



# Data Champions Training

Dr David Tarrant | @davetaz | The Open Data Institute

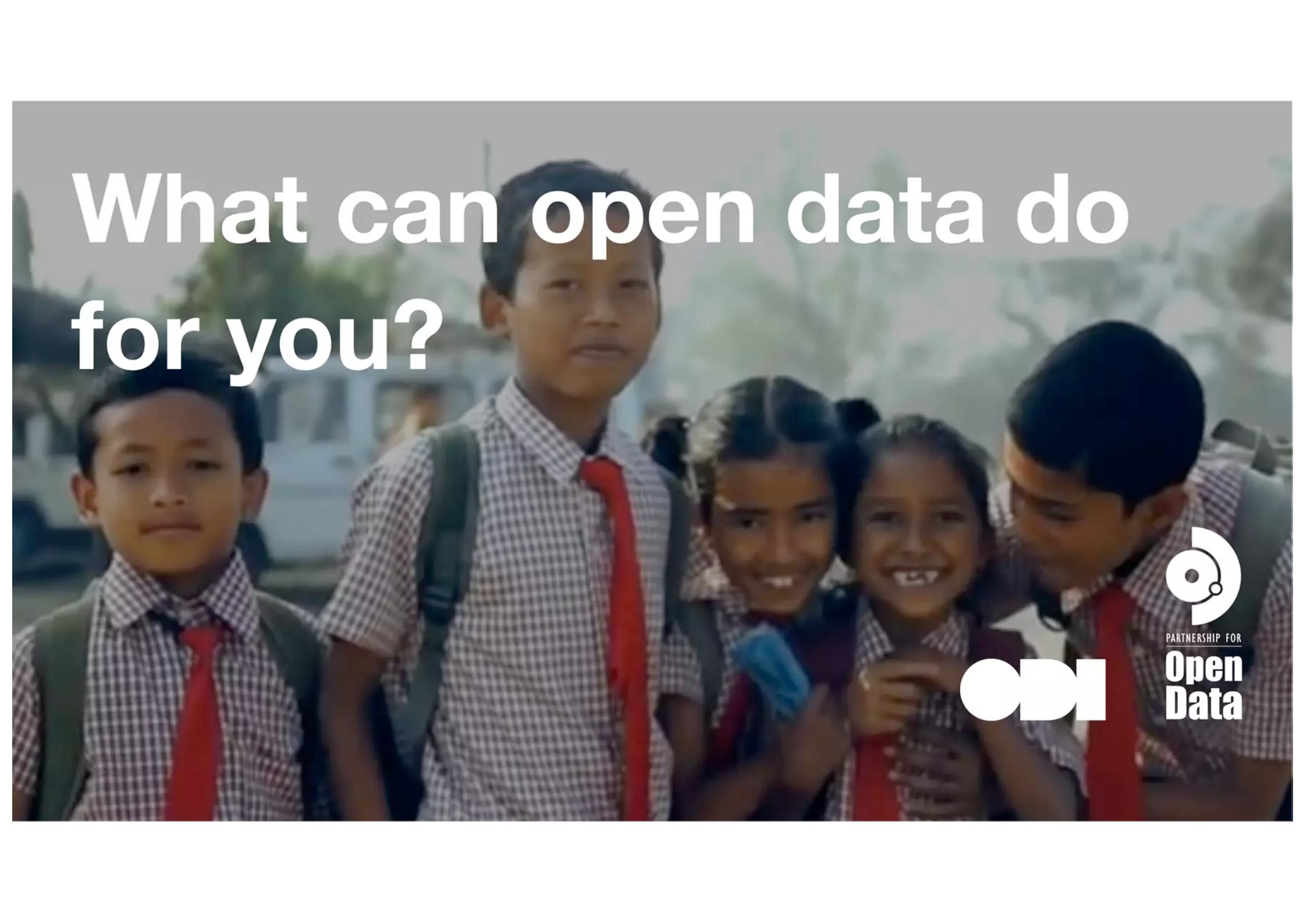
<http://training.theodi.org/Malaysia/2017>



 Content created by  
The Open Data Institute

We connect, equip and inspire  
people around the world  
to innovate with data





# What can open data do for you?



# Introductions

Your name

What do you do?

What was your first experience of open data?



