textos	indice
		reseña
		cita_de_reseña
		titular_promocional
		comentario_autor
		descripcion_lector
		descripcion_medios
		descripcion_libreria
		texto_solapa
		extracto
		primer_capiltulo

Recomendaciones del Informe

¿Cómo incorporar el autor, el precio y la fecha de publicación de algunos libros?

Supongamos que la editorial ha recibido en el Informe la siguiente recomendación:

> Para alcanzar el Nivel 0 (Inicial) complete y/o revise los siguientes datos:
> Tiene 1 (20%) libros sin autor.
> Tiene 1 (20%) libros precio.
> Tiene 1 (20%) libros sin fecha de publicación/edición.

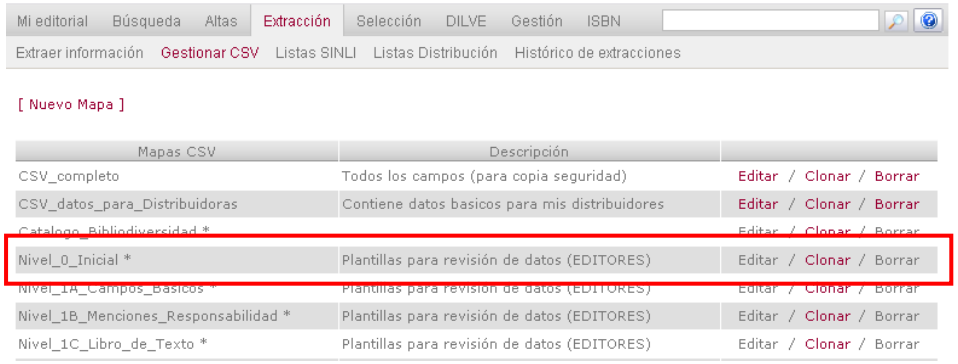
A continuación se describen los pasos necesarios para **detectar** los libros afectados **y completar la información** mencionada.

1 Extracción de los datos

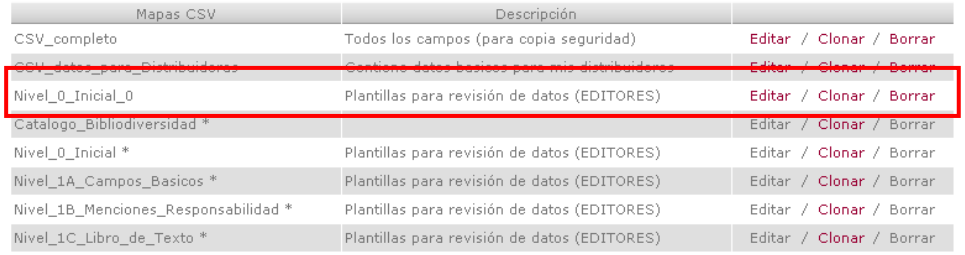
Cada uno de estos Mapas CSV permitirá la extracción de subconjuntos, más o menos, reducidos de datos. Volvemos a insistir en que el mejor procedimiento para modificar los datos es el trabajo en archivos diferenciados por niveles. Modificar todos los datos de todos los libros en una única hoja Excel, dependiendo del nº de títulos y de campos, puede resultar, además de engorroso fuente de errores indeseables.

Preparar el mapa CSV

Una vez dentro de la aplicación (www.dilve.es), el usuario editorial encontrará en la sección **EXTRACCIÓN > GESTIONAR CSV** los mapas CSV preconfigurados (o listas de campos), se distinguen porque su nombre lleva un asterisco (*) y solo pueden clonarse.



A continuación vamos a duplicar el mapa denominado Nivel_0_Inicial (que incluye los campos `isbn13`, `titulo`, `autor`, `fecha_public` y `precio_sin_iva`)[1] mediante el botón Clonar.



Con esta operación tendremos disponible el mapa CSV durante el proceso de extracción de datos.

[1] Por defecto se creará el archivo con el ISBN 13 con guiones, en estos mapas CSV puede hacer las modificaciones que le interesen.

