

Open data in a day

<http://training.theodi.org/InADay>

Dr David Tarrant

@davetaz
theODI.org



Aim

Enable you to help build an open,
trustworthy data ecosystem at your
organisation.

The team



**Dr Jeni
Tennison**
CEO



**Sir Nigel
Shadbolt**
Chairman



**Sir Tim
Berners-Lee**
President

Founded in 2012, the Open Data Institute (ODI) is an international, independent and not-for-profit organisation based in London, UK.

Our mission

We work with
companies and
governments to
build an open,
trustworthy data
ecosystem.

Our vision

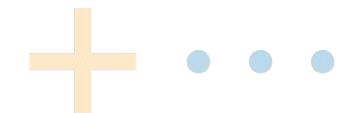
We want people,
organisations and
communities to
use data to make
better decisions
and be protected
from any harmful
impacts.

Agenda:

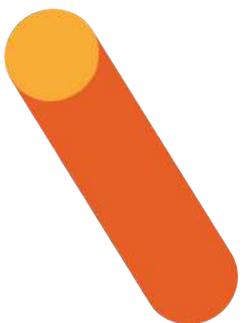
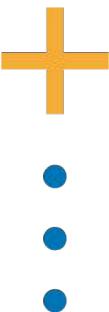
- 09:30** What open data is and isn't
- 11:15** Case studies in open and transparent ecosystems
- 12:45** **LUNCH**
- 13:30** Licenses and certificates
- 15:00** Accessing, cleaning and validating data
- 17:00** **END**

Introductions

- Your name
- What do you do?
- What was your first experience of open data



What is data?



Definition of data

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



Definition of data

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

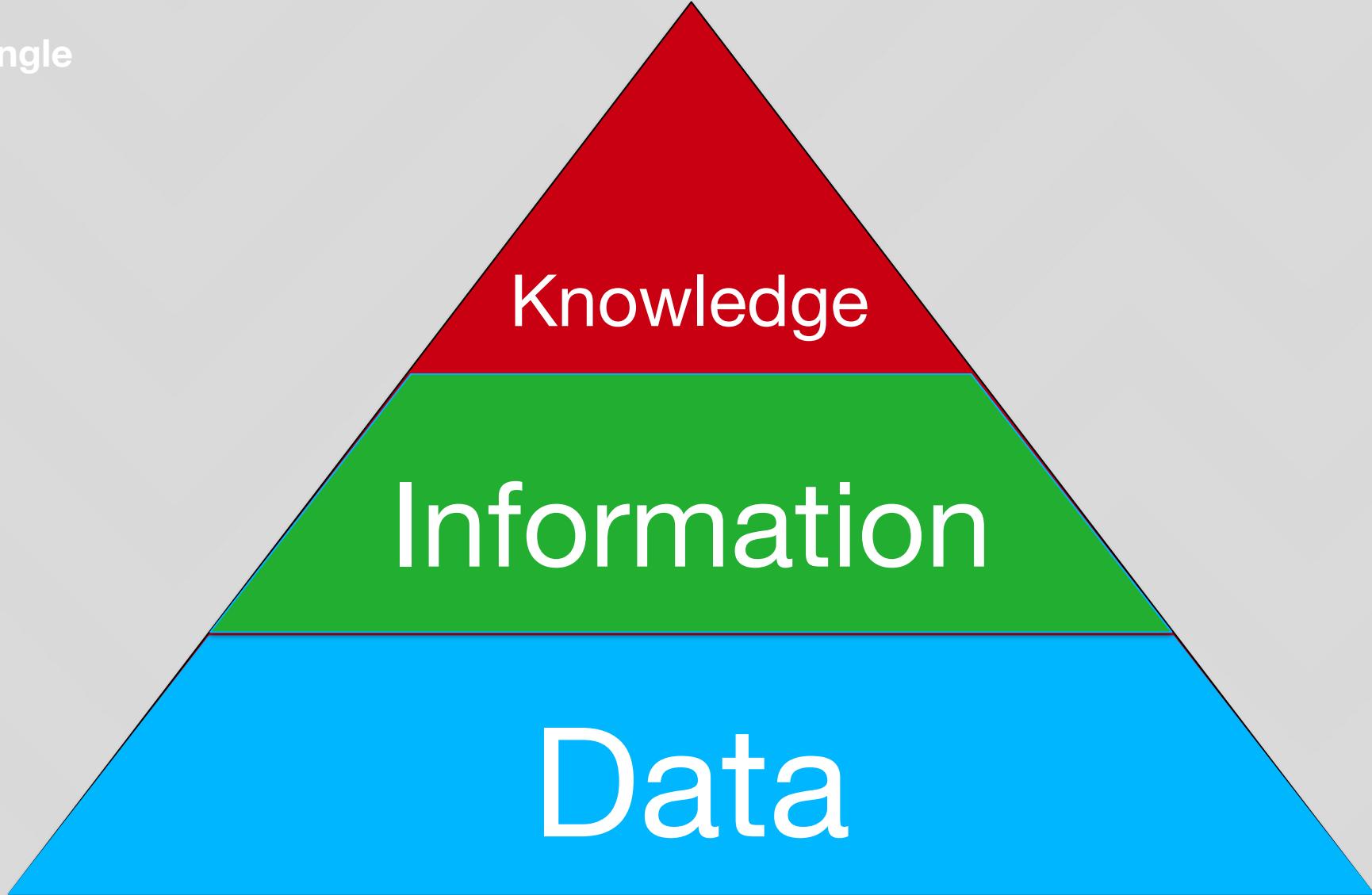


Definition of data

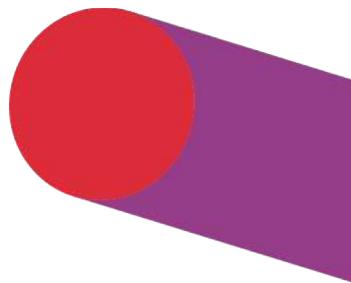
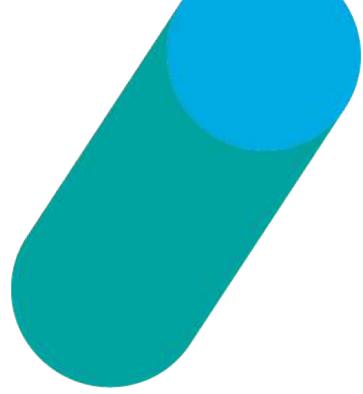
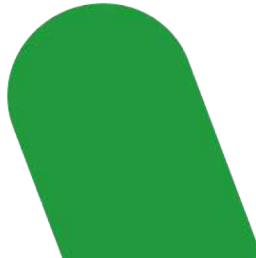
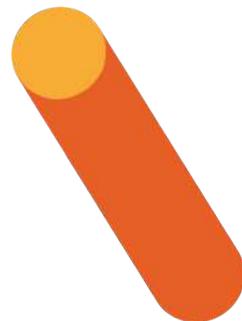
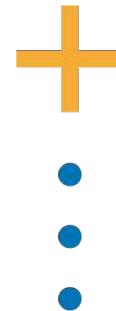
The lowest level of abstraction from which information and then knowledge are derived.



DIK(W) Triangle



What is open data?



Open Definition

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

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

Open Data Institute
Retired November 2014

Data that anyone can
access, use and share.



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



“

*Data is the new raw
material of the
digital age*

”

Francis Maude (Former MP)



©C

Our theory of change

How value is
created from data



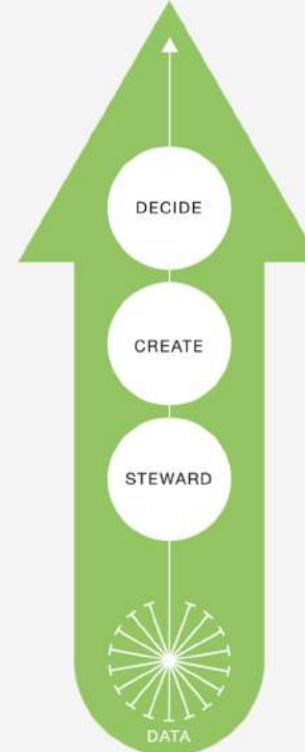
Our theory of change

Scenario:

