

Mapa de los archivos de la Comunidad de Madrid Map of the archives of the Community of Madrid.

Francisco Javier Saiz Rodrigo

Tecnologías documentales

fjsrodrigo@gmail.com

Antonio Godino Muñoz

Tecnologías documentales

antonio.godino@gmail.com

María-Jesús Colmenero-Ruiz

Universidad Carlos III de Madrid

mcolmene@bib.uc3m.es

Resumen

El 9 de julio de 2015 se aprobó en España la Ley 18/2015 por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público.

Su objeto es la incorporación al ordenamiento jurídico español de los cambios introducidos por la Directiva 2013/37/UE del Parlamento Europeo y del Consejo, de 26 de junio de 2013, en el régimen de reutilización de documentos del sector público.

Dicha ley ha ampliado su ámbito de aplicación a los archivos, bibliotecas y museos con el fin de favorecer el uso de sus recursos en el interés de los ciudadanos.

Con este fin, y con la intención de difundir y poner en valor los fondos archivísticos se presenta la primera fase, incluyendo únicamente los archivos de la Comunidad de Madrid, de una aplicación con información geolocalizada de archivos españoles. La necesidad de presentar una gran cantidad de información de una forma esquemática, clara y eminentemente visual en la web ha hecho de la geolocalización una herramienta de uso creciente para este fin. Esta aplicación presenta una ficha desplegable con los

datos del archivo y las características de los fondos y pretende servir como base para generar rutas turísticas que representen valor añadido a otras instituciones y, finalmente, al usuario.

Palabras clave: archivos, datos abiertos, reutilización de la información pública, Comunidad de Madrid, geolocalización, datos culturales

Abstract:

On July 9, 2015 was approved in Spain the Law 18/2015 amending Law 37/2007 of 16 November, on reuse of public sector information. Its object is the incorporation into Spanish law of the changes introduced by Directive 2013/37 / EU of the European Parliament and of the Council of 26 June 2013 in this area. This law has expanded its scope to archives, libraries and museums in order to promote the use of its resources in the interest of citizens.

To this end, and with the intention of spreading and to value the archival funds, an application with geographic information Spanish archives is presented, including only files of the Community of Madrid. The need to present a lot of information in a schematic, clear and highly visual way in the web has made the geolocation tool increasingly used for this purpose. This application presents a scrollable tab with the file data and characteristics of the archives. It's intended to serve as a basis for generating tourist routes that represent value to other institutions and finally to the user.

Keywords: archives, public information reuse, Community of Madrid, geolocalization, cultural data.

Introducción

La Unión Europea (UE) adoptó, a finales del año 2003, la Directiva 2003/98/CE relativa a la reutilización de la información del sector público¹. Esta iniciativa de la UE, conocida también como *PSI Directive* en inglés, pretendía proporcionar un marco legislativo común para todos los países que la componen en relación con la reutilización de los documentos gestionados por los organismos del sector público, dado que en aquél momento se daban “*considerables diferencias de un Estado miembro a otro en relación con las normas y prácticas de explotación de los recursos de información del sector público*”. La reutilización de la información supone permitir el acceso a terceros a la información que los organismos del sector público utilizan para realizar su misión, en formatos abiertos que faciliten su uso con fines distintos de aquellos para los cuales fueron inicialmente concebidos, solos o en combinación con otros, para generar nuevos servicios de valor añadido incluso con fines comerciales. El objetivo perseguido era eliminar las barreras existentes, evitando que se dieran condiciones o prácticas discriminatorias, monopolios o falta de transparencia.

Por aquellas fechas la UE estaba promoviendo con sus políticas la evolución del modelo económico existente hacia otro basado en una Sociedad de la Información, estructura económico-social que se preveía impulsora de la economía y el empleo basada en la explotación y gestión de la información. Precisamente los contenidos digitales estaban teniendo un papel generador de ese ansiado empleo y la UE veía en

¹ Directiva 2003/98/CE del Parlamento Europeo y del Consejo, de 17 de noviembre de 2003, relativa a la reutilización de la información del sector público. *Diario Oficial de la Unión Europea*, núm. 345, de 31 de diciembre de 2003, pp. 90-96. (DOUE-L-2003-82244). [Consulta: 16-08-2016]. Disponible en http://eur-lex.europa.eu/legal-content/ES/TXT/?uri=uriserv:OJ.L_.2003.345.01.0090.01.SPA y en <https://www.boe.es/buscar/doc.php?id=DOUE-L-2003-82244>.

la información del sector público una materia prima importante para la generación de diversos productos y servicios de contenidos digitales dado que “*el sector público recoge, produce, reproduce y difunde una amplia gama de información relativa a numerosos ámbitos, por ejemplo información social, económica, geográfica, meteorológica o turística y sobre empresas, patentes y educación*”. Para poder generar esos nuevos productos y servicios para todos los ciudadanos de la comunidad, es decir, con una vocación transfronteriza de mercado único, debían darse unas condiciones mínimas legislativas que esta directiva pretendió establecer. España no incorporó la directiva en su legislación hasta el año 2007² a pesar de que la fecha tope impuesta para hacerlo era junio de 2005, lo que le supuso amonestaciones y multas de la EU por su incumplimiento, al igual que a otros países pues pocos cumplieron con el plazo dado.

Lo cierto es que la directiva no consiguió del todo sus objetivos porque, entre otras cosas, no comportaba obligación alguna de autorizar la reutilización de los documentos, dejándolo a criterio de cada Estado miembro. Sin embargo, el argumento que la inspiró, la consideración de que los contenidos digitales iban a tener una gran importancia en el futuro mercado digital, fue haciéndose una realidad a la par que se hacían evidentes las dificultades que la UE tenía para competir en este campo³: sólo se había logrado una armonización de mínimos y las dificultades transfronterizas limitaban el desarrollo de servicios transnacionales. En 2009 la

² España. *Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público*. BOE nº 276, 17/11/2007, nº 19814, pp. 47160 – 47165. [Consulta: 16-08-2016]. Disponible en <https://www.boe.es/diario_boe/txt.php?id=BOE-A-2007-19814>

³ La OCDE ya lo recoge en su informe del año 2008, *OECD Recommendation of the Council for enhanced access and more effective use of public sector information* (Recomendación Ministerial de la OCDE sobre mejora del acceso a la información del sector público y utilización más eficaz de la misma). [C(2008)36]. [Consulta: 17-08-2016]. Disponible en <<http://www.oecd.org/internet/ieconomy/40826024.pdf>>, documento citado en la Agenda Digital Europea de 2010.