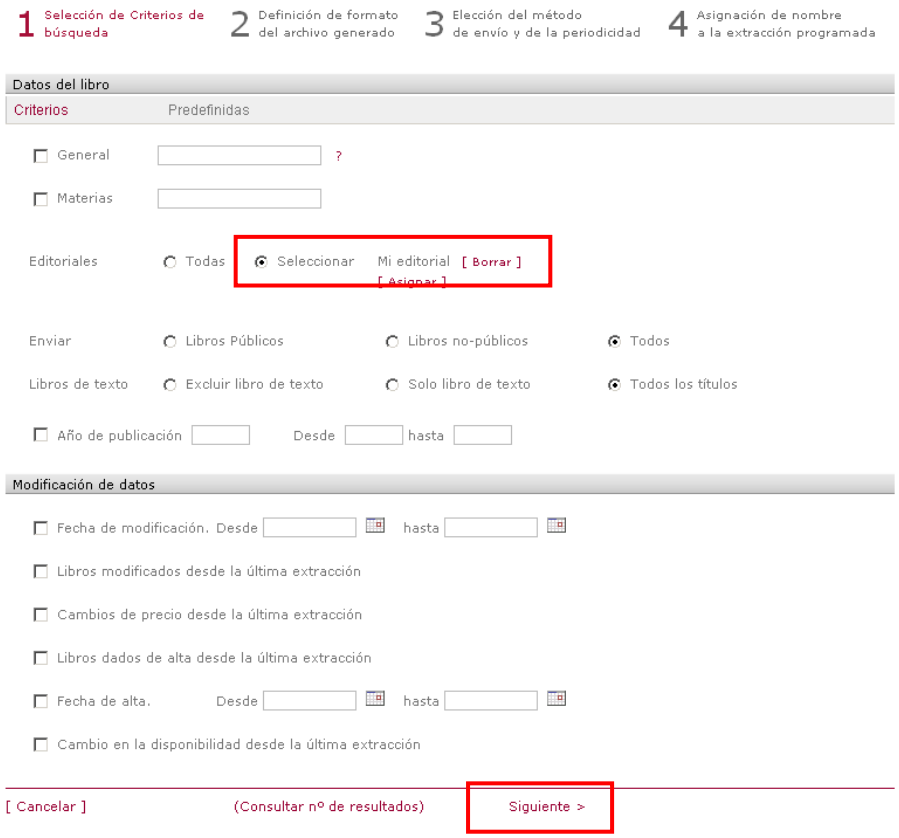
Así mismo, se incluye en todas las plantillas al final la columna `libro_publico` para evitar que la última columna tenga valores vacíos por los problemas que esto genera en la carga.

Extracción de datos

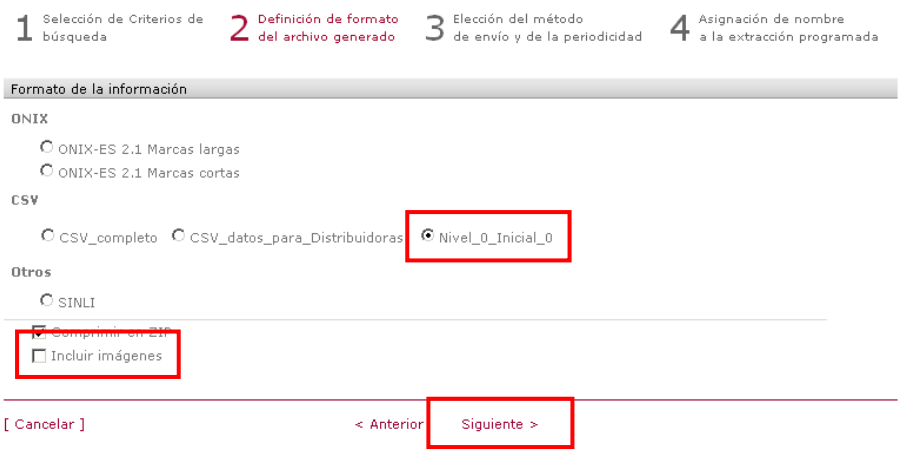
Para extraer los datos el usuario debe crear una programación. En la sección **EXTRACCIÓN > EXTRAER INFORMACIÓN** se muestran las programaciones creadas por usuario, las plantillas existentes y se crean nuevas programaciones:

