

Taller de iniciación a aplicaciones básicas para la visualización de datos de GVA Oberta

OpenRefine – Google Fusion Tables – Carto - ...

CURSO TRANSPARENCIA A TRAVES DE LOS DATOS 2016

7 de septiembre 2016

David Pardo – dapargi@upv.es



TALLER VISUALIZACION
- DATOS GVA OBERTA -

Para quién

- Cualquier ciudadano, incluidos profesionales de cualquier ámbito, interesado en iniciarse en visualizaciones con datos públicos abiertos y más en concreto de la Generalitat Valenciana, GVA
 - A partir de la descarga de archivos del conjunto de datos abiertos del Portal de datos abiertos de *GVA Oberta*.



- Sin competencias en programación, análisis de datos (*business intelligence*, minería de datos, etc.)
- Mínimas competencias digitales
- Avanzado: recomendables conocimientos mínimos en hojas de cálculo y estadística (fórmulas, funciones, tablas dinámicas, tipos de gráficos) y ... programación

Objetivos

- Estudiar de forma básica un set de herramientas gratuitas (algunas *open source*) que nos permitan realizar unas visualizaciones con datos públicos sin conocimientos específicos en programación, estadística, ...
 - Explorar tablas que permitan recoger datos
 - Para informarse / contar historias / periodismo de datos
 - Crear gráficos
 - Crear mapas
- Herramientas a utilizar: Open Refine, Google Fusion Tables (tablas dinámicas) y Carto (Cartodb)
- Conocer otras herramientas alternativas para la visualización de Datos Abiertos

Datos del sector público

Open Government Data o Public Sector Information

- Datos de titularidad pública: datos que recoge, producen, reproducen y difunden las Administraciones y organismos públicos
- Atractivos: fiables, completos, de calidad ???
- REUTILIZACIÓN:
 - Justificación económica: sector infomedario (€€€, trabajo)
 - Calidad de vida: nuevos productos y servicios al ciudadano.
 - Participación, transparencia y rendición de cuentas: control ciudadano, prensa, organizaciones civiles
 - Mejora eficiencia de la Administración
- Nuevo marco legal: Normas y leyes RISP y de Transparencia, Acceso a la información pública, buen gobierno, ...
 - ... en el ámbito estatal (2007 y 2011) y autonómico (GVA, 2015)
 - Normas Interoperabilidad, ciudades inteligentes, ...

Clasificación de los datos abiertos



Publica en la web (con cualquier formato: no estructurada – no procesable-) y bajo una licencia abierta

Estructurado
No propietarios
Con URIs que permitan su enlace
Enlazados a otros datos



<http://5stardata.info/es/>

8 + 2 PRINCIPIOS DEL OPEN (PUBLIC) DATA (del Open Government Data):

1. **COMPLETOS:** más completa posible (todos los datos), base de datos, y descrita semánticamente
2. **PRIMARIOS:** de la fuente de los datos (mayor nivel granularidad posible, ni modificados ni agrupados)
3. **OPORTUNOS:** actuales, a tiempo (actualizado según su naturaleza)
4. **FACILIDAD DE ACCESO:** electrónico, al mayor nº de usuarios y propósitos
5. **LEGIBLES POR MÁQUINAS:** procesables, estructurado para procesado automático, ...
6. **NO DISCRIMINATORIOS:** para cualquiera, sin requisitos de registro (anónimo)
7. **NO PROPIETARIOS:** en formatos abiertos y estándares (ninguna entidad tiene control exclusivo)
8. **LIBRES DE LICENCIAS:** no está sujeta a ningún tipo de regulación de derecho de autor, patentes o registros marcas.
Se puede permitir licencias abiertas como mínimas restricciones de reutilización (razonables de privacidad, seguridad o privilegio)
9. **PERSISTENTES:** accesible a lo largo del tiempo
10. **COSTE DE USO:** gratuito o costes marginales (recogida, producción, reproducción y difusión)

[Eight principles of open government data \(Opengovdata.org\)](http://Opengovdata.org)

[TEN PRINCIPLES FOR OPENING UP GOVERNMENT INFORMATION \(Sun Light Foundation\)](#)

Visualización de datos

MAPA | Resultado electoral del 26J, municipio a municipio

Consulta en directo los resultados de cada municipio y descubre quién ha ganado en cada pueblo

Comparte 1127 Me gusta 27 comentarios



De qué mueren los valencianos?

