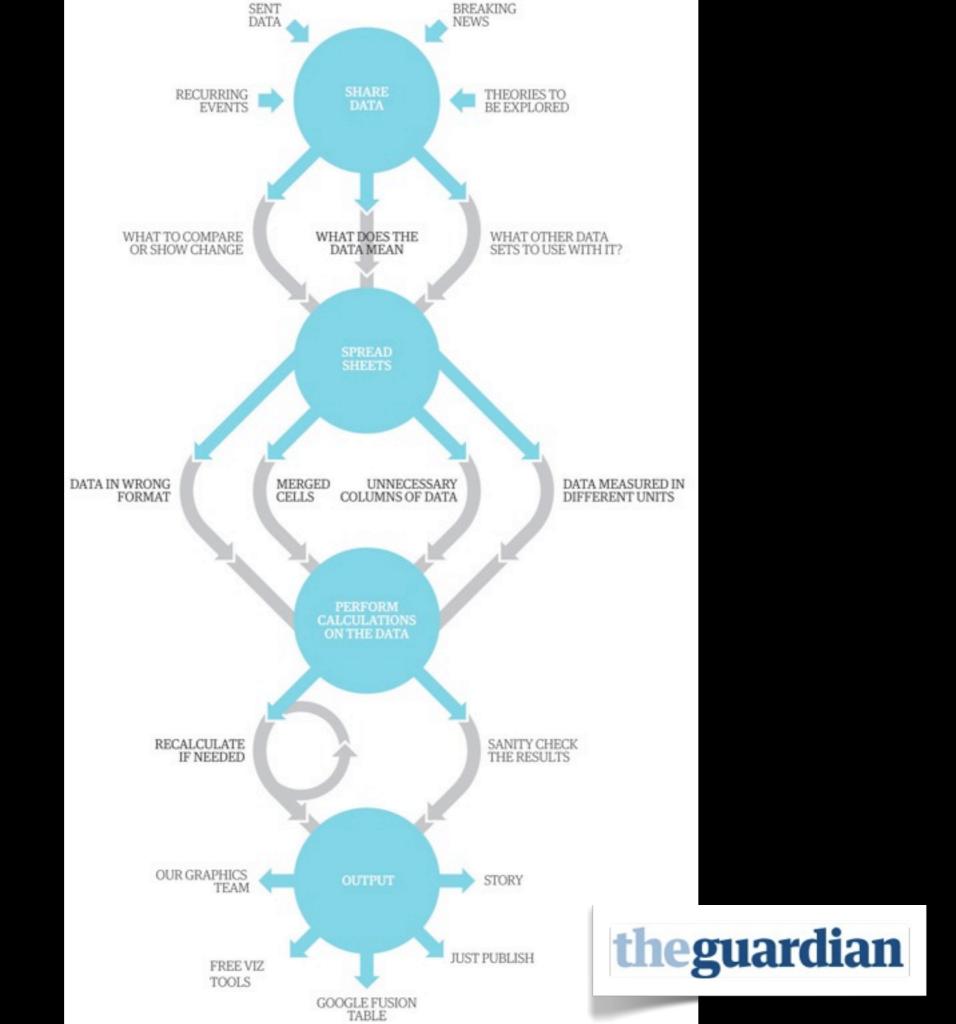
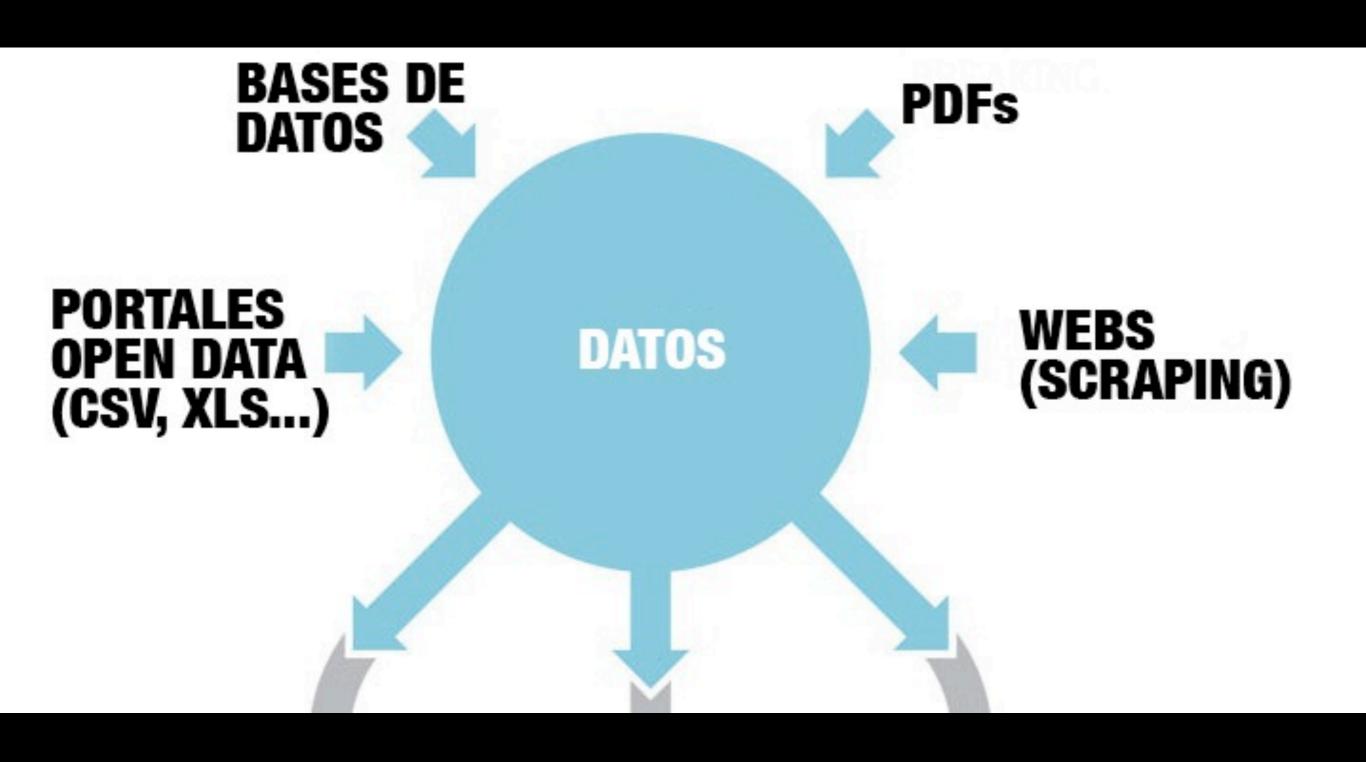
# Extrayendo datos de PDFs (sin programar)

