

División de Gestión y Difusión de la Información

XIII Jornada Expania

Santiago de Compostela, 19 de mayo de 2016

Trabajando con las utilidades de SFX

Caso práctico 3.2.1 Carga y activación de un grupo de libros

José Luis Galán Cabilla

Unidad de Biblioteca

ÍNDICE

1	Objetivo	1
2	Dificultad	1
3	Descripción	1
4	Solución	1
4.1	Preparación del fichero de carga	1
4.2	Guardar el fichero en formato TSV	2
4.2.1	Guardar en formato TSV en Calc	2
4.2.2	Guardar en formato TSV en Excel	2
4.3	Carga del fichero en el DataLoader	3
4.4	Comprobación de resultados	3
5	Informes	4
6	Conclusión	5

1 Objetivo

Utilización de **DataLoader** para la carga y activación de **portfolios** de un fichero en formato texto delimitado por tabulaciones (**formato TSV**).

2 Dificultad

Baja.

3 Descripción

Se ha adquirido una selección de libros electrónicos de la colección *Business and Management de ElgardOnline* y se desea incluir esta colección en SFX. El proveedor ha suministrado un **fichero Excel con los ISBN y los títulos adquiridos**.

4 Solución

4.1 Preparación del fichero de carga

Al abrir el fichero facilitado por el proveedor se puede observar que incluye los ISBNs con espacios entre sus grupos de números (por ejemplo, aparece 978 0 85793 228 0 en lugar 9780857932280). Utilizando la **función de SUSTITUIR de Calc y Excel** es fácil quitar estos espacios, pero no es necesario ya que SFX es capaz de reconocerlos con este formato. A nivel de ISBN e ISSN, SFX depura los caracteres que no son numéricos antes de cargarlos.

También se puede observar que algunos libros no tienen ISBN papel (columna I) pero sí elISBN (columna K), como por ejemplo la celda I20. Esto tampoco es problema pues se puede utilizar indistintamente el ISBN o elISBN (y el ISSN o elISSN) para la carga.

Para conservar el fichero original sin cambios, guardar una copia de trabajo con el nombre **321carga** (**321carga.ods** y **321carga.xlsx**) sobre la que se realizan los siguientes cambios:

- Rellenar las celdas vacías de los ISBN en papel de la *columna I* con los valores de sus elISBN de la *columna J*
- Cortar la *columna. J* y pegarla como *columna. A*
- Insertar una *columna B* en blanco y rellenarla con la palabra **ACTIVE**
- Eliminar las filas sin información, *filas 40-43* y *fila 1*
- Para mayor claridad, aunque no es necesario¹, eliminar también las *columnas. C-D*.



A	B	C	D	E
978 0 85793 229 7	ACTIVE	Handbook of Multilevel Finance		Ahmad, E. Brosio, G.
978 0 85793 267 9	ACTIVE	Handbook of Research Methods and Applications in Economic Geography		Karlsson, C. Andersson, M. Normal
978 0 85793 781 0	ACTIVE	The Evolution of Central Banking and Monetary Policy in the Asia-Pacific		Hossain, A.A.
978 0 85793 783 4	ACTIVE	Central Banking and Monetary Policy in Muslim-Majority Countries		Hossain, A.A.
978 0 85793 793 3	ACTIVE	Handbook of Research Methods and Applications in Transport Economics and Policy		Nash, C.
978 1 78100 026 7	ACTIVE	Industrial Policy in Developing Countries	Failing Markets, Weak States	Altenburg, T. Lütkenhorst, W.
978 1 78195 453 9	ACTIVE	The Economics of International Integration, Second Edition		Jovanović, M.N.
978 1 78254 494 4	ACTIVE	Companion to the Political Economy of Rent Seeking		Congleton, R.D. Hillman, A.L.
978 1 78254 744 0	ACTIVE	The Encyclopedia of Central Banking		Rochon, L.-P. Rossi, S.
978 1 78347 237 6	ACTIVE	The New Economics of Income Distribution	Introducing Equilibrium Concepts into a Contested Field	Sell, F.L.