Valencia registró en 2014 un total de 7.274 defunciones, 3.453 de hombres y 3.821 de mujeres, con una tasa de 923 cada 100.000 habitantes y cuyas principales causas de mortalidad fueron diferentes tipos de tumores y enfermedades del sistema circulatorio.

5 Cosas que tu banquero no quiere que sepas este artículo revela secretos que le pueden ayudar a ganar dinero.

Valencia registró en 2014 un total de 7.274 defunciones, 3.453 de hombres y 3.821 de mujeres, con una tasa de 923 cada 100.000 habitantes y cuyas principales causas de mortalidad fueron diferentes tipos de tumores y enfermedades del sistema circulatorio.

Así se contempla en un informe de la "Mortalidad según causas 2014" que desde hoy incorpora el portal web de la Oficina de Estadística del Ayuntamiento de Valencia, cuya fuente de información es el Movimiento Natural de la Población publicado por el Instituto Nacional de Estadística (INE).

En concreto, fallecieron en ese año 2.157 personas por enfermedades del sistema circulatorio, con una tasa cada 100.000 habitantes de 273,98, y de los que 922 fueron hombres y 1.235 mujeres.

Los tumores fueron la causa de muerte hace dos años de 2.114 personas, con una tasa cada 100.000 habitantes de 268,52, con 1.241 hombres

ELECCIONES MUNICIPALES 2015



4 Me gusta 18 agosto 2016 19:03

Valencia registró en 2014 un total de 7.274 defunciones, 3.453 de hombres y 3.821 de mujeres, con una tasa de 923 cada 100.000 habitantes y cuyas principales causas de mortalidad fueron diferentes tipos de tumores y enfermedades del sistema circulatorio.

Así se contempla en un informe de la "Mortalidad según causas 2014" que desde hoy incorpora el portal web de la Oficina de Estadística del Ayuntamiento de Valencia, cuya fuente de información es el Movimiento Natural de la Población publicado por el Instituto Nacional de Estadística (INE).

En concreto, fallecieron en ese año 2.157 personas por enfermedades del sistema circulatorio, con una tasa cada 100.000 habitantes de 273,98, y de los que 922 fueron hombres y 1.235 mujeres.

Los tumores fueron la causa de muerte hace dos años de 2.114 personas, con una tasa cada 100.000 habitantes de 268,52, con 1.241 hombres

Visualización de datos

- (*data visualization*) es el proceso de búsqueda, interpretación, contrastación y comparación de datos que permite un conocimiento en profundidad y detalle de los mismos de tal forma que se transformen en información comprensible para el usuario (*Wikipedia*)
- Que el usuario (ciudadano) adquiera los conocimientos necesarios para comprender correctamente los datos facilitados
 - Aporten otra manera alternativa a la oficial de mostrar la información pública más cercana al ciudadano
 - Saquen información relevante de los datos públicos y pueda ser comunicada de forma atractiva
 - Motiven o molesten a la institución
 - Mantengan o impulsen la mejora de alguno de sus servicios
 - Denunciar la falta de datos o su formato
 - Datos abiertos mín. ★★
 - Y cumplan 8 principios del open government data

Etapas visualización de datos

- **Compilar:** Búsqueda y recogida datos
- **Interpretación:** Contextualizar datos / investigación (de dónde, fiables, descripción datos y categorías)
- **Limpieza** errores y **adecuación** archivos para las tablas / gráficos / mapas finales:
 - Limpieza:
 - Renombrar encabezados y valores
 - Eliminación de datos vacíos y espacios en blanco (antes, entre y en el final)
 - Unión de datos distintos que son el mismo valor (tildes, may/minúsculas, espacios, letras cambiadas)
 - Detección de valores anormalmente altos o bajos (ordenar)
 - Valores duplicados
 - Preparar tablas con los datos necesarios y limpios (en función de lo que se desea contestar y visualizar)
 - Cambiar formato,
 - Dividir valores
 - Creación de columnas basadas en otras
 - Exploración y Filtrado / facetas
 - Eliminación de filas (datos) y columnas no necesarias
- **Exportación** y combinación o Mezcla de datos
- **Comunicación:** Conversión a gráficos o mapas (visualización) y narración de historias / personalizar / humanizar ...