David Cabo (@dcabo)





#### Imagen escaneada



## Generado electrónicamente (buscable)

#### **Tablas**

DEPUTE CHIN PROVENCIAL DE SEVELLA	is the total of the accomplete, to perform the process asserted in a separate little and the latest assertation on the latest assertation on the latest assertation and the latest asse
the contraction can be disposed on all sections 74.7 p. de	Stories Institution
It Deputits Advois Developme breaked revisions	Applicate Strage Vision
per la loy diffe, or let un regio, de babble, de la cosp little, or I de disti, fragiladore de las financias fragisco (cost, or pre-	Cop-Tade Spinol-Faltin Follow J. Ryson
And the final party of the party of an extension of posterior party.	C-1 Regions at a decidado que mater.
folio per la Cinettia Ferincatio, la Cape Stadiones	detailed from an object of the factor of the factor
Assignment Meligum in Coper front in a light	P-Athleto Designite Crescell
some Manipula Articles, or the Manipula spiritude pair in Reachignites for in Franklinean de la Compressión sens.	Determine the section (A)
1 Tolor, as Automotive	Autor Servicia Mesopali
976/0-1	their late lage a homeochilded are it support to
Registro & controls	non-Processes.
Makes district	After the street on a serie is congress.
If any or receive a fatality is produced in	In processors reported to the control beginning to the
the upon, 150° Eap 1961, do 1 do abot y 16° y up. dol 8.31. 1,100 No. 18 do accompany, page historia salar, pagents have	100
at the complete of the product of the contract	Clark Milligene Change Contract
North Hand.	2 - Agree & Companion
And the Bellipse Collins Copy Chair . Stands . Synth Petron Store	A - Sometime Tox
Copy (Note Reports Retail Police RCP)	1 - Strottment French / States File
de	Pathodylogist, 5-17. Non-season (8)75
t - Papero di actividate per regitto	Otto Mice Carbocal
Пакон ето В но сверчи з Провеля Рочноск.	3 - Dischae's desire Co LWCP core; a Water name.
Control Surgery Congr.	
Antichation Montpe	Testina colo cida a prima
Party to have a consequently on all stage of Party	One desp Earl Of St.
an France Sc	1 - Jerothante Colle - Totacin olgonide I
After a polymer condition to proper to the	Table 1.80 curts
and Department of the	Ones desta (Constr.C)
to believe a best believe once observed	No Careta, Indoor, Agenton, places in presions you
Class th highes Street	1 - Destinate of Total Seeks 1984
<ol> <li>Region in Auto-participants</li> </ol>	Tricked Resources, Ser Service.
A Brown trace Water	E-matters of territorial etc.
1 Street Section Visional Chicago Section Section 1	Tenachenne Cont
Professionary (M) International (1994) was	C - The industries prophies a Particular to a problem. No
Citie dies Carriers service Londrie-	Character & contacts & opening
B Dec York ( decites / ip 1 8708 speci.	NO.
aptions nation	SET Agent E.HT-open
By Currier, server, Expense, Joseph in personne y who	Toronto a
Rapin.	Wilder I
to Marchaner on Spiral Geograph Colleges Strates September 1 pp. 5-20	Region in obstance
	Million Control Co. Adhard S. 2001 C.
1 - Meethway Van Carry 1 Winney	Madday - 251-363.
Contractor Contractor	to seek, To have held an extension to complete
and functionally and an extended to the	Children To the secretary assessment over consider
to indicate in Lauritation in Security.	THE RANGE OF STREET
to broken a contractor or begunes.	Souther Engineer
DE MENT	Oppose Space Personal State
I. Participano.	To Report to an installer our makes
1. Succeibbre	singles and property figure in factors.
Prigon J.	P. Africk - Teorge's County
Decree 2 Mar-201	Street App. In Albertaille. Pall
Market Williams	Agent Participe Manager
TI PROPERTY AND A THROUGH AS AND RESIDENCE OF	The facility of a boson of billions are of cooper to fit
do are. This has been an in which it from \$100.0	And Francisco

## Imagen escaneada



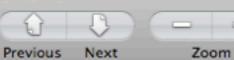
## Generado electrónicamente (buscable)

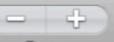
#### Tablas

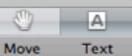


### Tablas

- 1. Copiar y pegar a Excel no funciona bien
- 2. Programas de pago: deskUNPDF, Nitro, Acrobat Pro...
- 3. Servicios web gratuitos:
  - cometdocs.com
  - pdftoexcelonline.com
  - zamzar.com













