

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 3.2.3 Carga de thresholds locales

José Luis Galán Cabilla

Unidad de Biblioteca

ÍNDICE

| | | |
|----------|---|----------|
| 1 | Objetivo | 1 |
| 2 | Dificultad | 1 |
| 3 | Descripción | 1 |
| 4 | Solución | 1 |
| 4.1 | Conversión de fondos al formato de thresholds | 1 |
| 4.1.1 | Paso 1- Escribir inicio de los thresholds | 1 |
| 4.1.2 | Paso 2 - Crear threshold de fecha desde | 2 |
| 4.1.3 | Paso 3 - Crear thresholds de fecha hasta | 3 |
| 4.1.4 | Paso 4 - Crear thresholds compuestos | 4 |
| 4.1.5 | Paso 5 - Rellenar celdas vacías | 4 |
| 4.1.6 | Paso 6 - Convertir thresholds a valores | 5 |
| 4.1.7 | Paso 7 - Borrar filas y columnas sobrantes | 5 |
| 4.1.8 | Paso 8 - Reordenar columnas | 5 |
| 4.2 | Guardar el fichero en formato TSV | 6 |
| 4.2.1 | Guardar en formato TSV en Calc | 6 |
| 4.2.2 | Guardar en formato TSV en Excel | 6 |
| 4.3 | Carga del fichero en el DataLoader | 7 |
| 4.4 | Comprobación de resultados | 7 |
| 5 | Informes | 8 |
| 6 | Conclusión | 9 |

1 Objetivo

Utilización de **DataLoader** para cargar los **thresholds locales** de un **portfolio**.

2 Dificultad

Media.

3 Descripción

Se ha adquirido una selección de revistas de la colección *Backfiles Economics* de *Elsevier* y se dispone de una hoja de cálculo con información de la cobertura temporal suscrita. Se desea cargar y activar estas revistas en SFX con su información de cobertura (*thresholds* locales).

Fichero origen:

- *Calc*: 323origen.ods
- *Excel*: 323origen.xlsx

4 Solución

4.1 Conversión de fondos al formato de thresholds¹

Previamente a poder cargar y activar las revistas hay que convertir su cobertura temporal al formato especial de objetos de *Perl* utilizado por SFX para definir los *thresholds de fecha*. Se dispone de las fechas de cobertura desde y hasta de cada revista por lo que se debe construir un *threshold compuesto* de un *threshold de fecha desde* y un *threshold de fecha hasta* unidos por el operador **&&** (AND). Para ello vamos a utilizar la función **CONCATENAR** celdas que ofrecen *Calc* y *Excel*.²

Para conservar el fichero original sin cambios, guardar una copia de trabajo con el nombre *323carga* (*323carga.ods*, *323carga.xlsx*) sobre la que realizar los cambios necesarios.

4.1.1 Paso 1³- Escribir inicio de los thresholds

En la celda vacía *B29* escribir el principio de la expresión de los *thresholds de fecha desde*:⁴

`$obj->parsedDate('>='`

En la celda vacía *B30* escribir el principio de la expresión de los *thresholds de fecha hasta*:

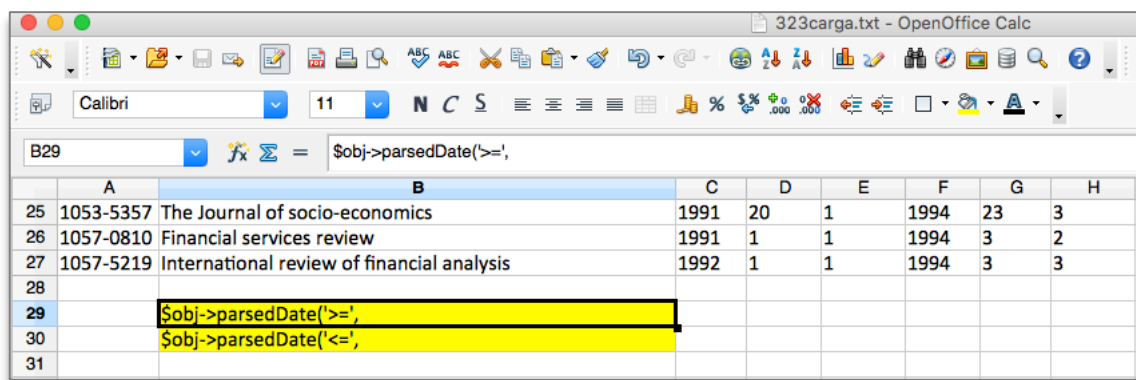
`$obj->parsedDate('<='`

¹ Una explicación detallada en *Using Excel to Create Local Thresholds* (Guía, p. 175-181).