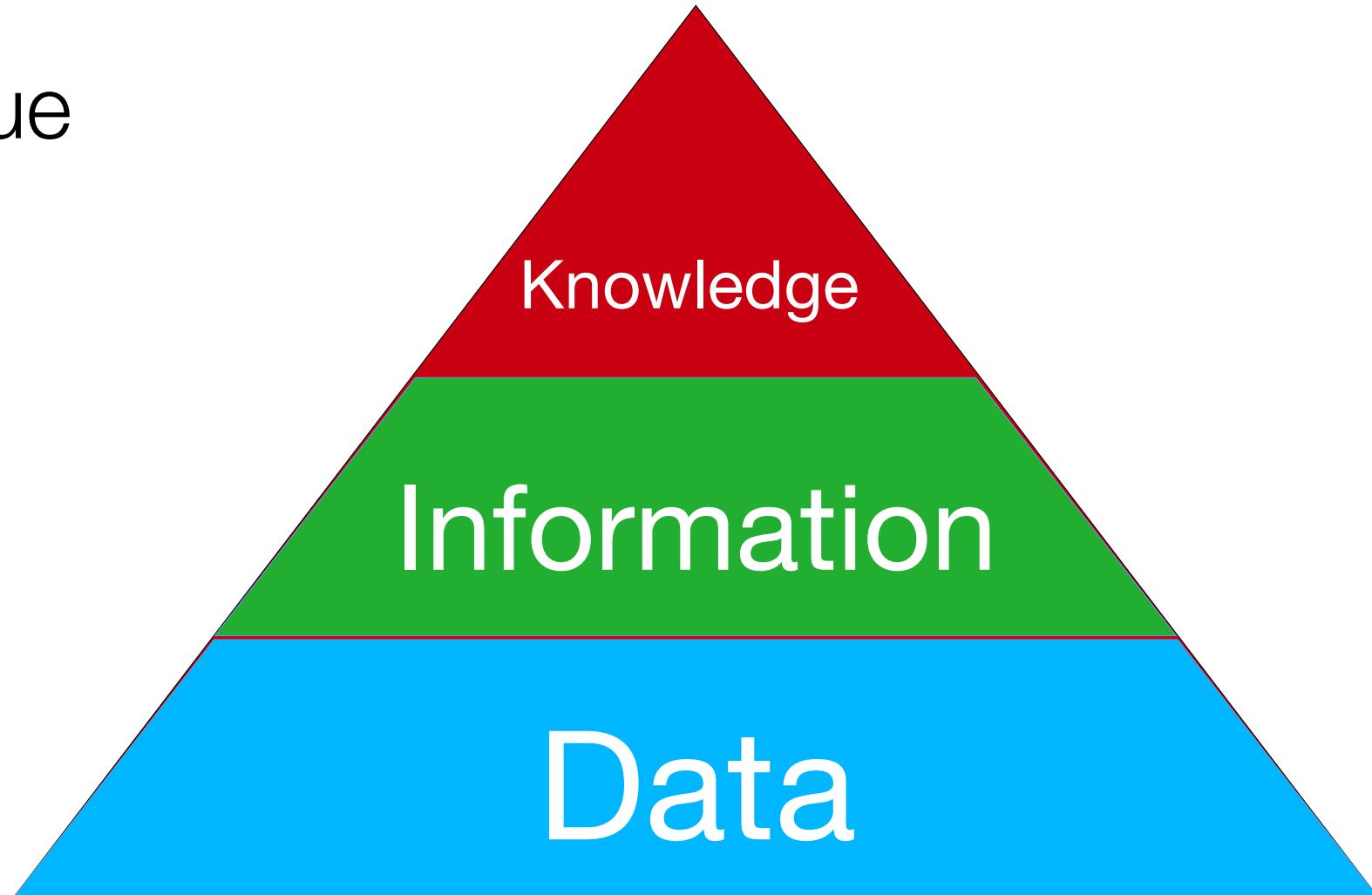
# Exercise

What is Data?



**odi** Content created by  
The Open Data Institute

Value



Data is the new raw material of the digital age



A photograph of an oilfield at sunset or sunrise, showing numerous pumpjacks in various stages of their stroke. The sky is a warm orange and yellow. In the bottom right corner, there is a small white logo consisting of the letters "ODI" in a bold, sans-serif font, with a small circle above the "O".

ODI

We connect, equip and inspire  
people around the world  
to innovate with data



# Exercise

What is Open Data?



