

Introducing Open Data

David Tarrant · davetaz@theodi.org · [@davetaz](https://twitter.com/davetaz)

Provide a good foundation in the principles of open data, and explore examples of open data usage.

Outcomes

Define open data

Describe a number of key open data stories

Evaluate the challenges facing open data

Analyse the future of open data on a global scale



Exercise

What is Data?



Definition of Data (1)

A collection of **facts, information** and
statistics that can be analysed to
develop new knowledge



Definition of Data (2)

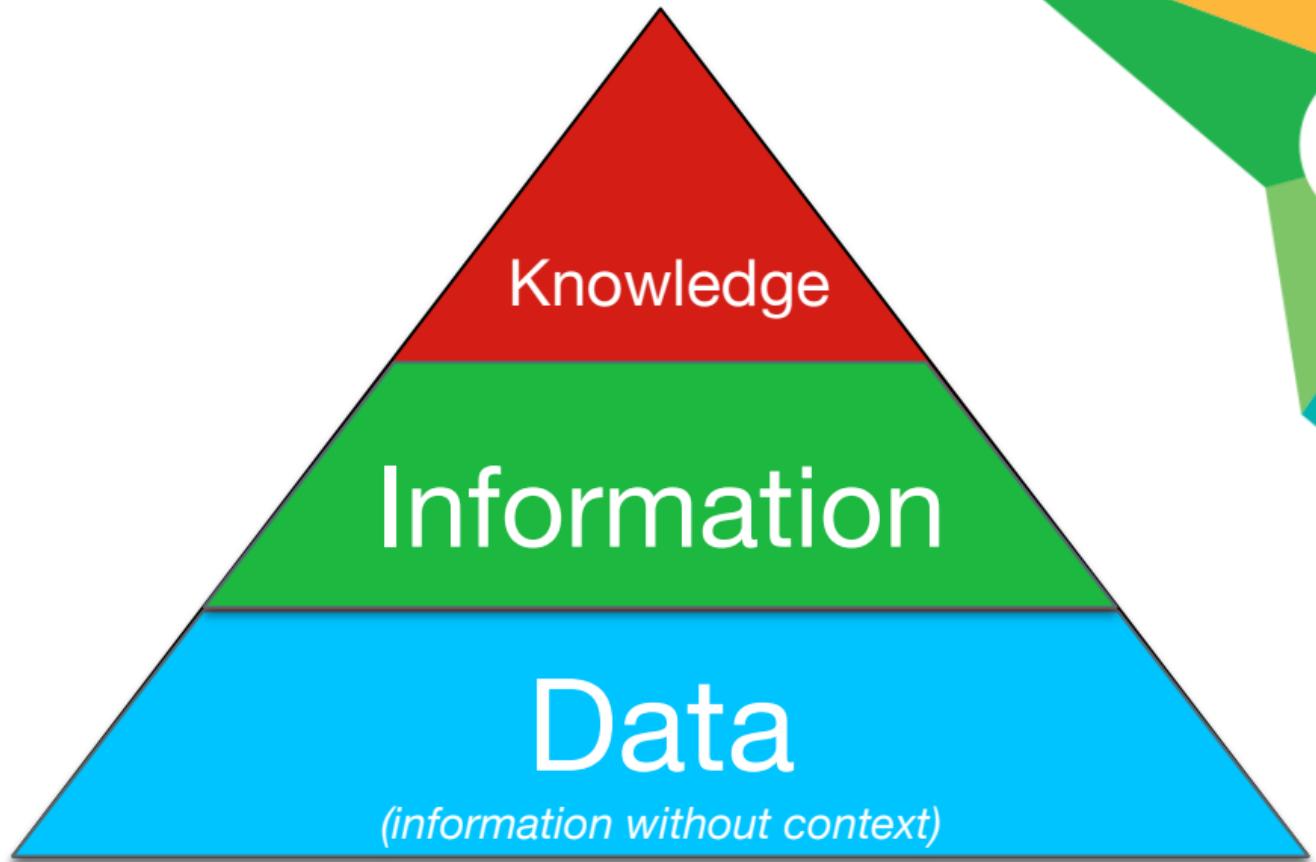
A collection of numbers assigned as values to quantitative variables and/or characters assigned as values to qualitative variables



Definition of Data (3)

The **lowest level** of abstraction
from which information and then
knowledge are derived.





Exercise

What is Open Data?



Option A

Open data is data that is made **available** by **organisations**, **businesses** and **individuals** for **anyone** to **access**, **use** and **share**.

- Open Data Institute

Introduced November 2014



Option B



Open means **anyone** can **freely access, us
modify and share for any purpose** (subject, at
most, to requirements that **preserve provenance**
and **openness**).

