

Tema 2

Introducción a sistemas informáticos para la extracción, tratamiento y visualización de datos,

TALLER DE PERIODISMO DE DATOS



Proyecto
OpenCourseWare-UNIA
(ocw.unia.es)



Autor
Pablo Martín Muñoz



ÍNDICE

Introducción

Objetivos

Contenidos

1. Definición del marco de trabajo
2. Ideas generales sobre hojas de cálculo
3. La base de la extracción de datos de PDFs
4. De los datos a las visualizaciones
5. Construye un mapa con datos
6. Insertar visualizaciones en otras webs

Ideas clave

Referencias



INTRODUCCIÓN

Con esta sesión pretendemos mostrar una serie de **prácticas y herramientas digitales para el trabajo con datos**. Se mostrarán anotaciones básicas para la extracción de datos de documentos, algunos aspectos relacionados con el tratamiento de datos y las posteriores visualizaciones de los mismos en formato gráfica y mapa. Por último se indicará las diferentes formas de exportar el trabajo para ser publicado en cualquier sitio web.



OBJETIVOS

El objetivo principal de esta unidad será mostrar algunas herramientas digitales de trabajo con datos y sus características básicas y comunes a otras. Con ello pretendemos despertar en el alumnado esa inquietud hacia el mundo de los datos desde acciones eminentemente prácticas y que contribuyan a una mejora en sus capacidades de estudio y análisis de bases de datos.

Además queremos introducir en conceptos como las visualizaciones de los datos, una forma de conectar mejor con las personas facilitando la comprensión de los datos y poniendo de manifiesto el sentido estético del trabajo.



CONTENIDOS

1. Definición del marco de trabajo

La importancia de este punto radica en los nuevos paradigmas de trabajo que se abren estando Internet integrado en nuestros espacios de trabajo.

De manera simplista, los trabajadores de la información necesitan software y ficheros con datos y/o información.

Asumido lo anterior podemos encontrarnos diferentes combinatorias. Éstas son:

- Software instalado en un PC
- Software accesible en Internet
- Ficheros guardados en un PC
- Ficheros guardados en Internet

En esta sesión formativa se intentará trabajar, de forma natural, con todas las opciones anteriores. Esto no es algo excepcional pero sí es conveniente tomar conciencia de ello. Al final es la forma en la que actualmente se trabaja en la red.

las fotos de mi último viaje, la aplicación de correo que uso, actas de reuniones, mis correos electrónicos, las canciones que me gustan, el reproductor de películas, mi emisora de radio favorita, los artículos de mi blog



2. Definición del marco de trabajo

Las hojas de cálculo son un tipo de documento que nos permiten manipular datos numéricos y/o alfanuméricos.

La representación de los datos se realiza en forma de filas y columnas (formato tabular) lo que hace que cada dato esté dispuesto en una celda (conexión entre fila y columna).

Cada celda puede contener un valor numérico, alfanumérico o una fórmula que opere con los valores de otras celdas.

Debido a estas cuestiones, el uso de las hojas de cálculo está muy extendido en todas las áreas ya que permiten un trabajo simple/medio/complejo con los datos para obtener resultados de forma rápida y automatizada.

Existen ciertos formatos (ods, xls, ...) propios de los diferentes tipos de software de gestión de hojas de cálculo pero se recomienda el uso del formato CSV (valores separados por comas) para la exportación de los mismos una vez que se ha realizado el trabajo con ellos.

Dado el formato tabular de las hojas de cálculo es automático transformar un documento en su análogo CSV, el cual perderá todo el estilo y solo conservará los datos.

En un formato CSV cada fila representa una fila de la hoja de cálculo y cada columna se representa por su valor más una coma indicando que termina una columna y empieza otra.

Países	Población	Área (km²)	Capital
España,	46704314,	504645,	Madrid
Portugal,	10427301,	92212,	Lisboa
Italia,	60782668,	301338,	Roma

Todo software de hojas de cálculo puede abrir sin problemas los ficheros CSV y presentarlos en formato tabular para su posterior procesamiento.

Para obtener un fichero CSV de una hoja de cálculo basta con abrir dicha hoja de cálculo con el software que se use y guardarla como "CSV" en la opción determinada del programa. Con ello se obtiene una copia estandarizada de los datos en un formato abierto.



El objetivo de esta sesión formativa no es indagar en el funcionamiento de las hojas de cálculo pero sí se valoró como necesario el presentarlas dado el amplio uso que de ellas se realiza en diferentes ámbitos como forma de mostrar y operar con conjuntos de datos.



3. La base de la extracción de datos de PDFs

Los PDFs son el formato de ficheros más extendido de Internet. Fue concebido como un tipo de fichero para manejar papel electrónico. Esto es, un sistema que permita la misma visualización del contenido con independencia de la plataforma usada para su visualización.

Dado el gran número de plataformas y software disponibles, el uso de un formato de papel electrónico proporciona grandes ventajas.

Pero el mal uso o el uso indiscriminado del formato PDF lo ha convertido en un dolor de cabeza para los profesionales que trabajan con datos dada la gran implantación del mismo. Y es que el PDF no es un formato de representación de datos.