Comisión Europea hizo una revisión de la Directiva en la que evaluó su nivel de aplicación e impacto, además de las dificultades a superar⁴. Entre las medidas que sugería adoptar era una nueva revisión de la misma “*a más tardar en 2012, cuando se disponga de más datos sobre las repercusiones, los efectos y la aplicación de la misma, []. Se planteará entonces la conveniencia de introducir modificaciones legislativas si se sigue obstaculizando el desarrollo del potencial de reutilización*”.

Por estos motivos se incluyó en la Agenda Digital para Europa⁵ la reutilización de la información del sector público dentro de una de las áreas clave de actuación (Acción clave 1, apartado “Un mercado único digital dinámico”, campo “Apertura del acceso a los contenidos”), proponiendo una reforma de la directiva original para el año 2012, en particular en lo relativo a su ámbito de aplicación y los principios de tarificación del acceso y el uso.

La revisión de la directiva se publicó finalmente un año después, en 2013⁶. La nota de prensa de la Comisión Europea al dar la noticia de su adopción⁷ señalaba que la misma daba lugar a la constitución de “*un derecho genuino a la reutilización de la información pública; todos los datos públicos no cubiertos por una de las excepciones*

⁴ Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones - Reutilización de la información del sector público: revisión de la Directiva 2003/98/CE – [SEC(2009) 597]. COM/2009/0212 final. [Consulta: 17-08-2016]. Disponible en <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52009DC0212>.

⁵ Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Una Agenda Digital para Europa. COM(2010)245 final. [Consulta: 17-08-2016]. Disponible en <http://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX:52010DC0245>

⁶ Directiva 2013/37/UE del Parlamento Europeo y del Consejo, de 26 de junio de 2013, por la que se modifica la Directiva 2003/98/CE relativa a la reutilización de la información del sector público. [Consulta: 17-08-2016]. Disponible en <<http://eur-lex.europa.eu/eli/dir/2013/37/oj>>.

⁷ Comisión Europea (2013), *Commission welcomes Parliament adoption of new EU Open Data rules*. Memo 13-555. [Consulta: 17-08-2016]. Disponible en <http://europa.eu/rapid/press-release_MEMO-13-555_en.htm>.

serán reutilizables”⁸ destacando además que “*todas las instituciones de la UE se han comprometido a liberar todo el potencial de la mina de oro de los datos abiertos*”⁹.

Precisamente el movimiento de los datos abiertos había ido adquiriendo empuje creciente, especialmente desde 2010, presionando para que los datos públicos se publicaran en formatos abiertos y reutilizables de forma gratuita o a coste marginal. La UE ya se había posicionado a favor de esta opción¹⁰, establecido esta modalidad para la publicación de sus datos¹¹ y creado un portal de datos abiertos.¹²

La Directiva revisada incorpora una novedad de especial interés para nuestro campo profesional pues amplía su ámbito de aplicación a las instituciones culturales como las bibliotecas, incluidas las universitarias, los museos y los archivos, con el fin de favorecer el uso de sus recursos en el interés de los ciudadanos. con el fin de favorecer el uso de sus recursos en el interés de los ciudadanos. No obstante, por el momento, se les aplican algunas excepciones y las reglas son menos estrictas para este tipo de datos, en la línea de la directiva original de 2003.

El 9 de julio de 2015 se aprobó en España la Ley 18/2015 por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público. Su objeto era la incorporación al ordenamiento jurídico español de los cambios introducidos por la última revisión de la Directiva. Dicha ley ha ampliado su ámbito de aplicación a los archivos, bibliotecas y museos con el fin de favorecer el uso

⁸ “*a genuine right to re-use public information; all public data not covered by one of the exceptions will become re-usable*, en el original.

⁹ “*all EU institutions are committed to unlock the full potential of the open data goldmine*”, en el original.

¹⁰ Comisión Europea (2011). Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social europeo y al Comité de las Regiones. “Datos abiertos. Un motor para la innovación, el crecimiento y la gobernanza transparente”. COM/2011/0882 final. [Consulta: 18-08-2016]. Disponible en <<http://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52011DC0882>>.

¹¹ Comisión Europea (2011). Decisión de la Comisión de 12 de diciembre de 2011 relativa a la reutilización de los documentos de la Comisión (2011/833/UE). [Consulta: 18-08-2016]. Disponible en <<http://eur-lex.europa.eu/eli/dec/2011/833/oj>>.

¹² <http://data.europa.eu/euodp/en/data>

de sus recursos en el interés de los ciudadanos. En esta ocasión España no se demoró del todo en incorporarla a su legislación, dado que el deadline era a finales de año.

Basándonos en esta apertura de los datos de los archivos impulsada por la legislación hemos intentado hacer una primera aproximación de reutilización de datos de archivos para la generación de otros servicios de información de valor añadido con ellos.

Descripción de la propuesta de reutilización de datos de archivos

El análisis masivo de datos se está convirtiendo en una técnica habitual de análisis a partir de la disponibilidad de grandes volúmenes de datos digitales. Del mismo modo que Internet es un factor de desintermediación, el análisis de información en grandes volúmenes, de diversas fuentes, a gran velocidad y con una flexibilidad sin precedentes es sin duda un factor clave en cualquier acción de futuro: análisis que tardaban antes varios días en realizarse ahora es un factor determinante para la creación de modelos predictivos y para la toma de decisiones.

Permite encontrar relaciones entre datos que, a primera vista, parecen no tener nada en común y puede beneficiar a las empresas, las organizaciones y el público en general, porque les permite detectar patrones, tendencias y correlaciones para poder tomar decisiones informadas. Es una herramienta muy útil para detectar correlaciones, para realizar diagnósticos y pronósticos fiables en numerosas áreas de gestión, y que es imposible detectar al analizar conjuntos de datos más pequeños.