# Open Definition (v1)

A piece of data or content is open if **anyone** is **free to use**, **reuse**, and **redistribute** it - subject only, at most, to the requirement to attribute and/or share-alike.

- Open Knowledge

Retired August 2014



 Content created by  
The Open Data Institute

# Open Data Institute (v1)

Open data is information that  
is **available** for **anyone** to **use**,  
for **any purpose**, at **no cost**.

- Open Data Institute  
Retired November 2014



# Open Data Institute (v2)

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



# Open Data Institute (v3)

Data that **anyone** can **access**,  
**use** and **share**.

- Open Data Institute  
Introduced November 2014



data.gov.uk

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***.



mampu.learndata.info





# Enabling innovation

Dr David Tarrant | @davetaz | The Open Data Institute

<http://training.theodi.org/Malaysia/2017>



 Content created by  
The Open Data Institute

We connect, equip and inspire  
people around the world  
to innovate with data



# What happened next? (Health – 2013)

UK Health Service (NHS) releases anonymised prescription data detailing which drugs have been prescribed per practice region.

What happened next?





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

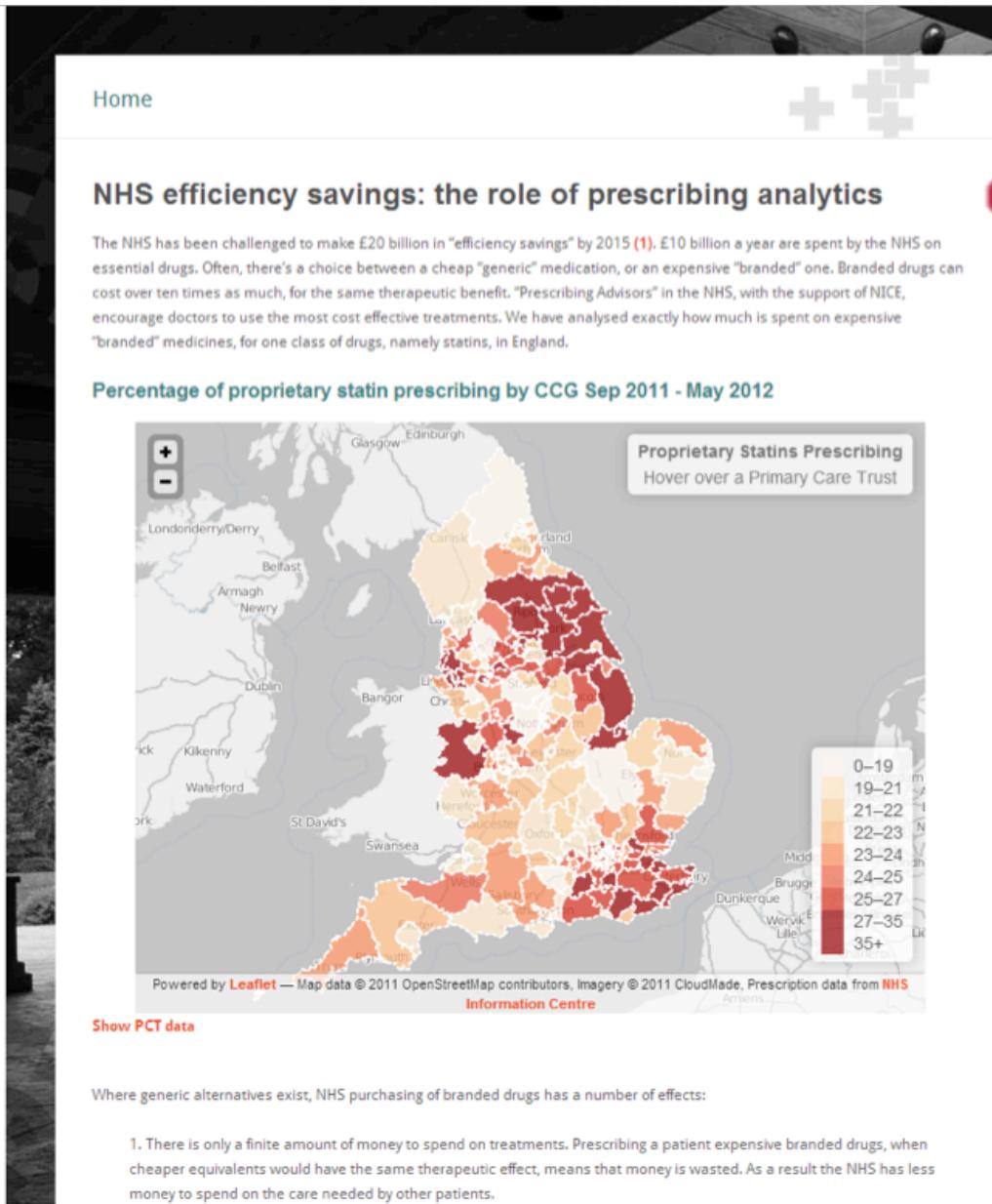




Home  
Detailed statin analysis  
Why prescriptions matter  
FAQ  
Methodology  
Contact us  
About this site

About Prescribing Analytics  
Prescribing Analytics is a joint venture of a group of UK technologists and NHS doctors, who believe in the power of data and technology to help the NHS.

Drop us an email at  
[info@prescribinganalytics.com](mailto:info@prescribinganalytics.com)



# What happened next? – (Transport 2012)

Transport for London and National Rail UK release live train times position, signalling and running timetable data as open data.

Over 600,000 live data points a second.

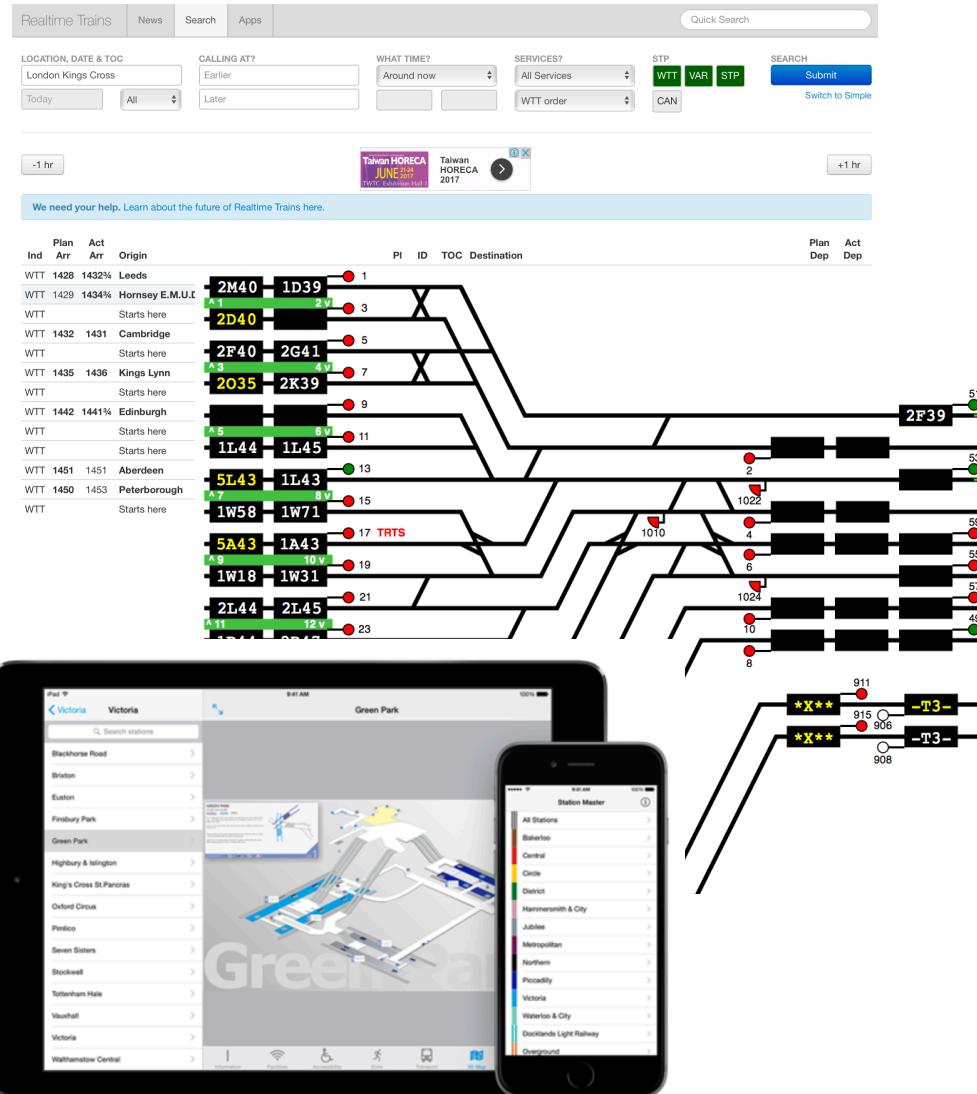
What happened next?