² El procedimiento y la sintaxis de la función **CONCATENAR** es la misma en ambas aplicaciones.

³ Cada paso se ha documentado en una hoja distinta de los ficheros *323carga.ods* y *323carga.xlsx*.

⁴ Ver el apartado 2.1.5 *Thresholds* del documento del Taller.



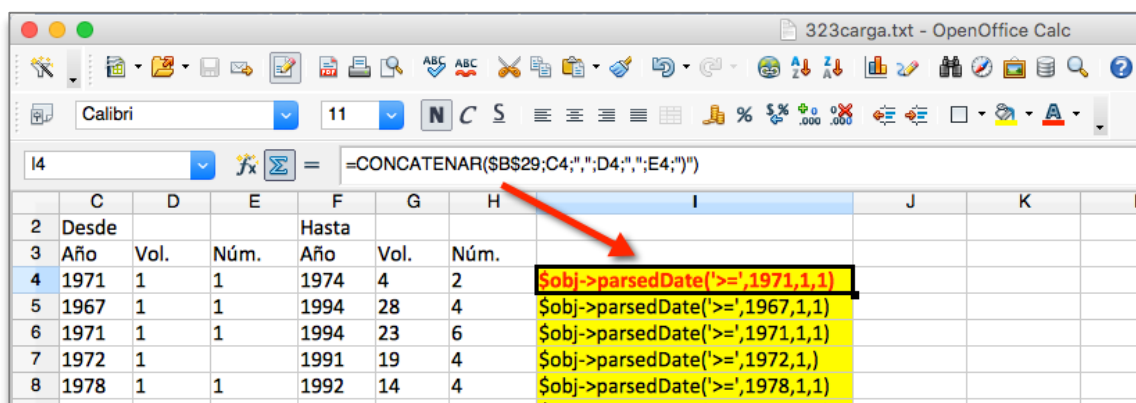
4.1.2 Paso 2 - Crear threshold de fecha desde

En la celda vacía I4 escribir la siguiente función:

=CONCATENAR(\$B\$29;C4;" ";D4;" ";E4;"")

Tras escribir la función y pulsar retorno aparece lo siguiente en la celda I4⁵:

\$obj->parsedDate('>=',1971,1,1)



Las funciones en *Calc* y en *Excel* comienzan por el símbolo igual (=) seguido del nombre de la función y sus argumentos entre paréntesis. Dentro de los paréntesis los argumentos se separan por punto y coma (;). En el caso de la función *CONCATENAR* los argumentos son los textos que se desean unir. A su vez estos textos pueden ser referencias a celdas (absolutas o relativas) o cadenas de caracteres encerradas entre comillas dobles ("").

El significado de cada componente de la función de la celda I4 es el siguiente:

=CONCATENAR

- (Inicio de los argumentos de la función
- \$B\$29** Referencia absoluta a la celda con inicio de *threshold* desde
- ; Separador de argumentos de la función
- C4** Referencia relativa a la celda con el año de inicio
- ; Separador de argumentos de la función
- "," Coma que separa los argumentos del *threshold*
- ; Separador de argumentos de la función
- D4** Referencia relativa a la celda con el vol. de inicio
- ; Separador de argumentos de la función

⁵ Si al pulsar retorno sigue apareciendo la fórmula en la celda se debe al formato de celda. Cambiar el formato de la celda a *General* y editar la fórmula quitando y volviendo a poner el igual del principio de la fórmula.