- Summary of Open Definition (v2.0)

Introduced August 2014



Option C



DATA.GOV.UK
Opening up Government

Open data is data that is **published** in an **open format**, is **machine readable** and is **published** under a **license** that allows for **free reuse**.

- data.gov.uk

Accessed November 2014





WIKIPEDIA
The Free Encyclopedia

Option D

Open data is the **idea** that certain data should be **freely available** to **everyone** to **use** and **republish** as they wish, **without restrictions** from **copyright, patents** or other **mechanisms of control**.

- Wikipedia

Accessed November 2014



Data is the raw material of the new industrial revolution

- Francis Maude, UK MP



Justifications

Trust and

Transparency

Enabling the
economy

Key Stories

Enabling transparency

Creating an income stream

Cutting costs

Improving services

Saving our planet



Enabling transparency

Exercise

Creating an income stream

Cutting costs

Improving services

In your groups, come up
with a short list of your
favourite open data stories

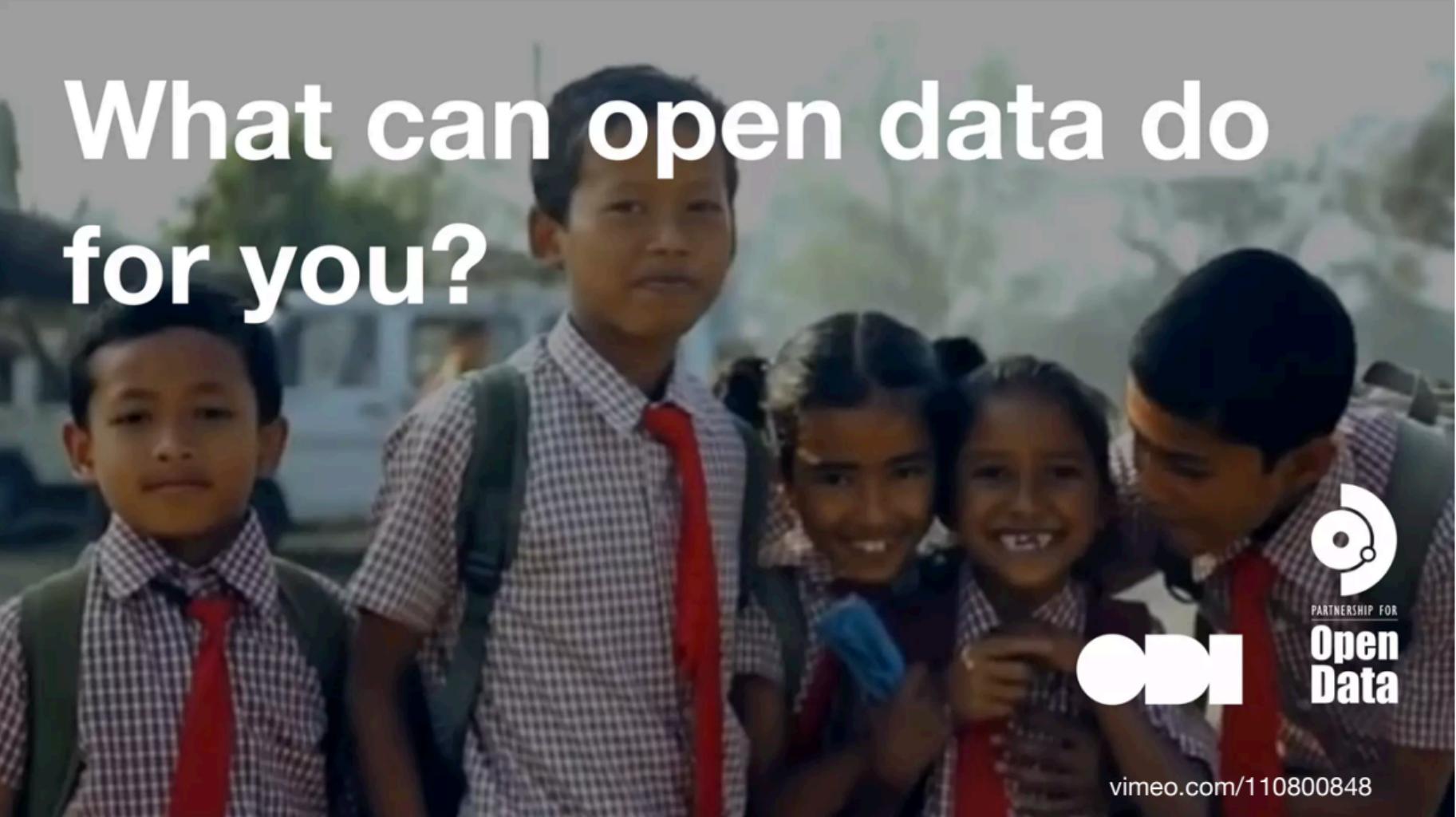
Saving our planet

(bonus points for one in each area)

7 minutes



What can open data do for you?