Sidebar



Search

#### SUBVENCIONES APROBADAS CON CARGO AL FONDO EUROPEO DE PESCA (FEP) ABRIL - JUNIO 2008

Nº EXPEDIENTE	BENEFICIARIO	NIF/CIF	Importe aprobado
151CAND2006	FRANCISCO ORTA RODRIGUEZ	29.777.792-Z	50.000,00
151CAND2007	FERNANDO MARQUEZ RODRIGUEZ	29.779.259-D	50.000,00
151CAND2009	ANTONIO RODRIGUEZ GUZMAN	29.491.982-W	52,500,00
151CAND2010	BLAS MARTIN LAFONT	29.725.240-V	49.976,81
151CAND2011	RAFAEL CANILLA FERNANDEZ	25.699.445-G	52.500,00
151CAND3001	RAFAEL MORALES PEREZ	75.736.236-L	71.454,51
151CAND3002	SEBASTIAN DIAZ DIAZ	24.752.638-S	59,890,58
151CAND3003	EL ALAMI ACHANDIR	1040947-J	52.189,01
151CAND3004	ANTONIO ARIAS PORTALES	24.761.027-D	31.733,74
151CAND3005	JOSE RAMON PEREZ PEREZ	75.737.034-N	48.359,68
151CAND3006	ANTONIO VARO GALLARDO	31.193.091-P	20.004,63
151CAND3007	EL MOKHTAR FARTAN	1299272W	59.998,24
151CAND3010	ANTONIO GUIROLA TAMAYO	31.199.616-R	33.869,77
151CAND3011	ANTONIO MANUEL JUNQUERA ROMERO	31.185.967-Z	14.162,62
151CAND3012	JUAN LUIS DOMINGUEZ QUINTERO	75.736.255-S	9.556,67
151CAND3013	JUAN MARTINEZ SANCHEZ	31.192.243-B	29.565,56
151CAND3014	JOSE RISCART PIPIO	31.312.145-Z	33.097,65
151CAND3016	JOSE MANUEL FLORES CARRASCO	29.467.257-W	9.058,96

#### MEDIDA 2.1: ACUICULTURA

Nº EXPEDIENTE	BENEFICIARIO	NIF/CIF	Importe aprobado
211CAND0001	BLENNIUS, S.L.	F-11311537	37.328,00
211CAND0002	PESQUERIAS LUBIMAR, S.L.	B-84603349	181.014,13
211CAND0003	PISCIFACTORIA ISLA MAYOR, S.A.	A-28029247	622.691,32
211CAND0004	PISCIFACTORIA AGUADULCE, S.L.	B-04301586	311.345,66
212CAND0001	PISCIFACTORIA DE TREBUJENA, S.A.	A-11232824	73.532,05
212CAND0002	CULTIVOS MARINOS DEL SUR, S.A.	A-41134487	12.974,27
212CAND0003	FITOPLANCTON MARINO, S.L.	B-11543667	34.627,00
212CAND0005	ESTEROS DE CANELA, S.A.	A-21168000	142.009,17
212CAND0006	PISCIFACTORIA DE TREBUJENA, S.A.	A-11232824	262.400,78
212CAND0007	ACUINOVA, S.L.	B-29199379	491.642,62
212CAND0008	PREENGORDE DE DORADAS PARA MARICULTURA, S.L.	B-04287975	63.032,92
212CAND0010	LANGOSTINOS DE HUELVA, S.L.	A-10028207	53.001,59

#### MEDIDA 2.3: TRANSFORMACION Y COMERCIALIZACION DE PESCADO

Nº EXPEDIENTE BENEFICIARIO NIF/CIF Importe aprobac	do
--	----



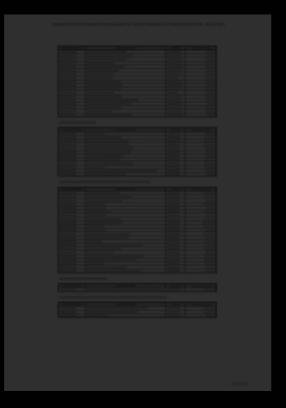


## Imagen escaneada



## Generado electrónicamente (buscable)

#### **Tablas**





Select

Text

#### DIPUTACIÓN PROVINCIAL DE SEVILLA

De conformidad con lo dispuesto en el artículo 75.7 y en la Disposición Adicional Decimoquinta (novedad introducida por la Ley 8/07, de 28 de mayo, de Suelo), de la Ley 7/85, de 2 de abril, Reguladora de las Bases de Régimen Local, se procede a la publicación de las Declaraciones de Intereses formuladas por los Diputados Provinciales, los Cargos Directivos y Funcionarios con Habilitación de Carácter Estatal de la Diputación Provincial de Sevilla, en los términos aprobados por la Resolución de la Presidencia de la Corporación núm. 2.727/07, de 18 de julio:

N°Rgtro.: 1.

Registro de intereses.

Elecciones: 22-Mayo-2011.

Mandato.: 2011/2015.

El presente documento se formaliza en cumplimiento de los arts., 75.7 Ley 7/85, de 2 de abril y 30 y ss. del R.D. 2.568/86, 28 de noviembre, manifestando quien suscribe bajo su responsabilidad, que los datos aportados son ciertos.

Nombre: Manuel.

Apellidos: Rodríguez Guillén.

Cargo Electo ... Diputado ... Partido Político. IULV-CA ...

Registro de actividades que realiza.

Distinta a la de su cargo en la Diputación Provincial.

1\*.— Actividad.— Descripción: Concejal.

Entidad: Pedrera.

Publica

Ambito Territorial: Municipal.

Puede dar lugar a incompatibilidad con el cargo de Diputado Provincial.

Afecta o está relacionado con el ámbito de competencias de la Diputación

Le proporciona o puede proporcionarle ingresos económicos.

Clase de Ingreso

Dietas por asistencia X

II.— Registro de bienes patrimoniales.

2.568/86, 28 de noviembre, manifestando quien suscribe bajo su responsabilidad, que los datos aportados son ciertos.

Nombre: José Leocadio. Apellidos: Ortega Irizo.

Cargo Electo Diputado Partido Político: P. Popular.

Registro de actividades que realiza.

distinta a la de su cargo en la Diputación Provincial.

1\*.— Actividad.— Descripción: Concejal.

Entidad: Ayto. de Pilas

Publica

Ámbito Territorial: Municipal

Puede dar lugar a incompatibilidad con el cargo de Diputado Provincial.

Afecta o está relacionado con el ámbito de competencias de la Diputación

Le proporciona o puede proporcionarle ingresos económicos.

Clase de Ingreso

Dietas por asistencia X

II.— Registro de bienes patrimoniales.

A.— Bienes inmuebles.

Identificación: Vivienda Ubicación: Pilas.

Fecha adquisición: 9- 01

Valor catastral: 34.272,01

Otros datos: Ganancial.

B.— Otros bienes y derechos (+ de 3.000,00 euros).

a) Bienes muebles.

 Identificación: Coche Fecha de adquisición: 2008 Valor: 14.000 euros.

Otros datos: Kia CEE.D.

Z.— Identificación: Coche Fecha de adquisición: 2009 Valor: 3.000 euros.

Otros datos: Citroen C3.

 b) Cuentas, valores, depósitos, planes de pensiones y análogos.