Herramientas visualización 1/2

ESCRAPEAR:

- [Portia](#)
- [Import.io](#)
- Extensiones navegador: [Chromescraper](#), [Data Miner](#)

LIMPIEZA Y EXPLORACIÓN DE DATOS:

- [Open Refine](#) (antes Google Refine):
- [Data Wrangler](#)
- Hojas de cálculo:
 - [Excel](#) (Office)
 - [Hojas de cálculo de Google](#),
 - [Calc](#) (LibreOffice)

Herramientas visualización 2/2

Gráficos y mapas (además de hojas de cálculo: Excel, de Google, Calc y Fusion Tables)

- [Carto](#)
- [Inforgr.am](#)
- [Datawrapper](#)
- [Mapbox](#)
- [Quadrigram](#)
- Relaciones, grafos y jerarquías: [Gephi](#), [VIS](#), [FoamTree](#), [Onodo](#)
- Presupuestos: [OpenBudgets](#), [OpenSpending](#), [Visual Government](#)
- INFOGRAFÍAS: [Infogr.am](#), [Canva](#), [Piktochart](#), [Venngge](#), [Easel.ly](#), [Visually](#)
- Herramientas de Business Intelligence (gráficos, mapas, dashboards,...):
 - [Qlik](#): [Qlik Sense](#) y [QlikView](#)
 - [Tableau](#)
 - [Pentaho](#)
 - [RapidMiner](#)
 - [BIME](#)

Herramientas para ejercicios del taller

Herramientas principales para ejercicios taller

- Open Refine (antiguo Google Refine): limpieza y exploración datos
- Google Fusion Tables: limpieza, exploración y visualización (gráficos y mapas) de datos
- Cartodb → Carto Builder : mapas

Fuente de datos:

- Fuente principal
gva Oberta - Portal de transparencia de la Generalitat Valenciana:
Datos Abiertos
- Fuente secundaria: INeBase – Instituto Nacional de Estadística.

OpenRefine

- <http://openrefine.org/>
- Anteriormente Google Refine (oct, 2012)
- Actual versión: *OpenRefine 2.6-rc2 Release Candidate 2*
- Herramienta open source para limpieza de información:
 - Importar información en varios formatos
 - Explorar información fácilmente,
 - Crear **facet**s para filtrar de forma simple o múltiple
 - **Clustering o agrupaciones:** Encontrar automáticamente agrupaciones de diferentes valores que pueden ser representaciones alternativas de la misma cosa y unirlos en un único valor
 - Aplicar **transformaciones** a celdas: quitar espacios al principio o al final, dos espacios consecutivos, cambiar min/mayúsculas, convertir a número o texto o fecha,
 - Filtrar y dividir la información con expresiones regulares
 - Permite guardar los cambios realizados en un proyecto y exportarlos en formato JSON en un archivo txt para aplicarlos a otros archivos iguales.
- Programa en Java que se ejecuta en local en un navegador (no hace falta conexión). Sino se ejecuta automáticamente por la dirección <http://127.0.0.1:3333/>
- Tipos de archivos: TSV, CSV, *SV, Excel (.xls, xlsx), JSON, XML, RDF –XML, Google documentos.



Google Fusion Tables

Tablas dinámicas de Google



Aplicación experimental de Google para obtener, visualizar (gráficos y mapas) y compartir tablas de datos

Acceso desde *Google Drive* en aplicación Tablas dinámicas o desde tables.googlelabs.com

Datos se almacenan en Google Drive.

Permite:

- Crear e importar tablas.
- Crear tablas/vistas: Editar, explorar, filtrar
- Tarjetas (cards): Visualizar información en tarjetas (cards):
- Mapas (es posible geolocalizar automáticamente o indicar la columna para situar).
 - Configurar estilos de puntos (iconos únicos, basados en un columna, por intervalos), polígonos, líneas
 - Configurar leyenda o ventana de información .
- Resúmenes
- Gráficos
 - De dispersion (scatter)
 - Líneas
 - Áreas
 - Columnas y barras (histogramas)
 - De sectores (tarta)