# Transport - 2012

Over 700 Apps now rely on open transport data including:

- City Mapper (Millions of users)
- Google Maps (Millions of users)
- Real time trains (500,000 users)
- Station master (10,000's users)



# What happened next? – (Welfare 2014)

Small business leases 2 acres of disused waste land in serene  
Kampung Sungai Limau, Lunas, Kedah.

What happened next?



# Shortcomings at Kg Sg Limau (2014)

Still undeveloped, lot of empty land

Most household earning below RM1500 p/m

Huge families with big number of childrens

Traditional way of living, polarized

Number of single mothers and widows

Most working 2 jobs and struggles to cope

Lack of English, spoken and written



Content thanks to Gugen Navamani  
ANB Agro Trainers

# Hiring 100% Local Villagers For Farm Operation

90% local Malays

8% Indians

2% Chinese

20% Veterans, 80% Youths

Age 18 to 68



Content thanks to Gugen Navamani  
ANB Agro Trainers



Content thanks to Gugen Navamani  
ANB Agro Trainers



# Develop and Turn Waste Land For Agriculture





Content thanks to Gugen Navamani  
ANB Agro Trainers

# Post Harvest Management: Producing Dried Chillies



Created by  
Data Institute

# Language Empowerment Spoken English Class



# Business model

Investors invest in business to buy Chilli plants

Initial investment: RM 37500 (5000 Chilli plants)

9 month growth period



# Business model

Initial investment: RM 37,500 (500 Chilli plants)

Return on investment: RM 51,750

Profit margin: **38%**



# What happened next? – (Education 2015)

Burkina Faso (3<sup>rd</sup> poorest country in world) releases open data on school amenities, gender distribution and pass rate for one region in the country.

What happened next?



# What happened next? (Crime data)

US and UK release crime data detailing where and when a crime occurs and any resultant action.

Data is updated at 15 minute intervals, so is almost live.

What happened next?



# What happened next? (Crime data)

Answer:

People discovered they had no idea on how much crime was local to them.

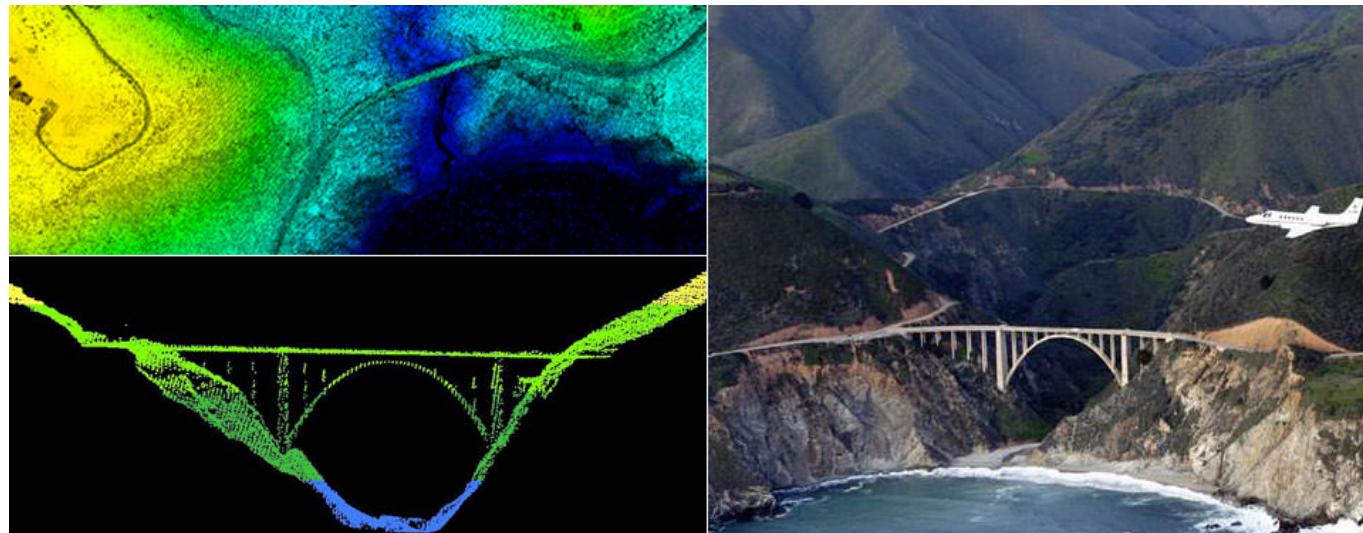
Perception and myth in different from fact.