Y finalmente también es posible dar una respuesta visual en la representación de los datos. Gracias a esta forma de interpretar visualmente la información se ha podido predecir terremotos, conocer las tasas de paro, reducir índices de criminalidad o hacer seguimiento de epidemias, por nombrar algunos ejemplos.

Este tipo de análisis sobre grandes conjuntos de datos es lo que se conoce conocido como “big data”.

Figura 1A Big data. Open Government. Technologies. Management Information. Analysis Research. Science. Web Semantic



En nuestro caso, nos propusimos utilizar datos publicados sobre archivos españoles, combinándolos con otros de forma que permitieran hacerlos más visibles a los ciudadanos. Para hacer una primera aproximación se ha tomado una pequeña muestra para trasladar nuestro ejemplo, restringiéndonos a los datos de los archivos públicos de la Comunidad de Madrid, combinándolos con sus coordenadas espaciales para su geolocalización en un mapa, creando una ruta que sirviera de base para propuestas turísticas, culturales o simplemente informativa. En nuestro caso es posible concentrar

ubicar y configurar una ruta con los más de 10000 archivos censados en el Ministerio de Cultura¹³. [Consulta: 18-08-2016].

El resultado permitirá desarrollar aplicaciones de interés general. La base del desarrollo está en la capacidad de conectar diferentes variables y actores que intervienen en el producto final. Por un lado, la posibilidad que tienen los archivos de dar visibilidad a sus fondos, de hacerse en ver mediante aplicaciones web que están dando la oportunidad de ubicarlos físicamente y de conocer sus características a un golpe de clic. De esta forma será posible acercarlos al ciudadano, quien podrá, por ejemplo, confeccionarse sus rutas por los diferentes tipos de archivos que existen. Se trata de una gestión inteligente de la información digital, que puede ayudar a tomar decisiones que redunden en la mejora de las condiciones de vida de la gente e incrementen la rentabilidad de los modelos de negocio.

La tecnología asociada a la reutilización de la información nos permite convertir los datos semiestructurados de la web de la administración en un activo y nos abre nuevas vías de explotación de la información.

Los datos publicados en abierto de la ubicación del archivo en sí tienen ya un valor informativo pero su difusión a través de la búsqueda en una base de datos limita su posible impacto. Si los representamos sobre un mapa, un plano, adquieren otro significado. La geolocalización de objetos es una de las herramientas más comunes y útiles en la web, especialmente porque proporciona mucha información en poco

¹³ La geolocalización de todos los archivos españoles mostrando su situación geográfica y datos básicos se abordó como una propuesta práctica en las clases de uno de los autores de este artículo usando básicamente el mismo procedimiento. Ese mapa general es posible consultarlo en el blog que se creó a tal fin en <https://archivosespanoles.wordpress.com/ubicacion/>. La descripción de esta actividad está publicada en Colmenero- Ruiz, María-Jesús. "Visualizing archives: Spanish archives map". *Netcom* [Online], 29-3/4, 2015, [Consulta: 19-08-2016]. Disponible en <<http://netcom.revues.org/2143>>.

espacio, visual y sobre un tipo de representación gráfica, el mapa, que resulta muy familiar y, por tanto, más atractiva. Si sobre ese plano trazamos además rutas de interés a las que se les aporta otra información de utilidad, el resultado es aún mejor. En esencia, el trabajo a realizar para conseguirlo supone estructurar esos datos, integrarles coordenadas, encontrar las relaciones entre las distintas fuentes que conforman el conjunto de la información y, finalmente, añadir información adjunta como características, fotos, zonas de interés de la localidad del archivo y cualquier dato que aporte valor a la ruta. Es un trabajo de procesamiento de datos, modelado, análisis y representación de la información.

Como veremos en el siguiente apartado, el procedimiento utilizado para conseguirlo ha supuesto recolectar de forma asistida la información de los archivos en la web para más tarde obtener sus coordenadas geográficas de Google a partir de su dirección postal para, por último, utilizando una herramienta GIS como CartoDB, obtener vistas filtradas de los archivos o incluso rutas recomendadas.

Descripción del procedimiento realizado

Figure 2A (<http://censoarchivos.mcu.es/CensoGuia/portada.htm>)



Se describe la metodología empleada en el flujo de trabajo seguido desde la búsqueda de la información, hasta su posterior recuperación, descripción, clasificación y difusión.

En primer lugar, nos situamos en la web del ministerio donde están publicado el censo de los archivos y en concreto en el formulario de búsqueda del directorio.

Figure 2B (<http://censoarchivos.mcu.es/CensoGuia/directorioarchivos.htm>)

Búsqueda sencilla en el Directorio de Archivos

CRITERIO DE BÚSQUEDA

Búsqueda general:

Nombre del archivo:

Categoría: Sin Categoría ▼

SubCategoría: Sin Subcategoría ▼

UBICACIÓN GEOGRÁFICA DEL ARCHIVO

Área Geográfica: Todas las áreas ▼

País: Todos los países ▼

CC.AA./1ª División: Todas las comunidades ▼

Provincia/2ª División: Todas las provincias ▼

Municipios: Todos los municipios ▼

Núcleo de población menor:

Limpiar Buscar

La idea es poder utilizar todo tipo de aplicaciones de software libre para realizar este proceso, por lo que para la realización de la primera parte de la práctica se ha montado un servidor Linux (Ubuntu) y se ha instalado la librería BeautifulSoup en el entorno de ejecución Python¹⁴. El proceso que vamos a lanzar para recuperar los

¹⁴ <https://www.python.org/>

datos sería los mismos que te daría los resultados de cualquier tipo de búsqueda filtrada por el formulario del buscador del Ministerio de manera manual.

Figure 2C. (Listado de archivos correspondientes a la búsqueda filtrada por la comunidad de Madrid. 878 resultados)

Listado de Archivos

Ordenador Por : [Fecha Modificación](#) ▾ [Nombre Archivo](#)

[Atrás](#)


Results 1 - 10 from 878


[Prim./Prev.] [1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [7](#), [8](#), [9](#), [10](#) [Sig./Ult.]