Las programaciones se configuran a través de 4 pantallas:

- En la primera se define el conjunto de libros a extraer: todo el catálogo de nuestra editorial.
- En la segunda pantalla se define el conjunto de campos mediante el mapa CSV creado en el apartado anterior: mapa Nivel_0_Inicial_0.
- En la tercera pantalla se especifica el destino del archivo resultante: nuestra dirección de correo electrónico.
- En la cuarta pantalla se asigna un nombre descriptivo a la programación creada.
- Por último es necesario ejecutar la programación para que los datos lleguen a nuestro buzón de correo electrónico.

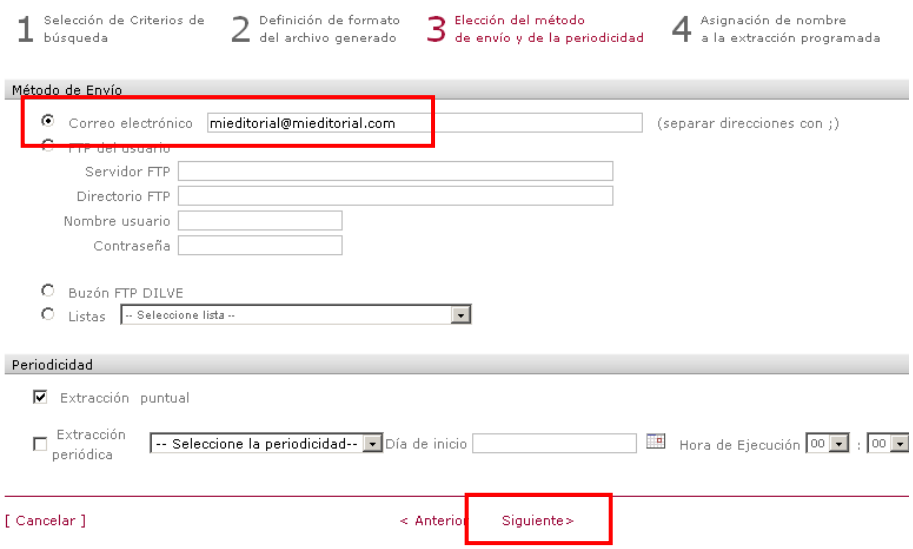


Por defecto aparece seleccionada nuestra propia editorial y se pasa a la pantalla nº 2.

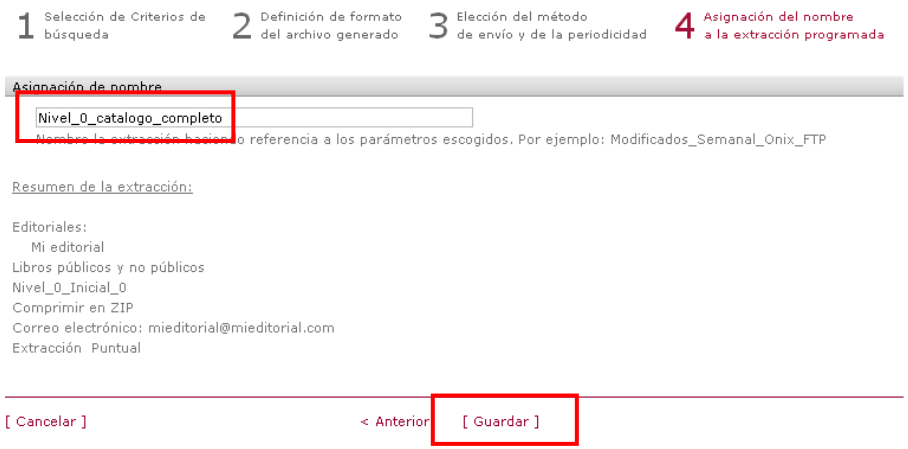


Seleccionar el formato CSV creado en el capitulo anterior y pasar a la pantalla nº 3.

