

El futuro y el impacto del open data: Mentiras, medias verdades y otros hechos incómodos

desideDatum

**WE MAKE
DATA-DRIVEN
ORGANIZATIONS**



(Open) Data



Transparencia



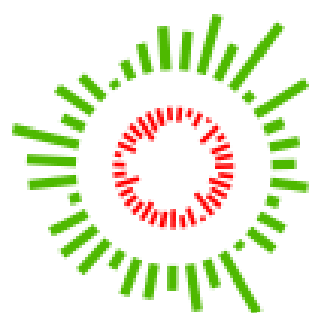
Gobierno Abierto

Implicación ciudadana

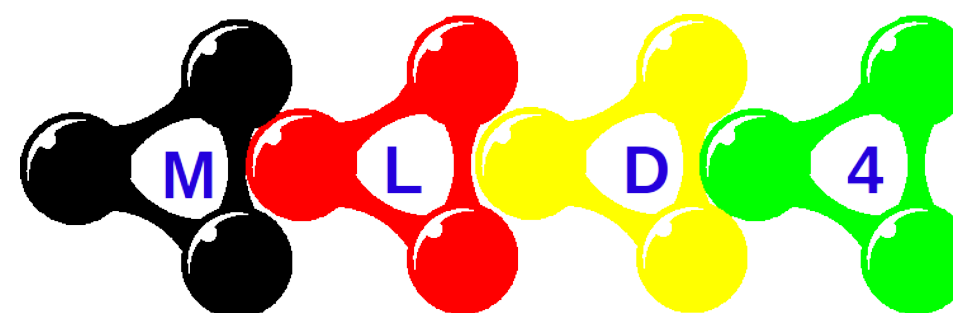


Smart City

El futuro y el impacto del open data: Mentiras, medias verdades y otros hechos incómodos



<http://okfn.es>



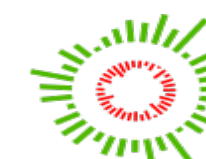
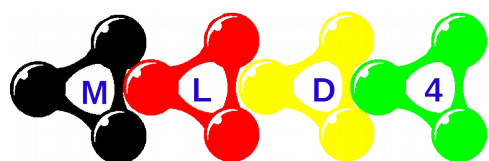
<http://meloda.org>



<http://iniciativabarcelonaopendata.cat>



<http://urjc.es>



Conocemos

Que es

Objetivos

¿Por que open data?

- *Acelerar la investigacion*
- *Mejorar la sociedad*
- *Reducir la corrupción y la evasión fiscal*
- *Crear empleo*

Impacto

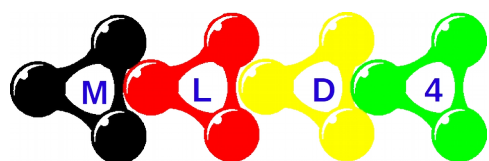
Mentiras

Medias verdades

Hechos incómodos

Futuro

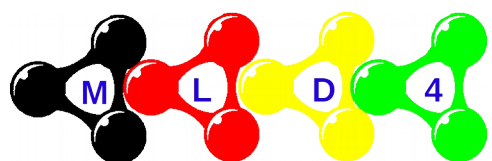
Conclusión



Para tímidos y lejanos

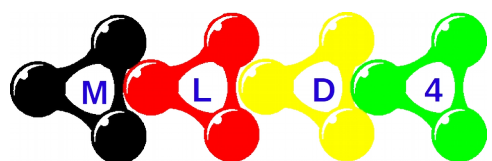
<http://bit.ly/MUGI-opendata>

Twitter polls: @aabella



Minicuestionario

- RISP
- 5 stars
- Licencias Creative commons
- Linked data
- NTI-RISP
- DCAT-AP
- Meloda



¿Cuántos de Vd han utilizado open data (sabiéndolo) ?

¿Para qué?

<http://bit.ly/MUGI-opendata>

¿Cual fue la impresión?



Introducción

¿Chistes sobre Open data?



Haztelo mirar

Que es el Open data

Open data (y contenidos) pueden ser libremente

- utilizados
- modificados,
- compartidos

Por cualquiera

Con cualquier propósito



Open gata

Datos no LPI

No LOPD para no
personales (salvo
tributarios)

Podemos dar
consentimiento a publicar
dato privado

Titularidad del ciudadano

¿Entidad de gestión?

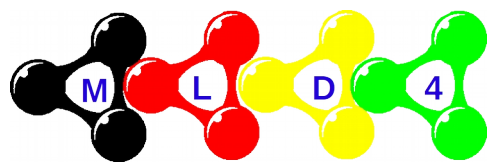
Por defecto open
excepción secretos
oficiales



By Loliloli (Own work) [Public domain], via Wikimedia Commons

Explicación

- RISP
- 5 stars
- Licencias Creative commons
- Linked data
- NTI-RISP
- DCAT-AP
- Meloda



Objetivos

Aprender algo

Tomar decisiones

Llegar al final

<http://bit.ly/MUGI-opendata>

Responder a pregunta P1



By Adrián Cerón (Own work) [CC BY-SA 4.0 (<http://creativecommons.org/licenses/by-sa/4.0>)], via Wikimedia Commons

¿Por que pedimos que publiquen?