ARCHIVO HISTÓRICO NACIONAL

Serrano, 115 28006 MADRID

Tel: 91.768.85.00/ Fax: 91.563.11.99


 <http://www.mecd.gob.es/cultura-mecd/areas-cultura/archiv...>


 ahn@mecd.es


ARCHIVO CENTRAL DE LA AGENCIA DE VIVIENDA SOCIAL DE LA COMUNIDAD DE MADRID

Basílica, 23 28020 MADRID

Tel: 91 580 91 00/ Fax: 91 580 96 36

 http://intranet.madrid.org/cs/Satellite?c=CM_InfPractica...

 http://www.madrid.org/cs/Satellite?c=CM_Agrupador_FP&cid...


 archivo.agenciavivienda@madrid.org

ARCHIVO CENTRAL DEL MINISTERIO DE AGRICULTURA Y PESCA, ALIMENTACIÓN Y MEDIO AMBIENTE. ÁREA DE AGRICULTURA Y ALIMENTACIÓN

Paseo Infanta Isabel, 1 28014 MADRID

Tel: 913473709 / Fax: 915975903


 <http://www.mapama.gob.es/>


 archivo@maqrma.es

ARCHIVO GENERAL DEL MINISTERIO DE AGRICULTURA Y PESCA, ALIMENTACIÓN Y MEDIO AMBIENTE. ÁREA DE MEDIO AMBIENTE

Plaza de San Juan de la Cruz, SN 28071 MADRID

Tel: 915976160 915975431

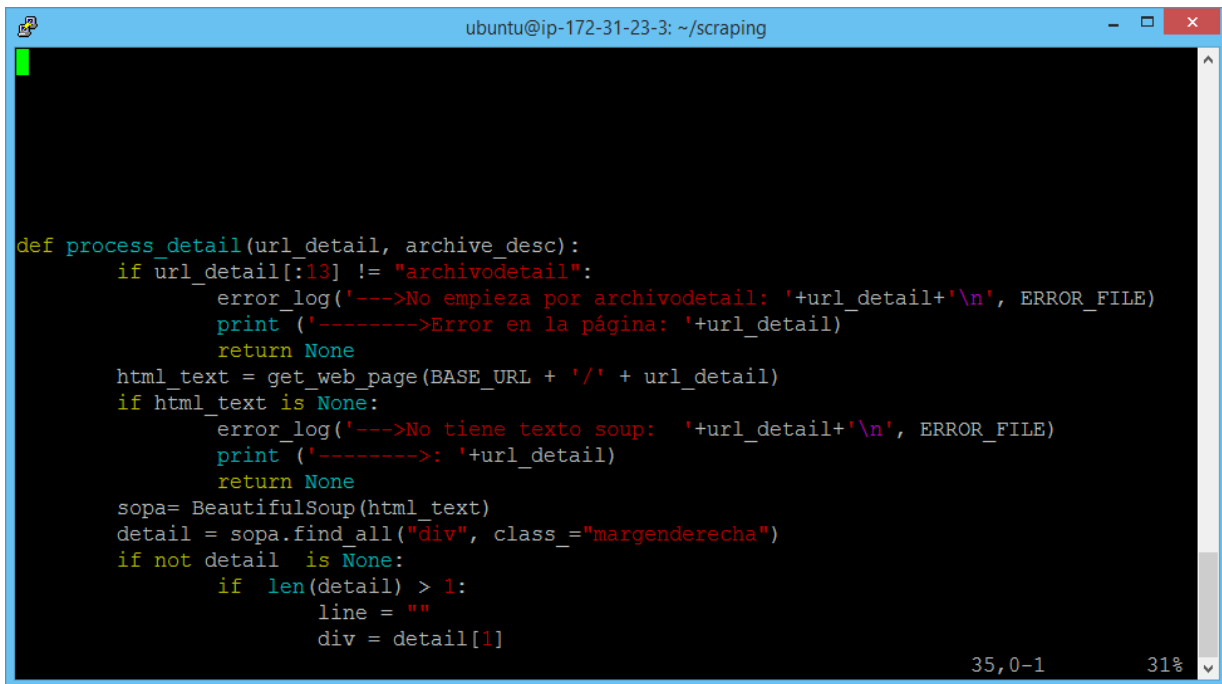
 <http://www.mapama.gob.es/>

 archivo@maqrma.es

En este caso en vez de lanzar 111 búsquedas Archivo por archivo para copiar los datos, lo que hacemos es preparar un script en Python, un lenguaje de programación cuya filosofía hace hincapié en una sintaxis que fue diseñada para ser leída con facilidad.

Python es un lenguaje denominado interpretado, lo que significa que no se necesita compilar el código fuente para poder ejecutarlo, lo que ofrece ventajas como la rapidez de desarrollo, la versatilidad y simplicidad. Lo que de manera manual podría llevarnos días, con un simple programa podemos realizar el proceso denominado scrapping a partir de la url del Ministerio de Cultura.¹⁵

Figure 2D

A screenshot of a terminal window titled 'ubuntu@ip-172-31-23-3: ~/scraping'. The terminal displays Python code for a function named 'process_detail'. The code checks if a URL detail starts with 'archivodetail', fetches the page content, and uses BeautifulSoup to find a specific class 'margenderecha'. It then prints the first element of the found list. The terminal output shows the code being executed, with some error messages and a return value of None. The terminal status bar at the bottom right shows '35,0-1' and '31%'.