En caso de estar completando los datos de la plantilla *Nivel 2A Cubiertas* será necesario marcar la opción “Incluir imágenes”. Si su tarea consiste en estandarizar los nombres de los archivos, le recomendamos que consulte el documento **Proceso de normalización de archivos de cubierta** disponible en la sección de Documentación de la web.



Elegir el método de envío (correo electrónico) y verificar que la dirección es correcta. Para terminar de crear la programación, pasar a la pantalla nº 4.



Asignar un nombre descriptivo de su contenido a la programación y Guardar el resultado.

La programación recién creada aparecerá en el listado:

Ejecutar la programación, en unos minutos (dependiendo del tamaño del catálogo) recibiremos en la dirección de correo el archivo con los datos solicitados.

2 Actualizar los datos

El archivo recibido, una vez descomprimido si es necesario, pude abrirse con Excel que proporciona herramientas muy útiles para el manejo de los datos.

	A	B	C	D	E
1	ISBN13_guiones	titulo	autor	precio_sin_iva	fecha_public
2	978-84-88015-03-7	Blanquito, peón de brega		8,65	1991
3	978-84-88015-06-8	El solitario y su multitud	Garoz Moreno, Rafael		199301
4	978-84-88015-18-1	Últimos poemas (1983-1992)	Gurruchaga, Joaquín	8,65	19951001
5	978-84-88015-26-6	El tiempo, el humo, el pasado	Gurruchaga, Joaquín	11,54	
6	978-84-88015-32-7	Mediobiografía: selección de poemas biográficos	Otero, Blas de	11,54	1997
7					

Compruebe que al abrir el fichero los datos de ISBN, fechas, etc. tienen el formato correcto. Si los datos presentan una estructura extraña consulte el apartado *Anexos*

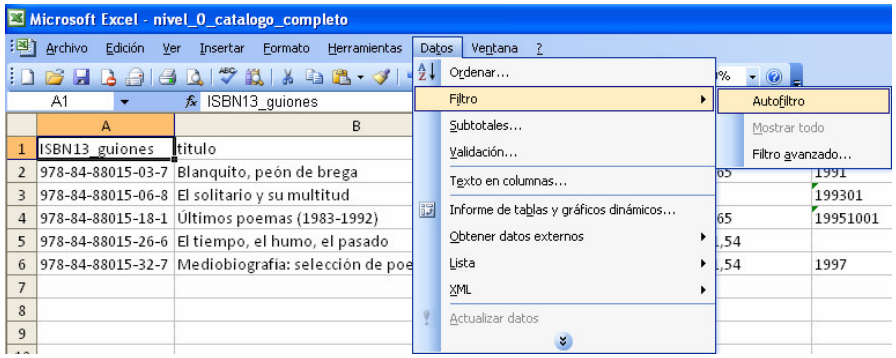
Abrir CSV con Excel al final de esta guía.

Es conveniente guardar el archivo en formato Excel (xls) en vez de CSV para no perder los formatos de las columnas cada vez que se abra.

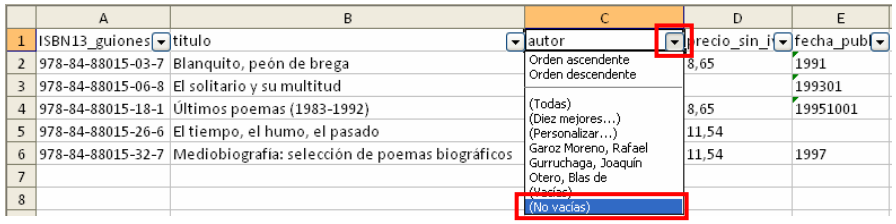
También es importante que los libros que no vayan a sufrir modificaciones no se envíen a DILVE (se falsearía la fecha de actualización), por ello es necesario eliminarlos de este archivo. A continuación se explica cómo hacerlo.