En este apartado se mostrarán sus mayores inconvenientes y se introducirá a una herramienta que puede aliviar ciertos problemas que con ellos se tenga.

Algunos de los problemas de los PDFs son:

- Dependencia del software con el que se creó (mayor o menor calidad de ficheros)
- Documentos escaneados (de lo analógico a lo digital se convierte en un PDF que contiene una imagen)
- Detección automática de datos en tablas
- Tratamiento de tablas complejas

Uno de los programas más conocidos (libre y gratuito) para la extracción de datos en tablas de PDFs es TABULA. Fue inicialmente desarrollado por Manuel Aristarán y mejorado por la comunidad.

TABULA acepta como entrada un fichero PDF basado en texto (no imágenes escaneadas) y devuelve los datos existente en tablas de ese PDF en formato CSV.

TABULA se fundamenta en el uso de técnicas de visión por computador para el reconocimiento de caracteres y su futura extracción.

El software existe para diversas plataformas y puede ser usado, modificado y compartido (licenciamiento libre).

Más información en: tabula.nerdpower.org

Nótese que el trabajo de extracción de datos de los PDFs no asegura un resultado



satisfactorio en todos los casos ya que existen demasiadas dependencias que complican el proceso. La misión de este apartado es animaros a realizar alguna extracción para valorar su idoneidad en un proceso más complejo o recurrir a otras técnicas manuales pero paralelas como el *crowdsourcing* (ver referencias finales).



4. De los datos a las visualizaciones

Para la inmensa mayoría de los seres humanos la comprensión de los datos se simplifica aplicando sobre ellos diversas técnicas de visualización de los mismos, tanto para facilitar su entendimiento como para destacar alguna característica o elemento diferenciador oculto en cierto conjunto de datos.

Desde la estadística en particular y desde la ciencia matemática en general siempre se ha tendido a realizar representaciones gráficas de conceptos abstractos difícilmente entendibles sin dichas gráficas.

Es por ello que el campo de las representaciones gráficas de números (datos) siempre ha tenido una gran importancia tanto en su uso como en su investigación.

Con la irrupción de la **Era del Dato** en la que actualmente nos encontramos dentro de la Sociedad de la Información, en donde toda entidad del sistema genera datos sin cesar, el estudio de las representaciones gráficas de los datos ha experimentado un tremendo auge y no son pocas las posibilidades con las que se cuenta.

A continuación se verán algunas posibilidades, desde las clásicas hasta las más actuales en la red.

- **Gráfico de barras (*bar chart*)**: Consta de dos ejes y representa en el eje Y los elementos a medir (agrupados o no) y en el eje X el valor medido.
- **Gráfico de columnas (*column chart*)**: Consta de dos ejes y representa en el eje X los elementos a medir (agrupados o no) y en el eje Y el valor medido.
- **Gráfico de tarta o donut (*pie/donut chart*)**: Representa de forma circular los porcentajes del elemento a medir en torno al total que es la tarta/donut completo.
- **Gráfica de dispersión (*scatter plot/chart*)**: Basado en el sistema cartesiano, permite representar los valores de dos variables para un conjunto de datos. Los datos se muestran como una colección de puntos tomando valores en el eje vertical (variable1) como en el horizontal (variable2).
- **Gráfico de barras apiladas (*stacked bars graph*)**: Similar a los gráficos de columnas pero en el eje Y se muestran valores de una variable agrupados en torno a una categoría. El objetivo de la visualización es intentar comparar *las partes de un todo*.
- **Gráfico de flujo (*stream graph*)**: Visualización muy recomendable para datos continuos tales como series temporales (en lugar de usar barras apiladas).
- **Circle packing**: Círculos anidados que permiten representar jerarquías y comparar los valores. Esta visualización es particularmente eficaz para



mostrar la proporción entre los elementos a través de sus áreas y su posición dentro de una estructura jerárquica.

- **Treemap:** Permite representar las jerarquías de datos y la proporción entre los elementos. Los diferentes niveles jerárquicos crean agrupaciones visuales a través de la subdivisión en rectángulos en proporción al valor de cada elemento. Estas visualizaciones son útiles para representar las diferentes proporciones de estructuras de datos jerárquicos anidados.

Para la creación de las visualizaciones se usarán dos aplicaciones web, por lo que no será necesario instalar ningún software en los equipos personales. Las aplicaciones web son:

- RAW [app.raw.densitydesign.org]
- DATA WRAPPER [datawrapper.de]

Antes de empezar con algunos ejemplos de uso de las herramientas comentadas sería conveniente identificar algunos aspectos comunes de toda herramienta de creación de visualizaciones.

Básicamente siempre se pueden diferenciar 3 apartados fundamentales: inserción de datos, configuración y adaptación, y exportación de la visualización.

- **Inserción de datos:** Se trata de introducir los datos que servirán de materia prima al sistema para la creación de la visualización.
- Pueden existir diversas formas de inserción de datos pero las más comunes son mediante texto (formato CSV) o mediante la subida de un fichero (CSV, ODS, XLS, etc).
- **Configuración y adaptación:** Una vez cargados los datos se deberá elegir la gráfica que mejor represente el conjunto de datos con el que se trabaja. Además de ello hay que definir las variables a mostrar, los colores en los que se visualizarán, los textos y algunas cuestiones más.
- **Exportación:** Finalizada la visualización es tiempo de exportarla para usarla en otros entornos, ya sea obteniendo la imagen resultante u obteniendo su código embebido (código HTML) para ser integrada en otras webs.