“ ” Coma que separa los argumentos del *threshold*
 ; Separador de argumentos de la función
 E4 Referencia relativa a la celda con el núm. de inicio
 ; Separador de argumentos de la función
 “ ” Paréntesis de cierre de los argumentos del *threshold*
) Fin de los argumentos de la función

Ahora se traslada la fórmula al resto de celdas de la *columna I* mediante copiar o arrastrar. Al copiar se puede observar como el inicio de la expresión de los *thresholds* no cambia ya que hemos utilizado una **referencia de celda absoluta**. Por el contrario, tal como se desea, la información de los años, volúmenes y números cambian en cada fila ya que se utilizan **referencias de celda relativas**.

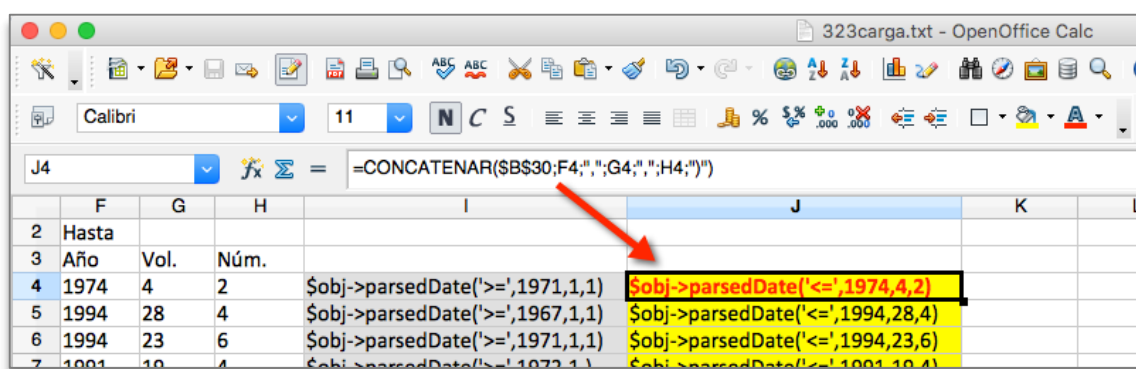
4.1.3 Paso 3 - Crear *thresholds* de fecha hasta

En la celda vacía *J4* escribir la siguiente función:

=CONCATENAR(\$B\$30;F4;" ";G4;" ";H4;"")

Tras escribir la función y pulsar retorno aparece lo siguiente en la celda *J4*:

\$obj->parsedDate('<='1974,4,2)



Como en el paso anterior, trasladar la función al resto de celdas de la *columna J* mediante copiar o arrastrar. Al copiar se puede observar como el inicio de la expresión de los *thresholds* no cambia ya que se ha utilizado una **referencia de celda absoluta**. Por el contrario, tal como se desea, la información de los años, volúmenes y números cambian en cada fila ya que se utilizan **referencias de celda relativas**.

4.1.4 Paso 4 - Crear thresholds compuestos

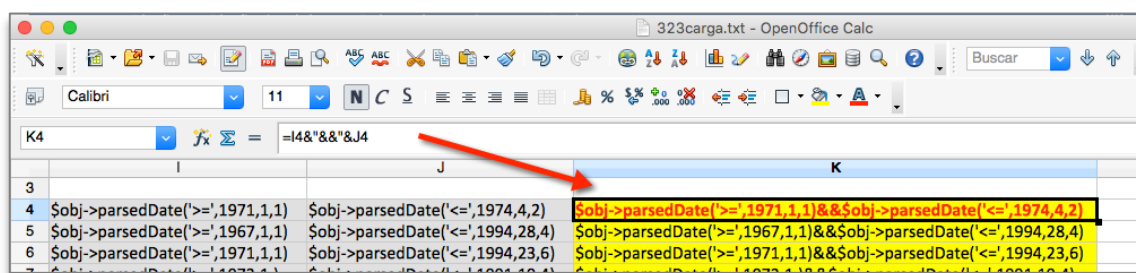
Ahora que se tienen los *thresholds* de fecha desde y hasta en las columnas I y J, se puede construir el *threshold* compuesto de fecha. Para ello se puede utilizar también la función *CONCATENAR* pero se va a utilizar el **operador &** (AND) que funciona de forma muy parecida.