Autofiltros en Excel

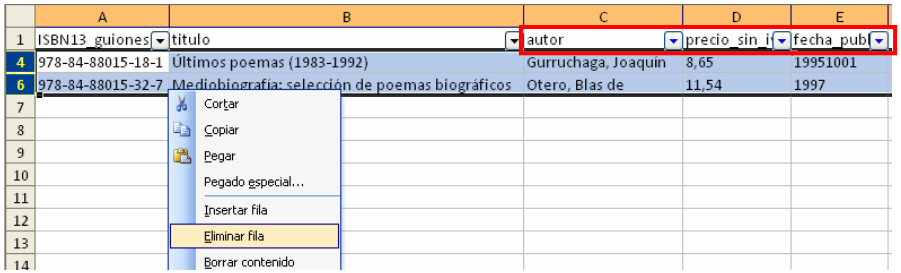
Seleccionar la opción de filtrado de Excel en el menú **DATOS > FILTRO > AUTOFILTRO**:



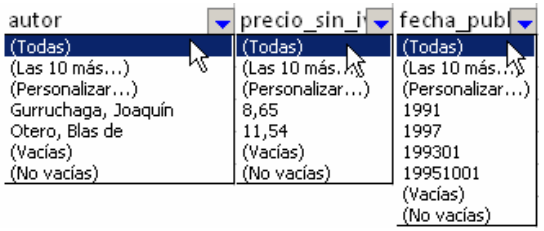
Para descartar los libros que tienen todos los datos es necesario seleccionar/filtrar las celdas 'NO VACÍAS' de todas las columnas que pueden tener celdas vacías:



Una vez aplicado el filtro a todas las columnas, quedarán visibles únicamente las filas de libros que deben eliminarse:



Una vez eliminados los libros se deshacen los filtros[2] eligiendo la opción 'TODAS'



como resultado quedará el listado de libros que tiene alguno de los campos mencionados vacío.

[2] Como puede observase en la imagen estos menús desplegables muestran, además de los datos existentes en toda la columna, otro tipo de opciones que pueden ser de utilidad a la hora de seleccionar y tratar datos.

Completar datos

Ahora la utilización del Autofiltro buscando celdas 'VACÍAS' columna por columna nos permitirá detectar y completar los datos necesarios, en el ejemplo empezaremos por el autor:

	A	B	C	D	E
1	ISBN13_guiones	titulo	autor	precio_sin_i	fecha_pub
2	978-84-88015-03-7	Blanquito, peón de brega		8,65	1991
3	978-84-88015-06-8	El solitario y su multitud			199301
4	978-84-88015-26-6	El tiempo, el humo, el pasado		11,54	
5					
6					
7					
8					

Orden ascendente
Orden descendente
(Todas)
(Diez mejores...)
(Personalizar...)
Garoz Moreno, Rafael
Gurruchaga, Joaquín
Otero, Blas de
(Vacías)
(No vacías)

Al aplicar el filtro VACÍAS Excel mostrará únicamente las filas (los libros) en las que dicha columna esté vacía:

	A	B	C	D	E
1	ISBN13_guiones	titulo	autor	precio_sin_iva	fecha_public
2	978-84-88015-03-7	Blanquito, peón de brega		8,65	1991
3					

Tras completar los registros y para continuar con el resto de datos deberá anular el Autofiltro seleccionando "(TODAS)" en la columna) y volver a seleccionar las celdas vacías del siguiente campo.

Repetir este proceso hasta completar toda la información:

	A	B	C	D	E
1	ISBN13_guiones	titulo	autor	precio_sin_iva	fecha_public
2	978-84-88015-03-7	Blanquito, peón de brega	Cela Trulock, Jorge	8,65	1991
3	978-84-88015-06-8	El solitario y su multitud	Garoz Moreno, Rafael	7,69	199301
4	978-84-88015-26-6	El tiempo, el humo, el pasado	Gurruchaga, Joaquin	11,54	1996

El programa Excel permite agilizar tareas como rellenar varias columnas con los mismos valores, selecciones, ordenaciones, etc. por lo que es muy aconsejable disponer de algún manual o libro de ayuda en el manejo de este programa.