What
happens
when we
hoard data –
the oil field

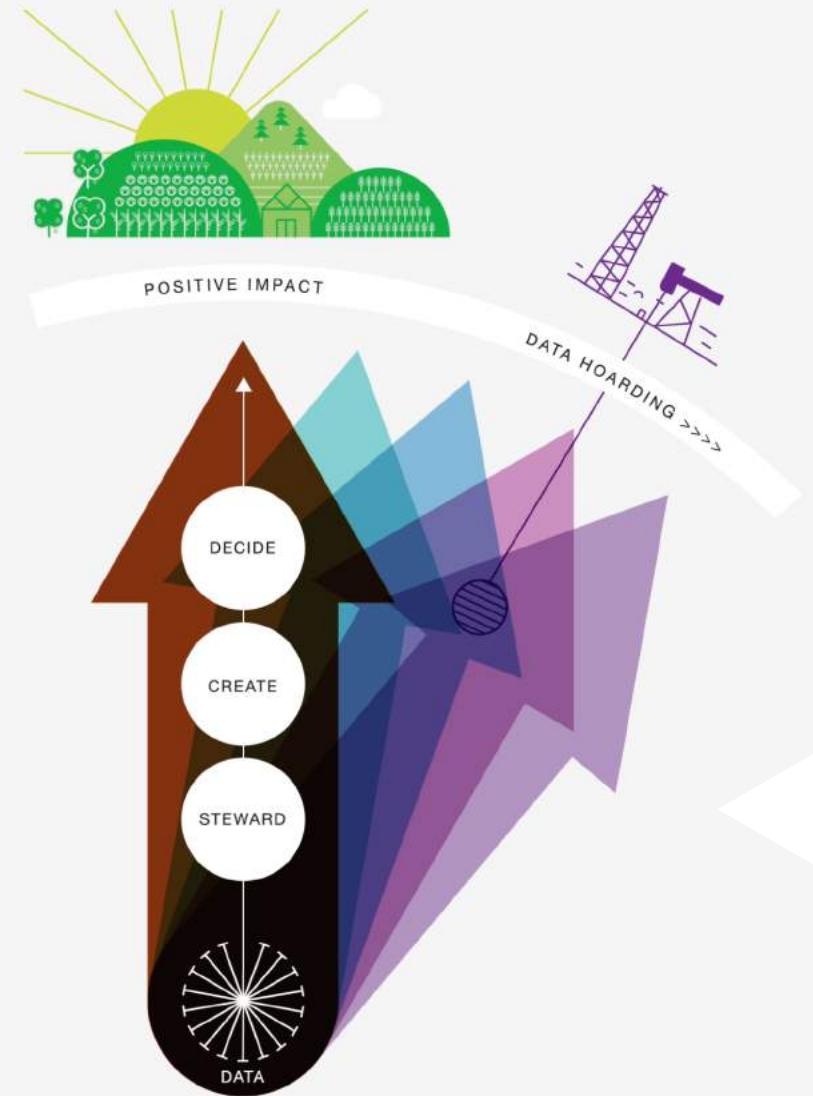


POSITIVE IMPACT



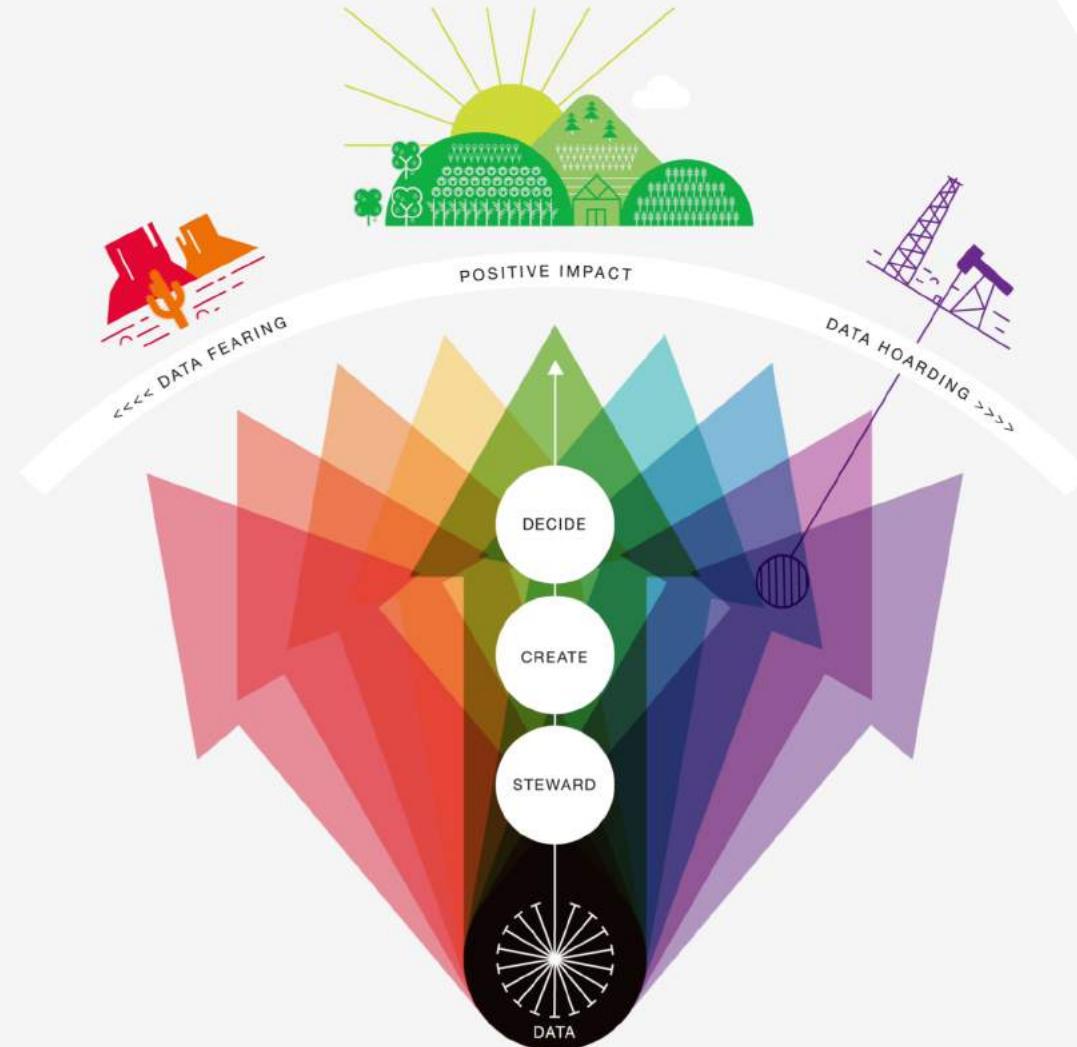
Scenario:

What
happens
when we
fear data –
the wasteland



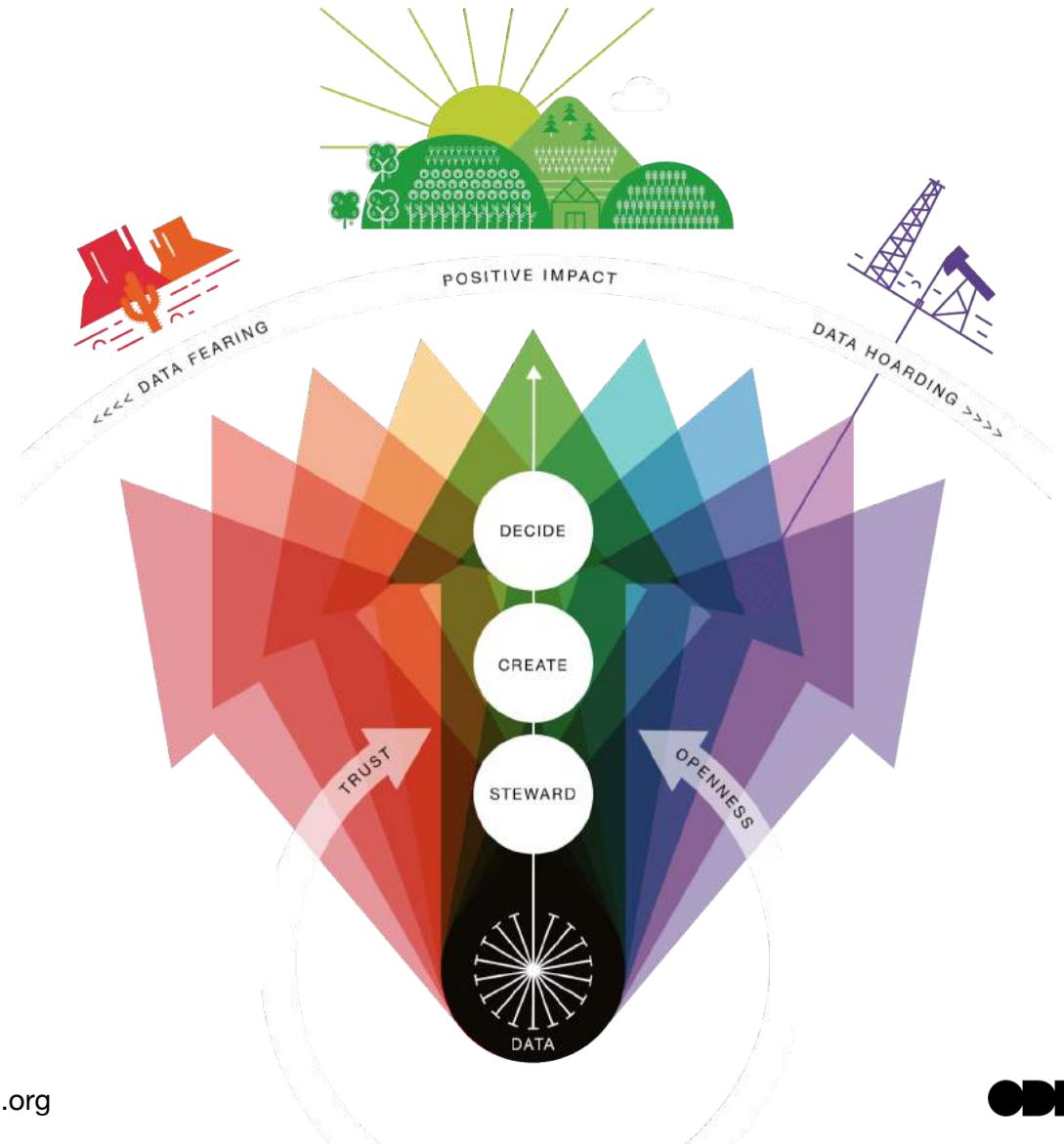
Our theory of change

We are one of many organisations working towards a good balance between encouraging and restricting how data is collected and used.



Focusing on positive impacts

What can we do to increase the trust and openness in data?