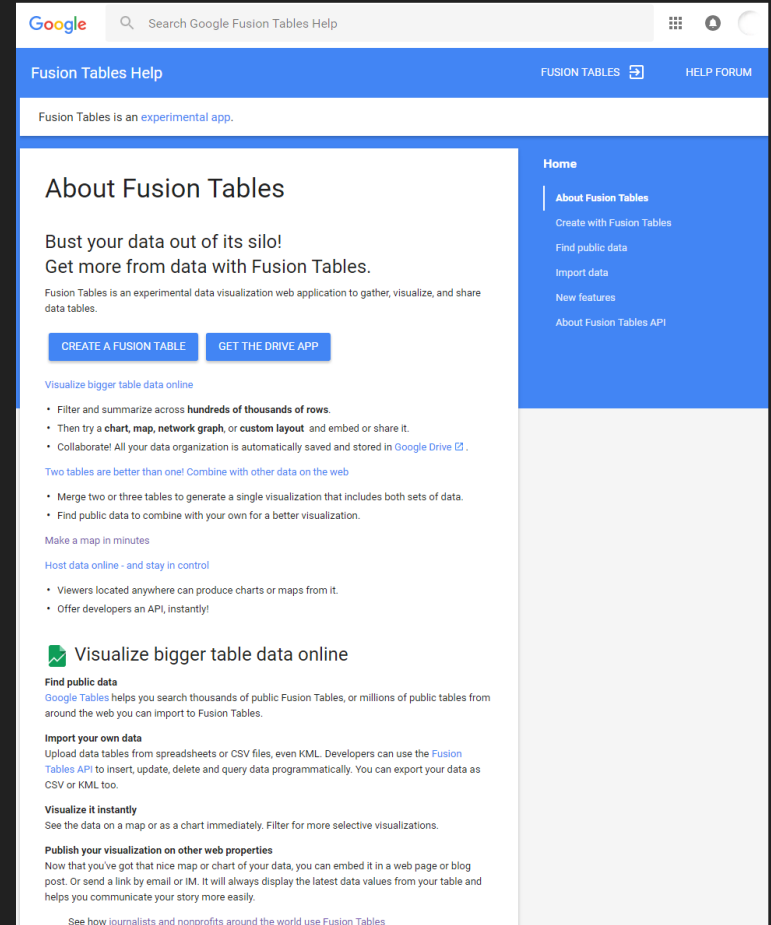
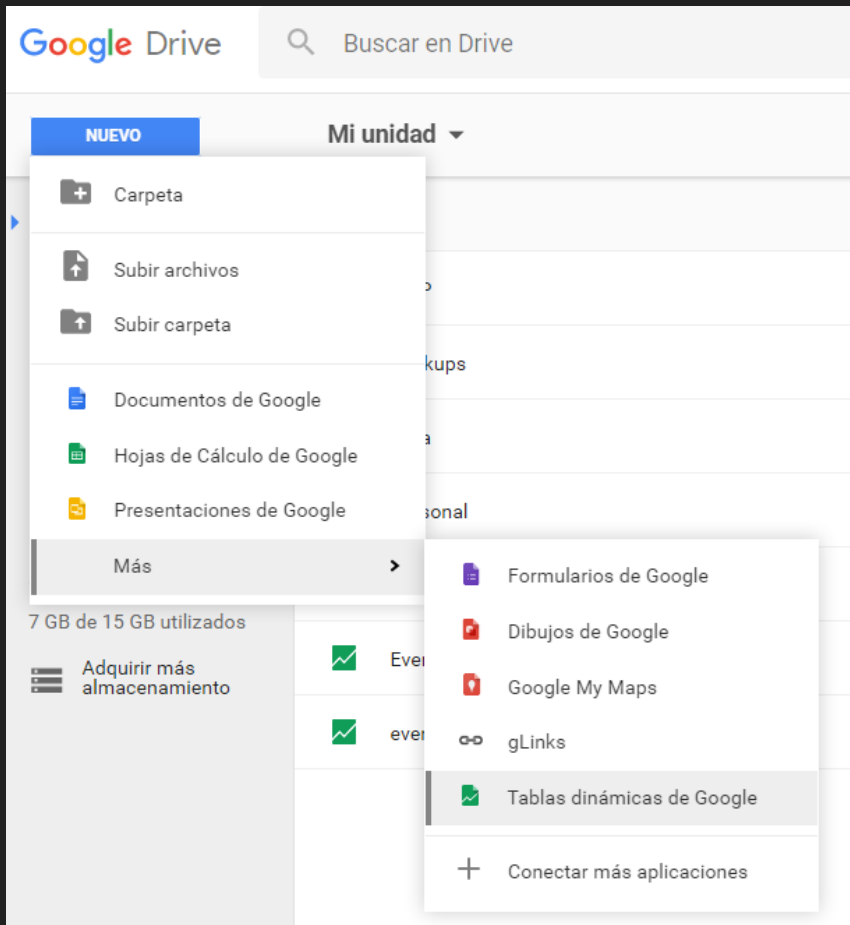
Google Fusion Tables