I ♥
Taxes

Acelerar la investigación

No repetir
experimentos

Contrastar
resultados

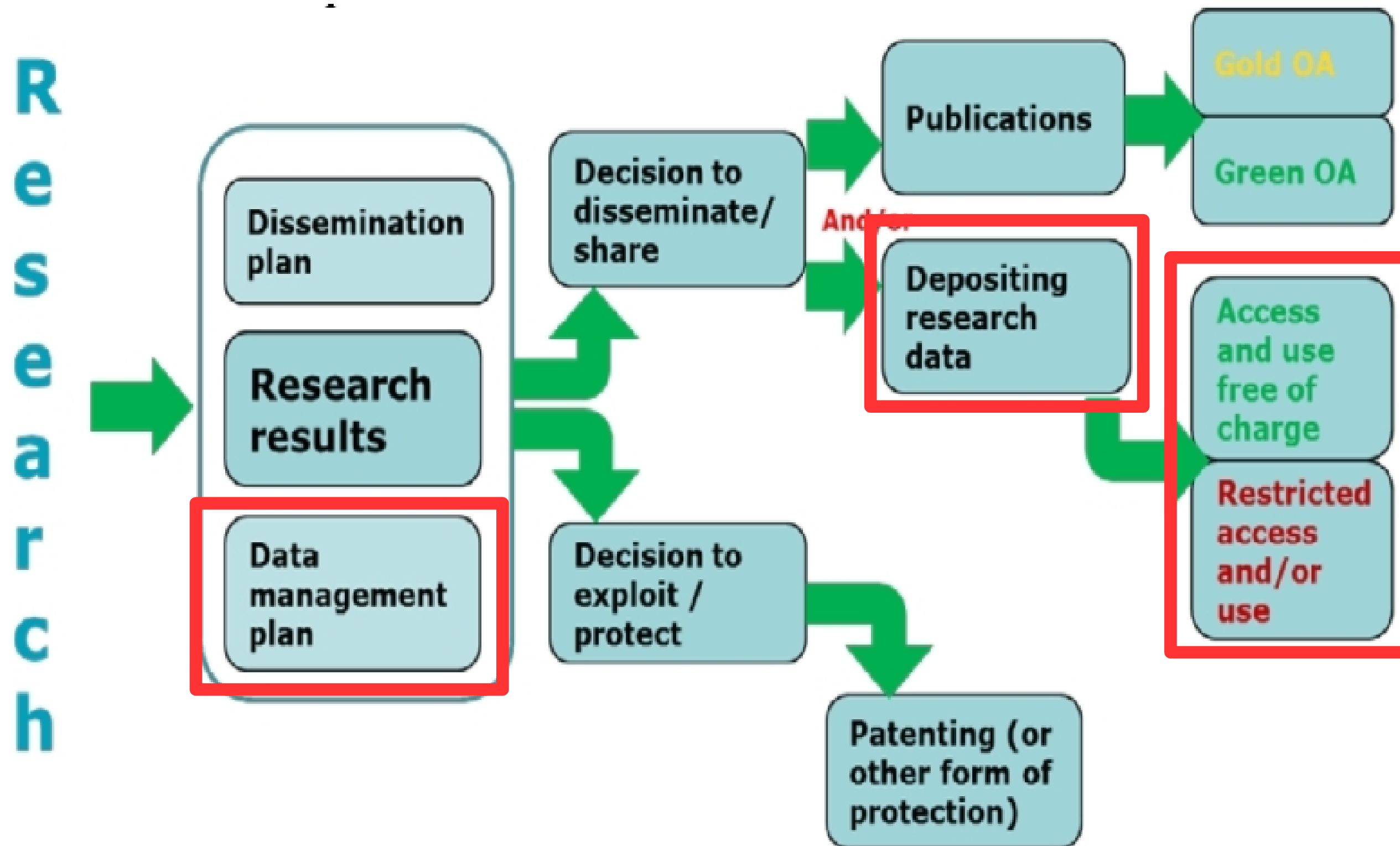
'A hombros de
gigantes'



By Zonda (Own work) [GFDL (<http://www.gnu.org/copyleft/fdl.html>) or CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)], via Wikimedia Commons

Acelerar la investigación

Difusión de resultados de investigación en Horizon 2020.
Voluntario 80.000 millones de €



Fuente: Guidelines on Open access to scientific and Research Data in Horizon 2020. Version 16 Diciembre 2013.

Acelerar la investigación

Financiado en incubadora europea

<http://xpressomics.com>

Buscador de temas genéticos

Normalizar datos existentes open data

Centros de investigación y usuarios

The screenshot shows the Xpressomics website header with the logo and navigation links: About, Pricing, Contact, Sign In, and Sign Up. The main content area features a 'Gene Expression Search Engine' section with the following text: 'Make unexpected discoveries by querying for differential expression from expert-curated and analyzed raw data. Currently available 38,525 profiles from 6,727 experiments'. Below this is a search input field with the placeholder 'Enter gene symbol or keyword' and a 'Search' button. A note below the search field reads: 'Prepend keywords with "#", use "!" to exclude matching profiles. Examples: p53 #cancer, ldh1 #yeast, #hypoxia #human, nqo1 gclc #toxicogenomics!'. At the bottom of the search section, it states: 'Data from: GEO, Array Express, Japanese Toxicogenomic Project, Connectivity Map, DrugMatrix' and 'Algorithm: DEMI'.