So what happened... NOTHING



# What happened next? (Environment – 2016)

UK environment agency release LIDAR data.



# What happened next? (Environment – 2016)

- Wind modeling & city planning
- Discovered a roman villa!
- Created minecraft maps!



# How?

The UK has released over 40,000 datasets.

Of that total maybe 10% are heavily used!

Just like....



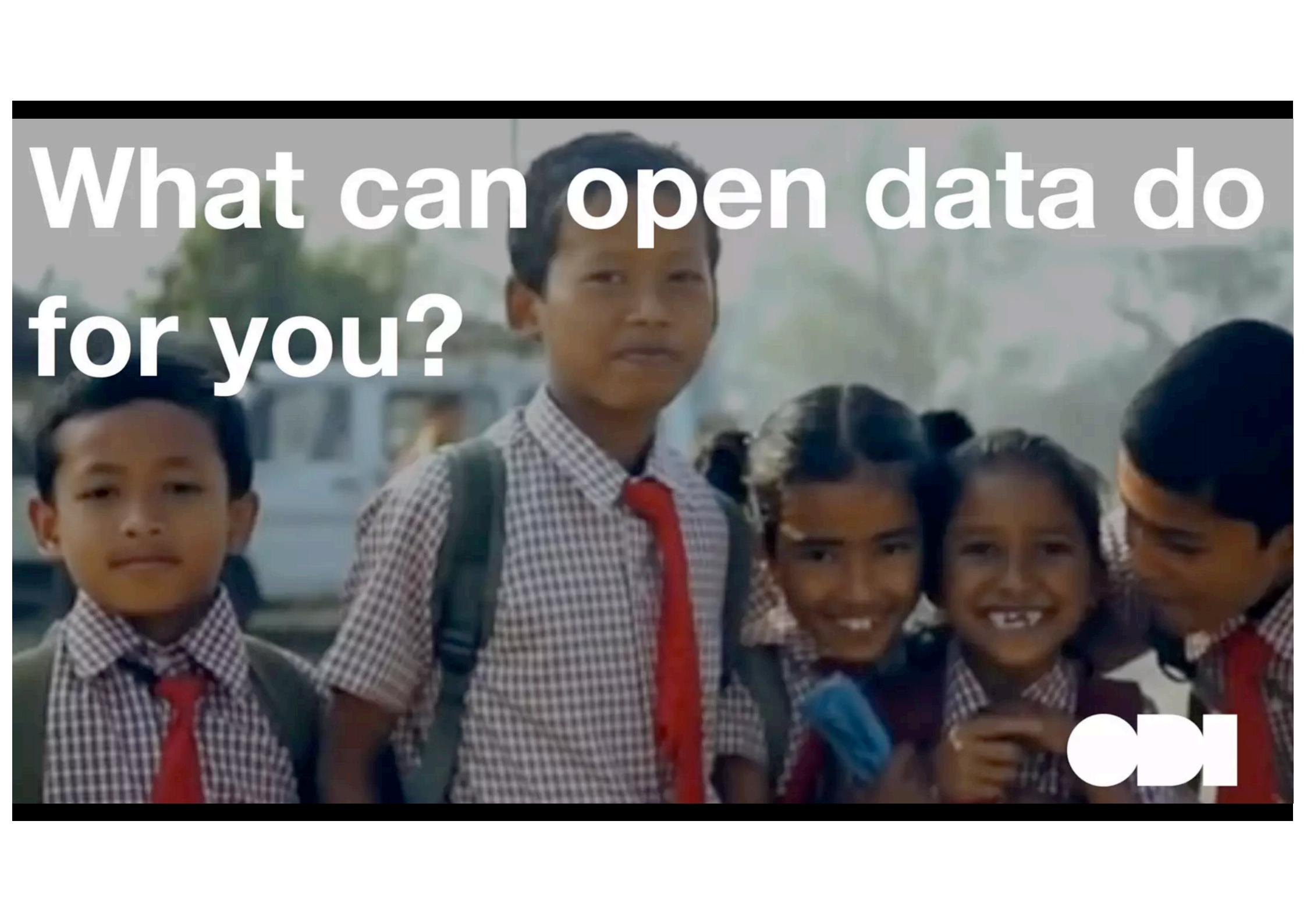


Roads help us navigate  
to a location.