A photograph of five school children, three boys and two girls, standing outdoors. They are all wearing matching school uniforms consisting of white shirts with red ties and grey trousers. The children are looking towards the camera with varying expressions of neutral, slightly smiling, and laughing. The background is a soft-focus outdoor setting, possibly a school campus.

vimeo.com/110800848

Transparency

NEWS POLITICS

[Home](#) | [World](#) | [UK](#) | [England](#) | [N. Ireland](#) | [Scotland](#) | [Wales](#) | [Business](#) | [Politics](#) | [Health](#) | [Education](#) | [Sci/Ec](#)

31 March 2011 Last updated at 12:56

Share   

Former MP Jim Devine jailed for 16 months over expenses

Ex-Labour MP Jim Devine has been jailed for 16 months for fraudulently claiming £8,385 in expenses.

Devine was last month found guilty of using false invoices for cleaning and printing work.

The ex-MP for Livingston, 57, is the third current or former MP to be jailed for fiddling their expenses, but was the first to stand trial.

The judge, Mr Justice Saunders, said he "set about defrauding the public purse in a calculated and deliberate way".



Devine was the first MP to stand trial over his expenses

Related Stories



<http://www.telegraph.co.uk/news/newstopics/mps-expenses/>



Income



CLIMATE | BASIC

Better Data. Better Decisions.

The most advanced way to get insight into
your fields.



Weather



Tracker



Alerts



Scouting
& Notes

[LEARN MORE »](#)



Cutting Costs

Open Data found a £200m saving in the NHS budget

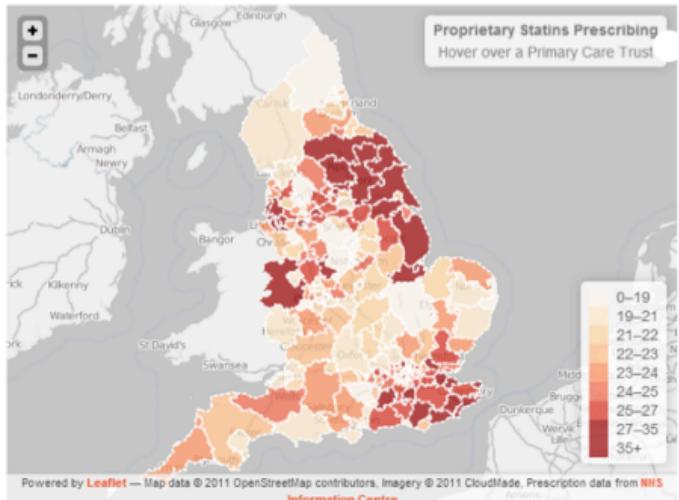


Home

NHS efficiency savings: the role of prescribing analytics

The NHS has been challenged to make £20 billion in "efficiency savings" by 2015 (1). £10 billion a year are spent by the NHS on essential drugs. Often, there's a choice between a cheap "generic" medication, or an expensive "branded" one. Branded drugs can cost over ten times as much, for the same therapeutic benefit. "Prescribing Advisors" in the NHS, with the support of NICE, encourage doctors to use the most cost effective treatments. We have analysed exactly how much is spent on expensive "branded" medicines, for one class of drugs, namely statins, in England.

Percentage of proprietary statin prescribing by CCG Sep 2011 - May 2012



Where generic alternatives exist, NHS purchasing of branded drugs has a number of effects:

1. There is only a finite amount of money to spend on treatments. Prescribing a patient expensive branded drugs, when cheaper equivalents would have the same therapeutic effect, means that money is wasted. As a result the NHS has less money to spend on the care needed by other patients.