Identificación: c/c

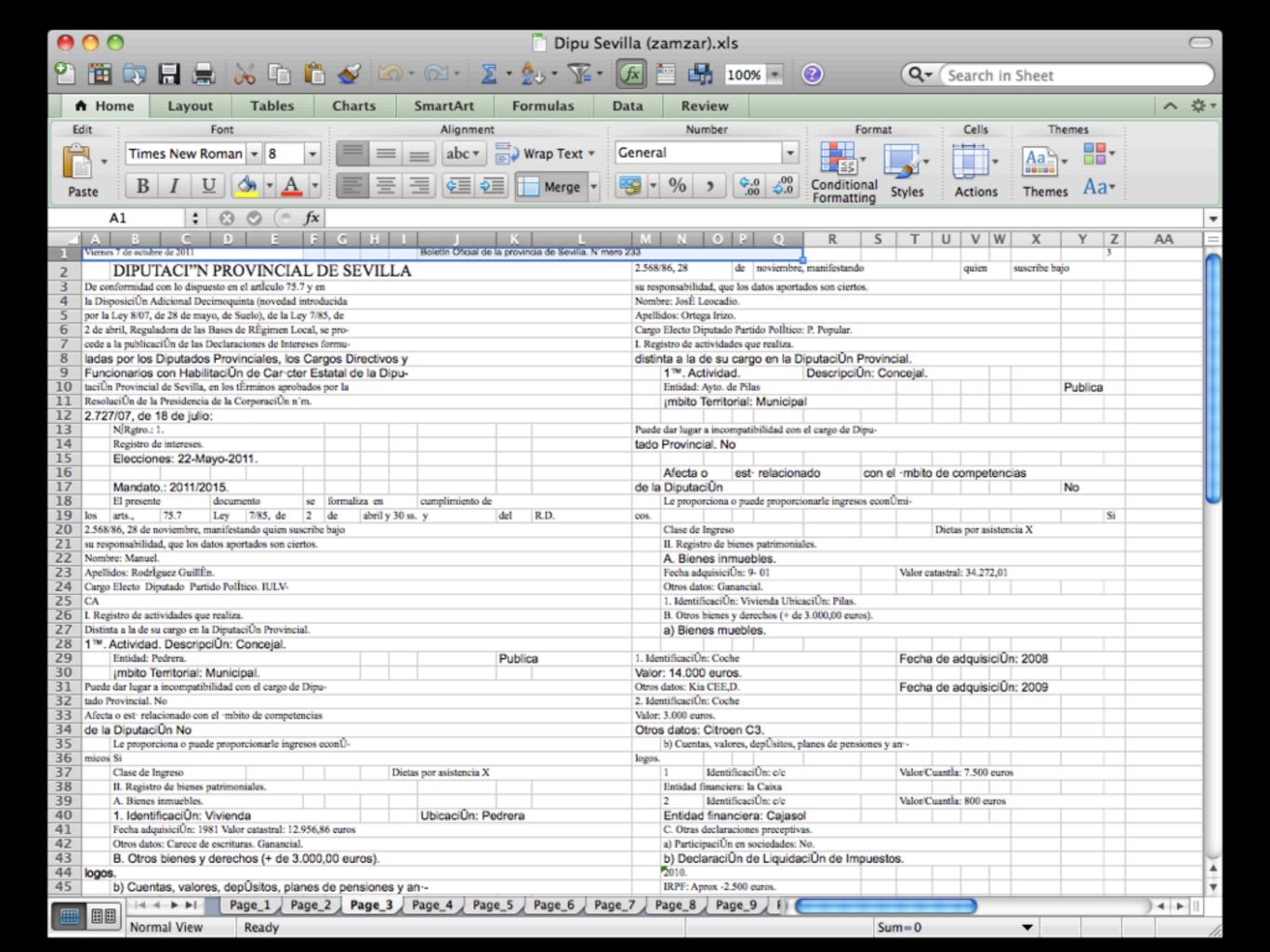
Valor/Cuantía: 7.500 euros

Entidad financiera: la Caixa









- Columnas, datos partidos en varias lineas, datos intercalados entre el texto...
- Las herramientas automáticas no suelen ser capaces de extraer la información
- Es necesario:
  - extraer el texto del PDF (xpdf / poppler)
  - y crear un programa específico (Ruby, Python, Perl... + expresiones regulares)



#### DIPUTACIÓN PROVINCIAL DE SEVILLA

De conformidad con lo dispuesto en el artículo 75.7 y en la Disposición Adicional Decimoquinta (novedad introducida por la Ley 8/07, de 28 de mayo, de Suelo), de la Ley 7/85, de 2 de abril, Reguladora de las Bases de Régimen Local, se procede a la publicación de las Declaraciones de Intereses formuladas por los Diputados Provinciales, los Cargos Directivos y Funcionarios con Habilitación de Carácter Estatal de la Diputación Provincial de Sevilla, en los términos aprobados por la Resolución de la Presidencia de la Corporación núm.-2.727/07, de 18 de julio:

N°Rgtro.: 1.

Registro de intereses.

Elecciones: 22-Mayo-2011.

Mandato.: 2011/2015.

El presente documento se formaliza en cumplimiento de los arts., 75.7 Ley 7/85, de 2 de abril y 30 y ss. del R.D. 2.568/86, 28 de noviembre, manifestando quien suscribe bajo su responsabilidad, que los datos aportados son ciertos.

Nombre: Manuel.-

Apellidos: Rodríguez Guillén.-

Cargo Electo ... Diputado ... Partido Político. IULV-

CA ...

I.-- Registro de actividades que realiza.

Distinta a la de su cargo en la Diputación Provincial.

1<sup>a</sup>.-- Actividad.-- Descripción: Concejal.

Entidad: Pedrera.

Ámbito Territorial: Municipal.

Puede dar lugar a incompatibilidad con el cargo de Diputado Provincial.

Afecta o está relacionado con el ámbito de competencias de la Diputación

Le proporciona o puede proporcionarle ingresos económicos.