Genomic search. Xpressomics

Acelerar la investigación

2015

Ayuda a valorar adecuadamente proyectos investigación científica

Usa bases de datos de Reino unido, EEUU y EU

The screenshot shows the 'My Collaborator' search interface. At the top, there are navigation links: 'My Collaborator', 'Advanced Search', 'Create Budget', and 'FAQs'. A purple button in the top right corner says 'Logout - Alberto Abella García'. The main content area is divided into two columns. The left column is titled 'DATA SOURCE' and has three radio button options: 'Open Data' (selected), 'Closed/Proprietary Data', and 'Mixed Data (Open and Closed/Proprietary Data)'. The right column is titled 'Area of Interest' and has three text input fields for 'Second Keyword (optional)' and 'Third Keyword (optional)'. Below these fields, there is a section titled 'Duration' with a list of radio button options: 'Normalised (1 year)' (checked), '<= 1 year', '> 1 year and <= 2 years', '> 2 years and <= 3 years', '> 3 years and <= 4 years', '> 4 years and <= 5 years', and '> 5 years'. A note below the duration options says 'Refine your search using Duration, Average Funding or Average Cost'.

My collaborator

Apps4EU 2014

GoOV discapacitados mentales en transportes públicos

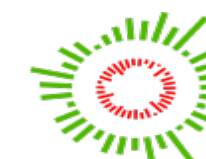
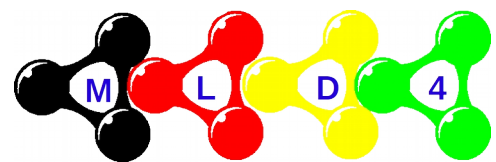
GoOV monitoriza al viajero

GoOV, le da soporte activo o pasivo

El cuidado le ve en su portal.



www.vinkfotografie.nl ©



Mejorar la sociedad

2011

Bikecityguide

Ganador Apps4EU

1500€

Nacido en Graz

300 ciudades

<http://www.bikecitizens.net/>



Los Angeles County

Publicación cualificación restaurantes en 2003

13.1 % descenso ingresos hospitalarios por intoxicación en restaurantes

J Environ Health. 2005 Mar;67(7):32-6, 56; quiz 59-60.

Impact of restaurant hygiene grade cards on foodborne-disease hospitalizations in Los Angeles County.

Simon PA¹, Leslie P, Run G, Jin GZ, Reporter R, Aquirre A, Fielding JE.

⊕ Author information

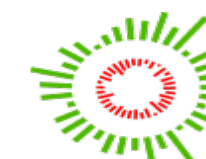
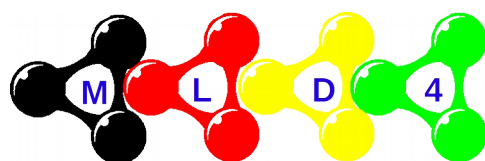
Abstract

Although health departments routinely inspect restaurants to assess compliance with established hygienic standards, few data are available on the effectiveness of these efforts in preventing foodborne disease. The study reported here assessed the impact on foodborne-disease hospitalizations in Los Angeles County of a restaurant hygiene grading system that utilized publicly posted grade cards. The grading system was introduced in January 1998. Hospital discharge data on foodborne-disease hospitalizations were analyzed for Los Angeles County and, as a control, for the rest of California during the period 1993-2000. Ordinary least-squares regression analysis was done to measure the effect of the grading program on these hospitalizations. After baseline temporal and geographic trends were adjusted for, the restaurant hygiene grading program was associated with a 13.1 percent decrease ($p < .01$) in the number of foodborne-disease hospitalizations in Los Angeles County in the year following implementation the program (1998). This decrease was sustained over the next two years (1999-2000). The results suggest that restaurant hygiene grading with public posting of results is an effective intervention for reducing the burden of foodborne disease.

Comment in

The paper "Impact of restaurant hygiene grade cards on food-borne-disease hospitalizations in Los Angeles County. [J Environ Health. 2005]
Impact of restaurant hygiene grade cards on foodborne-disease hospitalizations in Los Angeles County. [J Environ Health. 2005]

Source <http://www.ncbi.nlm.nih.gov/pubmed/15794461>



Reduciendo la corrupción y evasión fiscal

2005-2007 Canada

Publicación datos
desgravaciones fiscales
charities

3200 M\$ canadienses
fraude o posible fraude

eaves.ca

if writing is a muscle, this is my gym

home • about david • public speaking • publications • in the media • get in touch

April 14, 2010

Case Study: How Open data saved Canada \$3.2 Billion

70 Replies

Note: I'll be on TVO's The Agenda with Steve Paikin tonight talking about Government 2.0.