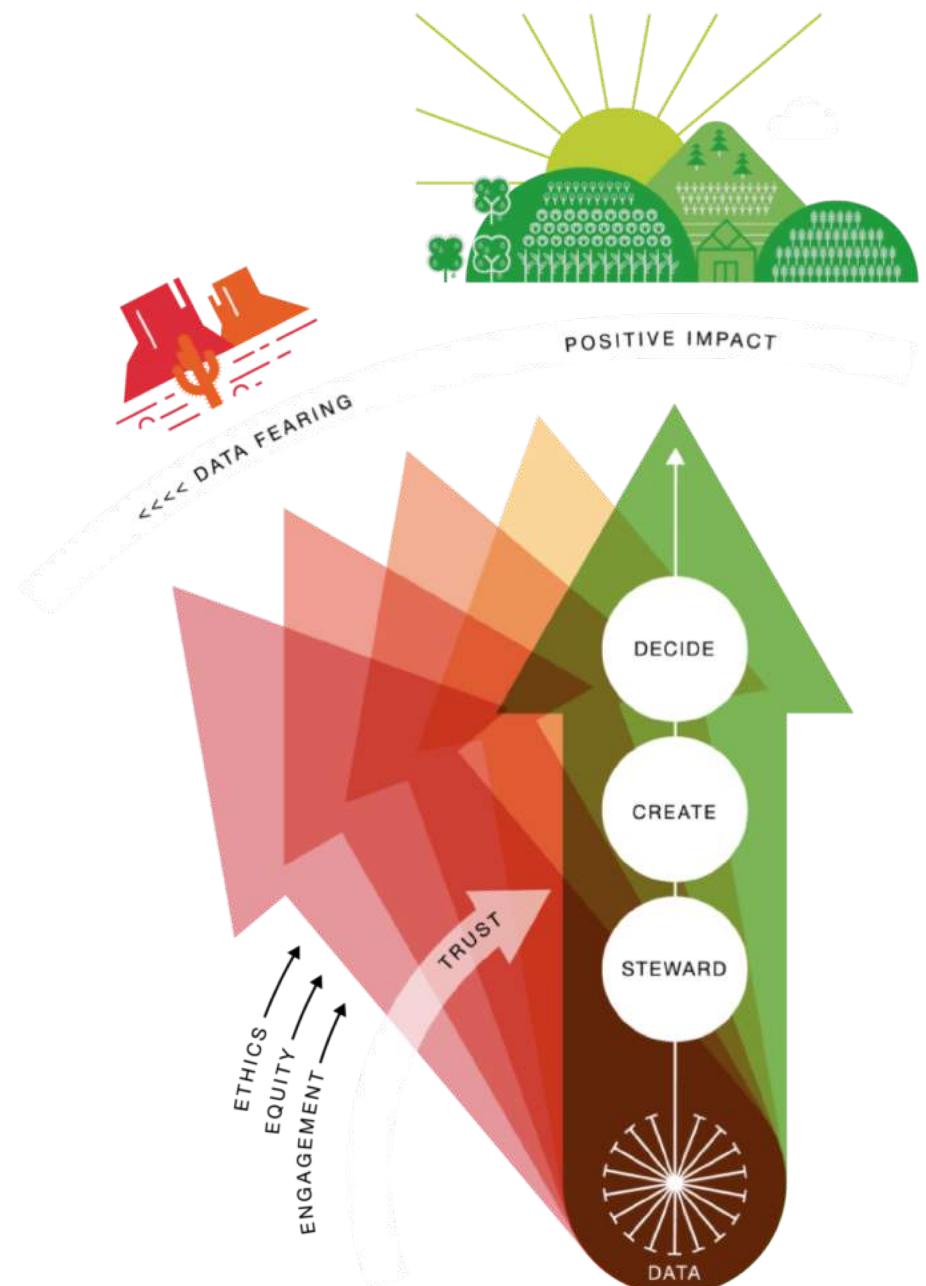


Increasing trust

There is a danger that unaddressed fears arising from legitimate concerns – such as who has access to data, the quality of the data and how it might be used – prevent us from realising its full benefits.

We advocate:

- building in **ETHICAL** considerations to data collection, management and use.
- ensuring that there is **EQUITY** around who accesses, uses and benefits from data.
- **ENGAGING** widely with people and organisations who are affected.

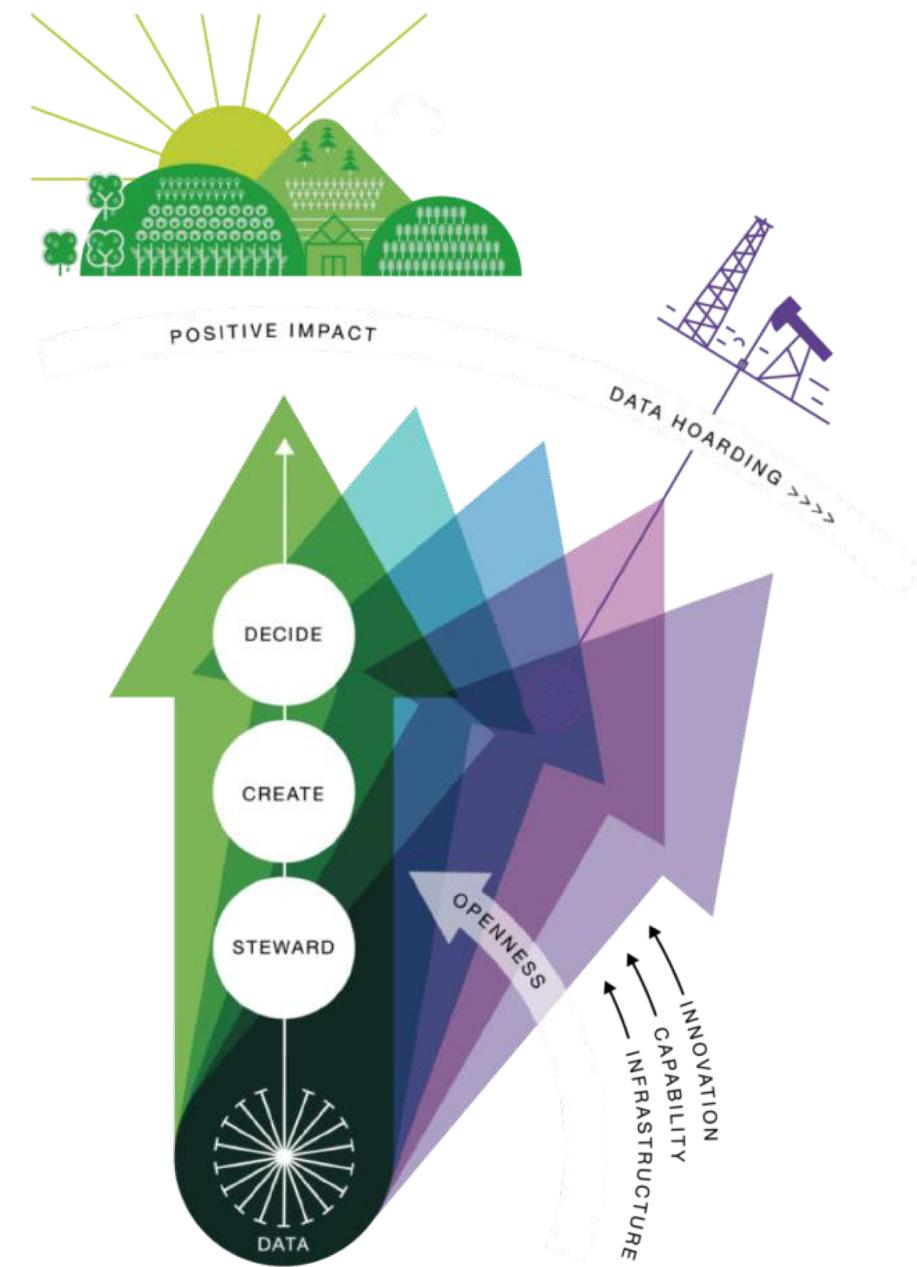


Increasing openness

There is a danger that in some places, organisations hoard data. They may use inappropriate business or funding models for data that limit the value that we, as a society, get from it.

We advocate for and support applying an **OPEN CULTURE** to the three pillars of how data is used within our economies:

- A data **INFRASTRUCTURE** that is as open as possible.
- Data literacy and **CAPABILITY** for all.
- Open **INNOVATION** for faster, shared benefit.



1. Infrastructure

Data is infrastructure

2. Equity

Data and rights

3. Engagement

Measuring success

4. Innovation

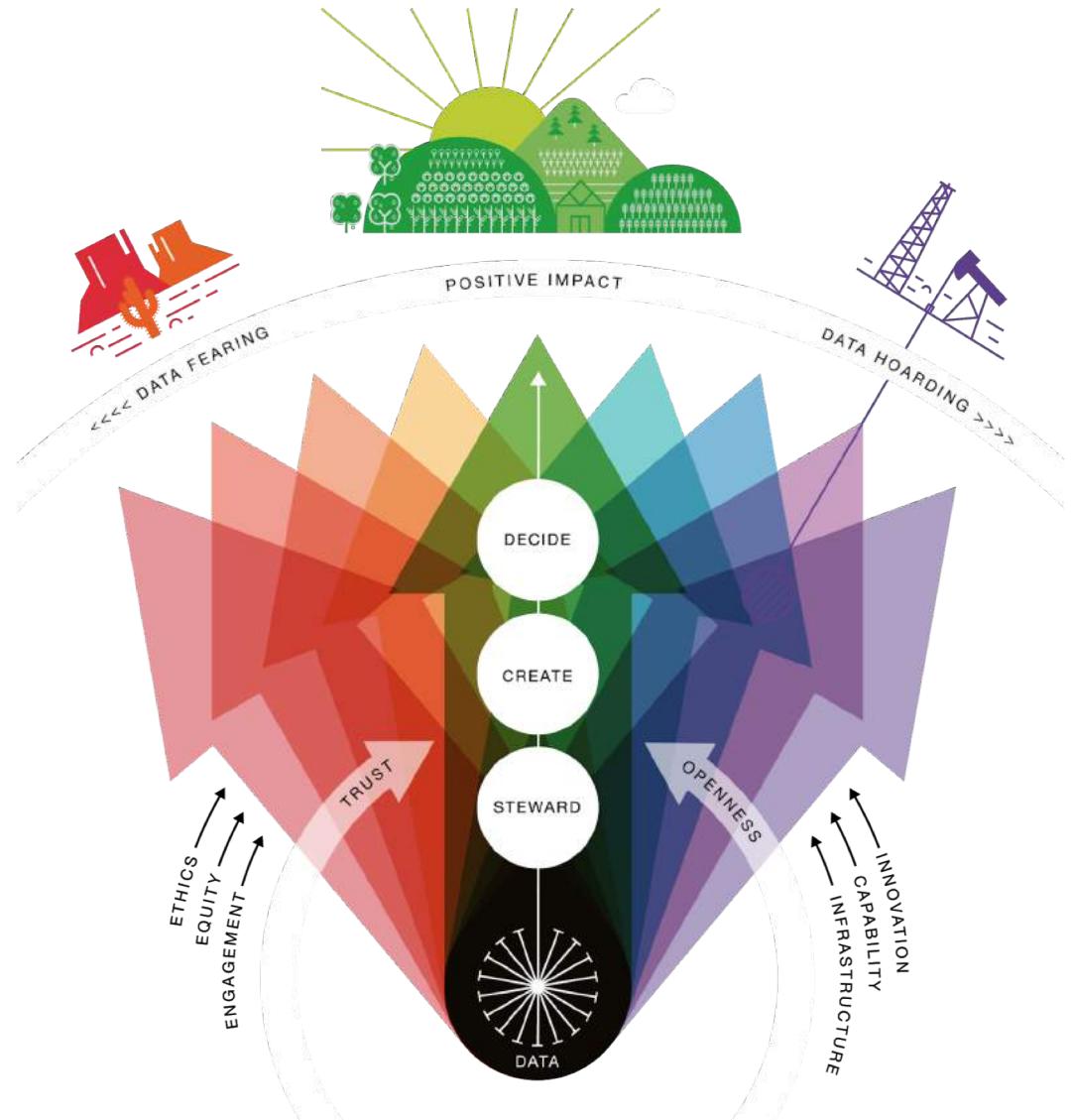
Growing economies

5. Capability

Making data usable

6. Ethics

Using data responsibly



Agenda:

- 10:00** What open data is and isn't
- 11:30** Case studies in open and transparent ecosystems
- 12:45** **LUNCH**
- 13:30** Licenses and certificates
- 15:00** Accessing, cleaning and validating data
- 16:00** **END**

1. Infrastructure

Data is infrastructure

2. Equity

Data and rights

3. Engagement

Measuring success

4. Innovation

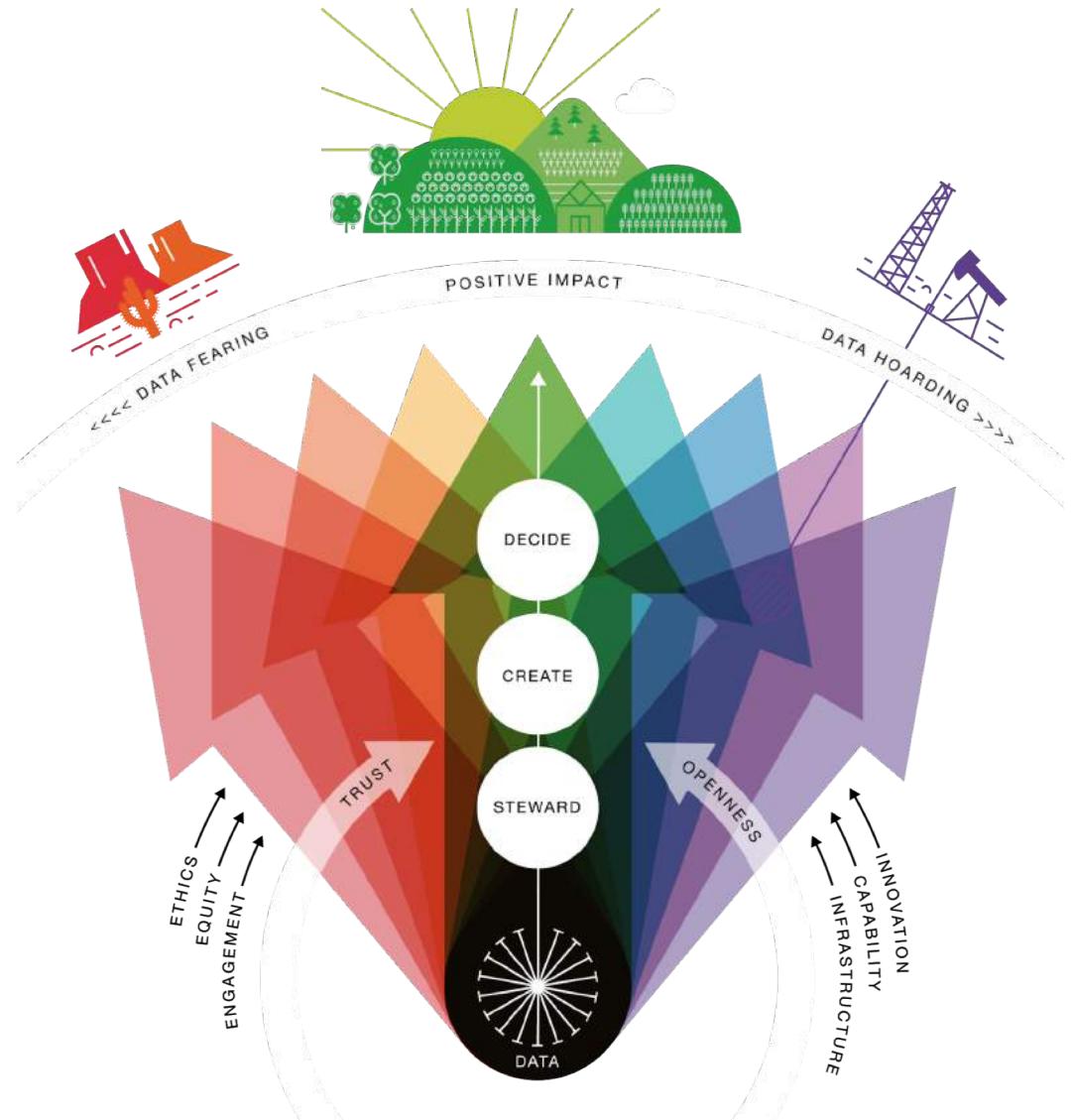
Growing economies

5. Capability

Making data usable

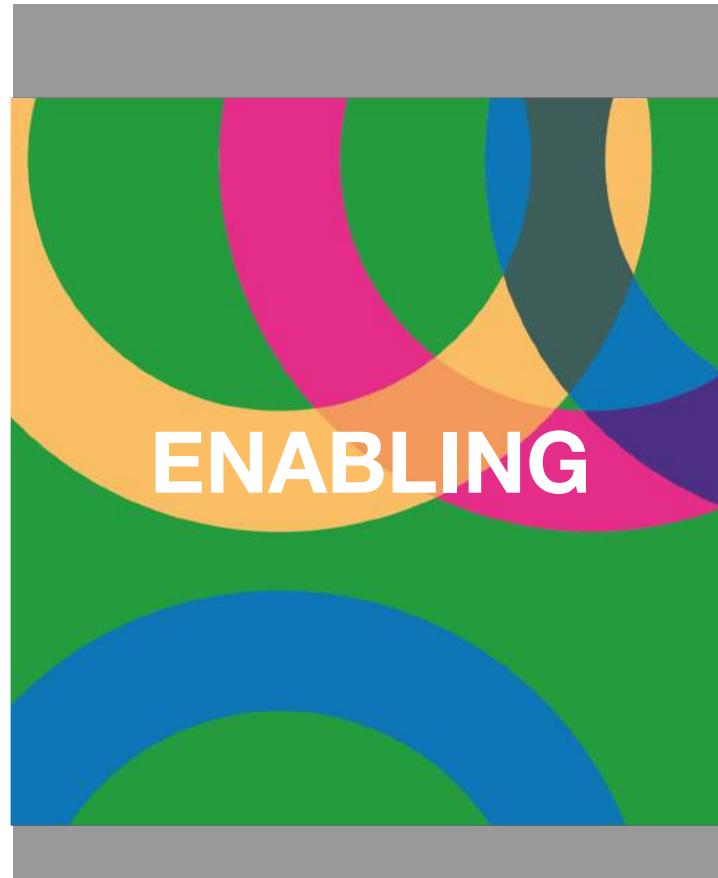
6. Ethics

Using data responsibly

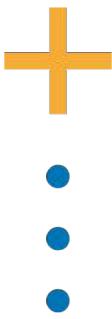


4. Innovation Growing economies

Justifications



What are your
favorite case
studies involving
open data?

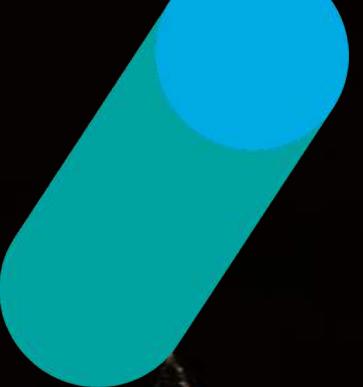
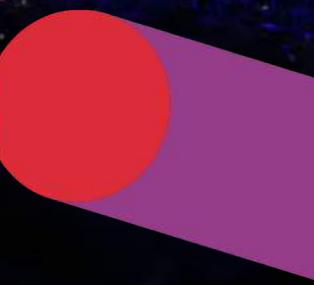


Open data that has changed the world*

Dr David Tarrant
@davetaz
theODI.org



*or at least a small part of it



Humanitarian Aid – Over 800 edits in 48 hours



OpenStreetMap

CC-by-SA www.itoworld.com

Map data www.openstreetmap.org 31 Jan 2010

Where to park (and income from fines)

Facts and Figures

Devon County Council



[Home](#)

Highways, Land and Buildings

[About the county](#)

On-street parking

[The people](#)

[Income from Civil Parking Enforcement Activities](#)

[Social issues and deprivation](#)

The number of parking spaces in Devon.

[The economy](#)

[January 2015](#)

[Education](#)

[August 2017](#)

[Housing](#)



[The environment](#)

[Transport](#)

Civil Parking Enforcement Policy

The Civil Parking Enforcement Policy is available in [PDF format](#) or in [Markdown format](#).

Markdown is a text-only version and is provided to enable easier reuse of the policy, such as copying and pasting to a new document – it should not be relied on for accuracy and the PDF version remains the definitive version.