Vistos estas tres secciones de trabajo bastante común en este tipo de herramientas pasamos a mostrar, mediante ejemplos, cada una de las aplicaciones webs comentadas previamente para realizar las visualizaciones de los datos.

RAW

RAW es una herramienta web libre y gratuita desarrollada por Density Design Research Lab (Politécnica de Milán) para servir de puente entre hojas de cálculo y gráficas vectoriales.

Está construido usando como base D3js, una de las librerías gráficas más potentes y extendidas en el mundo web. Además el ser libre la convierte en una herramienta altamente configurable y modificable.

- **Inserción de datos:**



Esta sección está preparada para que copies el contenido de un fichero CSV o directamente de una hoja de cálculo en la caja de texto que aparece en la interfaz.

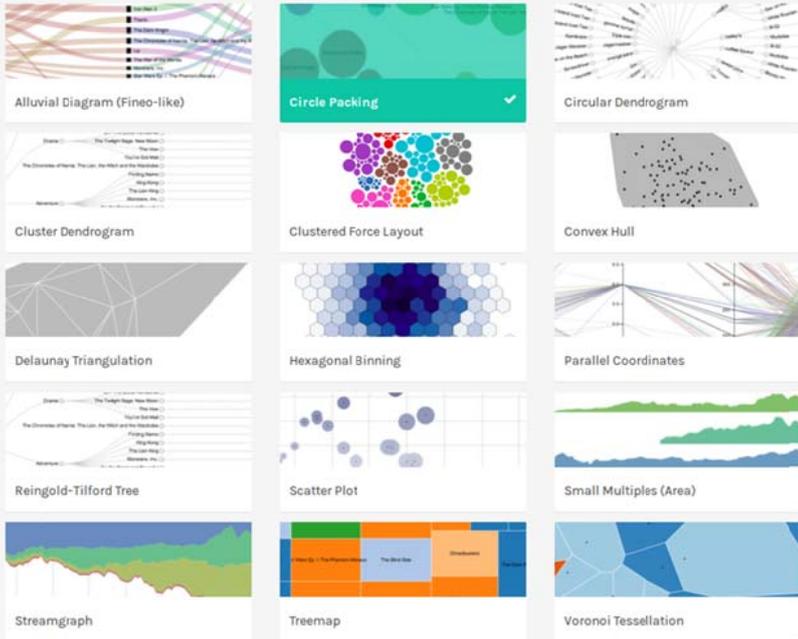
Tiene dos tipos de visualización una vez copiados los datos: en modo texto y en modo tabla. Además el sistema indicará si la importación se realizó correctamente.

- **Configuración y adaptación:**

Choose a Chart

Circle Packing

Nested circles allow to represent hierarchies and compare values. This visualization is particularly effective to show the proportion between elements through their areas and their position inside a hierarchical structure.
Based on <http://bl.ocks.org/mbostock/4063530>



Alluvial Diagram (Fleuro-like)

Circle Packing

Circular Dendrogram

Cluster Dendrogram

Clustered Force Layout

Convex Hull

Delaunay Triangulation

Hexagonal Binning

Parallel Coordinates

Reingold-Tilford Tree

Scatter Plot

Small Multiples (Area)

Streamgraph

Treemap

Voronoi Tessellation

En el siguiente paso podremos seleccionar que tipo de gráfica o *layout* para los datos queremos aplicar.

El sistema nos proporciona una pequeña miniatura de ejemplo del gráfico, un texto explicativo y un enlace para ahondar en el tipo de gráfica.

Map your dimensions

Movie string

Genre string

Production Budget number

Total Domestic Box Office number

Rating IMDB number

HIERARCHY
Drag numbers, strings, dates here

Genre string

SIZE
Drag numbers here

Production Budget number

COLOR
Drag numbers, strings, dates here

Rating IMDB number

LABEL
Drag numbers, strings, dates here

Customize your Visualization

DIAMETER

848

PADDING

5

SORT BY SIZE

COLOR SCALE

Ordinal (categories)

Search...

50	#b1c3e1
63	#b1c3e1
68	#b1c3e1
76	#b1c3e1

Una vez marcada que tipo de gráfica se va a usar es necesario asociar algunas variables del conjunto de datos a los parámetros de la gráfica para su construcción (serie, jerarquía, tamaño, color, etiqueta, tamaño, ...)

Además se indicarán algunos parámetros de estilo adicionales tales como tamaños del gráfico y colores de los elementos a mostrar.

- **Exportación:**

Download

Choose type

Filename

Download

Embed Code

```
<svg xmlns="http://www.w3.org/2000/svg" height="841" width="848"><g transform="translate(10,10)"><circle style="fill-opacity: 0; stroke: rgb(221, 221, 221); stroke-opacity: 1;" r="418.00000000000006" transform="translate(418,418)" class="node node--root"/></circle><circle style="fill: rgb(191, 195, 105); fill-opacity: 1; stroke: rgb(231, 231, 231); stroke-opacity: 0;" r="208.7481356041458" transform="translate(263.1708419064643,556.6298949233396)" class="node node--lna2"/></circle></g></svg>
```

Copy the code above into your HTML to embed the visualization



Este es el paso final antes de poder compartir la visualización. RAW, una vez realizado el trabajo, nos permite descargarlos como imagen, como gráfico vectorial o como modelo de datos en JSON para usarlo en otra aplicación. Además nos permite copiar el código HTML que genera la visualización para pegarlo directamente en otra web o blog.