Tablas dinámicas de Google



Crear una tabla dinámica desde Google Drive o

<https://tables.googlelabs.com/>

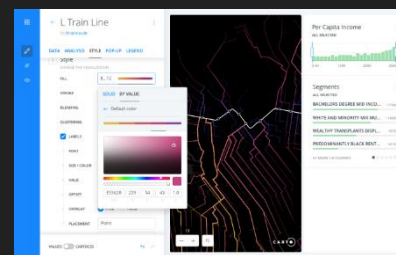
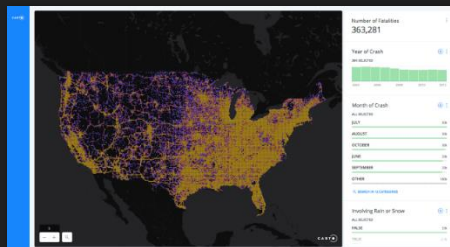
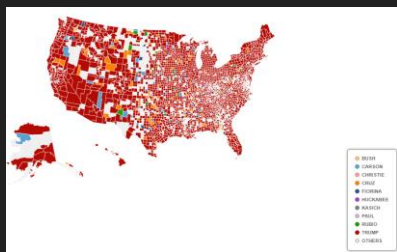


Carto (Cartodb / Carto Builder)



- CARTO (antes CartoDB)
- Cartodb:
 - Visualización y análisis de datos geográficos
 - Creación de tabla de datos (*Datasets*) y Mapas (*Maps*)
 - Conocimientos en SQL
 - Visualización de datos (para periodistas de datos no programadores) en gráficos y mapas
- CARTO Builder (+intuitiva, análisis, predicción)

Análisis datos (business intelligence) para la optimización de decisiones y predicción de consumo (en toda clase de empresas, en especial en servicios financieros para la prevención de riesgos de inversiones en áreas específicas).



Dades Obertes

gva Oberta - Portal de transparencia de la Generalitat Valenciana: Datos Abiertos

The screenshot shows the 'dades Obertes' portal interface. At the top, there is a navigation bar with the logo and the text 'PORTAL DE TRANSPARÈNCIA DE LA GENERALITAT VALENCIANA'. Below this, there are several menu items: 'CONJUNTS DE DADES', 'TEMES', 'PARTICIPA', 'QUÈ SÓN LES DADES OBERTES?', and 'GVA OBERTA'. The main content area is titled 'CONJUNTS DE DADES' and features a search bar and a list of datasets. The left sidebar contains two sections: 'Temes' (Topics) and 'Etiquetes' (Tags), each with a list of categories and their respective counts. The 'Temes' section includes categories like 'Medi ambient (196)', 'Educació (43)', 'Turisme (23)', 'Legislació i justícia (14)', 'Urbanisme i infraestructures (11)', 'Salut (8)', 'Ocupació (5)', 'Societat i benestar (4)', 'Indústria (4)', and 'Transport (3)'. The 'Etiquetes' section includes 'ICV (169)', 'Cubierta terrestre (156)', 'Institut Cartogràfic València (103)', 'Uso del suelo (62)', 'Vuelo fotogramétrico (59)', 'Edificios (57)', 'fotograma (55)', 'Fototeca Valenciana (55)', 'IDEE_INSPIRE (43)', and 'estacions (35)'. The main list of datasets includes: 'Centres educatius 2011 - 2012' (with download options: CSV, JSON, PDF, XML), 'Contaminació Atmosfèrica i Ozó - Mitjanes Diàries 2011' (with download options: CSV, JSON, XML, PDF), 'Centres educatius 2009 - 2010' (with download options: CSV, JSON, PDF, XML), 'Resum d'escolarització 2012 - 2013' (with download options: CSV, JSON, PDF, XML), 'Vuelo fotogramétrico del año 1988 del Parque Natural de la Albufera' (with download option: HTML), 'Vuelo fotogramétrico del año 1990 de la Marjal de la Safor' (with download option: HTML), 'Vuelo fotogramétrico del año 2003 de la provincia de Castellón' (with download option: HTML), 'Alumnat matriculat en Conservatoris Professionals de Música 2006 - 2007' (with download options: CSV, JSON, PDF, XML), and 'Centres educatius 2010 - 2011' (with download options: CSV, JSON, PDF, XML).