1x

Rosuvastatin Calcium

1.2x

Simvastatin

11x

Atorvastatin

2nd Largest Drug Cost (prior to June 2012)





Improving services

1 Tube network

13 Bus Companies with 100s of routes

11 Train companies connecting urban areas

1 integrated payment system



Communicate and support

NEWS UK

South West Trains @SW_Trains • 9m

Home **UK** **Africa** **Asia** **E**
England | Northern Ireland

23 September 2014 Last upda

Rail passenger information

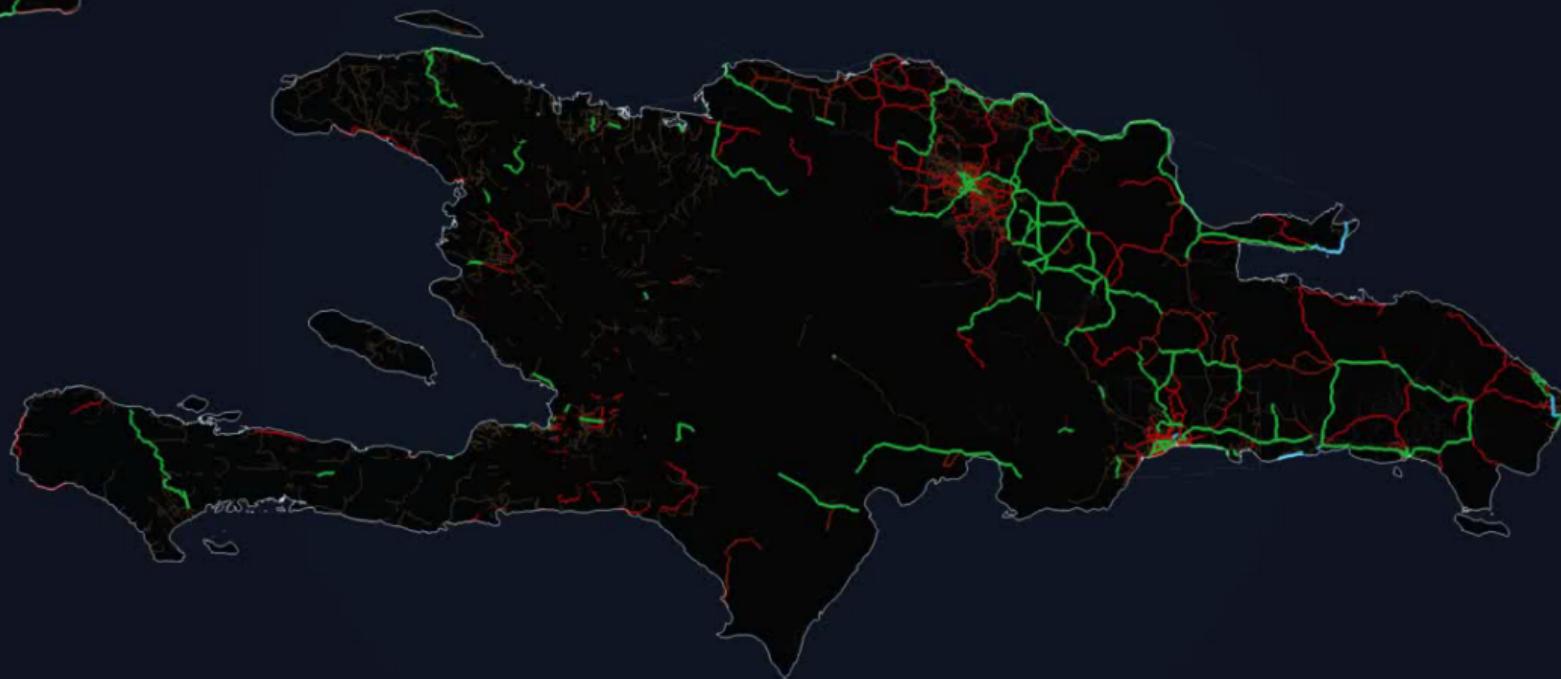
"Passengers need information as quickly as possible - ideally before leaving home. **Only 17% knew about the disruption before arriving at the station."**

"Passengers now receive information from a range of sources, so train companies must ensure that staff at stations and on trains are ahead of the information game."



Saving our planet

itō!