DATA WRAPPER

DATA WRAPPER es una herramienta web libre y gratuita desarrollada por ABZV (Instituto alemán de formación en periodismo) para realizar visualizaciones de datos. Estas visualizaciones cuentan además con ciertas interacciones que se pueden definir en la herramienta.

Al igual que RAW, el ser libre la convierte en una herramienta altamente configurable y modificable.

- **Inserción de datos:**

1 Upload Data | 2 Check & Describe | 3 Visualize | 4 Publish & Embed

Upload Your Data

Copy & paste your data here

How it works..

We made getting data into Datawrapper as easy as possible. If you're working in Excel or OpenOffice, just select your data (including header row/column) and paste it in the following text field. [Learn more about how to upload your data.](#)

Sample Data

If you simply want to play around with Datawrapper, here's a list of some example datasets:

- Line chart: [Youth Unemployment in Europe](#)
- Line chart: [US Trade with United Kingdom](#)
- Line chart: [Marriages in Germany](#)
- Bar chart: [Fearless Felix: How far did he fall](#)
- Bar chart: [Net borrowing of Germany](#)
- Bar chart (grouped): [Women in German Parliament](#)
- Choropleth map: [Urban population and other country metrics](#)
- Choropleth map: [Unemployment in U.S.](#)

Esta sección está preparada para que copies el contenido de un fichero CSV o directamente de una hoja de cálculo en la caja de texto que aparece en la interfaz. Además permitirá subir directamente un fichero CSV de datos para operar con él.

Muestra en la parte lateral algunos textos de ayuda que pueden venir bien si surgen dudas.

- **Configuración y adaptación:**

The screenshot shows a data visualization tool interface with four steps: 1 Upload Data (checked), 2 Check & Describe (highlighted in red), 3 Visualize, and 4 Publish & Embed. Below the steps is a table with columns A, B, and C. The table contains 13 rows of movie data. To the right of the table is a 'Describe your data' panel with options for 'First row as label' and 'Credit the source', along with input fields for organization name and source link.

	A	B	C
1	Movie	Genre	Production Budget
2	Avatar	Action	425,000,000
3	The Blind Side	Drama	35,000,000
4	The Chronicles of Narnia: The Lion, the Witch and the Wardrobe	Adventure	180,000,000
5	The Dark Knight	Action	185,000,000
6	ET: The Extra-Terrestrial	Drama	10,500,000
7	Finding Nemo	Adventure	94,000,000
8	Ghostbusters	Comedy	30,000,000
9	The Hunger Games	Thriller/Suspense	80,000,000
10	Iron Man 3	Action	200,000,000
11	Jurassic Park	Action	63,000,000
12	King Kong	Adventure	207,000,000
13	The Lion King	Adventure	78,300,000

Con los datos ya cargados es hora de revisar que todo se subió correctamente. Para ello tenemos un visualizador en modo tabla de los mismos.

Además podremos describir los datos indicando si éstos cuentan como cabecera de información en la primera fila y añadiendo a los mismos ciertos datos como créditos de la fuente.

1 Upload Data ✓ 2 Check & Describe ✓ 3 Visualize 4 Publish & Embed

UP rating

Production Budget Total Domestic Box Office Rating IMDB

Movie	Rating
The Dark Knight	9
Raiders of the Lost Ark	8.7
The Lion King	8.4
Up	8.3
Finding Nemo	8.1
Avatar	8
Jurassic Park	8
Monsters, Inc.	8
ET The Extra-Terrestrial	7.9
Ghostbusters	7.8
Iron Man 3	7.6
The Blind Side	7.6
Titanic	7.6

Resize to: 600 x 400 (e.g. 480x320 480x480 640x360 640x480)

1 Select chart 2 Refine the chart 3 Tell the story

Guide the audience to read the chart

TITLE
UP rating

DESCRIPTION

Highlight the most important elements (optional)

- select element -

Up

Where did you get the data from?

SOURCE NAME SOURCE URL

name of the organisation URL of the dataset

Back Publish

En el siguiente paso podremos seleccionar que tipo de gráfica y estilos para los datos queremos aplicar.

Junto a la selección de la gráfica podremos además afinar el resultado de la visualización y añadir campos que nos permitan contar una historia en torno a los datos que se visualizan como es añadir textos y marcar elementos como destacados.

- **Exportación:**

1 Upload Data ✓ 2 Check & Describe ✓ 3 Visualize ✓ 4 Publish & Embed

Link to this visualisation
<http://cf.catawrappier.de/JVaTR/1/>

Embed into your website

Just copy & paste the following code into your website.