Ejercicios Taller

1. **Agenda de eventos de la Comunidad Valenciana:** geolocalización directa del dataset de eventos de la CV.
2. **Día del Alzheimer:** obtención de datos para artículo periodístico tras limpieza de datos de datasets y su visualización
3. **Centros educativos de la Comunidad Valenciana:** geolocalizar en función del tipo (público / concertado / privado) a través campo dirección o al mezclar con otras tablas de auxiliares con datos geográficos.
4. **Escuelas Oficiales de Idiomas:** geolocalizar y aportar datos de matriculación en determinados idiomas

Ejercicio 1: Agenda de Eventos

- Objetivos / Tarea:

Utilización rápida de Google Fusion Tables para crear en un mapa los eventos que se producen en la Comunidad Valenciana con datos ya georreferenciados.

- Herramientas: Google Fusion Tables (tablas dinámicas de Google)

- Conjunto de datos GVA: Agenda de Eventos (act. 29 julio 2016)

Ejercicio 2: Día del Alzheimer

- Objetivos:

Próximo 21 de septiembre es el Día internacional de Alzheimer.

Obtener y preparar datos para su visualización en gráficos y mapas sobre fallecidos de Alzheimer en las provincias de la Comunidad Valenciana para preparar un artículo.

- A. Datos y gráficos sobre fallecimientos por edad y sexo en el último año 2013 en cada provincia CV y en la CV (=2007)
- B. Comparar fallecidos del 2013 con los del 2007 (primer año del que se tienen datos en el Portal dades Obertes de la GVA (2007 a 2013) en cada provincia
- C. Resumir la información de 2007 y 2013 en un gráfico barras apilado (stack) y después, por medio de una tabla dinámica auxiliar "CV - Provincias KML.csv" mezclar las tablas para obtener la posición geométrica de las provincias
- D. Mapa de líneas con la evolución del Alzheimer en los últimos 20 años en la CV (INE)

- Tareas: Limpiar la tabla original y prepararla para explorar sus datos en Open Refine y crear posteriores visualizaciones en Google Fusion Tables.

- Conjunto de datos GVA: Datos de mortalidad 2007 a 2013

- Municipios - Causas CV (sin Alzheimer)
- Provincias - Causas INE-CCAA (Alzheimer G30)
- Municipios - Casusas CIE10 (Alzheimer: 051)

- Datos auxiliares

- Datos del INE de Alzheimer en los últimos 20 años en la CV
- Archivo con los datos geométricos de las provincias y su población

Ejercicio 3: Centros educativos

- Herramientas: Open Refine y CARTO
- Objetivos:
 - Localizar en pantalla los Centros Educativos de primaria y secundaria (público, privado o concertado) de la CV por tipología (público, privado o concertado)
 - Geolocalización de Carto por dirección o mezcla registro de Centros de Conselleria d' Educació
- Conjunto de datos GVA: Centros educativos 2015-2016 (de educación Infantil, Primaria y Secundaria) (aprox. 1838)
- Datos auxiliares (para mezclar)
 - "Registro Centros Docentes CV Julio 2016 - Conselleria Cultura – SoloLocalizacion" Fuente: Conselleria d' Educació (aprox. 3628)

Ejercicio 4: Escuelas Oficiales de Idiomas

- Herramientas: Open Refine y CARTO
- Objetivos:
 - Situar las EOI de la CV
 - Extra: Recoger datos sobre la matriculación en EOI 2015-2016 (mostrar los matriculados en inglés a partir del archivo).
- Conjunto de datos GVA:

Escuelas Oficiales de Idiomas CV - Julio 2016 (de Registro de Centros Docentes de Conselleria d'Eduació)

“Datos de matriculación en EOI 2015-2016” → “Extra - Resumen matriculado-EOIs-2015-2016-csv.xls”