9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25

January 2010

vimeo.com/9182869

OpenStreetMap

CC-by-SA www.itoworld.com

Map data www.openstreetmap.org 31 Jan 2010

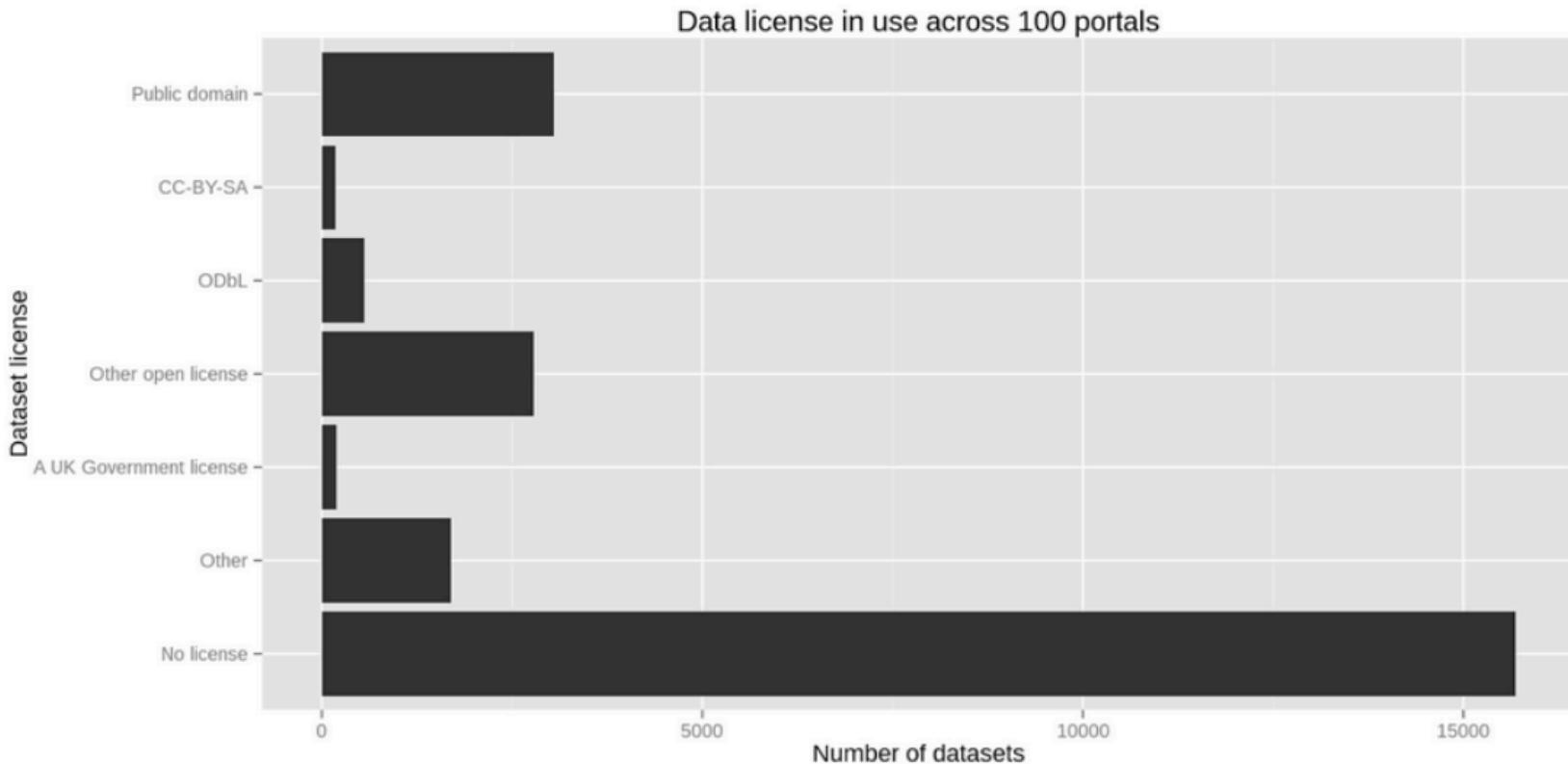
The challenges...



Knowing what open data is

	Open Definition Open Knowledge Foundation	OMB Memo, 2013 The White House Sylvia Burwell et al.	Data.Gov.UK	Antonio Acuña	"DBpedia: A Nucleus for a Web of Open Data" Sören Auer et al.	Open Data Institute (ODI) Open Data Institute	LinkedGov LinkedGov	McKinsey James Manyika et al.	Open Data Now Joel Gurin	Open Data Barometer Tim Davies	The World Bank The World Bank
Free	✓	✓			✓	✓		✓			
Negligible Cost								✓			
Publicly Available	✓	✓				✓		✓	✓		
Re-usable	✓			✓		✓					✓
Can be Redistributed	✓				✓						✓
Non-exclusive (No Restrictions from copyright, patents, etc.)	✓				✓	✓			✓		✓
Structured for Usability		✓	✓					✓		✓	✓
Requires "Open" License			✓			✓	✓			✓	✓
Non Personally Identifiable							✓				
Produced during business operation							✓				
Belongs to the Taxpayer (when not in violation of laws/privacy)							✓				
Accessible in Bulk									✓		

Open data is hardly ever appropriately licensed.