Clase de Ingreso

Dietas por asistencia X

Publica

Si

TT -- Pagistro do bionos notrimonialos

2.568/86, 28 de noviembre, manifestando quien suscribe bajosu responsabilidad, que los datos aportados son ciertos.-

Nombre: José Leocadio. Apellidos: Ortega Irizo.-

Cargo Electo Diputado Partido Político: P. Popular.

I.-- Registro de actividades que realiza.-

distinta a la de su cargo en la Diputación Provincial.-

1<sup>a</sup>.-- Actividad.-- Descripción: Concejal.-

Entidad: Ayto. de Pilas

Publica-

Ámbito Territorial: Municipal-

Puede dar lugar a incompatibilidad con el cargo de Diputado Provincial.

No-

No-

Si-

Afecta o está relacionado con el ámbito de competenciasde la Diputación

Le proporciona o puede proporcionarle ingresos económi-cos.

Clase de Ingreso

Dietas por asistencia X-

II.-- Registro de bienes patrimoniales.-

A.-- Bienes inmuebles.-

1.-- Identificación: Vivienda Ubicación: Pilas.-

Fecha adauisición: 9- 01

Valor catastral: 34.272,01

Otros datos: Ganancial.-

B.-- Otros bienes y derechos (+ de 3.000,00 euros).-

a) Bienes muebles.

1.-- Identificación: Coche Fecha de adquisición: 2008-

Valor: 14.000 euros.

Otros datos: Kia CEE,D.-

2.-- Identificación: Coche Fecha de adquisición: 2009-

Valor: 3.000 euros.

Otros datos: Citroen C3.-

b) Cuentas, valores, depósitos, planes de pensiones y aná--

1.-- Identificación: c/c

Valor/Cuantía: 7.500 euros-

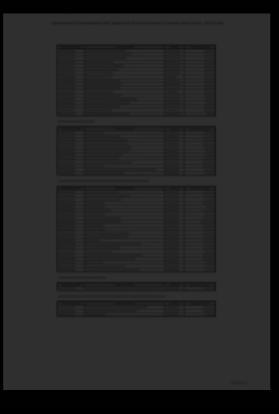
Entidad financiana: la Caiva

## Imagen escaneada



#### Generado electrónicamente (buscable)

#### **Tablas**



## lmágenes

- El texto original no está disponible: no podemos buscar en él, ni copiarlo
- Software para reconocer carácteres (OCR)
- La fiabilidad depende de la calidad de la entrada (limpieza, resolución, tipo de letra)
- Bastante trabajo:
  - Hay que revisar el resultado
  - Difícil de automatizar para muchos ficheros

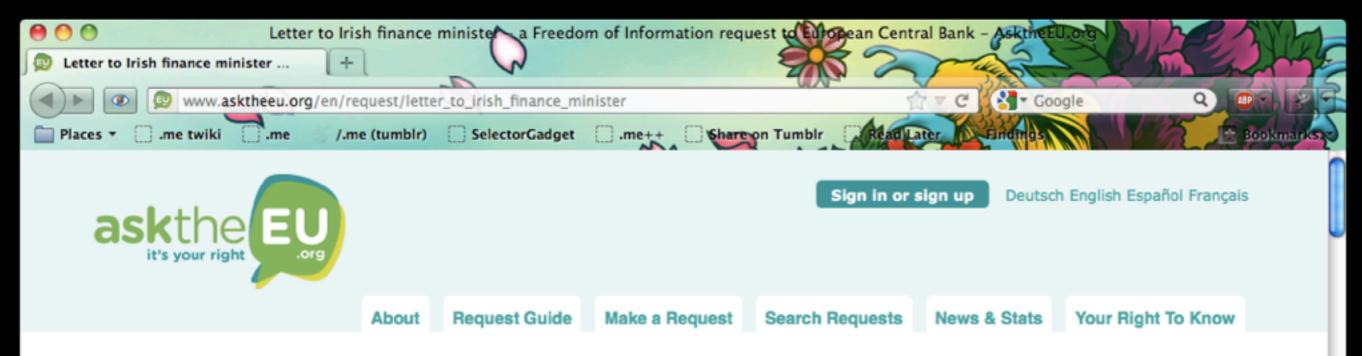
### OCR

#### **Open Source**

- Tesseract + OCROpus
- Usado por Google
- No es fácil de instalar y utilizar

#### **Comerciales**

- AbbyyReader
- Adobe Acrobat Pro
- Solid
- ...



#### Letter to Irish finance minister

Gavin made this access to information request to European Central Bank



Waiting for an internal review by European Central Bank of their handling of this request.

From: Gavin

9 December 2011

Dear European Central Bank,

Under the right of access to documents in the EU treaties, as developed in Regulation 1049/2001, I am requesting documents which contain the following information:

 Any and all communications from the ECB addressed to the Irish Finance Minister (or his direct office) in the month of November 2010.