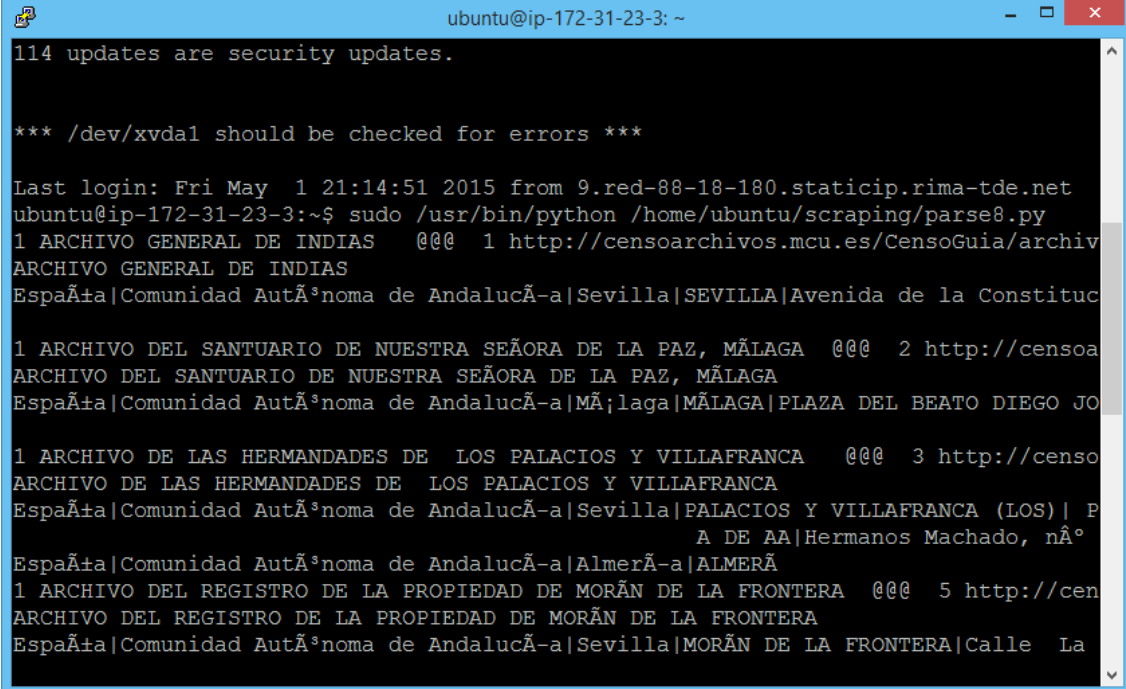
Tiene que quedar claro que no estamos robando datos, lo que estamos haciendo es simplificar un proceso de copia y pega de datos abiertos y públicos. No es una actividad ilegal y el uso es legítimo.

El script se recorre las 878 páginas de resultados de la web analizando la página de detalle de cada archivo. El resultado de la ejecución de este script es el volcado en un fichero de texto (además de un eco en pantalla) de los datos de interés de cada archivo. El fichero obtenido tiene una línea por cada archivo con datos como

¹⁵ <http://censoarchivos.mcu.es/CensoGuia/archivos.htm?generalArchivo=&nombreArchivo=&categoria=1&subcategoria=&areaSelect=1&paisSelect=2&comunidadSelect=1&provinciaSelect=0&municipioSelect=0&poblacionmenor=#>

nombre, dirección, localidad, provincia, url, etc. Además de un volcado en pantalla de cada línea como podemos ver en la imagen inferior.

Figure 2E (Fichero resultante de la extracción de datos de la web del ministerio. El script va volcando los datos en líneas de texto que serán recuperadas en un fichero para su posterior tratamiento)



```
ubuntu@ip-172-31-23-3: ~
114 updates are security updates.

*** /dev/xvda1 should be checked for errors ***

Last login: Fri May 1 21:14:51 2015 from 9.red-88-18-180.staticip.rima-tde.net
ubuntu@ip-172-31-23-3:~$ sudo /usr/bin/python /home/ubuntu/scraping/parse8.py
1 ARCHIVO GENERAL DE INDIAS   @@@ 1 http://censoarchivos.mcu.es/CensoGuia/archiv
ARCHIVO GENERAL DE INDIAS
España|Comunidad Autónoma de Andalucía|Sevilla|SEVILLA|Avenida de la Constituc
1 ARCHIVO DEL SANTUARIO DE NUESTRA SEÑORA DE LA PAZ, MÁLAGA   @@@ 2 http://censoa
ARCHIVO DEL SANTUARIO DE NUESTRA SEÑORA DE LA PAZ, MÁLAGA
España|Comunidad Autónoma de Andalucía|Málaga|MÁLAGA|PLAZA DEL BEATO DIEGO JO
1 ARCHIVO DE LAS HERMANDADES DE LOS PALACIOS Y VILLAFRANCA   @@@ 3 http://censo
ARCHIVO DE LAS HERMANDADES DE LOS PALACIOS Y VILLAFRANCA
España|Comunidad Autónoma de Andalucía|Sevilla|PALACIOS Y VILLAFRANCA (LOS)| P
A DE AA|Hermanos Machado, n.º
España|Comunidad Autónoma de Andalucía|Almería|ALMERÍA
1 ARCHIVO DEL REGISTRO DE LA PROPIEDAD DE MORÁN DE LA FRONTERA @@@ 5 http://cen
ARCHIVO DEL REGISTRO DE LA PROPIEDAD DE MORÁN DE LA FRONTERA
España|Comunidad Autónoma de Andalucía|Sevilla|MORÁN DE LA FRONTERA|Calle La
```

Los datos obtenidos en un fichero de texto son con los que vamos a trabajar en el siguiente proceso de refinado de la Información. Se usarán dos aplicaciones que son el programa Excel de la suite de Microsoft Office y OpenRefine¹⁶ (anteriormente llamada Google Refine). Si bien la tarea de análisis y limpieza de datos podría hacerse únicamente con la herramienta de OpenRefine pasaremos antes por Excel para hacer un primer tratamiento para posteriormente con los datos más estructurados conseguir las coordenadas de cada Archivo con la herramienta OpenRefine diseñada para entender la estructura y calidad de los datos, y permitir corregir determinados tipos de errores comunes en ellos.

¹⁶ Disponible a través de la dirección web <http://openrefine.org/>, donde pueden encontrarse bastante documentación para su uso.

Por tanto, lo que vemos ahora en las siguientes tablas son los datos obtenidos y volcados en una hoja de cálculo de Excel con las filas y columnas correspondientes al nombre del Archivo, el país, la provincia, la dirección física, las coordenadas, la identificación, titularidad y categoría, la subcategoría, la gestión, la historia de la Institución, la descripción del fondo y la URL correspondiente a la dirección web de la organización:

Tabla 1. Hoja de cálculo de Excel con los datos de los Archivos en filas y columnas.