Width: 600 px Height: 400 px

```
<iframe src="http://cf.catawrappier.de/JVaTR/1/"
frameborder="0" allowtransparency="true"
allowfullscreen="allowfullscreen"
webkitallowfullscreen="webkitallowfullscreen"
mozallowfullscreen="mozallowfullscreen">
```

Other things to do

+ Duplicate this chart

Show in Gallery

UP rating

Production Budget Total Domestic Box Office Rating IMDB

Movie	Rating
The Dark Knight	9
Raiders of the Lost Ark	8.7
The Lion King	8.4
Up	8.3
Finding Nemo	8.1
Avatar	8
Jurassic Park	8
Monsters, Inc.	8
ET: The Extra-Terrestrial	7.9
Ghostbusters	7.8
Iron Man 3	7.6
The Blind Side	7.6
Titanic	7.6

Note: It may take a couple of minutes until the re-direction from old chart uris to the newest version takes effect.



En este paso final se nos permite enlazar directamente a la visualización en el sistema compartiendo en enlace que se muestra. También nos permite copiar el código HTML que genera la visualización para pegarlo directamente en otra web o blog.

Por último permite crear otra visualización duplicando la actual, facilitando así el trabajo al usuario.



5. Construye un mapa con datos

Desde hace bastante tiempo y originariamente en el área de la navegación, el estudio de la ciencia cartográfica ha fascinado a la humanidad y ha facilitado el descubrimiento de nuevas áreas, técnicas, mundos, situaciones, etc.

Debido al complejo campo que se nos abre enunciando el concepto “ciencia cartográfica” hay que matizar que en este apartado intentaremos concretar cuáles son los conceptos básicos que se necesitan conocer a la hora de trabajar, de forma somera, con mapas.

Terminología básica:

- **Mapa:** Representación gráfica y métrica de una porción de territorio normalmente en una superficie plana. El concepto métrico añadido a un mapa nos permite calcular distancias o superficies, tal y como se realiza en el mundo real.
- **Proyección cartográfica:** Sistema que se encarga de definir la relación existente entre los puntos de la Tierra (superficie curva) y el mapa (superficie plana). Dichos puntos son conseguidos a través de una red de meridianos y paralelos en forma de malla.
- **Latitud y Longitud:** Valores (x,y) definidos en el eje cartesianos definido por los meridianos y paralelos de la Tierra. Esto son, valores en la malla plana que se define de la proyección del sistema curvo de la Tierra. (Estos son los puntos que se usarán como medidas representativas a la hora de colocar datos en los mapas)
- **Sistemas de información geográfica:** También conocidos como SIG (en español) o GIS (en inglés). Integran un conjunto de herramientas de diverso tipo para trabajar con datos provenientes del mundo real y poder representarlos de forma gráfica y métrica en mapas.

La cartografía puede ser muy compleja (realmente todo lo que queramos) pero vamos a intentar simplificar ciertos conceptos, posiblemente perdiendo rigor científico, con el fin de acercar al común el trabajo con mapas y la obtención de resultados en un tiempo de trabajo razonable.

Hay una herramienta que ha ayudado a democratizar el trabajo con mapas de forma notable. Esta es Google Maps [maps.google.com]. Desde su lanzamiento en 2005 ha evolucionado bastante y ha permitido tanto consumir como producir contenido relacionado con mapas, desde geolocalizar un negocio hasta encontrar rutas para llegar desde un origen hasta un destino.



En tal caso existen otras plataformas de acceso más libre y abierto a los datos. La referencia mundial principal es OpenStreetMap [openstreetmap.org]. Esta plataforma ha permitido que comunidades de ciudadanía activa entusiasmadas por los sistemas cartográficos puedan crear sistemas tecnológicos con conocimiento cartográfico solo al alcance de grandes corporaciones (como es el caso de Google).

Para sintetizar todos estos conceptos en un conocimiento práctico que nos permita realizar trabajos sencillos de geoposicionar datos en mapas vamos a usar una herramienta llamada TIME MAPPER desarrollada por la *Open Knowledge Foundation* [timemapper.okfnlabs.org] y basada en una herramienta llamada TimelineJS desarrollada por el KnightLab [timeline.knightlab.com]

TIME MAPPER es una herramienta web libre y gratuita desarrollada por la OKFN para servir de puente entre hojas de cálculo y líneas de tiempo geolocalizadas en un mapa.

Esta herramienta cubre dos áreas importantes y nos permite crear historias de forma muy directa: por un lado las líneas temporales para ordenar eventos en el tiempo y por otro la geolocalización de los mismos en un mapa para tener constancia exactamente de dónde se producen.