Otras plantillas

En el ejemplo anterior se ha tratado el Nivel 0, los datos iniciales para dar de alta un libro. En el Informe a las editoriales se hacen también recomendaciones para alcanzar el resto de niveles en cuanto a calidad de datos se refiere. Para simplificar el manejo de datos se han subdividido los niveles 1 y 2 en distintas plantillas descritas en el primer apartado.

En función de la plantilla que se esté utilizando existen una serie de recomendaciones:

- **1C Libro de texto.** Extraer únicamente los libros marcados como libro de texto seleccionando la opción “Sólo libro de texto” en el paso 1 de la extracción.
- **1D Obras Completas.** Hacer una selección previa de los ISBN de obras completas y de volúmenes utilizando las búsquedas o las selecciones almacenadas (ver *Guía de uso para editoriales*)
- **2A Cubiertas.** Extraer los archivos de cubierta solo si no dispone de ellos en su sistema. La inclusión de las cubiertas en la extracción genera un archivo de gran tamaño que puede no caber en su buzón de correo electrónico.

Al utilizar las plantillas preconfiguradas es posible que en algunas columnas los datos estén completos y no sea necesario enviar la información. Le recomendamos que elimine dichas columnas en el Excel para evitar el envío de datos innecesarios que ralenticen el proceso de carga.

3 Carga de datos

La actualización de los datos en DILVE se realiza de la forma habitual descrita tanto en el *Guía de uso para editoriales* como en el *Manual de campos CSV*.

Es necesario guardar el archivo Excel en formato

- “CSV (delimitado por comas) (*.csv)” en Windows o
- “CSV (para Windows)” en MacOS X.

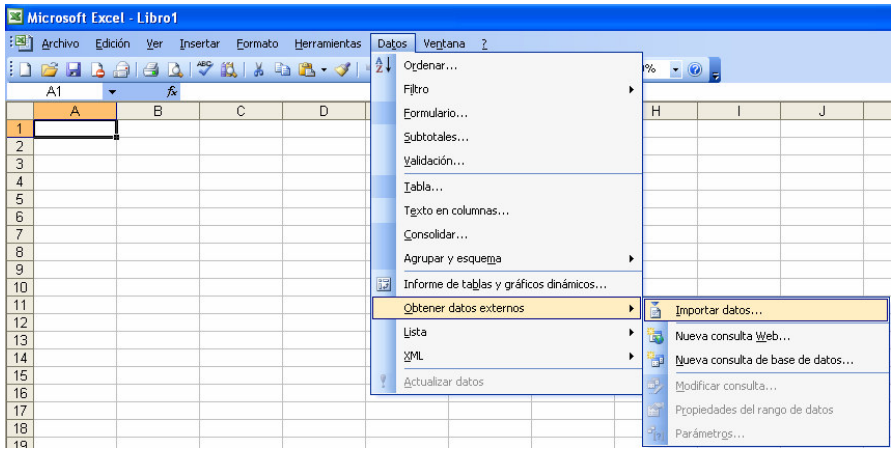
Una vez realizada la carga de datos en DILVE y comprobado el Informe de resultados, es posible volver a ejecutar la extracción programada en el paso 1 y comprobar los datos cargados.

Anexos

Abrir CSV con Excel

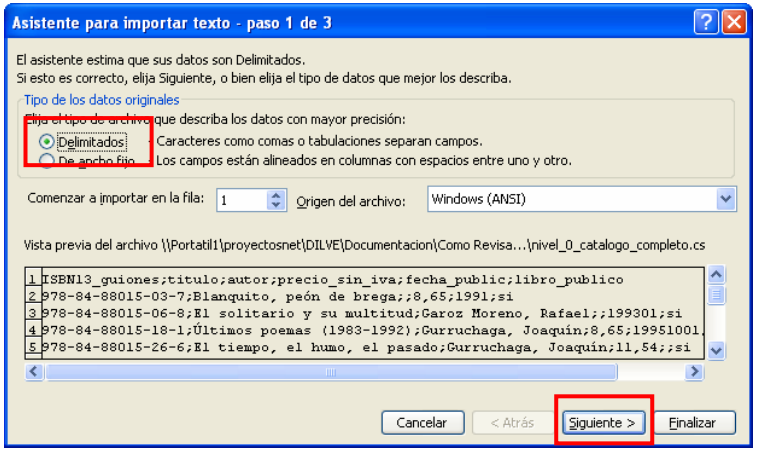
Para evitar que la edición del archivo CSV con Excel provoque cambios en los datos, como eliminación de ceros a la izquierda en algunos campos, formato del ISBN o las fechas, etc., a continuación se describe de forma detallada el proceso.