[Community Safety](#)

[Open data](#)

[Area profile](#)

[Community Performance](#)

The Great British Public Toilet Map



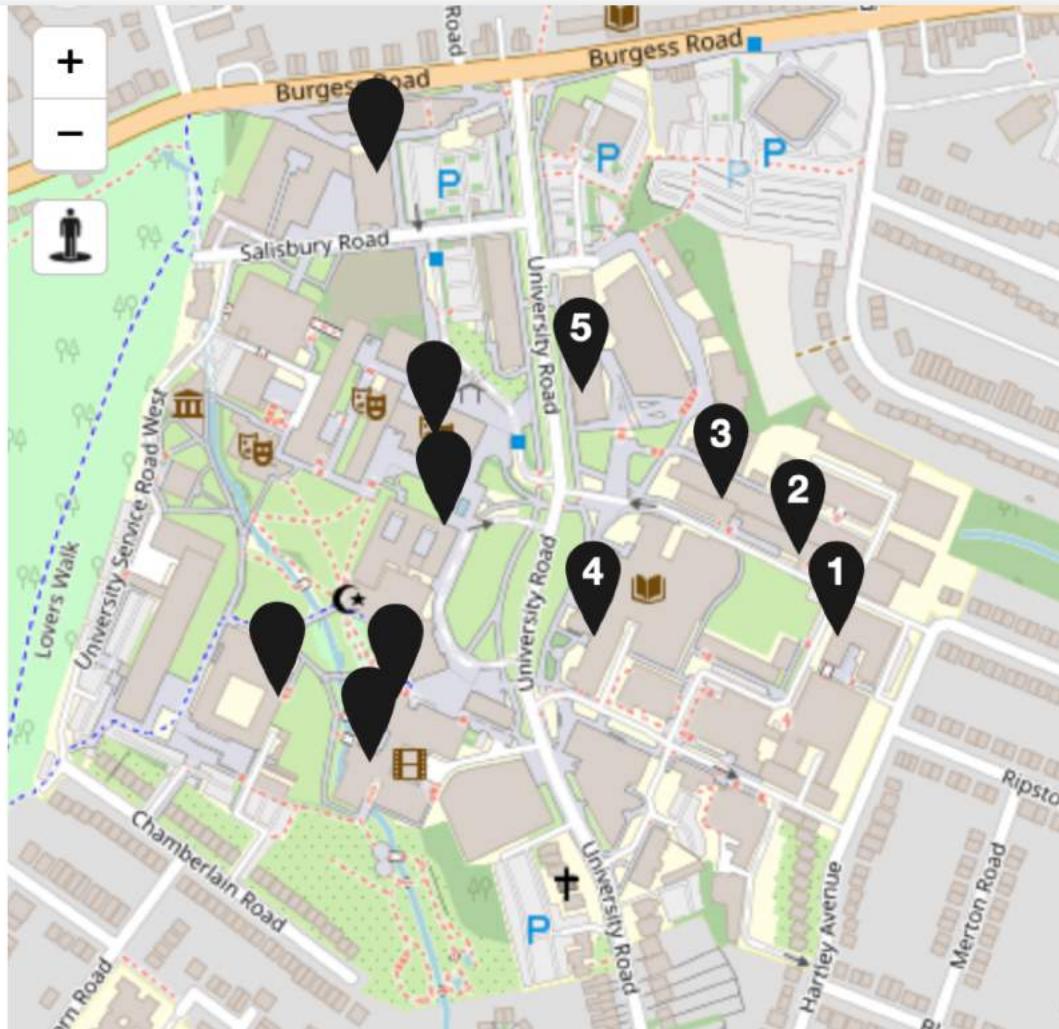
The Great British
Public Toilet Map

proudly sponsored by

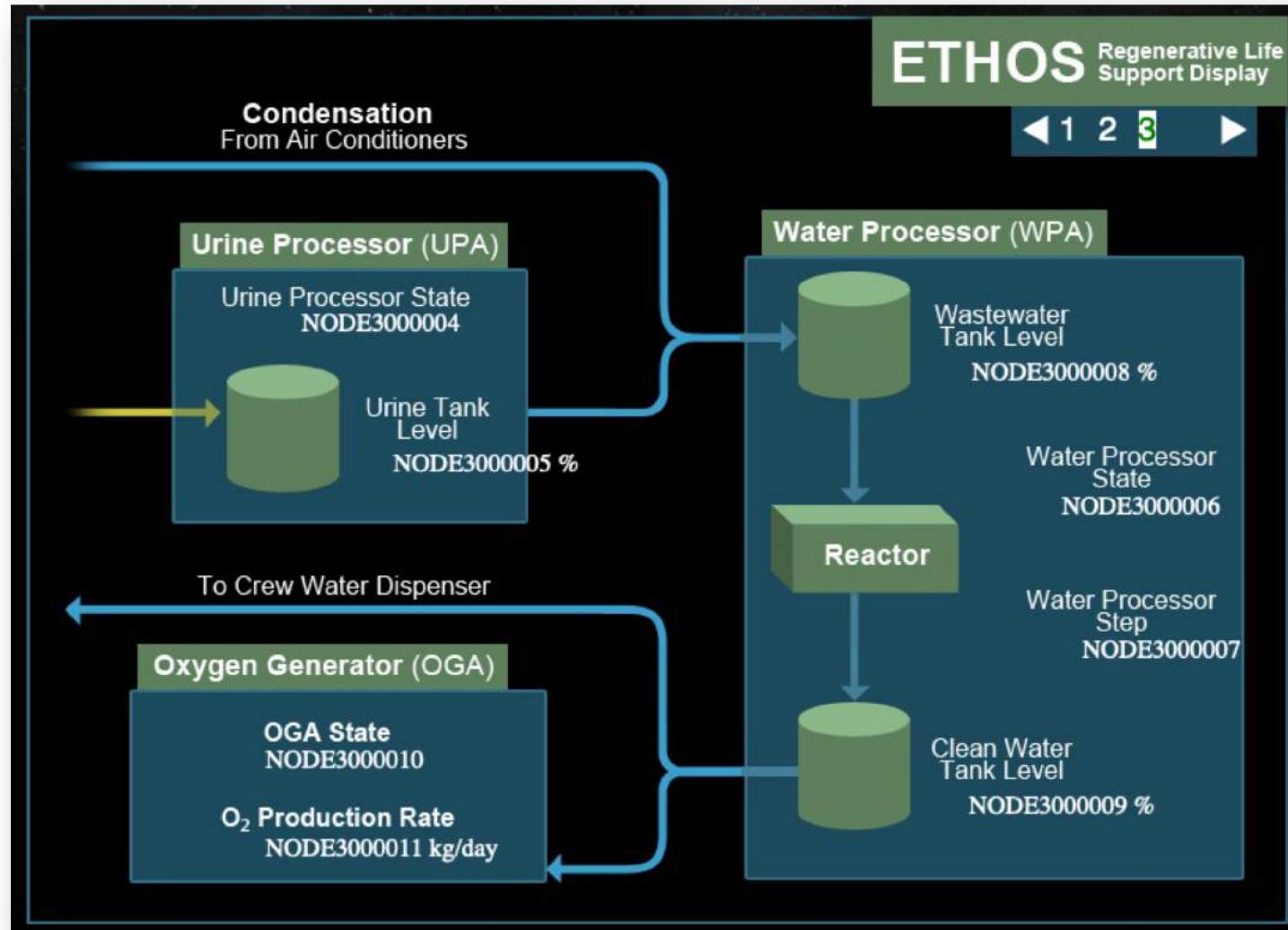
About the map

The Great British Public Toilet Map is the UK's largest database of publicly-accessible toilets, with over 11000 facilities.

It is run by **Public Convenience**, a joint venture between researchers Dr Jo-Anne Bichard and Gail Ramster (who created the map at the **RCA Helen Hamlyn Centre for Design**) and software development company **Neontribe** (who designed and built it).