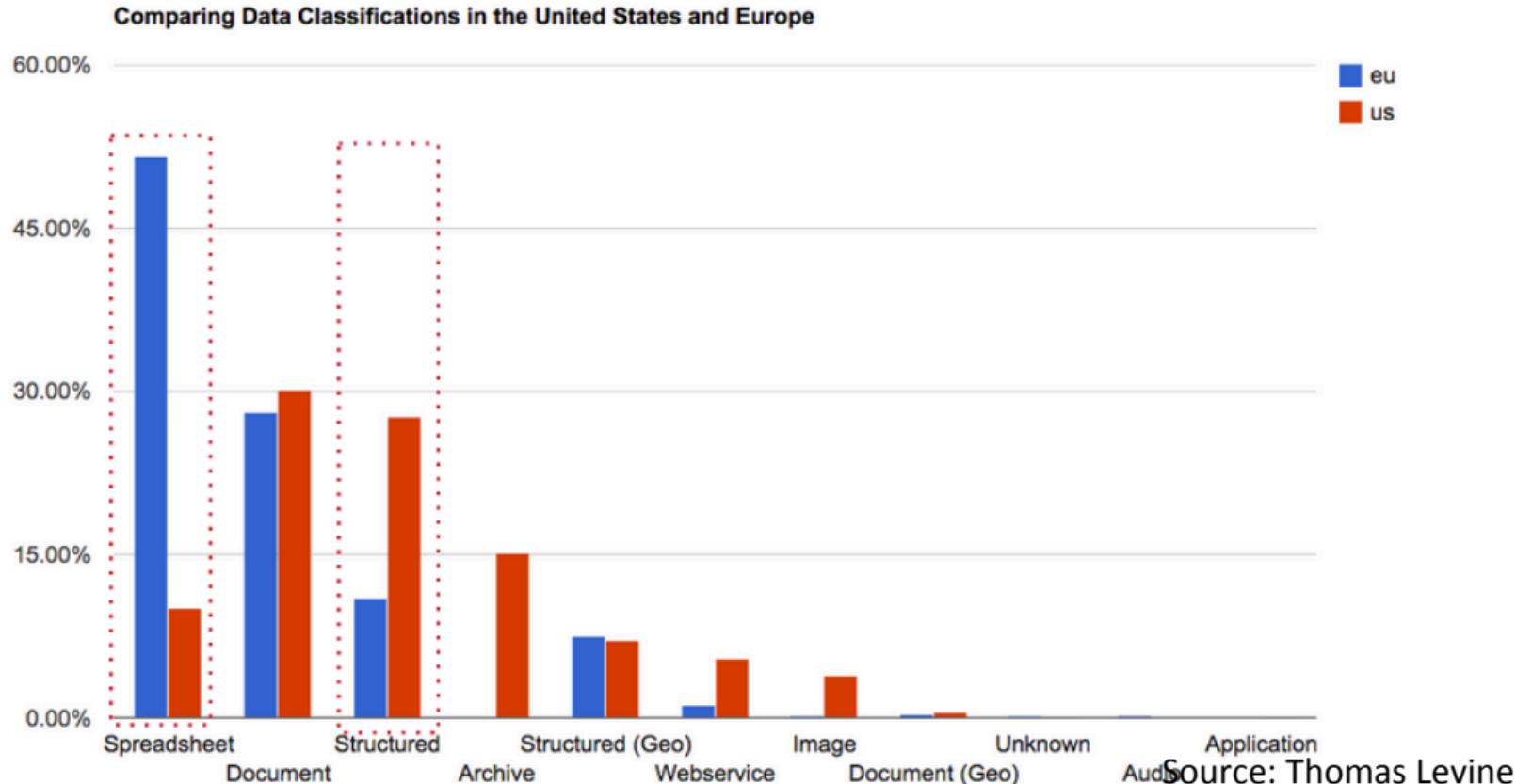
Hardly any datasets get updated.



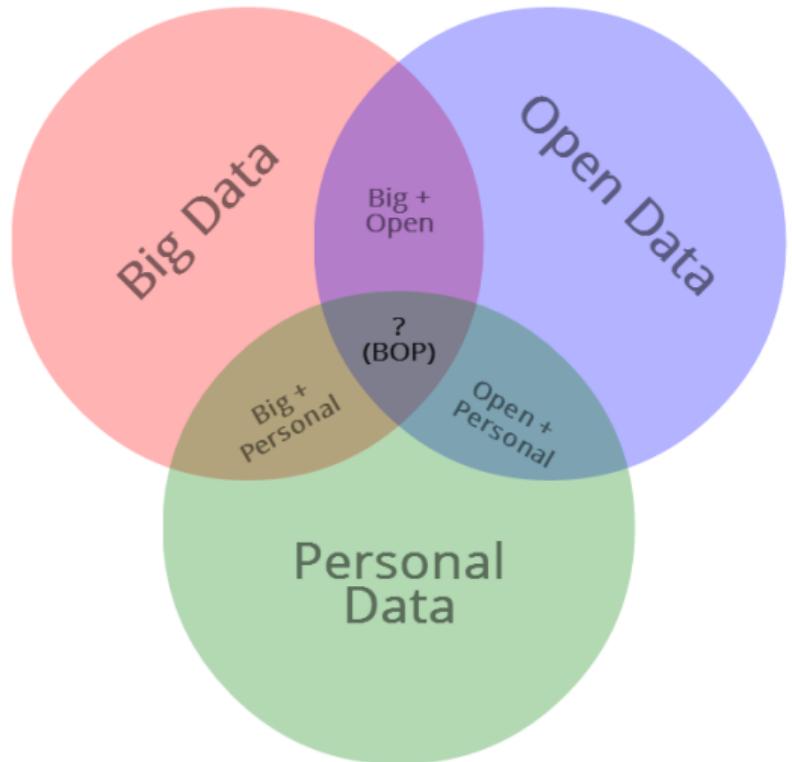
**Warning: This result was actually based
upon the wrong date field for updates.
I personally think you'll find a similar result
with the right date field being used.**