Why does open data matter? Rather than talk in abstract terms, let me share a well documented but little known story about how open data helped expose one of the biggest tax frauds in Canada's history.

It begins in early 2007 when a colleague was asked by a client to do an analysis of the charitable sector in Toronto. Considering it a simply consulting project, my colleague called the Canada Revenue Agency (CRA) and asked for all the 2005 T3010s - the Annual Information Returns where charities disclose to the CRA their charitable receipts and other information - in Toronto. After waiting several weeks and answering a few questions, the CRA passed along the requested information.

After spending time cleaning up the data my colleague eventually had a working excel spreadsheet and began to analyze the charitable sector in the Greater Toronto Area. One afternoon, on a lark, they decided to organize the charities by size of tax-receipted charitable donations.

At this point it is important to understand something about scale. The United Way of Greater Toronto



Follow me on Twitter: @daeaves

follow blog via email

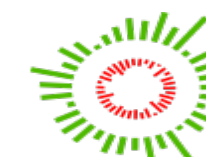
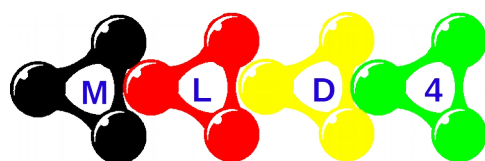
Enter your email address to follow this blog and receive notifications of new posts by email.

Join 12,226 other followers

5 most recent

[Open Mandate Letters](#)

Source <http://eaves.ca/2010/04/14/case-study-open-data-and-the-public-purse/>



Creando empleo

Apps4EU 2014

Crea un desafío en una ubicación turística con QR y otros.

Puede ser generada por usuarios

Tesoros ocultos

<http://geostep.me>



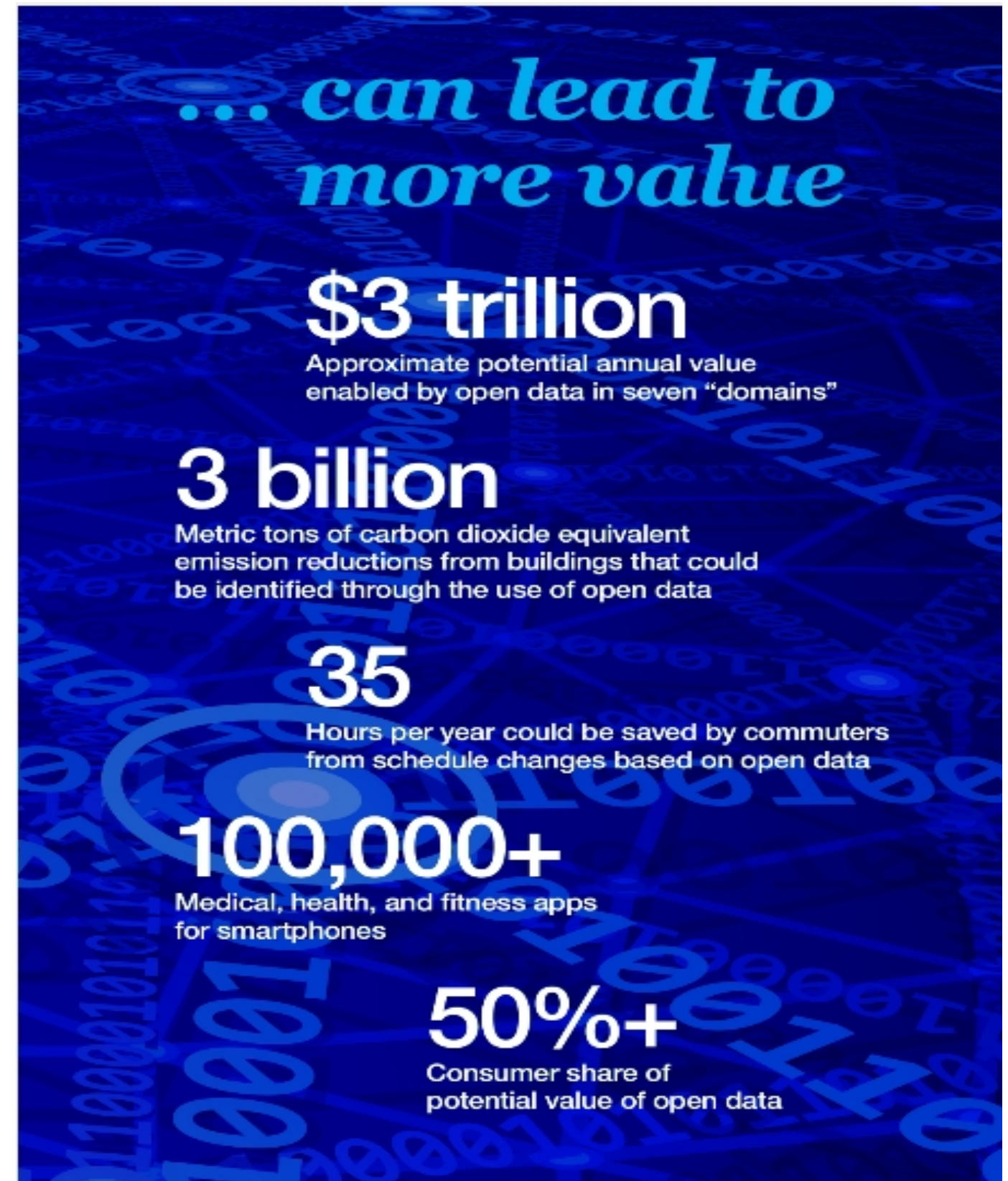
GEOSTEP



NEGOCIO DE LOS DATOS (INCLUIDOS NO ABIERTOS)