Astronauts going to the loo



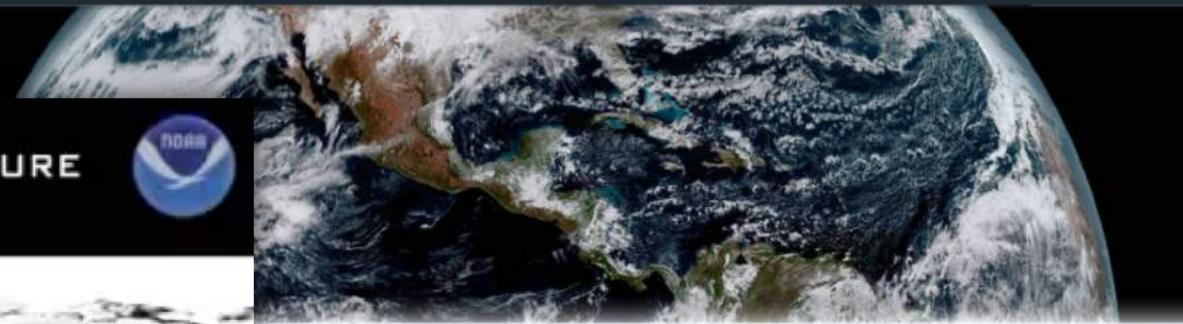
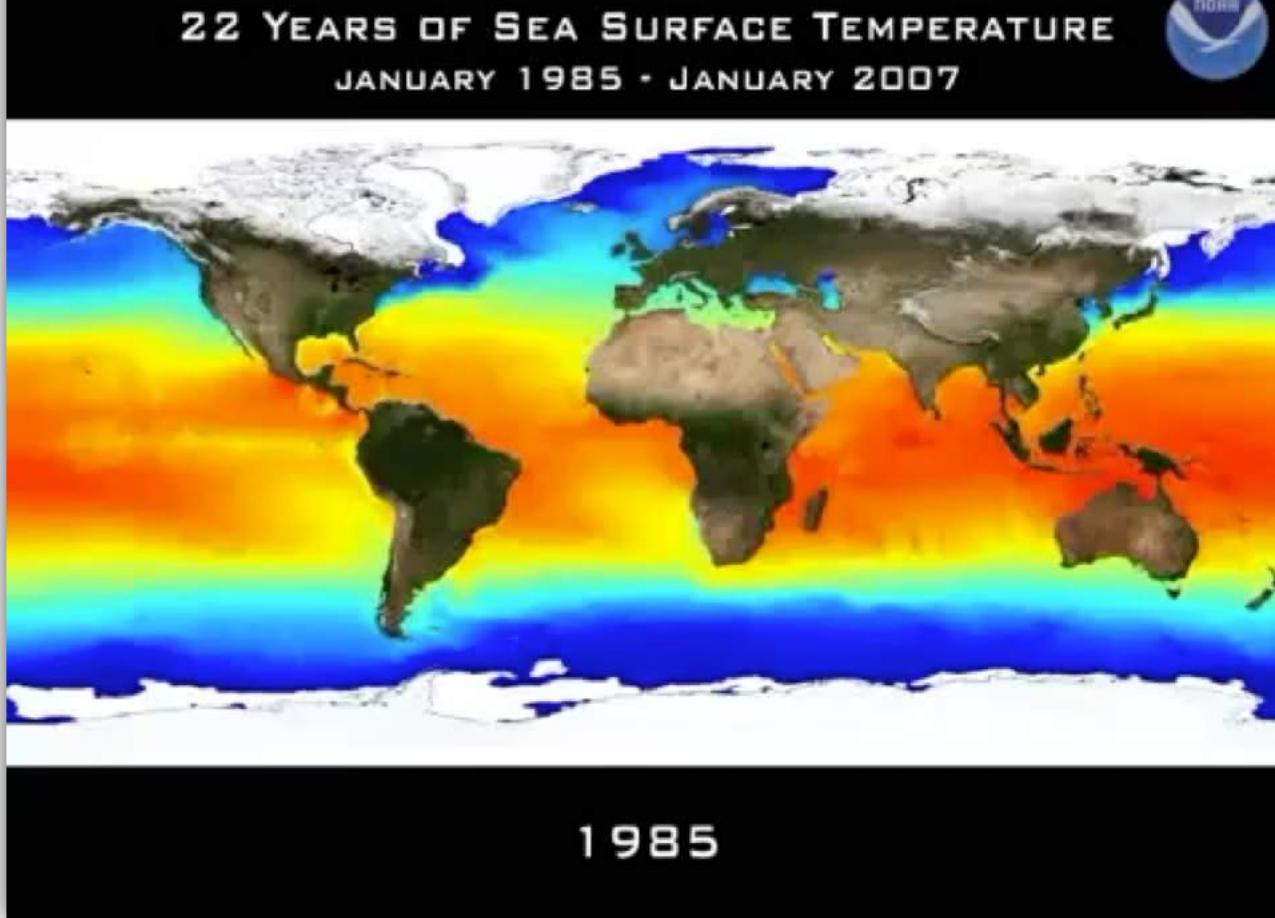
Keeping an eye on our planet



OneStop

NATIONAL OCEANIC AND
ATMOSPHERIC ADMINISTRATION

[Home](#) | [About Us](#) | [Help](#)



OneStop

A NOAA Data Search Platform

Search NOAA Data



Explore Popular Topics



Climate



Satellites



Fisheries



Coasts



Oceans

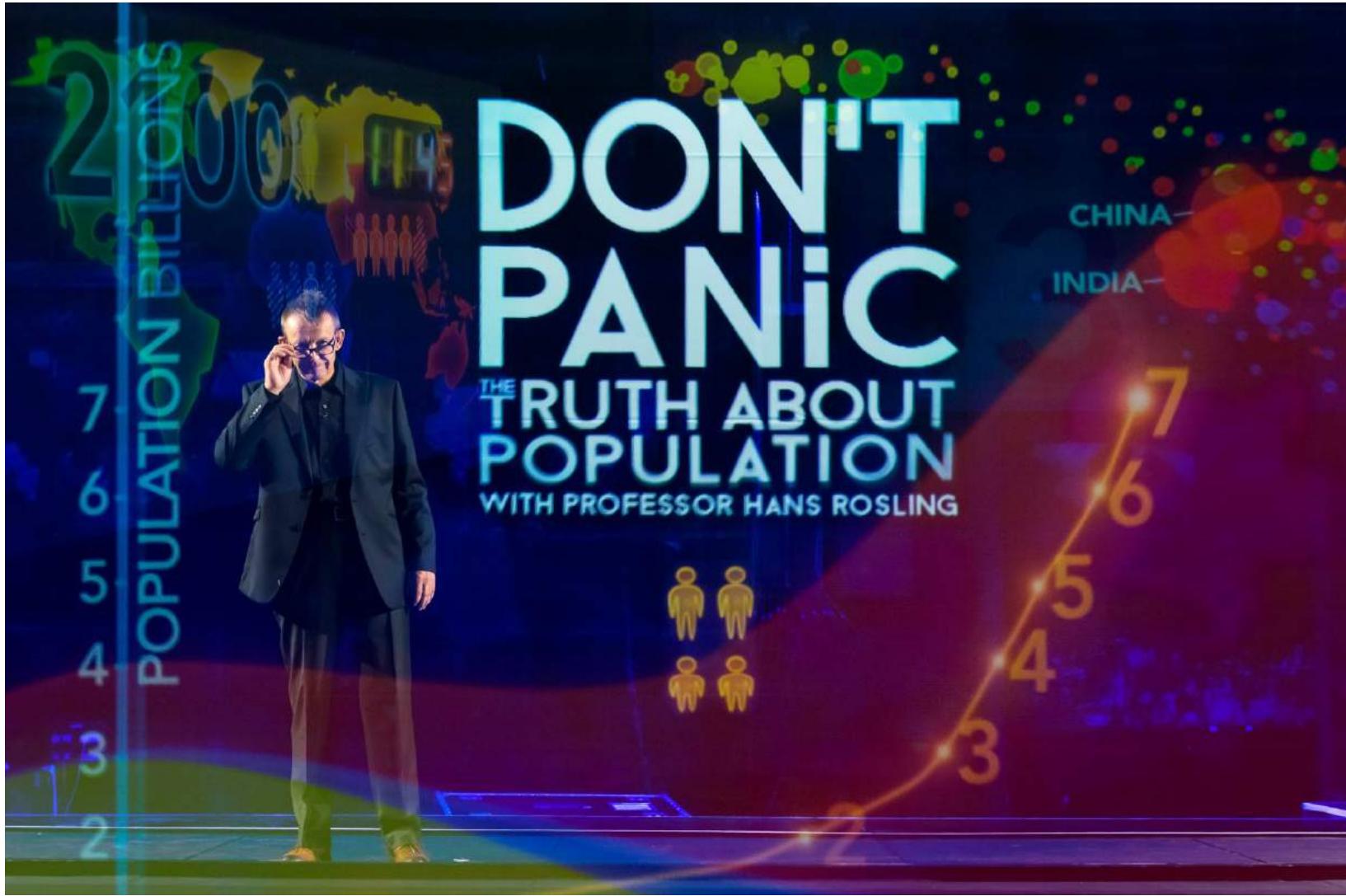
What do you know about the world?

What is the average number of children per family in Bangladesh?

- ◆ A: 2
- ◆ C: 4

- ◆ B: 3
- ◆ D: 5

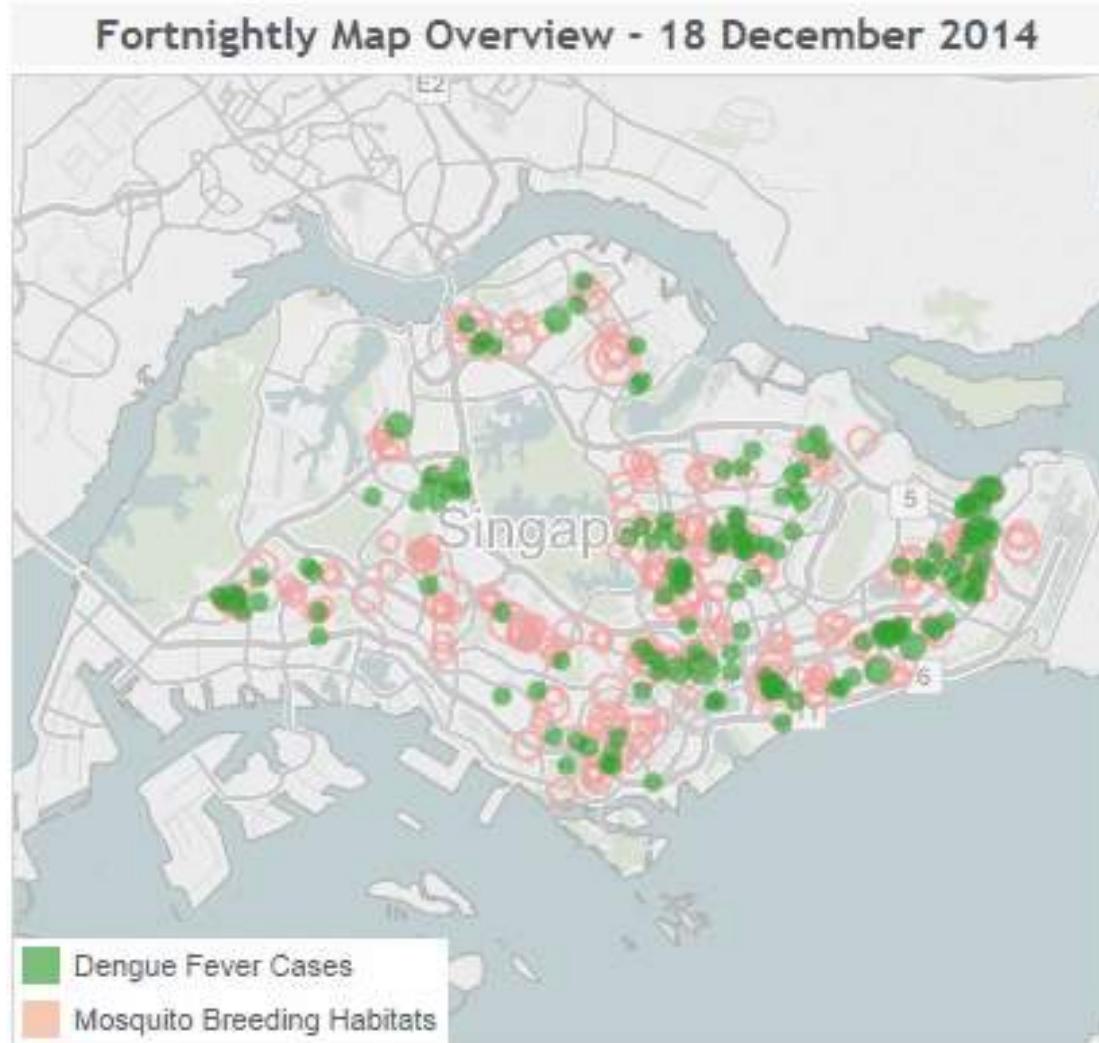




gapminder.org



Dengue maps in Singapore



Innovation in bin collection?