Yours faithfully,

Cavia Charidan

#### Follow this request

There are 2 people following this request

Follow by email

RSS feed of updates

#### Act on what you've learnt

Tweet this request

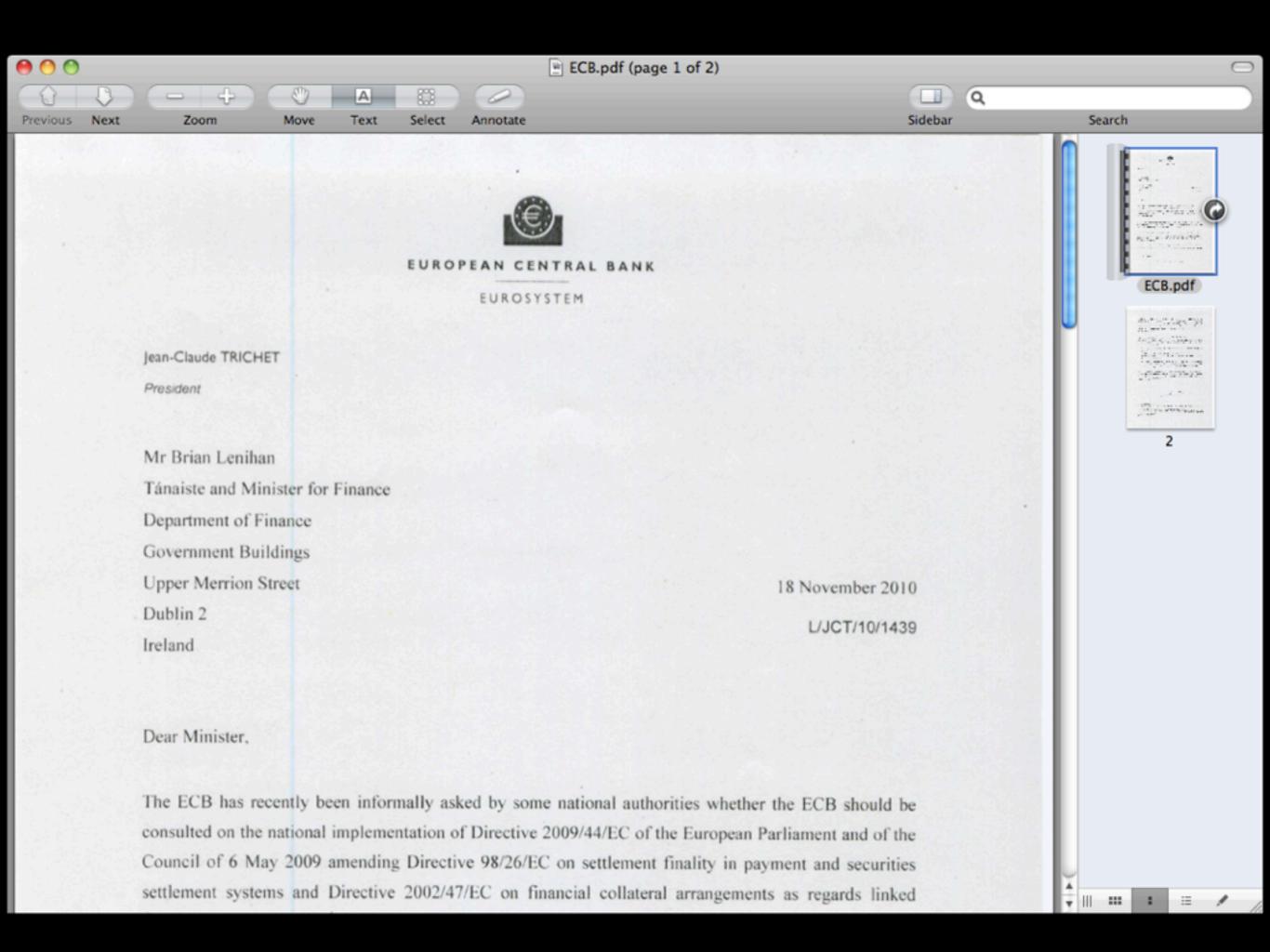
#### Similar requests

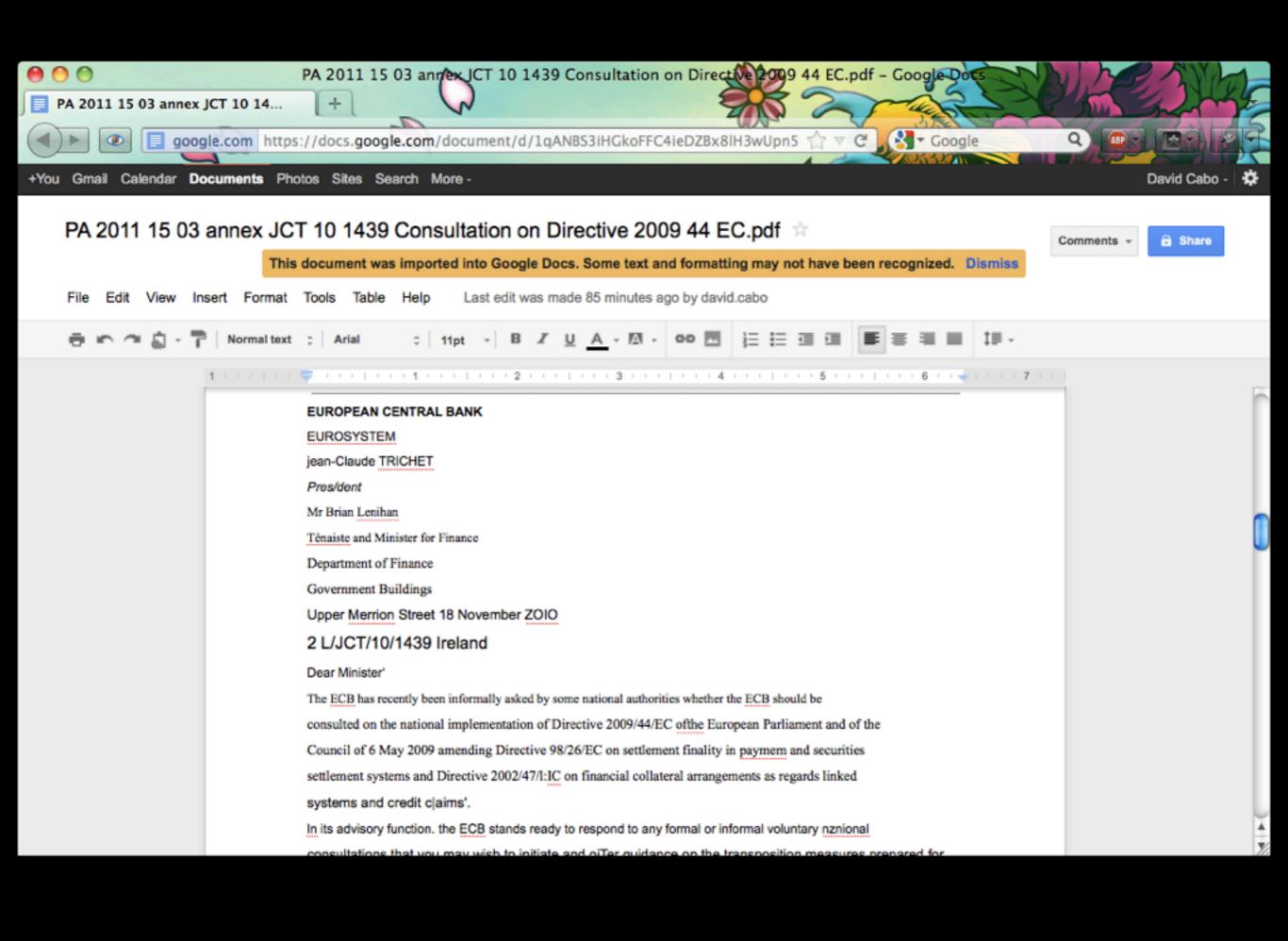
#### NTMA/NAMA communications

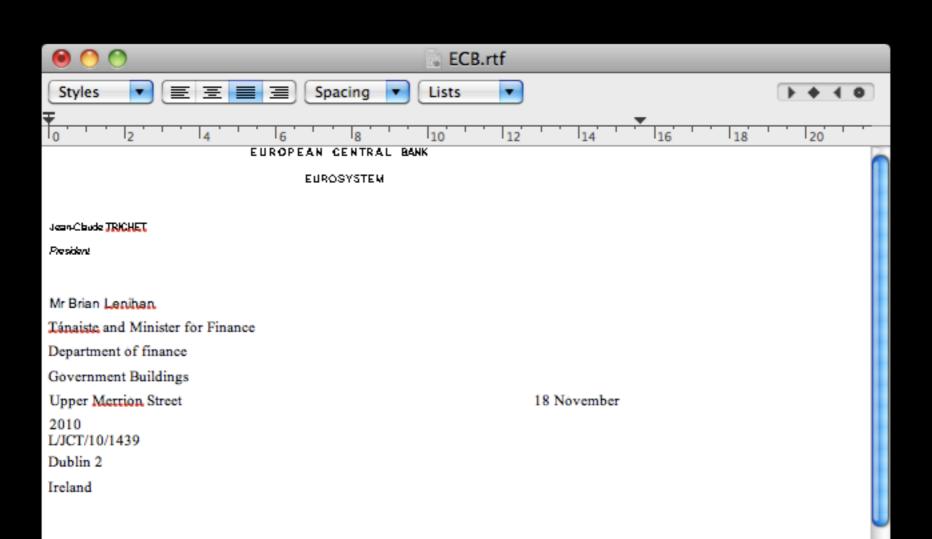
To Competition (COMP) by Gavin 9 December 2011

Communications between directive and Irish Government regardings VHI Insurance

To Competition (COMP) by Mark 15 December 2011







Dear Minister,

The ECB has recently been informally asked by some national authorities whether the ECB should be consulted on the national implementation of Directive 2009/44/EC of the European Parliament and of the Council of 6 May 2009 amending Directive 98/26/EC on settlement thality in payment and securities settlement systems and Directive 2002/47/EC on financial collateral arrangements as regards linked systems and credit claims.

In its advisory function, the ECB stands ready to respond to any formal or informal voluntary national consultations that you may wish to initiate and offer guidance on the transposition measures prepared for Ireland, even if there is no formal obligation to consult, which is the case.