NOMBRE DEL ARCHIVO	PAIS	PROVINCIA	DIRECCIÓN
ARCHIVO HISTÓRICO NACIONAL	España	Madrid	Serrano, 115 28006, Madrid
ARCHIVO GENERAL DE LA ADMINISTRACIÓN (ALCALÁ DE HENARES, MADRID)	España	Madrid	Paseo de Aguadores, 2 28804, Madrid
ARCHIVO HISTÓRICO DE PROTOCOLOS DE MADRID	España	Madrid	C/ Ramírez de Prado, 3 28045, Madrid
ARCHIVO CENTRAL DE LA CONSEJERÍA DE EDUCACIÓN DE LA COMUNIDAD DE MADRID	España	Madrid	Camino de Hormigueras, 152, 28031, Madrid
ARCHIVO DE LA CONFEDERACIÓN HIDROGRÁFICA DEL TAJO (MADRID, ESPAÑA)	España	Madrid	Avenida de Portugal, 81 28071, Madrid
ARCHIVO GENERAL DEL CUARTEL GENERAL DEL EJÉRCITO DEL AIRE (MADRID, ESPAÑA)	España	Madrid	Romero Robledo, 8 28008, Madrid
ARCHIVO CENTRAL DEL CUARTEL GENERAL DE LA ARMADA (MADRID, ESPAÑA)	España	Madrid	Montalbán, 2 28014, Madrid
ARCHIVO GENERAL E HISTÓRICO DE DEFENSA	España	Madrid	Paseo. Moret, 3 28008, Madrid
ARCHIVO DE LA OFICINA ESPAÑOLA DE PATENTES Y MARCAS (MADRID, ESPAÑA)	España	Madrid	Panamá, nº 1 28036, Madrid
ARCHIVOS PERSONALES DEL DEPARTAMENTO DE MANUSCRITOS, INCUNABLES Y RAROS DE LA BIBLIOTECA NACIONAL DE ESPAÑA	España	Madrid	Biblioteca Nacional de España - Paseo de Recoletos, nº 20-22 28071, Madrid
ARCHIVO GENERAL DE PALACIO (MADRID, ESPAÑA)	España	Madrid	Bailén 28071, Madrid
ARCHIVO CENTRAL DE LA SECRETARÍA DE ESTADO DE CULTURA	España	Madrid	Plaza del Rey, 1, 28071, Madrid
SECCIÓN ARCHIVO GUARDIA CIVIL DEL ARCHIVO GENERAL DEL MINISTERIO DEL INTERIOR	España	Madrid	C/ Guzmán el Bueno 110 28003, Madrid
ARCHIVO GENERAL DEL MINISTERIO DE FOMENTO	España	Madrid	Paseo de la Castellana, 67 28071, Madrid
SERVICIO DE REPRODUCCIÓN DE DOCUMENTOS DE ARCHIVOS ESTATALES (ESPAÑA)	España	Madrid	calle SERRANO 115, 28006, Madrid

Cada fila y columna corresponden a los datos de la búsqueda en la base de datos tal y como se haría de manera manual donde filtramos las opciones que deseamos que se busque. En este caso las opciones de filtrado vienen descritas en el script para recoger la información exacta.

Tabla 2. Continuación de la Tabla 1 con los datos extraídos de la búsqueda.

IDENTIFICACIÓN	TITULARIDAD	CATEGORIA	SUBCATEGORIA	GESTIÓN
ES.28079.AHN	Archivos de titularidad publica	Archivos Generales y Centrales		Administración Central
ES.28005.AGA	Archivos de titularidad publica	Archivos Generales y Centrales		Administración Central
ES.28079.AHPM	Archivos de titularidad publica	Archivos Notariales		Administración Autonómica/Territorial
ES.28079.ACCE	Archivos de titularidad publica	Archivos Generales y Centrales		Administración Autonómica/Territorial
ES.28079.ACHT	Archivos de titularidad publica	Otras categorías de Archivos	Archivos de Instituciones Hidrográficas y Portuarias	Administración Central
ES.28079.AICGEA	Archivos de titularidad publica	Archivos Militares (excluidos los Históricos Militares)		Administración Central
ES.28079.ACCGA	Archivos de titularidad publica	Archivos Militares (excluidos los Históricos Militares)		Administración Central
ES.28079.AGHD	Archivos de titularidad publica	Archivos Generales y Centrales		Administración Central
ES.28079.AOEPM	Archivos de titularidad publica	Otras categorías de Archivos	Archivos de Administraciones Públicas	Administración Central
ES.28079.	Archivos de titularidad publica	Archivos Personales y Familiares		Administración Central
ES.28079.AGP	Archivos de titularidad publica	Archivos Generales y Centrales		Administración Central
ES.28079.ACMCU	Archivos de titularidad publica	Archivos Generales y Centrales		Administración Central
ES.28079.SAGMINT	Archivos de titularidad publica	Archivos Generales y Centrales		Administración Central
ES.28079.ACMF	Archivos de titularidad publica	Archivos Generales y Centrales		Administración Central
ES.28079.SRDAE	Archivos de titularidad publica	Archivos de Instituciones Científicas, Culturales y de Investigación		Administración Central

La muestra es representativa del conjunto de 111 archivos que nos podemos encontrar en la Comunidad de Madrid.

Tabla 3. Columnas finales de la tabla correspondiente a los datos extraídos.

HISTORIA DE LA INSTITUCIÓN

El Archivo Histórico Nacional fue creado por R.D. de 28 de marzo de 1866 (Gaceta de Madrid, de 31 de Marzo) como 'Archivo público general del Reino', a instancias de la Real Academia de la Historia, con la finalidad de poner a disposición de los investigadores los archivos incautados por el Estado a las instituciones religiosas, en virtud de las leyes desamortizadoras del siglo XIX. Por otra parte, las reformas de la Administración Central sucedidas a lo largo de siglo XIX dieron lugar a la supresión o transformación de organismos del Antiguo Régimen (Consejos, Inquisición...) y, por tanto, a la acumulación en las dependencias ministeriales de un enorme volumen de documentación inútil desde el punto de vista administrativo. Todos esos documentos de gran valor histórico, que no reunían garantías de conservación y creaban problemas de espacio para el normal funcionamiento de las oficinas, fueron recogidos en diferentes etapas por el Archivo Histórico Nacional hasta constituir hoy el depósito histórico español con mayor volumen de documentación (40 km de estantería y más de 500.000 unidades de instalación). Es un archivo de carácter público, accesible desde el principio a la investigación y abierto para recibir fundamentalmente la documentación histórica producida por diferentes organismos de la Administración Central del Estado.