 **Darren Tarbard**
@tarbard

Follow ▾

I made a device that checks the council website and lights up to show you what colour bin to put out. I call it the Bindicator.



9:17 AM - 1 Jun 2018

6,002 Retweets 30,530 Likes



Crime in the USA

KEY

*Scale = 300 deaths



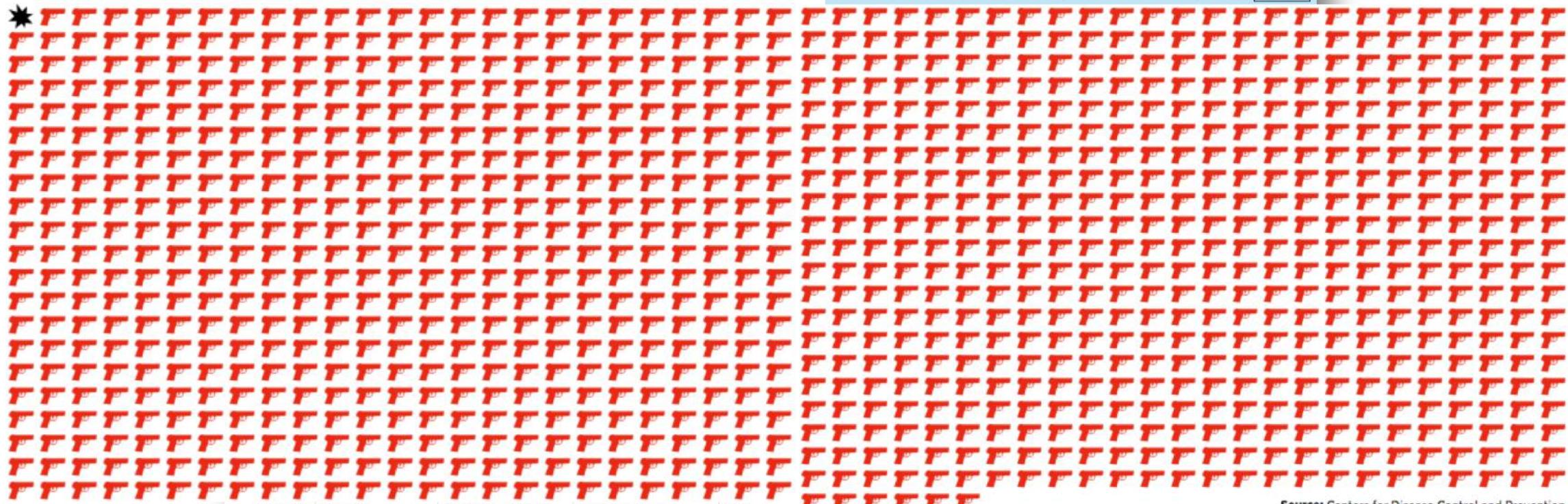
American Deaths by Terrorism

Total = 313



American Deaths by Firearms on U.S. Soil

Total = 316,545

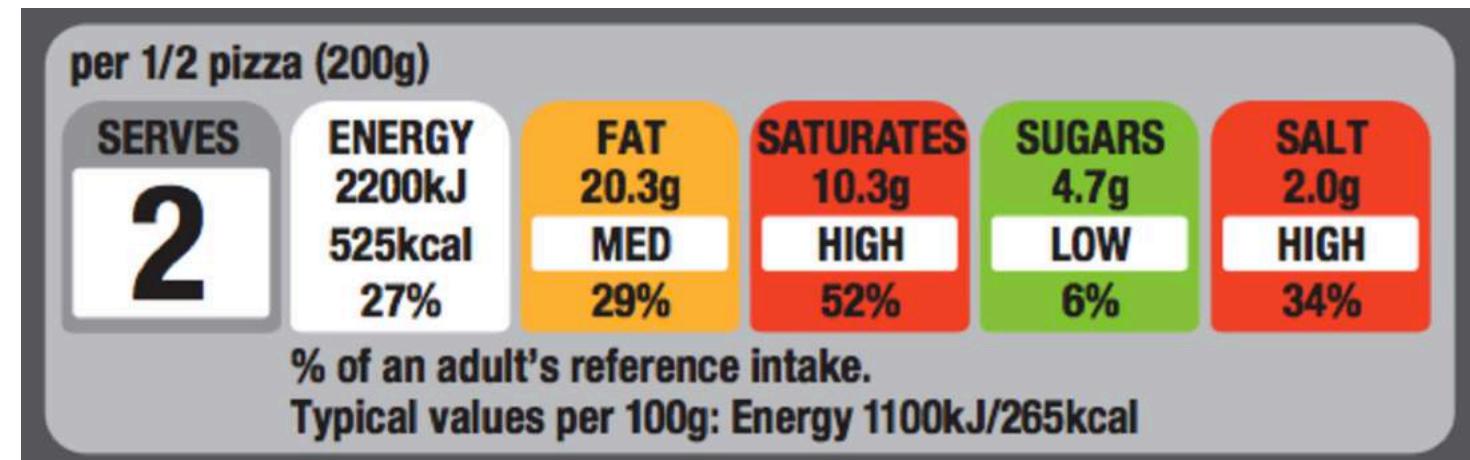


Source: Centers for Disease Control and Prevention

| | | | | | | |
|----------|----------|-------------------------|----------------------|------|-----------|--------------------------------|
| 11602135 | JC160018 | 02/20/2019 11:06:00 ... | 010XX E 45TH ST | 031A | | |
| 11602133 | JC160021 | 02/20/2019 11:05:00 ... | 062XX N KENMORE AVE | 0486 | | |
| 11602115 | JC160029 | 02/20/2019 11:05:00 ... | 052XX S WESTERN BLVD | 1812 | NARCOTICS | POSS: CANNABIS MORE THAN 30GMS |
| 11602117 | JC160019 | 02/20/2019 11:04:00 ... | 079XX S HALSTED ST | 2028 | NARCOTICS | POSS: SYNTHETIC DRUGS |

Showing Reported Crimes 1-100 out of 6,815,059

Making us healthier



Open Active reach

- **c.20%** of UK leisure operators publishing open data
- **170k** opportunities per month
- **Change4Life** activity finder has launched
- Run an **Accelerator**
- **Created standards** - Modelling Spec v2.0, Booking spec v1.0
- **Grown a community**, including Advisory Group, W3C, and champions



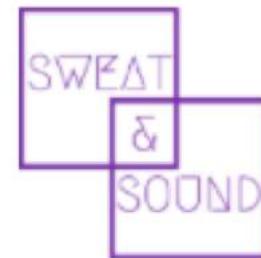
Sessions per month



Sessions per Month



The OpenActive accelerator



4m have signed up¹

98% state funded primary schools engaged²

93% of mothers aware of Change4Life trust it³

Sources: 1. PHE digital tracking 2. Edcoms survey with teachers (2016) 3. Kantar Public General Population Survey with mothers of 5 – 11s (2017 & 2015)

The screenshot shows the Change4Life website's 'Activity finder' section. At the top, there are navigation links: 'Skip navigation', 'Change4Life', 'Activities', 'Sports and activities', 'Activity finder'. Below the navigation is the 'change 4 life' logo. To the right of the logo are four icons: 'Food facts', 'Recipes', 'Activities', and 'Your child's weight'. A cartoon character of a person in an orange suit is on the right side of the header.

The main heading 'Activity finder' is in red. Below it is a search bar with fields for 'Activities' (set to 'All'), 'Location or Post Code' (set to 'London'), 'Range' (set to '8 m'), and 'Date' (set to 'All'). A yellow 'Search' button is to the right.

Below the search bar is a map of London with several yellow location pins. The map includes labels for various areas like Northwood, Harrow, Wembley, Greenford, Hayes, Ealing, Paddington, Westminster, Southwark, Lewisham, Greenwich, and Bexleyheath. A legend indicates that the pins represent 'Activity' locations.

Below the map is a table titled 'Find Activities BETA' showing three activity listings:

| Details | Distance | Cost | |
|---|-----------|-------|-------------------------|
| Toddler Time 7th December 18 - 10:30 to 12:00 Ironmonger Row Baths, 1 Norman Street | 0.3 miles | | Enquire |
| Athletics Courses 7th December 18 - 10:30 to 10:59 Paddington Rec Ground, Randolph Avenue | 3.8 miles | £6.95 | Enquire |
| Swim for Under 5's and Adults 7th December 18 - 10:30 to 12:00 Kensington Leisure Centre, Silchester Road | 5.0 miles | | Enquire |

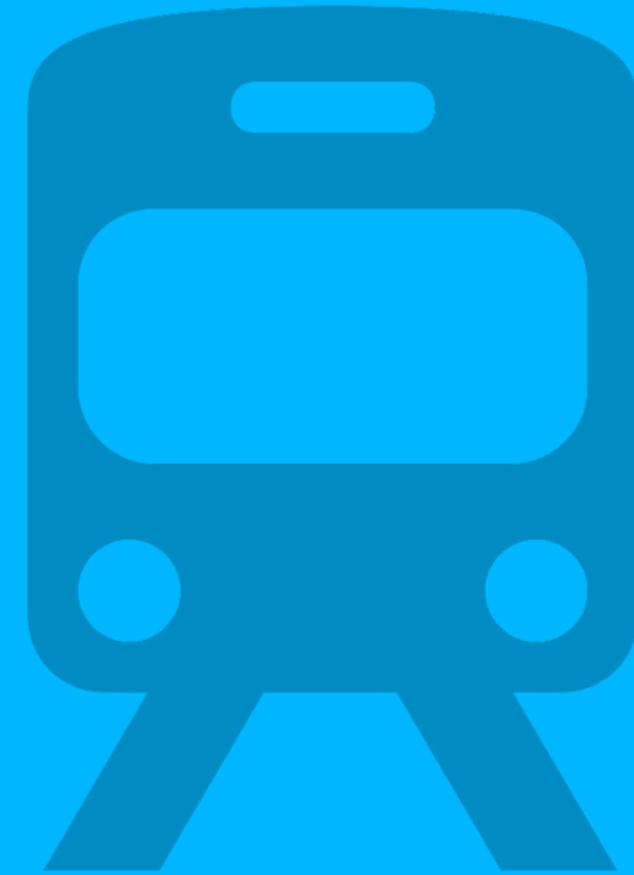
Over 700 Apps now rely on open transport data:

City Mapper (Millions of users)

Google Maps (Millions of users)

Real time trains (500,000 users)

Station master (10,000's users)



Station master – TFL winner

Earl's Court Exits

District Line

Carriage 5, Door 3 - Right hand side.