Article l(2) of Council Decision 98/415/EC of 29 June 1998 on the consultation of the European Central Bank by national authorities regarding draft legislative provisions" does not formally oblige national authorities to consult the ECB if the exclusive purpose of draft national legislative provisions is the transposition of EU Directives.

As you know; the ECB seeks to promote proactively a harmonised EU-wide implementation of

**CORTES GENERALES** 



Sidebar

ONGRESO DE LOS DIPUTADOS EGISTRO GENERAL DE ENTRADA





#### DECLARACIÓN DE BIENES Y RENTAS DE DIPUTADOS Y Echa: 07/09/2011-13:27:52

Nombre y apelli ANA MARIA PA		JULIAN			
Estado civil CASADO				I	Régimen económico matrimonial GANANCIALES
Fecha de elecci 09/03/2008	ón con	o parlamen	tario		Fecha de la presentación de la credencial en la Cámara 25/03/2008
Diputado	X	Senador		PONTEVEDRA	cripción por la que ha sido elegido/Asamblea Legislativa

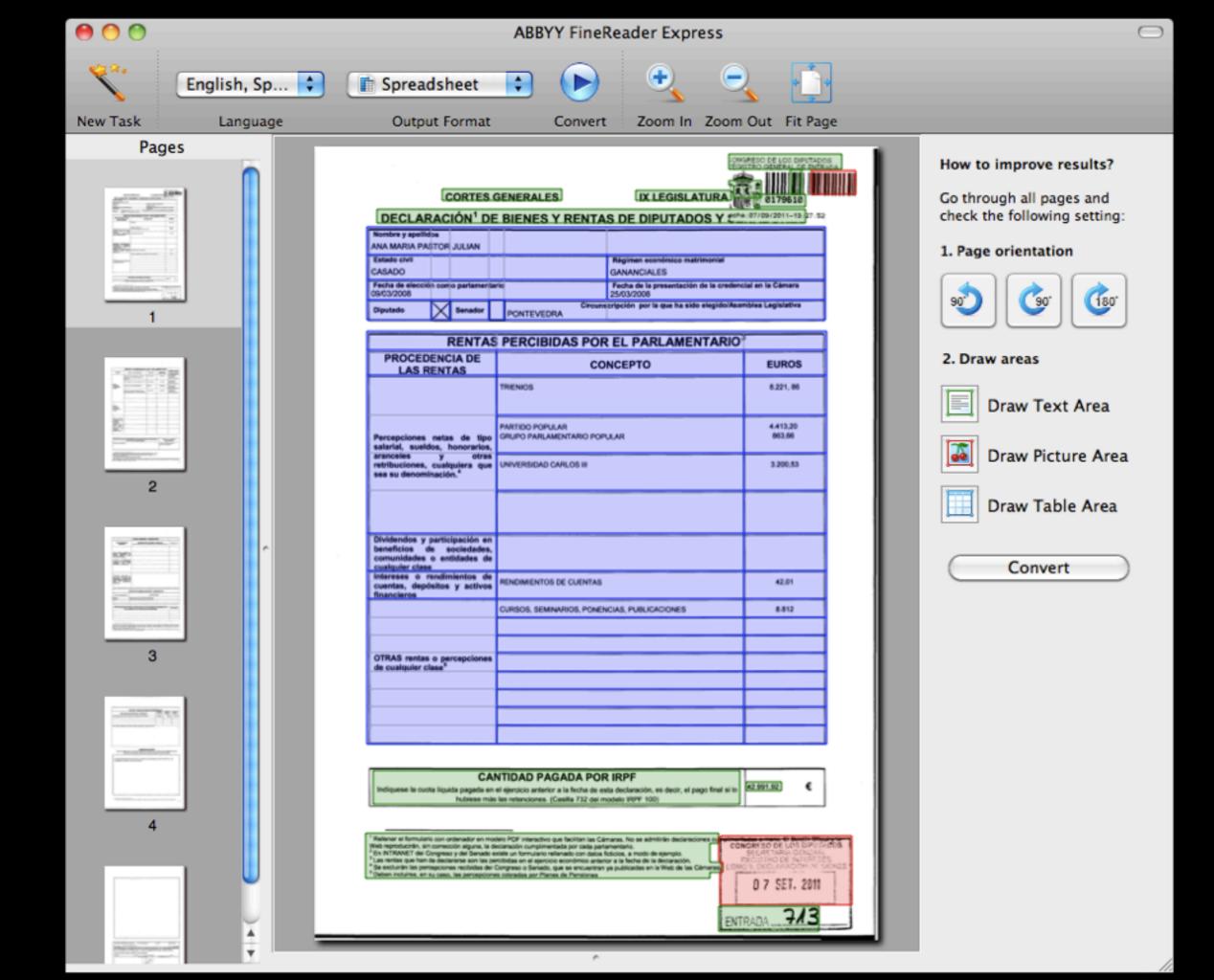
RENTAS PERCIBIDAS POR EL PARLAMENTARIO <sup>3</sup>			
PROCEDENCIA DE LAS RENTAS	CONCEPTO	EUROS	
	TRIENIOS	8.221, 86	
Percepciones netas de tipo salarial, sueldos, honorarios, aranceles y otras retribuciones, cualquiera que sea su denominación. <sup>4</sup>	PARTIDO POPULAR GRUPO PARLAMENTARIO POPULAR	4.413,20 863,66	
	UNIVERSIDAD CARLOS III	3.200,53	
Dividendos y participación en beneficios de sociedades, comunidades o entidades de			