URL

[archivodetail.htm?id=9](#)

Se han descartado datos que no consideramos de interés para poder representarlos y visualizarlos en un mapa, que es la finalidad de esta práctica.

Por ejemplo, el teléfono, los horarios de apertura, si tiene parking, el fax, los nombres de la persona de contacto, etc.

Una vez estructurado los datos en Excel llevamos el fichero resultante a la aplicación OpenRefine. Esta permite utilizar el API de Google para obtener las

coordenadas geográficas (latitud y longitud) que más tarde se utilizarán para ubicar los archivos en un mapa.

Figure 2D. Vista del proyecto abierto en la herramienta OpenRefine por columnas y filas con los datos importados del fichero resultante de Excel. Datos estructurados

« Start Over Configure Parsing Options

Column	NOMBRE DEL ARCHIVO	PAIS	COMUNIDAD	PROVINCIA	MUNICIPIO	DIRECCIÓN
1.	2 ARCHIVO HISTÓRICO NACIONAL	España	Comunidad de Madrid	Madrid	MADRID	Serrano, 115 28006, Madrid

Figure 2E. Vista del proyecto una vez han sido volcadas las coordenadas en una columna asociadas a cada Archivo.

A power tool for working with messy data.

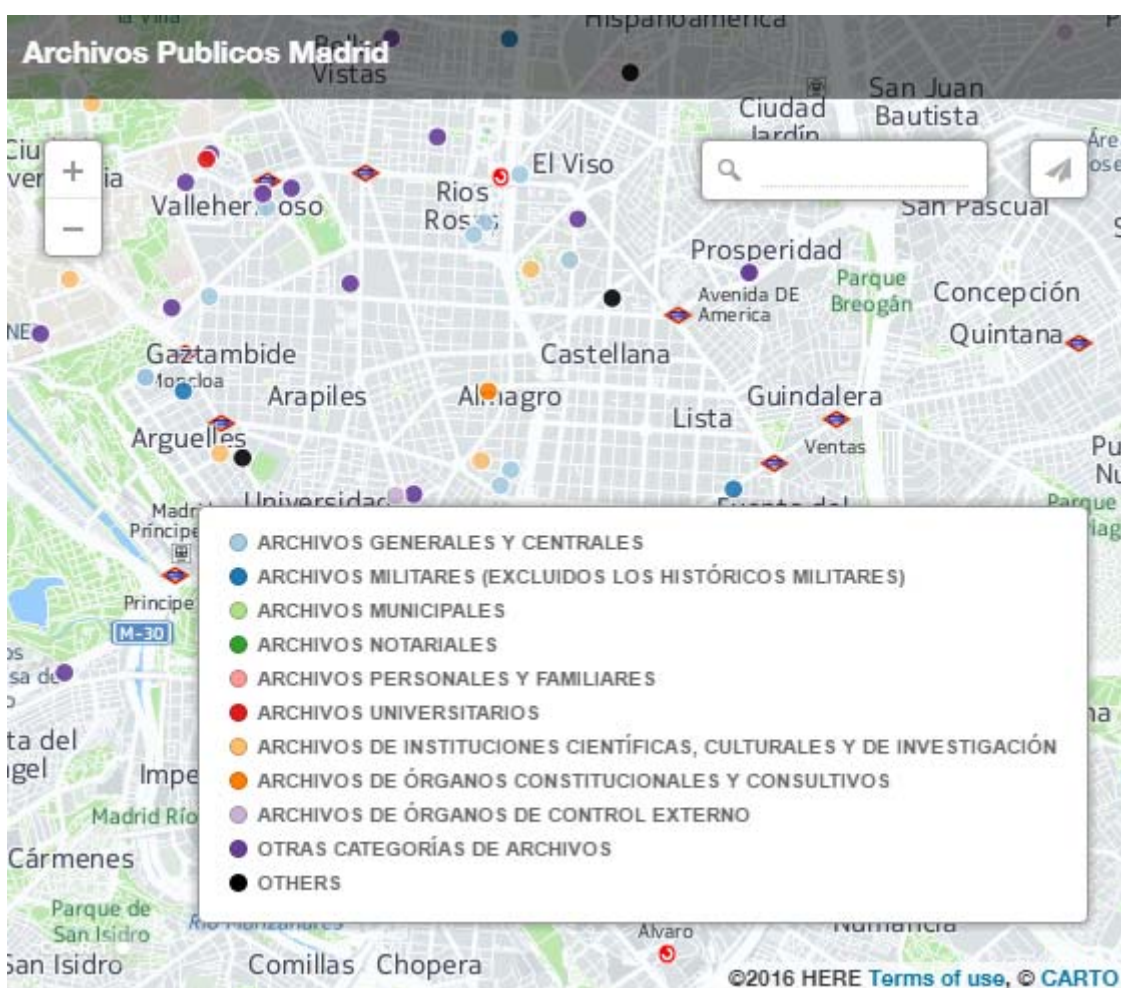
« Start Over Configure Parsing Options Project name Create Project »

LATITUD	LONGITUD	IDENTIFICACIÓN	TITULARIDAD	CATEGORIA	SUBCATEGORIA	GESTIÓN	HISTORIA DE LA INSTITUCIÓN	DESCRIPCIÓN DEL FONDO	URL
40.4411455	-3.6862931	ES.28079.AHN	Archivos de titularidad publica	Archivos Generales y Centrales		Administración Central	El Archivo Histórico Nacional fue creado por R.D. de 28 de marzo de 1866 (Gaceta de Madrid, de 31 de Marzo) como 'Archivo público general del Reino', a instancias de la Real Academia de la Historia, con la finalidad de poner a disposición de los investigadores los archivos incautados por el Estado a las instituciones religiosas, en virtud de las leyes desamortizadoras del siglo XIX. Por otra parte, las reformas de la Administración Central sucedidas a lo largo de siglo XIX dieron lugar a la supresión o transformación de organismos del Antiguo Régimen (Consejos, Inquisición...) y, por tanto, a la acumulación en las dependencias ministeriales de un enorme volumen de documentación inútil desde el punto de vista		archivodeta id=9

Los datos se nos mostrarán por filas y columnas de una manera similar a como lo podríamos ver en una tabla de cálculo de Excel. Lo que nos interesa aquí es el proceso por el cual vamos a obtener una columna con las coordenadas de cada archivo.