3.000 billones (Españoles) 'habilitados' por el open data / anualmente

3.000 M de toneladas de CO₂ no emitido



Source: Manyika et al, 2013.
Mckinsey global institute

Impacto



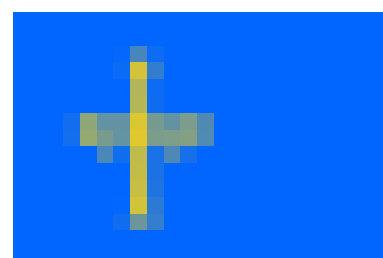
Algunos datos de impacto

1% del PIB → 130.700 M€ (PIRA 2000)

0.25% del PIB → 32.765 M€ (Dekkers et al, MEPSIR 2006)

1.7% del PIB → 222.190 M€ (Vickery 2013)

Esto en España supone: 23.400 M€ >



Diferentes tipos de análisis y condiciones de estudio



Algunos datos de impacto

En 2020

- 1.700 M€ ahorrados en EU AAPP
- 629 Mh no trabajadas y
- 16% energía no consumida.

(Carrara et al, 2015)

* Diferentes tipos de análisis y condiciones de estudio

España según MEPSIR (2.080M€) datos de muestreo
Red.es (400-500 M€)



Mentiras



Es de izquierdas. Es de derechas.

Datos de Febrero de 2013

OPEN nace mayoritariamente en el partido socialista

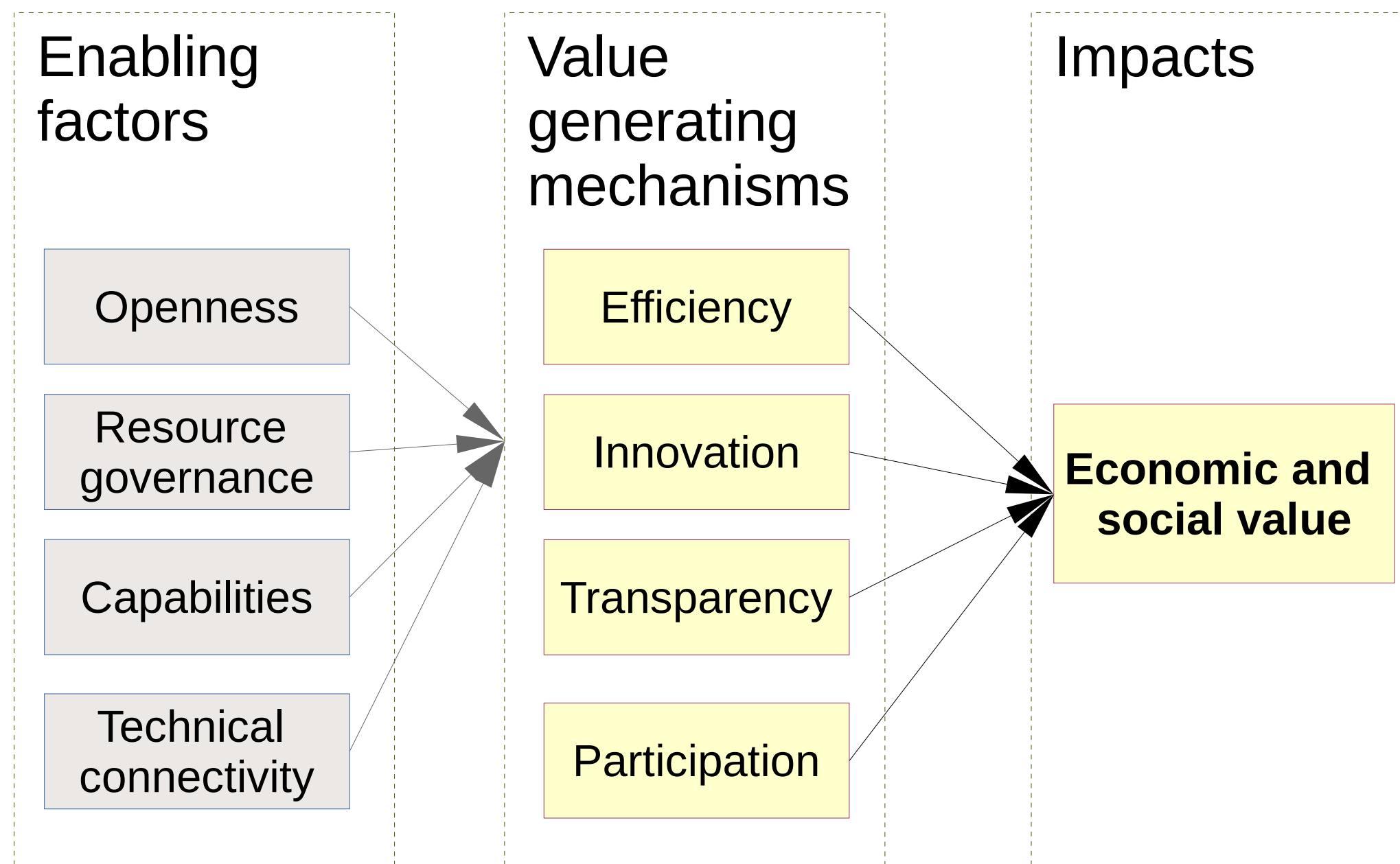
PSOE (solo o en compañía de otros) dio lugar al 80% (25) de las iniciativas