El funcionamiento de TIME MAPPER puede ser descrito mediante tres sencillos pasos:

- Crear hoja de cálculo y publicarla.
 - Integrarla en TimeMapper y configurarla.
 - Publicar y obtener código embebido que compartir.
-
- **Crear hoja de cálculo:**



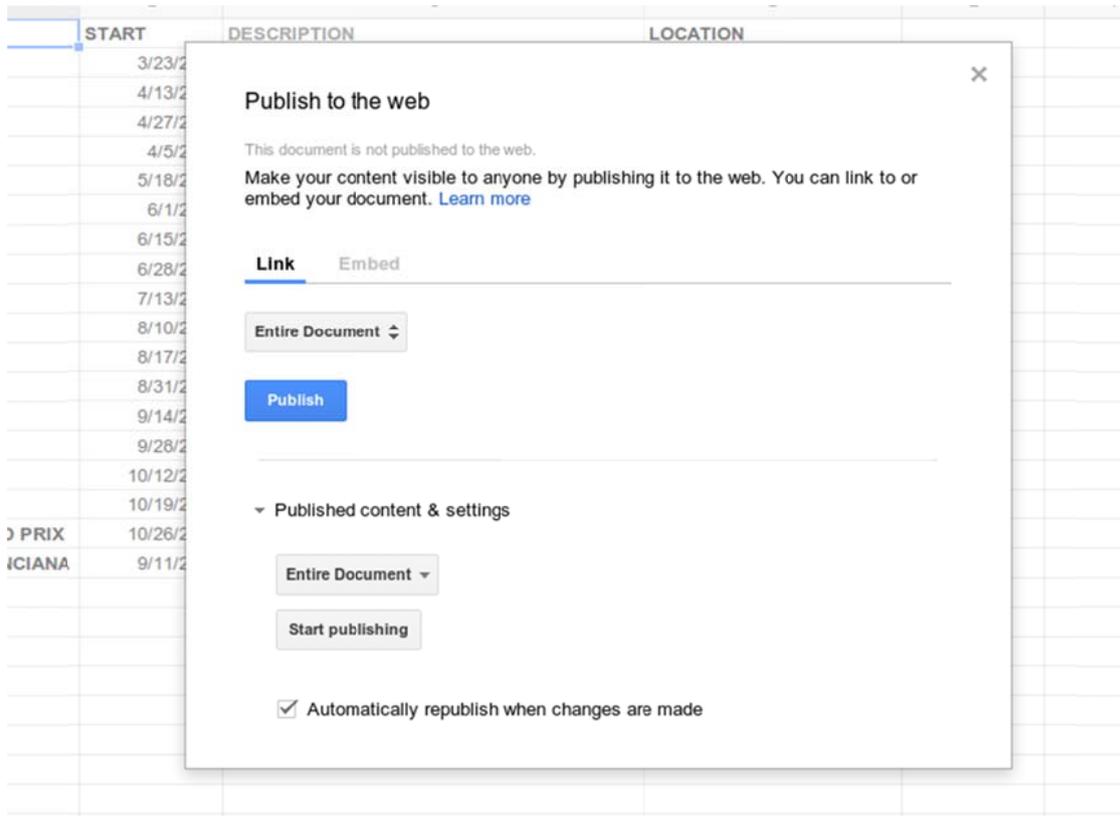
MotoGP Calendar and Places

TITLE	START	DESCRIPTION	LOCATION
COMMERCIAL BANK GRAND PRIX OF QATAR	3/23/2014	Losail Circuit Qatar	25.490581, 51.453963
RED BULL GRAND PRIX OF THE AMERICAS	4/13/2014	Circuit of the Americas Austin	30.138486, -97.635291
G.P. RED BULL DE LA REPÚBLICA ARGENTINA	4/27/2014	Autódromo Termas de Río Hondo Argentina	-27.494249, -64.857788
GRAN PREMIO bwin DE ESPAÑA	4/5/2014	Jerez Spain	36.683842, -6.126193
MONSTER ENERGY GRAND PRIX DE FRANCE	5/18/2014	Le Mans France	48.00611, 0.199556
GRAN PREMIO D'ITALIA TIM	6/1/2014	Mugello Italy	43.95, 11.4
GRAN PREMI MONSTER ENERGY DE CATALUNYA	6/15/2014	Circuit de Barcelona-Catalunya Catalunya	41.422533, 1.958677
IVECO DAILY TT ASSEN	6/28/2014	Assen Netherlands	52.992753, 6.564228
enl MOTORRAD GRAND PRIX DEUTSCHLAND	7/13/2014	Sachsenring Germany	50.924246, 6.951264
RED BULL INDIANAPOLIS GRAND PRIX	8/10/2014	Indianapolis Motor Speedway Indianapolis	39.716859, -86.295595
bwin GRAND PRIX ČESKÉ REPUBLIKY	8/17/2014	Automotodrom Brno Czech Republic	49.19506, 16.606837
HERTZ BRITISH GRAND PRIX	8/31/2014	Silverstone Great Britain	52.091477, -1.028061
GP TIM DI SAN MARINO E DELLA RIVIERA DI RIMINI	9/14/2014	Misano World Circuit San Marino	43.94236, 12.457777
GRAN PREMIO MOVISTAR DE ARAGÓN	9/28/2014	MotorLand Aragon Aragon	41.080304, -0.212602
MOTUL GRAND PRIX OF JAPAN	10/12/2014	Motegi Japan	36.532867, 140.183412
TISSOT AUSTRALIAN GRAND PRIX	10/19/2014	Phillip Island Australia	-38.48987, 145.203828
SHELL ADVANCE MALAYSIAN MOTORCYCLE GRAND PRIX	10/26/2014	Sepang Circuit Malaysia	2.759644, 101.733134
GRAN PREMIO GENERALI DE LA COMUNITAT VALENCIANA	9/11/2014	Comunitat Valenciana Valencia	39.404145, -0.334643