Con esta información ya cruzada es posible hacer la geolocalización. De manera similar los procesos previos guardamos los datos en un fichero que nos pueda permitir importarlo a una herramienta de publicación de mapas online. Se pueden usar, por supuesto, distintas herramientas para hacerlo. En nuestro caso elegimos la herramienta CartoDB, una herramienta de uso gratuito en sus prestaciones básicas para la visualización de mapas desarrollada por una empresa española¹⁷. Esta hoja se importa dentro de un mapa en obteniéndose el siguiente resultado:

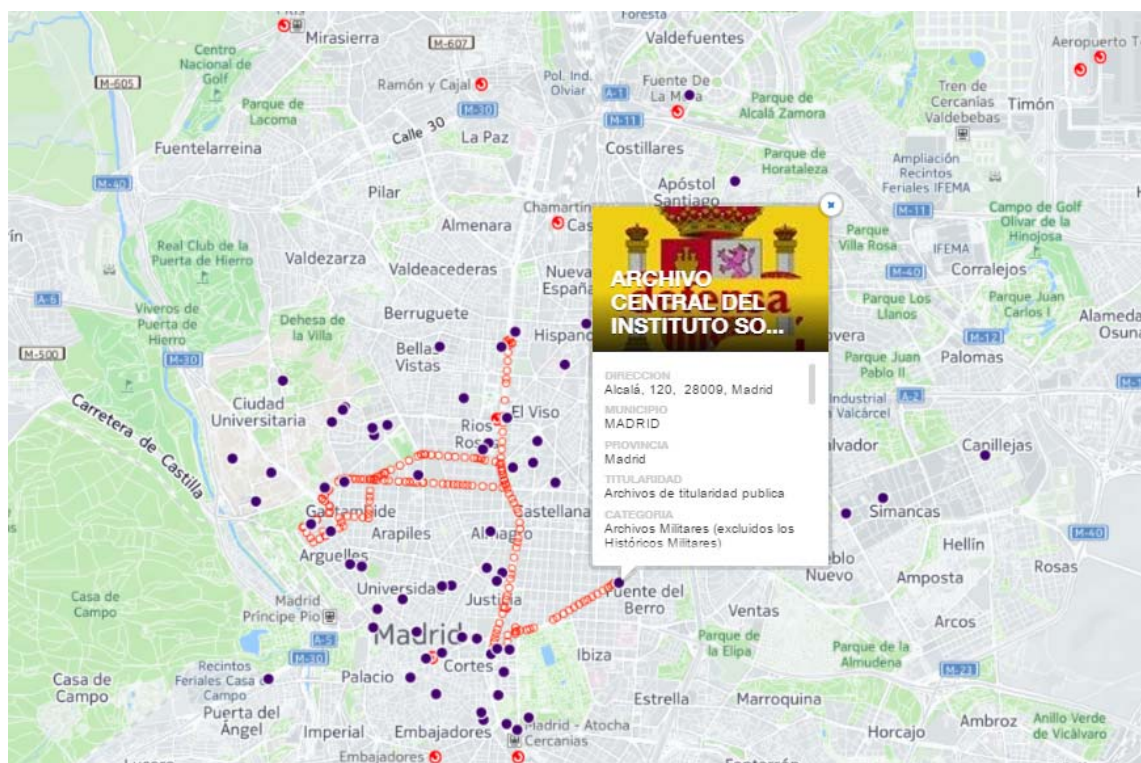
Figure 2F Vista de los datos integrados en la plataforma de geolocalización donde cada punto representa un tipo de Archivo.



¹⁷ Recientemente ha cambiado su nombre a CARTO. Está disponible en <https://carto.com>

Dentro de la tipología de archivo podemos seguir refinando la ruta interactiva haciendo que se vean aquellos que deseamos. En la siguiente foto podemos ver solo los archivos militares. La finalidad es que el usuario pueda tener opciones de visualización de los archivos de la Comunidad y que además cuando pulse sobre uno de los puntos que lo describen se pueda desplegar una ficha con los datos del arhivo y las características del fondo sus servicios y localización. A partir de aquí se pueden seguir representando itinerarios y rutas turísticas que den valor añadido a las Instituciones y faciliten el acceso a la Información al Usuario.

Figure 2G¹⁸ Vista en carto del mapa web sobre los archivos militares, vemos una ventana desplegada con la información sobreimpresionada de los datos de interés de la Institución.



Conclusiones

¹⁸ https://saizgodino.carto.com/viz/ff0f454a-becf-11e5-b11f-0e31c9be1b51/public_map

Los datos de los archivos públicos pueden reutilizarse de forma que se den productos de valor añadido a los ciudadanos, pero también permitan difundir la labor realizada por éstos. La propuesta presentada es únicamente una primera posibilidad, elaborada utilizando tecnologías de código abierto (Open Source) al alcance de todos pero que, a pesar de su simplicidad, proporciona un resultado mucho más atractivo y útil de una información valiosa, pero “escondida”, por la dificultad que añade la búsqueda en una base de datos en línea (desconocimiento del uso, desinterés si lleva algo de tiempo el obtener la información, etc...).

La idea es implementar Sistemas de Recuperación de Información¹⁹ mediante el conjunto de aplicaciones y elementos interrelacionados necesarios para verificar los diferentes procesos con los que satisfacer las necesidades de información de uno o más usuarios potenciales de cualquier organización.

Referencias

Colmenero- Ruiz, María-Jesús. “Visualizing archives: Spanish archives map”. Netcom [Online], 29-3/4, 2015. Consulta: 19-08-2016. Disponible en <<http://netcom.revues.org/2143>>.

¹⁹ Se puede ampliar Información en la web de los autores del trabajo.
<http://www.tecnologiasdocumentales.com/recuperación-información>