Cambio de gobierno → 3.8 Años media, sin cambio partido → 4.2 Años

PP detuvo el 54 % de las iniciativas open no uyas

Las iniciativas del PP tienen 6.7 años de vida, las del PSOE 3.2 años

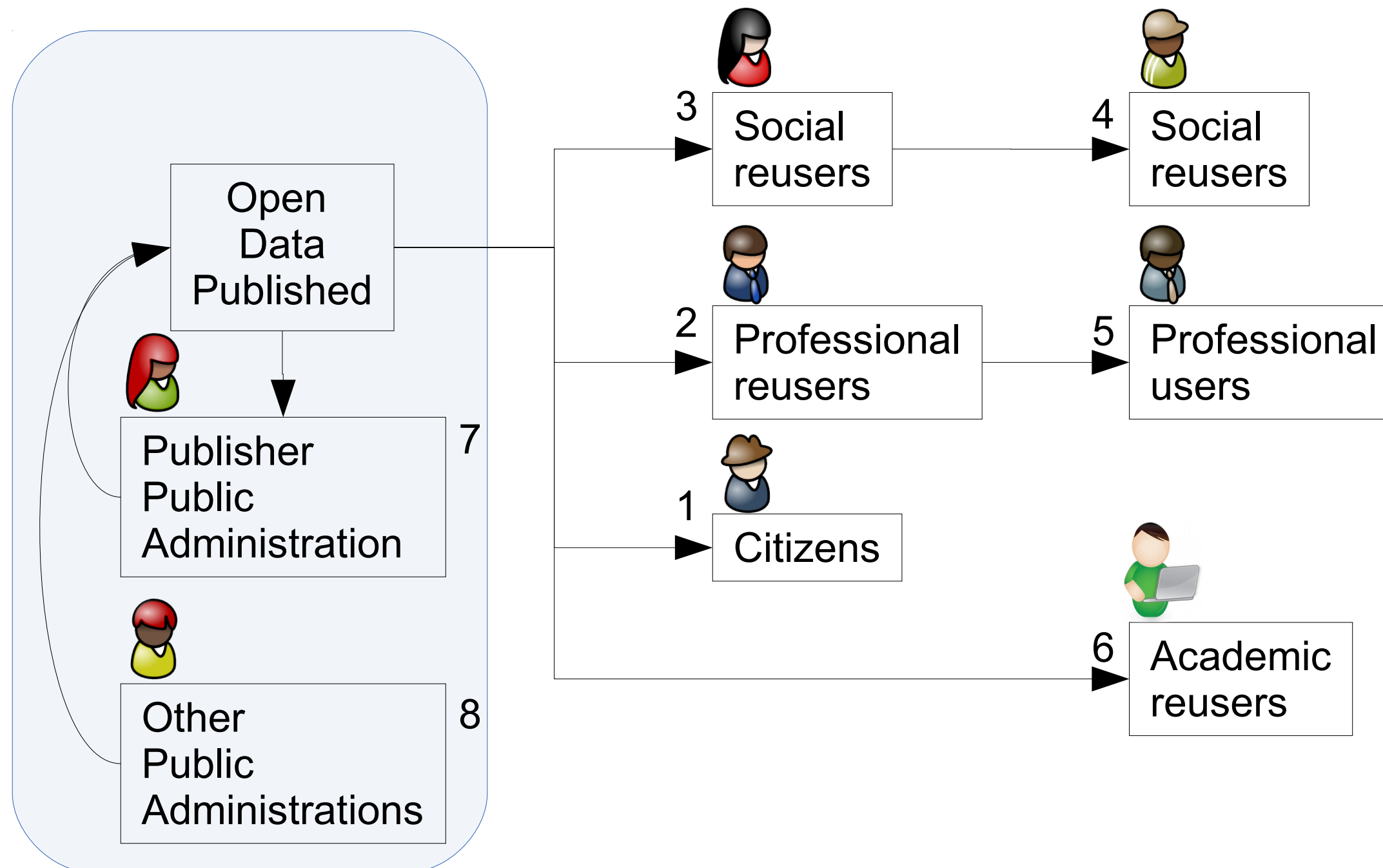
Sabemos como se produce el impacto



Conceptual model for OGD value generation (Jetzek, Avital and Bjorn-Andersen 2013)