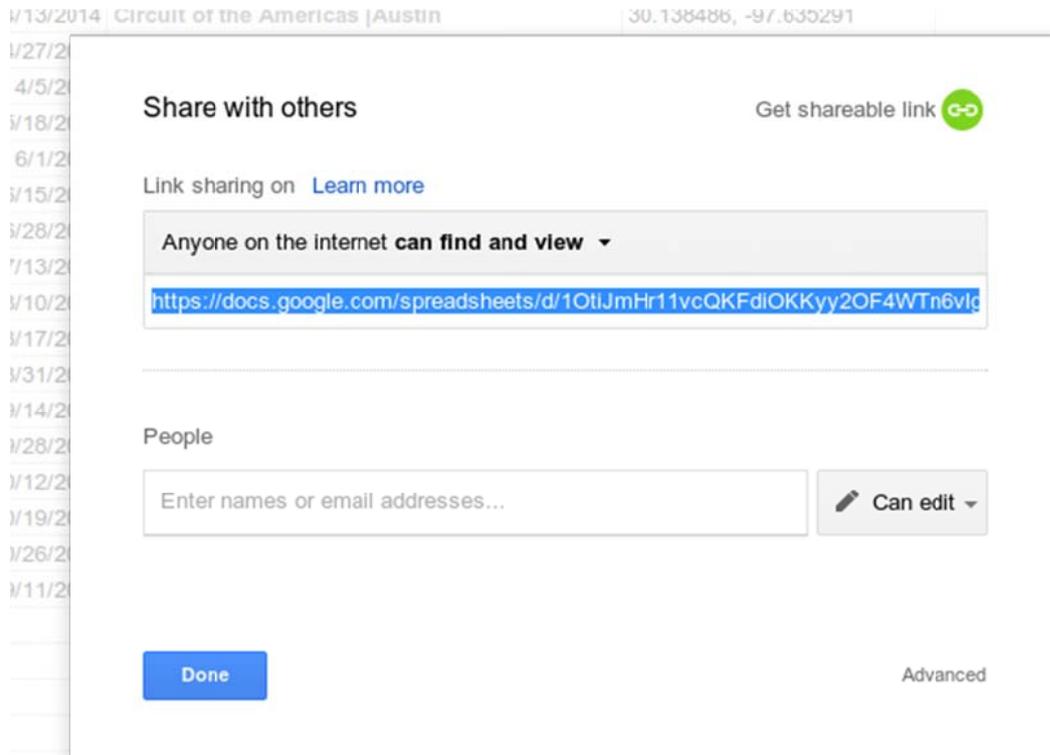
Debemos crear una hoja de cálculo que contenga los títulos de los acontecimientos, la fecha de inicio (y/o fin) en la que se produjo y la localización (latitud y longitud) del mismo.

Se hará uso para ello de Google Apps.

Para obtener la latitud y longitud de un lugar dado se puede usar www.latlong.net



Una vez insertado los datos en la hoja de cálculo de Google Apps hay que hacerla pública para que Time Mapper pueda acceder a los datos.
Para ello pulsamos en *Archivo > Publicar en la web*



Estando público el documento en Internet solo nos queda copiar el enlace público del documento. Para ello pulsamos sobre el botón *Compartir* (esquina superior derecha) y encontramos el enlace. Lo copiamos para a continuación pegarlo en Time Mapper para que éste acceda a los datos.

- **Integración y configuración:**



2. Connect and Customize

ALERT: you are **not signed in** so your timemap will be created '**anonymously**'.

If want to 'own' your timemap you should [sign in \(or sign-up\) now >>](#)
(Sign-up takes a few seconds with your [twitter account >>](#))

Find out more on anonymous vs logged in - [read FAQ below >>](#)

Data Source

 Select from Your Google Drive

(If nothing happens check you are not blocking popups ...)

<https://docs.google.com/spreadsheets/d/1OtiJmHr11vcQKFdiOKKy2OF4WTn6vlgx8>

Important: you must "publish" your Google Spreadsheet: go to File Menu in your spreadsheet, then 'Publish to the Web', then click 'Start Publishing'. See [the FAQ below](#) for more details

Title

MotoGP Calendar and Places

Accedemos a la dirección web de Time Mapper [timemapper.okfnlabs.org] y pulsamos en el botón azul de *comenzar ahora*.

Nos aparece un formulario con diferentes pasos. En el primero debemos de indicar dónde se encuentra el conjunto de datos. Pegamos la dirección que habíamos copiado previamente y listo.



Title

Type of Data View

- Timemap
- Timeline
- Map

Choose the visualization type of your data - TimeMap (Timeline and Map combined), Timeline or Map.

More Options

Ambiguous Date Handling

- mm/dd/yyyy - month first (US style)
- dd/mm/yyyy - day first (non-US style)

How to handle ambiguous dates like "05/08/2012" in source data (could be read as 5th August or 8th of May).
If you do not have any dates formatted like this then you can ignore this!

Start from

- First event
- Last event
- Today

Where on the timeline should the user start.