De esta forma, se dispone de los datos necesarios para crear un fichero de carga de **portfolios** para el **DataLoader**:

- Columna A *Primary key (ISBNs en este caso)*
- Columna B *ACTIVATION STATUS*

¹ El **DataLoader** ignora todas las columnas que siguen a las que se especifican al definir el contenido del fichero de carga, por lo que podemos conservar todas las que consideremos necesarias.

4.2 Guardar el fichero en formato TSV

MUY IMPORTANTE: los nombres de los ficheros que se utilizan en el *DataLoader* no pueden contener espacio y es aconsejable utilizar sólo caracteres alfanuméricos y guiones bajos.

4.2.1 Guardar en formato TSV en Calc

Para guardar una hoja de *Calc* en formato TSV hacer lo siguiente:

- Seleccionar **Archivo / Guardar como**
- Seleccionar la **ruta de destino del fichero**
- Dar un **nombre al fichero que no contenga espacios**
- Seleccionar **Texto CSV**
- Marcar **Editar configuración de filtros** para poder cambiar los separadores de coma a tabuladores
- Pulsar el botón **Guardar**
- Se abre una ventana informando que se puede perder el formato al guardar en archivo de texto CSV. Pulsar en **Mantener el formato actual**
- Se abre la ventana **Exportar a un archivo de texto**
- Seleccionar **UTF-8** como **codificación**
- Seleccionar **{Tabuladores}** como **delimitador de campo**
- **Dejar en blanco el campo delimitador de texto**
- Pulsar el botón **Aceptar**

4.2.2 Guardar en formato TSV en Excel

Para guardar una hoja de *Excel* en formato TSV hacer lo siguiente:

- Seleccionar **Archivo / Guardar como**
- Seleccionar la **ruta de destino del fichero**
- Dar un **nombre al fichero que no contenga espacios**
- Seleccionar **Texto delimitado por tabulaciones** en el desplegable de **tipo de archivo**.
- Pulsar el botón **Guardar**
- Pulsar **Aceptar** en la ventana que se abre avisando que se pueden perder algunas características del libro al guardarlo como texto delimitado por tabulaciones.
- Para evitar problemas de codificación² **abrir el fichero en el editor de texto y guardarlo como UTF-8**.

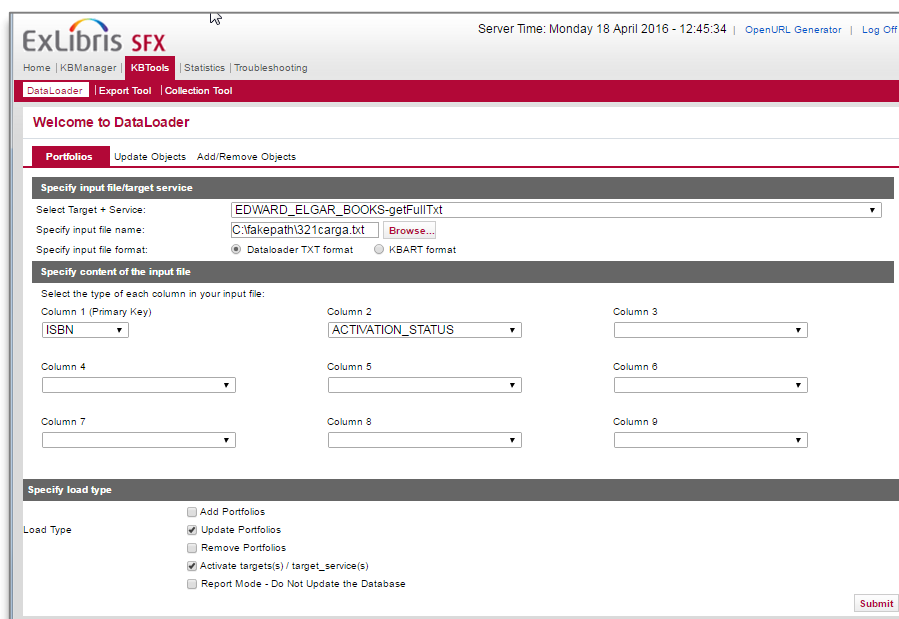
Fichero de carga en formato TSV: [321carga.txt](#)

² Excel guarda en codificación ANSI y en las versiones actuales no se puede cambiar la codificación al guardar.