Sabemos quien reutiliza los datos

Reuser categories

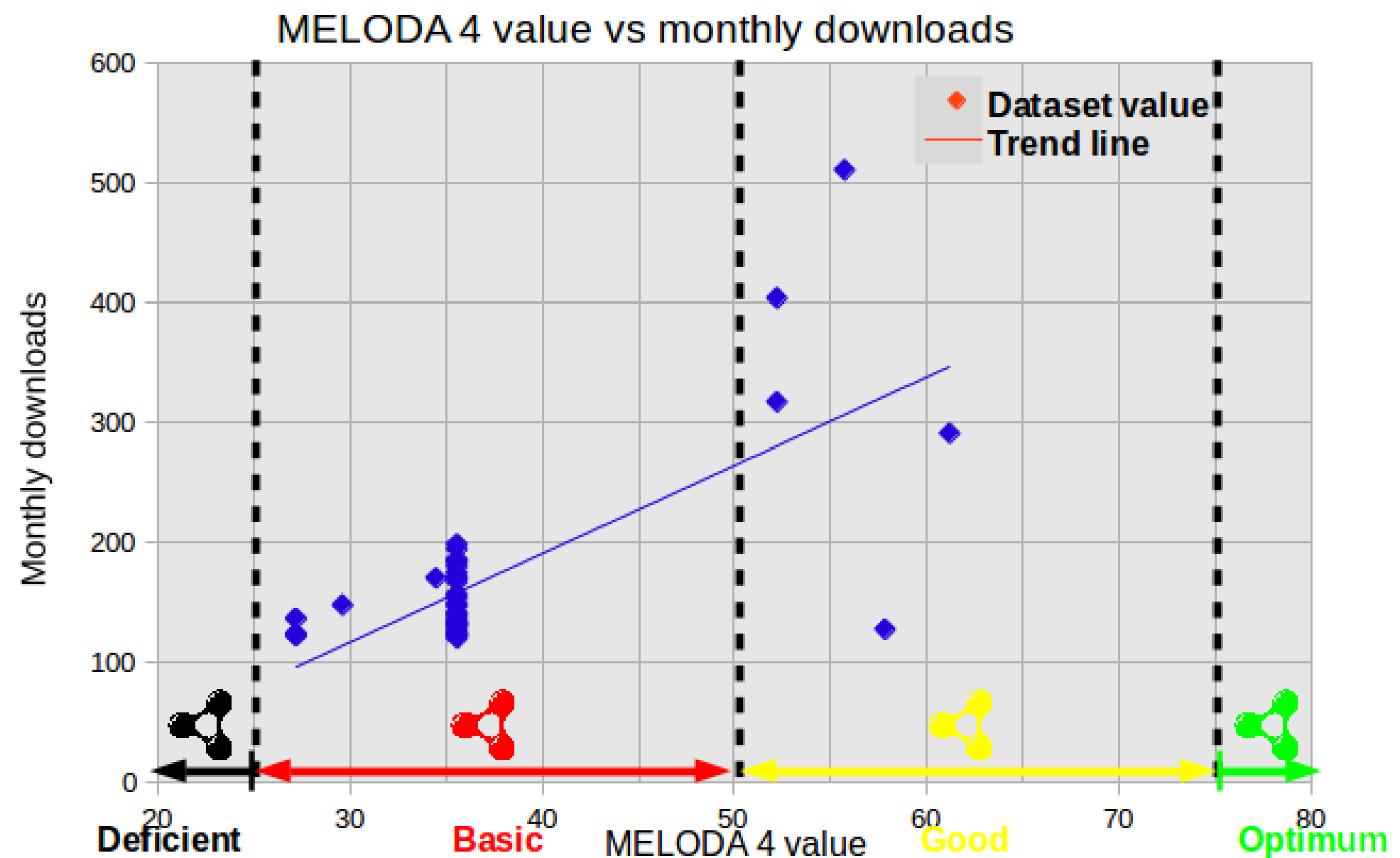


Medias verdades



Sabemos que es lo que más se demanda

Theme	samples	Downloads/day(average)
EDUCATION	2	8.9
GOVERNMENT	2	8.1
JUSTICE AND LEGAL	1	4.2
REGIONAL AND LOCAL	4	8.8
SOCIETY	27	5.0
TRANSPORT	7	5.0



El portal paneuropeo tiene todos los datos

>500.000 juegos de datos.

Usando DCAT-AP

71 portales (3 Spain)

Media más alta: 7 /día



Downloads on Portal 01/04/2016 - 30/06/2016

Label	Unique Downloads	Downloads
Content		
www.europeandataportal.eu	1380	1487
Data		
files.datapress.com	1986	2074
ec.europa.eu	889	1022
www.minrol.gov.pl	636	655
danepubliczne.gov.pl	240	246
crip.msit.gov.pl	77	77
www.eustat.eus	56	62
www.data.gouv.fr	51	60
datos.madrid.es	42	48
www.gov.uk	42	49
www.nfz.gov.pl	38	38
centrodedescargas.cnig.es	37	44
data.dublinked.ie	37	38
daten-hamburg.de	37	46
cohesiondata.ec.europa.eu	35	40
opendata.euskadi.eus	34	39

5 estrellas de Berners-Lee garantizan utilización

- Libera datos en la Web (cualquier formato) con licencia abierta
- Libera como datos estructurados
- Libera en un formato abierto no propietario
- Utiliza URI para identificar los datos para lo puedan enlazar
- Enlaza los datos a otros datos para proporcionar contexto

¿Excel es un formato abierto?