En la celda vacía K4 escribir la siguiente expresión:

=I4&"&&"&J4

Tras escribir la expresión y pulsar retorno aparece lo siguiente en la celda J4:

\$obj->parsedDate('>='1971,1,1)&&\$obj->parsedDate('<='1974,4,2)



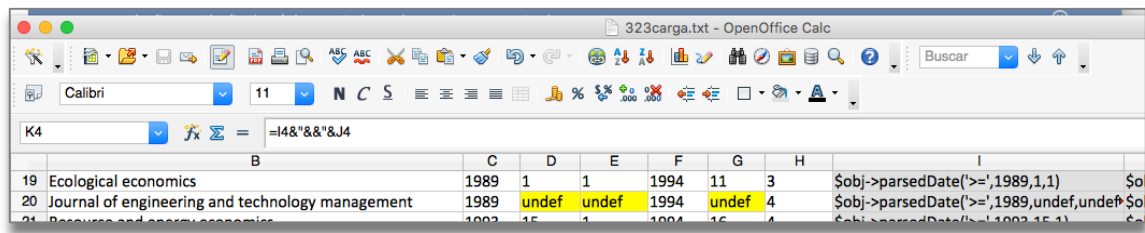
El significado de la expresión es el siguiente

| | |
|------|---|
| = | Inicio de la expresión |
| I4 | Referencia relativa a la celda I4 con los <i>thresholds</i> desde |
| & | Operador de unión |
| "&&" | Operador && para unir <i>thresholds</i> |
| & | Operador de unión |
| J4 | Referencia relativa a la celda J4 con los <i>thresholds</i> hasta |

Como en los dos pasos anteriores, trasladar la expresión al resto de celdas de la *columna K* mediante copiar o arrastrar. Al ser todas las **referencias a celdas relativas**, la información de los años, volúmenes y números cambia con cada fila.

4.1.5 Paso 5 - Rellenar celdas vacías

Algunas celdas de los volúmenes y años están vacías como, por ejemplo la celda E7. Cuando se desconoce algún argumento de los *threshold* se debe poner la expresión **undef** (valor indefinido).



Al escribir en cada celda vacía la expresión *undef* las celdas correspondientes de las columnas *I*, *J* y *K* varían reflejando el cambio.

4.1.6 Paso 6 - Convertir thresholds a valores

Aunque en cada celda se vea un valor, lo que hay en todas las celdas de las *columnas I, J y K* son funciones y expresiones. La **barra de fórmulas** de estas celdas muestra la fórmula y no el resultado. Si se borra alguna de las celdas o columnas que intervienen en el cálculo de estas fórmulas, se obtiene un error o un cambio no deseado en los resultados.

Para convertir las expresiones de la columna K a valores se utiliza la opción **pegado especial**. Seleccionar toda la columna K, situarse en la celda L4 y con el botón derecho del ratón seleccionar pegado especial y elegir *Texto sin formato* o *Valores* en la ventana que se abre. Ahora tanto en las celdas como en la barra de fórmulas se muestran los valores.

4.1.7 Paso 7 - Borrar filas y columnas sobrantes

Borrar las filas y columnas que no interesan:

- Filas 28 a 30 y 1 a 3
- Columnas C a K

4.1.8 Paso 8 - Reordenar columnas

Finalmente realizar los siguientes cambios:

- Insertar la *columna B* y rellenarla con *ACTIVE*.
- Mover la *columna C* con los títulos de las revistas a la *columna D*.

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

- Columna A *Primary key* (ISSNs en este caso)
- Columna B *Activation Status*
- Columna C *Threshold locales*

Además, se ha conservado la columna D con los títulos de las revistas. Esta columna es ignorada por el *DataLoader* y puede facilitar las labores de corrección en caso necesario⁶.

⁶ El *DataLoader* ignora todas las columnas que siguen a las que se especifican al definir el contenido del fichero de carga, por lo que se pueden conservar todas las que se consideren necesarias.

| | A | B | C | D |
|---|-----------|--------|---|----------------------------------|
| 1 | 0034-3331 | ACTIVE | \$obj->parsedDate('>=',1971,1,1)&&\$obj->parsedDate('<=',1974,4,2) | Regional and urban economics |
| 2 | 0038-0121 | ACTIVE | \$obj->parsedDate('>=',1967,1,1) && \$obj->parsedDate('<=',1994,28,4) | Socio-economic planning sciences |
| 3 | 0048-7333 | ACTIVE | \$obj->parsedDate('>=',1971,1,1)&&\$obj->parsedDate('<=',1994,23,6) | Research policy |
| 4 | 0090-5720 | ACTIVE | \$obj->parsedDate('>=',1972,1,1)&&\$obj->parsedDate('<=',1991,19,4) | Journal of behavioral economics |

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 codificación⁷ **abrir el fichero en el editor de texto y guardarlo como UTF-8**.