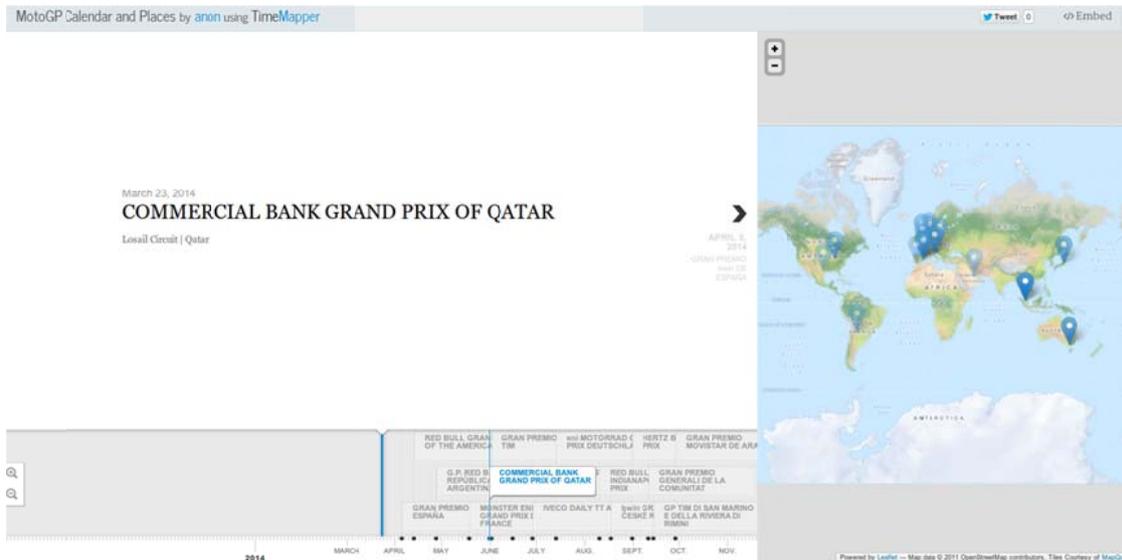
3. Let's Publish It!

[Publish »](#)

Una vez insertado los datos toca realizar algunas configuraciones. Para ello indicamos el Título de la visualización, seleccionamos *timemap* como tipo de vista, marcamos el formato de fecha y si queremos que los datos se visualicen de inicio a fin o al revés.

Una vez tengamos realizada la configuración se pulsa en *Publicar*.

- **Publicar y exportar:**



Ya tenemos creada la visualización de datos, tanto en el mapa (parte derecha) como en su línea de tiempo (parte inferior). Además se puede indicar título y descripción de cada acontecimiento para mostrarlo en la caja informativa.

Dicha url ya se puede compartir pero si queremos obtener el código embebido pulsaremos en el botón `</>Embed`

Embed Instructions

Copy and paste the following into your web page

```
<iframe src="http://timemapper.okfnlabs.org/anon/kjz1mo-motogp-calendar-and-places?embed=1" frameborder="0" style="border: none;" width="100%" height="780;"></iframe>
```

Simplemente copiando este código y pegándolo como código HTML en un blog o web ya tendríamos la visualización exportada.



6. Insertar visualizaciones en otras webs

Una vez realizado todo el trabajo será necesario publicarlo. Se pueden considerar dos tipos de formatos según lo que hemos visto:

- **Imagen:** Son relativamente fáciles de integrar en una web con independencia de su formato de fichero (JPG, PNG, SVG, ...). Dependiendo del sistema que estemos usando para publicar la noticia en la que se incluirá dicha visualización puede modificarse la nomenclatura pero buscando una opción de “añadir imagen” o similar tendríamos el trabajo finalizado.
- **Código embebido:** El código embebido siempre se encuentra en formato HTML (lenguaje básico de la Web). La mayoría de los sistemas de gestión de contenidos o de blogs (Drupal, Wordpress, Mezzanine o sistemas propietarios) nos permiten integrar dicho código embebido como parte de la noticia que estemos editando por lo que copiando el código embebido y pegándolo en el cuerpo de la noticia se tendría el trabajo finalizado.



IDEAS CLAVE

El **periodismo de datos** plantea dos solicitudes básicas en una era de saturación de información:

- El trabajo con los datos en sí para sacarle todo el partido mediante técnicas informáticas.
- La mejor comprensión de los mismos basándonos para ello en representaciones visuales.

En esta línea habrá que seguir trabajando y surgirán nuevos conceptos, técnicas o herramientas que hagan de esta disciplina un sistema en continua evolución.



REFERENCIAS

- Introducing Tabula - <https://source.opennews.org/en-US/articles/introducing-tabula/>
- “Crowdsourcing” en Wikipedia - <https://en.wikipedia.org/wiki/Crowdsourcing>
- “Cartografía” en Wikipedia - <https://es.wikipedia.org/wiki/Cartograf%C3%ADa>
- “Mapa” en Wikipedia - <https://es.wikipedia.org/wiki/Mapa>
- “Proyección cartográfica” en Wikipedia - https://es.wikipedia.org/wiki/Proyecci%C3%B3n_cartogr%C3%A1fica
- “Sistema de Información Cartográfica” en Wikipedia - https://es.wikipedia.org/wiki/Sistema_de_informaci%C3%B3n_geogr%C3%A1fica
- “Google Spreadsheets” - <https://docs.google.com/spreadsheets/u/0/>