4.3 Carga del fichero en el DataLoader

Para cargar el fichero hacer lo siguiente:

- Abrir el **DataLoader**
- Seleccionar la pestaña **Portfolios**
- Seleccionar el **Target Service** *EDWARD_ELGAR_BOOKS-getFullTxt*
- Seleccionar el **fichero de carga**
- Seleccionar *ISBN* para la **columna 1 (primary key)**
- Seleccionar *ACTIVATION STATUS* para la **columna 2**
- Marcar en **tipo de carga** *Update Portfolios* y *Activate targets(s) / target_service(s)*
- Pulsar el botón **Submit**



Esperar a que termine la carga.

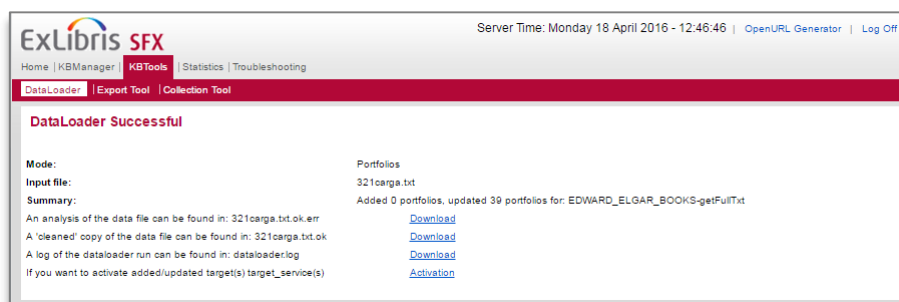
Existe también la posibilidad de realizar la carga en **Report Mode**, es decir, sin actualizar la base de datos y así poder comprobar previamente su resultado.

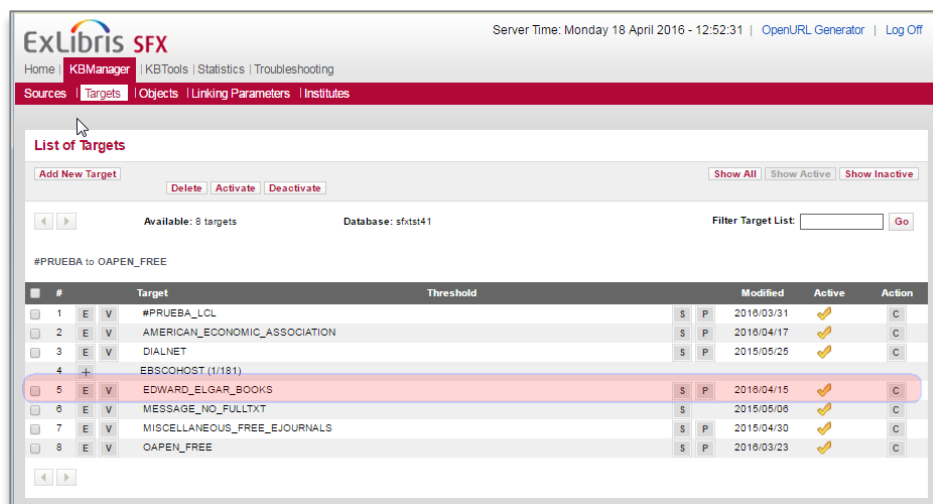
4.4 Comprobación de resultados

Una vez terminada la carga, el *DataLoader* presenta los resultados. En este caso parece que todo ha salido bien ya que se trataba de cargar y activar 39 portfolios y efectivamente esto es lo que dice el resumen de la carga:

Summary: updated 39 portfolios for: EDWARD_ELGAR_BOOKS-getFullTxt

Además se puede comprobar que efectivamente se han cargado y activado los libros consultando el *Target* y los *portfolios*.