Data helps us navigate  
to a decision.

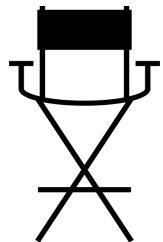
# Realising value



A photograph of a group of school children, likely of Asian descent, wearing matching uniforms consisting of white shirts with red ties and dark trousers. They are standing outdoors, possibly near a school building, with trees and a fence visible in the background. The children are looking towards the camera with neutral to slightly smiling expressions.

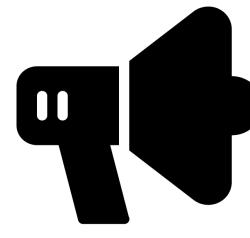
What can open data do  
for you?





## Leadership/Strategy

Open data supports the wider data strategy of the organisation



## Communications

Open data provides free media and increased reach



## Operations

Open data creates significant data access efficiency savings



## CSR/Foundation

Open data improves outreach and forges social partnerships



## Finance/Ventures

Open data produces viable equity ventures & business lines



## Marketing/Product

Open data enhances product features and customer experience



Content created by  
The Open Data Institute

# 1. I want a proof point for...

## 2. Relevant to a ...

Comms &  Transparency

Generating Income 

Efficiency Savings 

Improving Services 

CSR & Social Impact 

Decision Maker



Reaching new markets and removing the cloak of secrecy around new agritech products.



Incubate startups for new revenue sources, lower op. costs for services & higher reward than trad. acc.



Existing products can be linked together for a better sales channel & lowers ownership costs



Customer savings of between £15 and 58 million per year in time savings for transport customers



Allow community-built flood models that can save 20% on design and delivery costs

Colleague



CAN \$3.2b in charitable tax violations highlighted by citizens using OD



Est. 2007 using open farm data, acquired by Monsanto for \$930m in 2013



Save AU \$3.2m annually on F.o.I requests through proactive release



€2.6m per year in reduced staff costs from cross-dept. access to map data



NZ\$4m savings from OS tools & data in year 1 of rebuilding following earthquake

Partner



100m company records in over 100 jurisdictions allowing studies of beneficial ownership & control



£300k annual turnover from cleaning up UK transport data for other businesses



Identified £200m annual saving in NHS by switching to own-brand statins and better delivery routes



Joined up available transport data to enable choice. \$10m VC round and expanded to 29 cities



Emissions double reported figure. Evidence in parliament & planning debate

# Data management, formats and standards



# Data collection challenge

I would like you to collect the following information from within the room:

- Gender balance
- Agency representation
- Average family size
- *One other data point of interest to you*

I would like a spreadsheet each, you are not allowed to show anyone else your spreadsheet at any point.



# Exercise

Download [http://bit.ly/tz\\_source](http://bit.ly/tz_source)

- 1) Identify as many quality problems as you can with this dataset.
- 2) Create a new set of column titles that will help re-organise this whole dataset for better analysis.
- 3) Create a set of defined values that can exist in each column