Open data is rarely structured.



Challenges and Risks



Types of personal data

Open personal data

Data about people
not a person

Available to anyone

Has been anonymised

e.g. number of people attending
event, gender split, age ranges.
(bigger numbers are better!)

Available personal data

Data about a person
Available to the person only!

Often known as MiData

e.g. credit scores, energy and other
consumption data.

Personal data

Data about a person
which is neither open
nor available.

Might belong to you or
be collected by a
company.

Opportunities

Open Data

Brings transparency, open peer review

Big Data

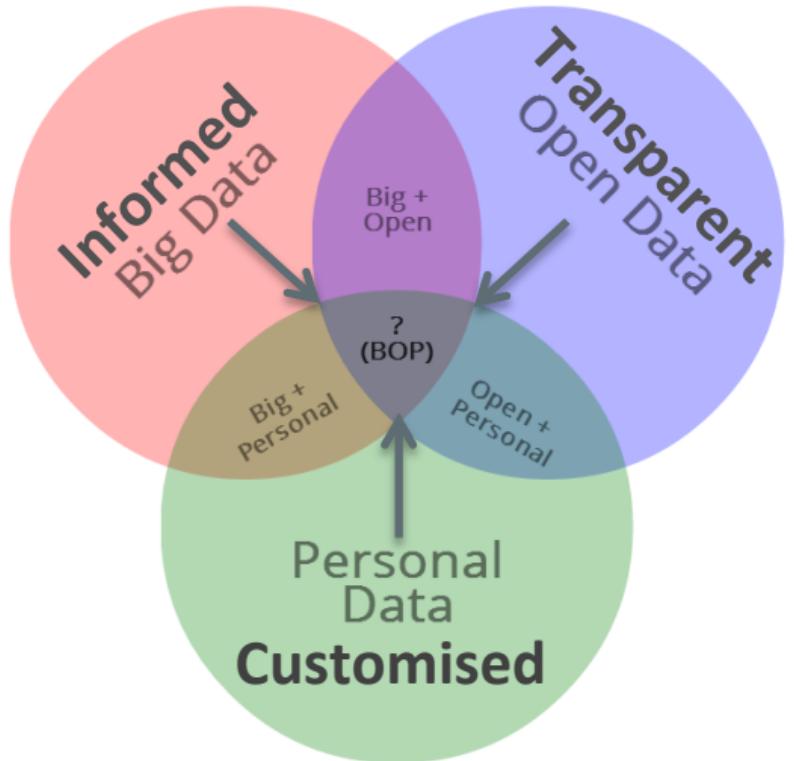
Brings the evidence

Personal Data

Makes it relevant



Opportunity



Why now?



Why now?