Fichero de carga en formato TSV: [323carga.txt](#)

⁷ 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**, en este caso *ELSEVIER_SD_BACKFILE-getFullTxt*
- Seleccionar el **subtarget + Service**, en este caso *ELSEVIER_SD_BACKFILE_ECONOMICS-getFullTxt*
- Seleccionar el **fichero de carga**
- Seleccionar *ISBN* para la **columna 1 (primary key)**
- Seleccionar *ACTIVATION STATUS* para la **columna2**
- Seleccionar *THRESHOLD* para la **columna 3**
- Marcar en **tipo de carga** *Update Portfolios* y *Activate targets(s) / target_service(s)*
- Pulsar el botón **Submit**

ExLibris SFX
Home | KBManager | KBTools | Statistics | Troubleshooting
DataLoader | Export Tool | Collection Tool

Server Time: Thursday 21 April 2016 - 00:02:45 | [OpenURL Generator](#) | [Log Off](#)

Welcome to DataLoader

Portfolios | Update Objects | Add/Remove Objects

Specify input file/target service

Select Target + Service: ELSEVIER_SD_BACKFILE-getFullTxt

Select subtarget + Service: ELSEVIER_SD_BACKFILE_ECONOMICS-getFullTxt

Specify input file name: C:\fakepath\323carga.txt [Browse...](#)

Specify input file format: ☒ Dataloader TXT format ☐ KBART format

Specify content of the input file

Select the type of each column in your input file:

Column 1 (Primary Key): ISSN

Column 2: ACTIVATION_STATUS

Column 3: THRESHOLD

Column 4:

Column 5:

Column 6:

Column 7:

Column 8:

Column 9:

Specify load type

Load Type

☐ Add Portfolios

☒ Update Portfolios

☐ Remove Portfolios

☒ Activate targets(s) / target_service(s)

☐ Report Mode - Do Not Update the Database

[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 24 *portfolios* y esto es lo que dice el resumen de la carga:

Summary: Added 0 portfolios, updated 24 portfolios for: ELSEVIER_SD_BACKFILE_ECONOMICS-getFullTxt

ExLibris SFX Server Time: Wednesday

Home | KBManager | **KBTools** | Statistics | Troubleshooting

DataLoader | Export Tool | Collection Tool

DataLoader Successful

Mode: Portfolios
Input file: 323carga.txt
Summary: Added 0 portfolios, updated 24 portfolios for: ELSEVIER_SD_BACKFILE_ECONOMICS-getFullTxt

An analysis of the data file can be found in: 323carga.txt.ok.err [Download](#)
 A 'cleaned' copy of the data file can be found in: 323carga.txt.ok [Download](#)
 A log of the dataloader run can be found in: dataloader.log [Download](#)
 If you want to activate added/updated target(s) target_service(s) [Activation](#)

Además se puede comprobar que efectivamente se han cargado y activado las revistas consultando el *Target* y sus *portfolios*. Como se puede ver, se han activado los 24 portfolios y todos tienen una marca en el la columna *Loc.* ya que tienen definidos *threshold locales*.

ExLibris SFX Server Time: Thursday 21 April 2016 - 01:12:58 | [OpenURL Generator](#) | [Log Off](#)

Home | KBManager | KBTools | Statistics | Troubleshooting

Sources | **Targets** | Objects | Linking Parameters | Institutes

List of Portfolios for ELSEVIER_SD_BACKFILE_ECONOMICS

[Add New Portfolio](#) [Activate All](#) [Deactivate All](#) [Delete](#) [Activate](#) [Deactivate](#) [Show All](#) [Show Active](#) [Show Inactive](#)

Show 20 portfolios per page

Find portfolio by Title Go Showing page 1 of 2 Database: sfxst41 [Back to Target](#) [Back to List of Services](#)