# Exercise (10 mintes)

Create a plan to publish this dataset as a usable dataset

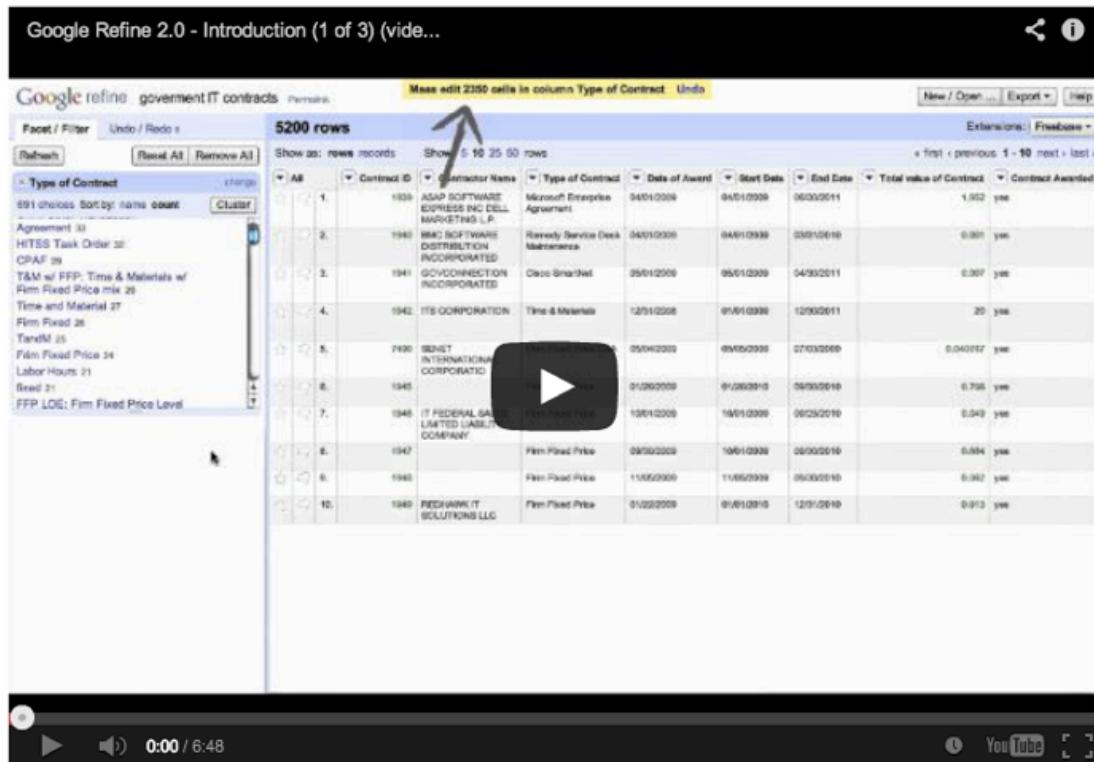


# Data cleaning



Knowledge  
for everyone

# Open Refine



Google Refine 2.0 - Introduction (1 of 3) (vide...

Google refine government IT contracts

Facet / Filter Undo / Redo ▾ Refresh Reset All Remove All

Mass edit 2350 cells in column Type of Contract Undo

5200 rows Show 5 10 25 50 rows

Contract ID Contractor Name Type of Contract Date of Award Start Date End Date Total value of Contract Contract Awarded

Contract ID	Contractor Name	Type of Contract	Date of Award	Start Date	End Date	Total value of Contract	Contract Awarded
1. 1938	ASAP SOFTWARE EXPRESS INC DELL MARKETING LP	Microsoft Enterprise Agreement	04/01/2008	04/01/2009	06/03/2011	1,952	yes
2. 1980	IBM SOFTWARE DEVELOPMENT INCORPORATED	Ranney Service Desk Maintenance	04/01/2009	04/01/2009	03/01/2010	0.001	yes
3. 1941	GOV-CONNECTION INCORPORATED	Cisco BranchNet	05/01/2008	05/01/2009	04/03/2011	0.007	yes
4. 1942	ITB CORPORATION	Time & Materials	12/01/2008	01/01/2009	12/03/2011	20	yes
5. 7199	SEREST INTERNATIONAL CORPORATION	ITIL Foundation	05/01/2009	05/05/2009	07/03/2009	0.04037	yes
6. 1945		ITIL Foundation	01/06/2009	01/06/2010	08/03/2010	0.705	yes
7. 1949	IT FEDERAL SALES LIMITED LIABILITY COMPANY	ITIL Foundation	10/01/2009	10/01/2009	06/29/2010	0.049	yes
8. 1947		Firm Fixed Price	09/09/2008	10/01/2008	09/03/2010	0.004	yes
9. 1940		Firm Fixed Price	11/05/2008	11/05/2009	05/03/2010	0.002	yes
10. 1940	RED HAWK IT SOLUTIONS LLC	Firm Fixed Price	01/02/2009	01/01/2010	12/01/2010	0.012	yes

YouTube 0:00 / 6:48

<http://openrefine.org>

Instances of refine are available via the menu at  
<http://training.theodi.org/InADay>