Policy Drivers



Technical Standards



Best Practice Guidelines



Open Data

Best Practice Guidelines



Open Knowledge
Foundation



Technical Standards



Policy Drivers



Policy paper

G8 Open Data Charter and Technical Annex

Published 18 June 2013

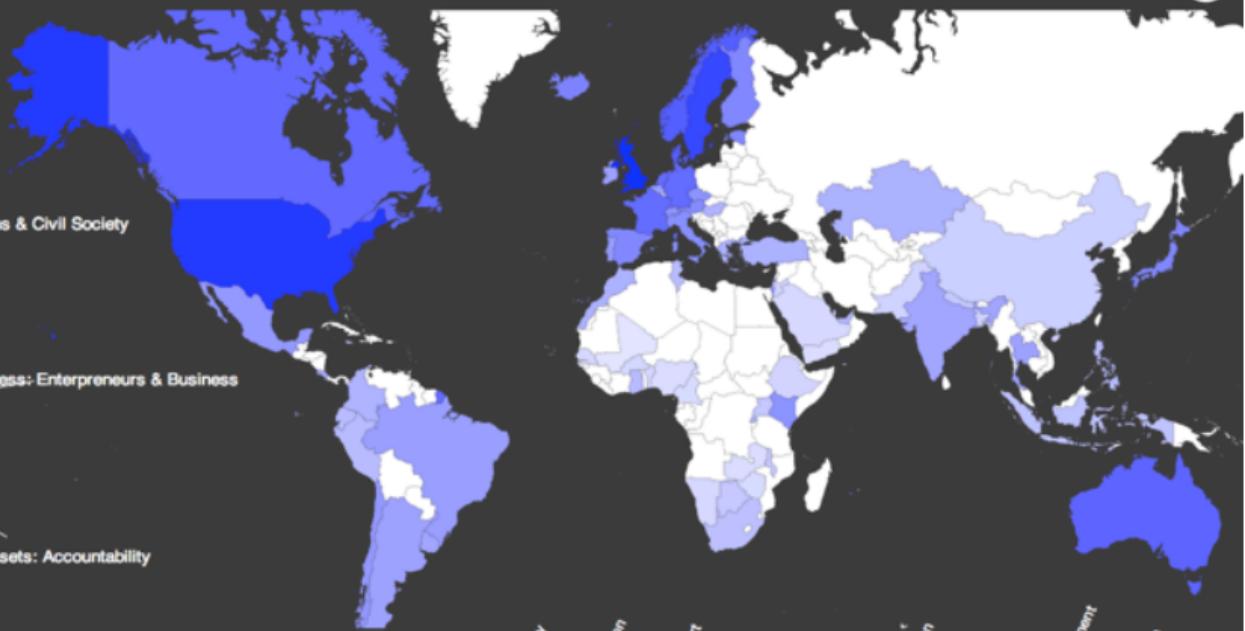
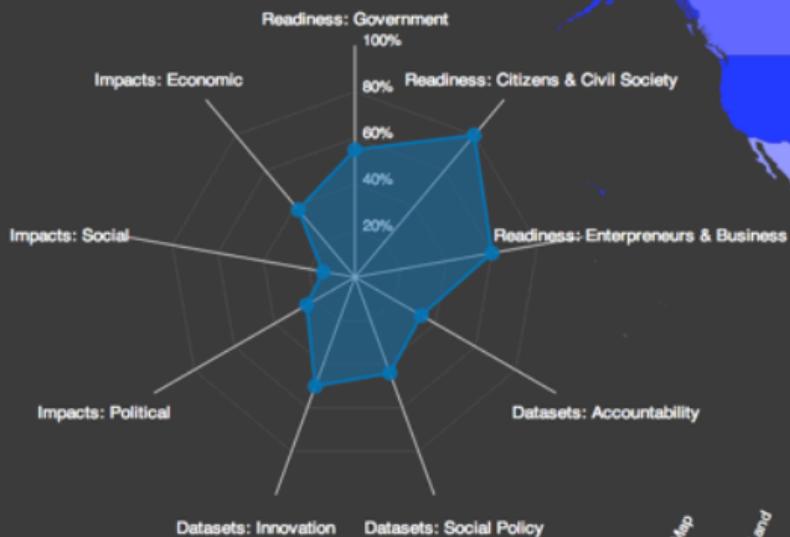
Contents

1. Principle 1: Open Data by Default
2. Principle 2: Quality and Quantity
3. Principle 3: Usable by All
4. Principle 4: Releasing Data for Improved Governance
5. Principle 5: Releasing Data for Innovation
6. Technical annex



Switzerland \$\$\$\$

22/77



License: CC-BY-SA | Visualisation by David Tarrant | [Full report](#) | [Get the data](#)



<http://www.opendataresearch.org/barometer>



<http://vimeo.com/97404977>

Data is a means, not an end.



In your country

Discuss:

1. What is the end? /
Why are you opening data?
2. What is a potential application of open data?



Outcomes

Define open data

Describe a number of key open data stories

Evaluate the challenges facing open data

Analyse the future of open data on a global scale



A wide-angle, nighttime photograph of the London Olympic Stadium. The stadium's iconic retractable roof is open, revealing a massive digital screen that displays the text "THIS IS FOR EVERYONE" in large, white, sans-serif capital letters. The stadium is surrounded by a dense crowd of people, and the surrounding urban environment is visible in the background.

THIS IS FOR EVERYONE

The biggest evolution of the web,
since the web itself



Thank-You

David Tarrant · davetaz@theodi.org · [@davetaz](https://twitter.com/davetaz)