En primer lugar crear un archivo nuevo en Excel y seleccionar DATOS > OBTENER DATOS EXTERNOS > IMPORTAR DATOS...



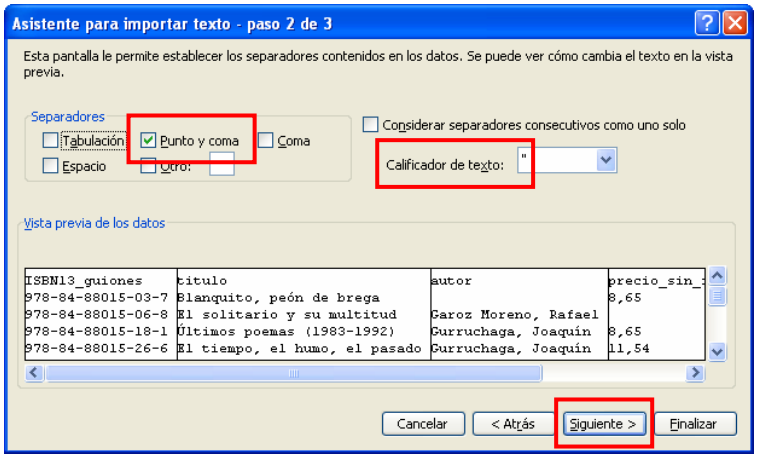
Seleccionar el archivo CSV recibido de DILVE y hacer clic en Importar. La importación se realiza en los siguientes 3 pasos.

Paso 1: Indicar que los datos están delimitados por un separador:

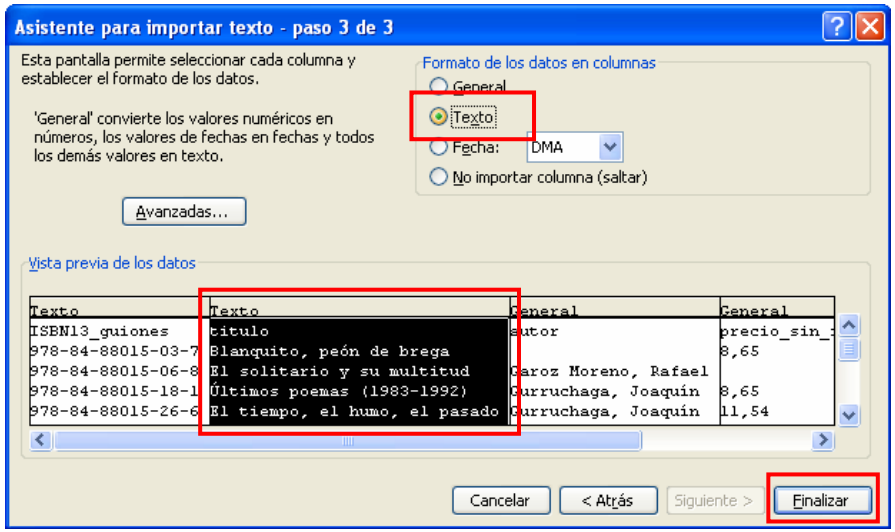


Paso 2: Indicar cuál es el separador (punto y coma) y el Calificador de texto (comillas si no se ha modificado al crear el Mapa CSV).

En la Vista previa de los datos se muestra la separación de campos que se va a realizar lo que nos permite comprobar que los valores seleccionados hasta ahora son correctos.



Paso 3: Para evitar que Excel interprete automáticamente el contenido de los datos es imprescindible indicar que todas y cada una de las columnas contienen datos en formato texto:



Por último el programa pregunta dónde colocar los datos (primera celda de la hoja vacía): Aceptar y Guardar el archivo.