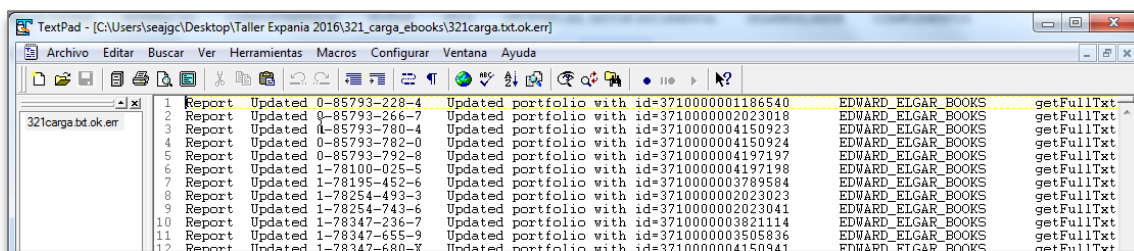




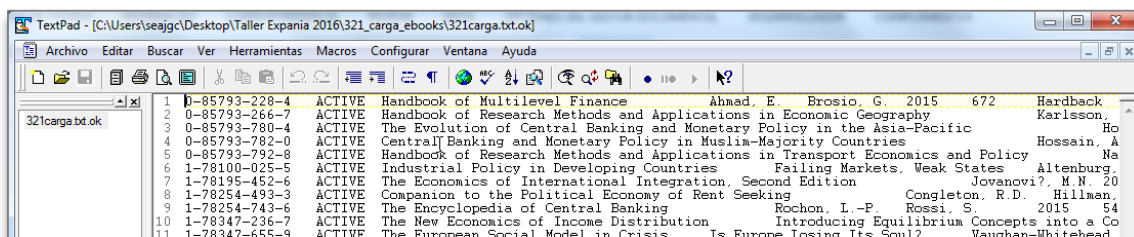
5 Informes

Como resultado de la carga se generan los siguientes informes:

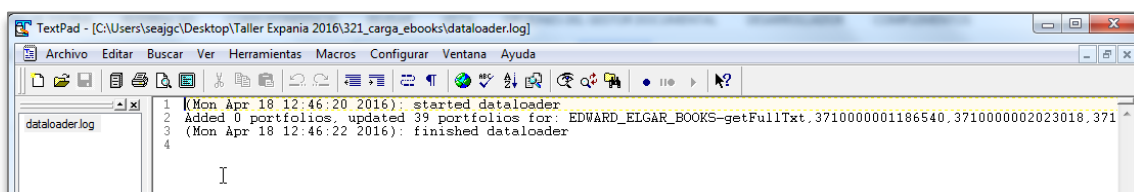
- Un análisis del resultado de la carga (**321carga.txt.ok.err**)



- Una copia del fichero de datos de carga (**321carga.txt.ok**)



- Un fichero log del *DataLoader* (**dataloader.log**) que incluye los datos principales de la carga: fecha y hora de inicio de la carga, resumen de resultados, *Portfolio ID* de los registros activados, la fecha y hora de fin de la carga



6 Conclusión

El *DataLoader* facilita la carga y activación masiva de *portfolios* de un mismo *Target Service* mediante la preparación de los datos en ficheros de texto en *formato TSV*. Estos ficheros deben cumplir las siguientes condiciones:

- El nombre del fichero de carga no puede contener espacios
- Tiene que estar codificado en **UTF-8**
- La primera columna debe contener una **clave primaria** (ISBN e ISSN normalmente)
- Las columnas no definidas en la configuración de la carga son ignoradas por el *DataLoader*