Hechos incómodos

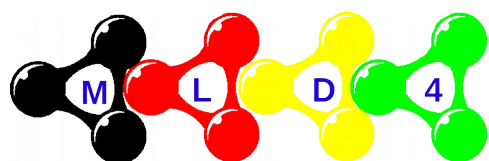


Linked data soluciona problemas interconexión

Linked data:

- 1) Utilizar URIs para identificar los recursos publicados en la Web
- 2) Aprovechar el HTTP de la URI para que la gente pueda localizar y consultar (es decir, referenciar) estos recursos.
- 3) Proporcionar información útil acerca del recurso cuando la URI haya sido referenciado.
- 4) Incluir enlaces a otras URI relacionadas con los datos contenidos en el recurso, de forma que se potencie el descubrimiento de información en la Web.

Nos tendremos que poner de acuerdo en las ontologías (mecanismos de codificación)



El open data acelera el crecimiento

ODS Challenge : entre x5 y x10 la inversión a 3 años

¿Cómo?

- Reducción de la asimetría de información
- Aumento de la eficiencia y
- Efectos de red

Fuente

<http://autoactualidad.autoscout24.es/plan-pive/comparativa-plan-pive>



Los portales de datos abiertos son un éxito

EMT Madrid Open data API reporta Más de un millón de llamadas / día



Theme	samples	Downloads/day (average)
EDUCATION	2	8.9
GOVERNMENT	2	8.1
JUSTICE AND LEGAL	1	4.2
REGIONAL AND LOCAL	4	8.8
SOCIETY	27	5.0
TRANSPORT	7	5.0

Fuente: Elaboración propia en base a fuentes del ayuntamiento de Barcelona

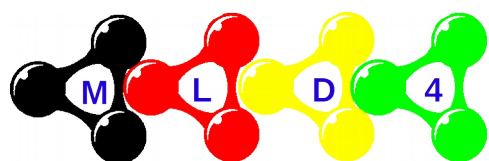
Futuro: Por donde empezamos



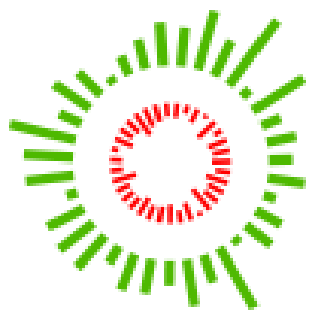
**La interoperabilidad es el
mayor problema,
posiblemente...**

....Y todavía no sabemos solucionarlo

....Y va a peor.



Si quieres llegar lejos hazlo con otros activistas



<http://okfn.es>

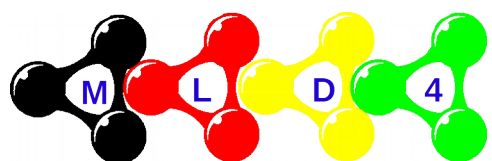


<http://iniciativabarcelonaopendata.cat>



<http://openkratio.org>

OpenKratio



Si quieres llegar lejos hazlo con otros empresas



<http://asolif.es>

Federación Nacional de
Empresas de Software Libre

Si quieres llegar lejos hazlo con otros Administraciones públicas



novagob

<http://novagob.org>



FEDERACION ESPAÑOLA DE
MUNICIPIOS Y PROVINCIAS

<http://femp.es>



<http://redciudadesinteligentes.es>

Si quieres llegar lejos hazlo con otros

116 iniciativas

IODC 6-7 Octubre,
Madrid



Conclusión



Por que hay que prmoover los datos abiertos

- 1.- Acelera la investigación y la innovación
- 2.- Reduce la corrupción y la evasión fiscal
- 3.- Mejora la sociedad
- 4.- Crea empleo

El mayor desafío del open data (y no open)
Interconectar datos

Tendencias en open data

- 1) Real-time y Geolocalización
- 2) Accesos vía API
- 3) Interconexión → Éxito

<http://bit.ly/MUGI-opendata>

¿Preguntas?

@aabella

Mail profesional : alberto.abella@desideDatum.com

Mail Académico : alberto.abella@meloda.org



By Nevit Dilmen (Own work) [GFDL (<http://www.gnu.org/copyleft/fdl.html>) or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)], via Wikimedia Commons



Desideratum [deb-sid-eh-rab-tub-m]
(from latin, "something wanted or needed")

Datum [dah-tub-m] (from latin,
"a single piece of information,
as a fact, statistic, or code")

desideDatum.com

@aabella

Mail profesional : alberto.abella@desideDatum.com

Mail Académico : alberto.abella@meloda.org