Jump to page Jump

Available: 24 portfolios

| # | Service | Object | Title | Threshold | Mod. | Loc. | Act. | Actions |
|---|---------|------------|--|---|------------|------|------|---------|
| 1 | E V | getFullTxt | 0034-3331 Regional and urban economics | Sobj->parsedDate(">=",1971,1,1)&&Sobj->parsed... | 2016/04/21 | | ✓ | C |
| 2 | E V | getFullTxt | 0038-0121 Socio-economic planning sciences | Sobj->parsedDate(">=",1967,1,1) && Sobj->parse... | 2016/04/21 | | ✓ | C |
| 3 | E V | getFullTxt | 0048-7333 Research policy | Sobj->parsedDate(">=",1971,1,1)&&Sobj->parsed... | 2016/04/21 | | ✓ | C |
| 4 | E V | getFullTxt | 0090-5720 Journal of behavioral economics | Sobj->parsedDate(">=",1972,1,1)&&Sobj->pa... | 2016/04/21 | | ✓ | C |
| 5 | E V | getFullTxt | 0165-0572 Resources and energy | Sobj->parsedDate(">=",1978,1,1)&&Sobj->parsed... | 2016/04/21 | | ✓ | C |

5 Informes

Como resultado de la carga se generan los siguientes informes:

- Un análisis de los resultados de la carga(323carga.txt.ok.err)

File Path: ~/Downloads/323carga.txt.ok.err

```

1 Report Updated 0034-3331 Updated portfolio with id=111060797595002
2 ELSEVIER_SD_BACKFILE_ECONOMICS getFullTxt
3 Report Updated 0038-0121 Updated portfolio with id=111060797596003
4 ELSEVIER_SD_BACKFILE_ECONOMICS getFullTxt
5 Report Updated 0048-7333 Updated portfolio with id=111060797596005
6 ELSEVIER_SD_BACKFILE_ECONOMICS getFullTxt
7 Report Updated 0090-5720 Updated portfolio with id=111060797596006
  
```

- Una copia del fichero de carga ([323carga.txt.ok](#))

```

File Path : ~/Downloads/323carga.txt.ok
323carga.txt.ok
1 | 0034-3331 ACTIVE $obj->parsedDate('>=' ,1971,1,1)&&$obj->parsedDate('<=' ,1974,4,2)
2 | Regional and urban economics
3 | 0038-0121 ACTIVE $obj->parsedDate('>=' ,1967,1,1) && $obj->parsedDate('<=' ,1994,28,4)
4 | Socio-economic planning sciences
5 | 0048-7333 ACTIVE $obj->parsedDate('>=' ,1971,1,1)&&$obj->parsedDate('<=' ,1994,23,6)
6 | Research policy

```

- 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

```

File Path : ~/Downloads/dataloader.log
dataloader.log
1 | (Thu Apr 21 00:59:15 2016): started dataloader
2 | (Thu Apr 21 00:59:15 2016): started translate_eissn_and_issn
3 | (Thu Apr 21 00:59:15 2016): Translated 0 rows from eISSN to ISSN
4 | Added 0 portfolios, updated 24 portfolios for:
5 | ELSEVIER_SD_BACKFILE_ECONOMICS-getFullTxt,111060797595002,111060797596003,111060797596005,
6 | 111060797596006,111060797596011,111060797597017,111060797597019,111060797597020,
7 | 111060797597021,111060797597022,111060797597023,111060797597025,111060797597029,
8 | 111060797597030,111060797598037,111060797598039,111060797598041,111060797598045,
9 | 111060797599047,111060797599048,111060797599050,111060797599052,111060797599053,
10 | 111060797599054
11 | (Thu Apr 21 00:59:15 2016): finished dataloader

```

6 Conclusión

Aunque requiere algo de trabajo y práctica, utilizando las funciones de concatenación y unión de las hojas de cálculo se pueden componer los *threshold locales* de los *portfolios* de forma masiva. Estas mismas funciones las podemos utilizar para crear, por ejemplo, enlaces o tablas en HTML (casos prácticos [3.2.4](#), [3.2.5](#) y [5.2.2](#)).