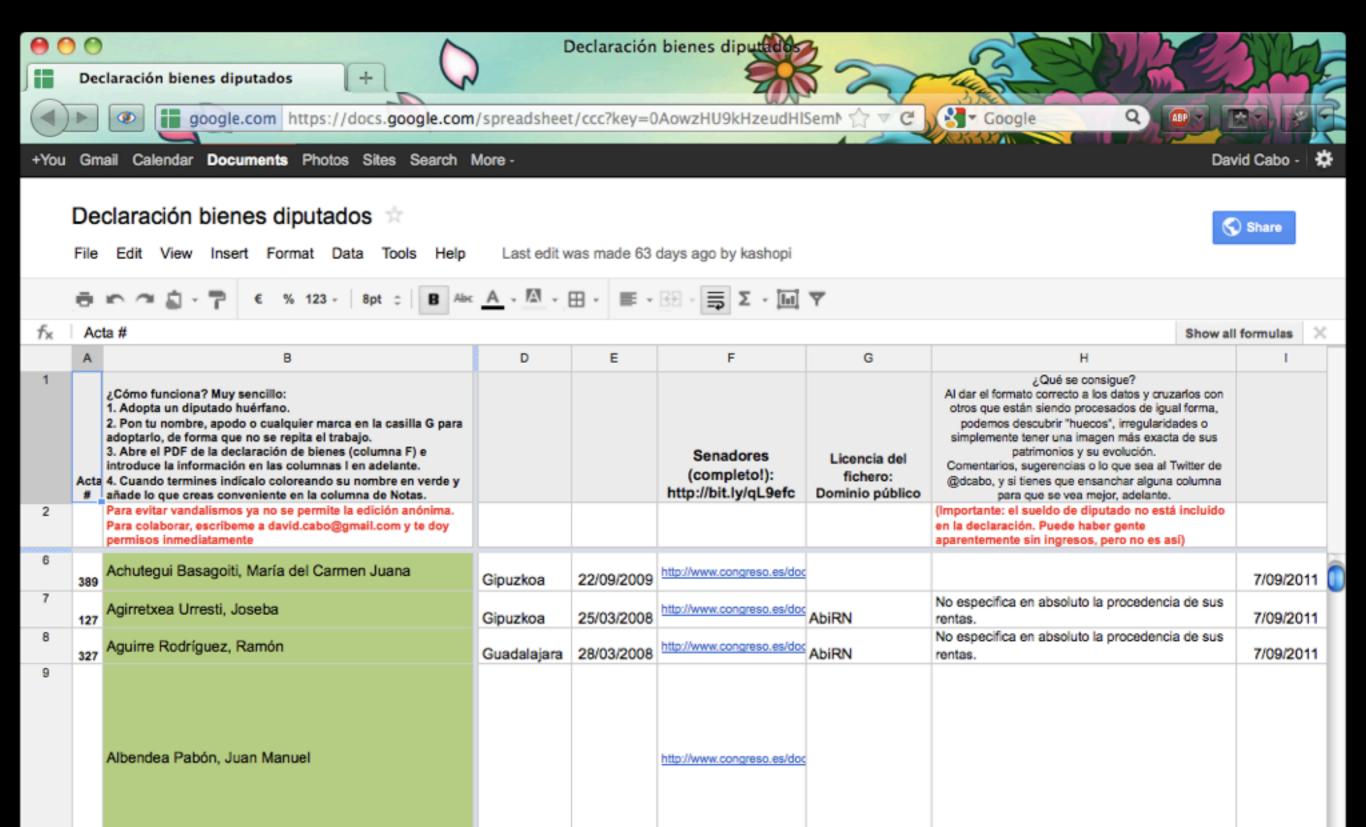












Sevilla

Zaragoza

19/03/2008

25/03/2008

AbiRN

ramontuitin

http://www.congreso.es/doc

Alegria Continente, María Pilar

40

23/08/2011

9/07/2011

) 4 b

## ¿Preguntas?

David Cabo (@dcabo)