Lift

Carriage 3, Door 3 - Right hand side.

Way out (Stairs)

Carriage 5, Door 1 - Right hand side.

Central West Acton Lines

CENTRAL (EALING BROADWAY BRANCH)
WESTBOUND
PLATFORM 1

Platform Accessibility Information

Fully accessible on foot. There is step access to the platforms.

Some effort required via wheelchair. There is step access to the platforms.

Some effort required via stroller/baby buggy. There is step access to the platforms.

No access to this platform via wheelchair.

PICCADILLY LINE
EASTBOUND
PLATFORM 3

Gap Information

Gap between the train and the platform: 102mm

Step down to the train from the platform: 187mm

Additional Information

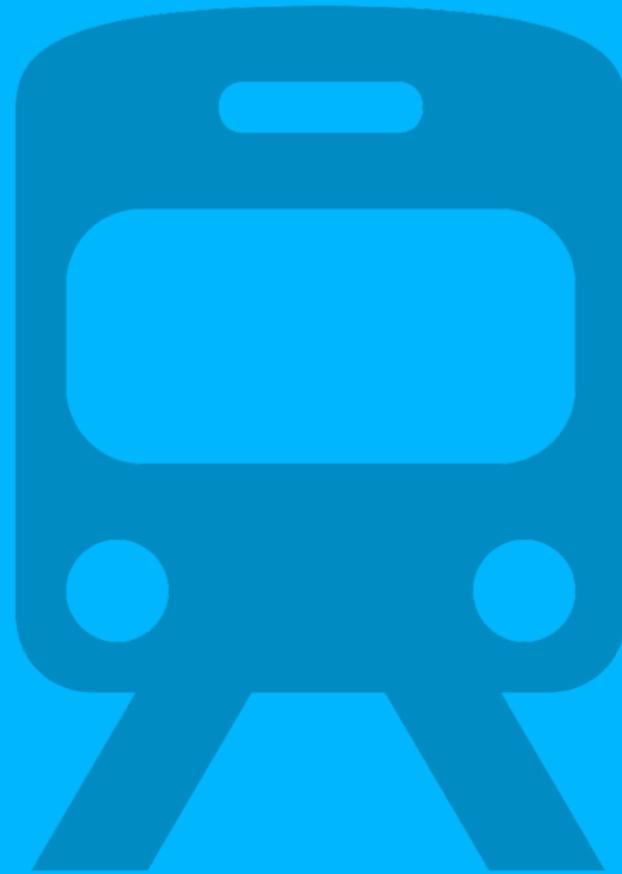
There are 36 steps in total between the platform and ticket barriers, split into two sets of 18.

Accessible without escalators Yes

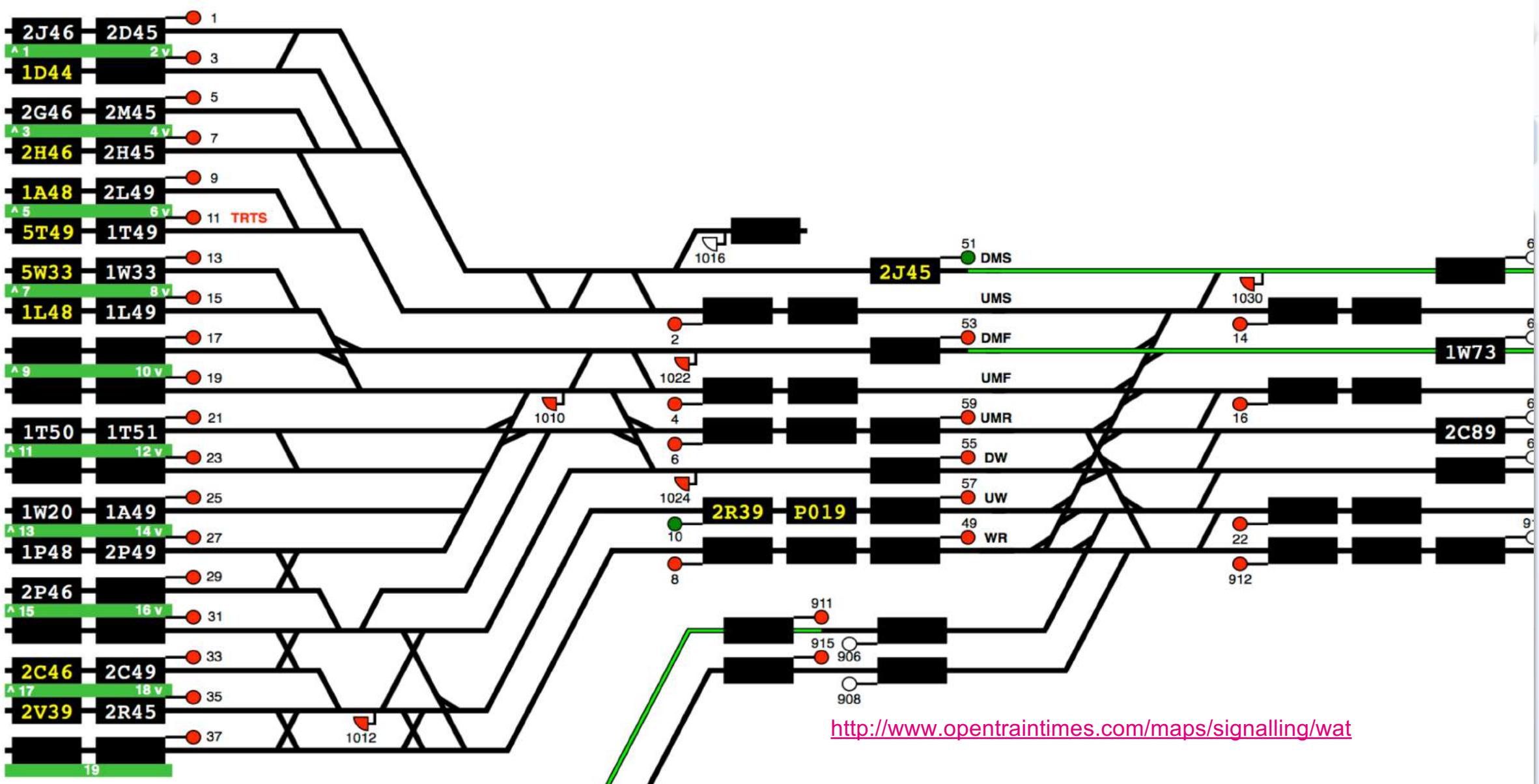
£130m

(and rising)

The yearly benefit to
London of open transport
data from Transport for
London



<https://tfl.gov.uk/info-for/media/press-releases/2017/october/tfl-s-free-open-data-boosts-london-s-economy>



| | | | | |
|--|-------|--------------------------|---------|-------------------------|
| | 16:28 | Windsor & Eton Riverside | On time | Details |
| | 16:30 | Portsmouth Harbour | On time | Details |

Car emissions under real driving conditions

EQUA® INDEX Aq ☁ AIR QUALITY

Clear filters

Show 10 entries Search:

| Make ▲ | Model ▲ | Fuel Type ▲ | Model Year ▲ | Engine Size L ▲ | Power Bhp ▲ | Drive Train ▲ | Driven Wheels ▲ | Transmission ▲ | Euro Stage ▲ | EQUA Aq Rating ▲ |
|-----------|----------|-------------|--------------|-----------------|-------------|---------------|-----------------|----------------|--------------|------------------|
| Audi | A8 | Diesel | 2013 | 3.0 | 247 | 4WD | 4 | Automatic | Euro 5 | H |
| BMW | 1 Series | Diesel | 2012 | 2.0 | 141 | RWD | 2 | Manual | Euro 5 | H |
| BMW | X3 | Diesel | 2012 | 2.0 | 181 | 4WD | 4 | Manual | Euro 5 | H |
| BMW | X3 | Diesel | 2013 | 2.0 | 141 | RWD | 2 | Manual | Euro 5 | H |
| BMW | X5 | Diesel | 2013 | 3.0 | 376 | 4WD | 4 | Automatic | Euro 5 | H |
| Chevrolet | Cruze | Diesel | 2013 | 1.7 | 128 | FWD | 2 | Manual | Euro 5 | H |
| Fiat | 500L | Diesel | 2013 | 1.6 | 105 | FWD | 2 | Manual | Euro 5 | H |
| Fiat | 500X | Diesel | 2015 | 1.6 | 118 | FWD | 2 | Manual | Euro 6 | H |
| Ford | Galaxy | Diesel | 2005 | 1.9 | 130 | FWD | 2 | Manual | Euro 3 | H |
| Hyundai | i30 | Diesel | 2013 | 1.6 | 108 | FWD | 2 | Manual | Euro 5 | H |

Showing 1 to 10 of 64 entries (filtered from 783 total entries)

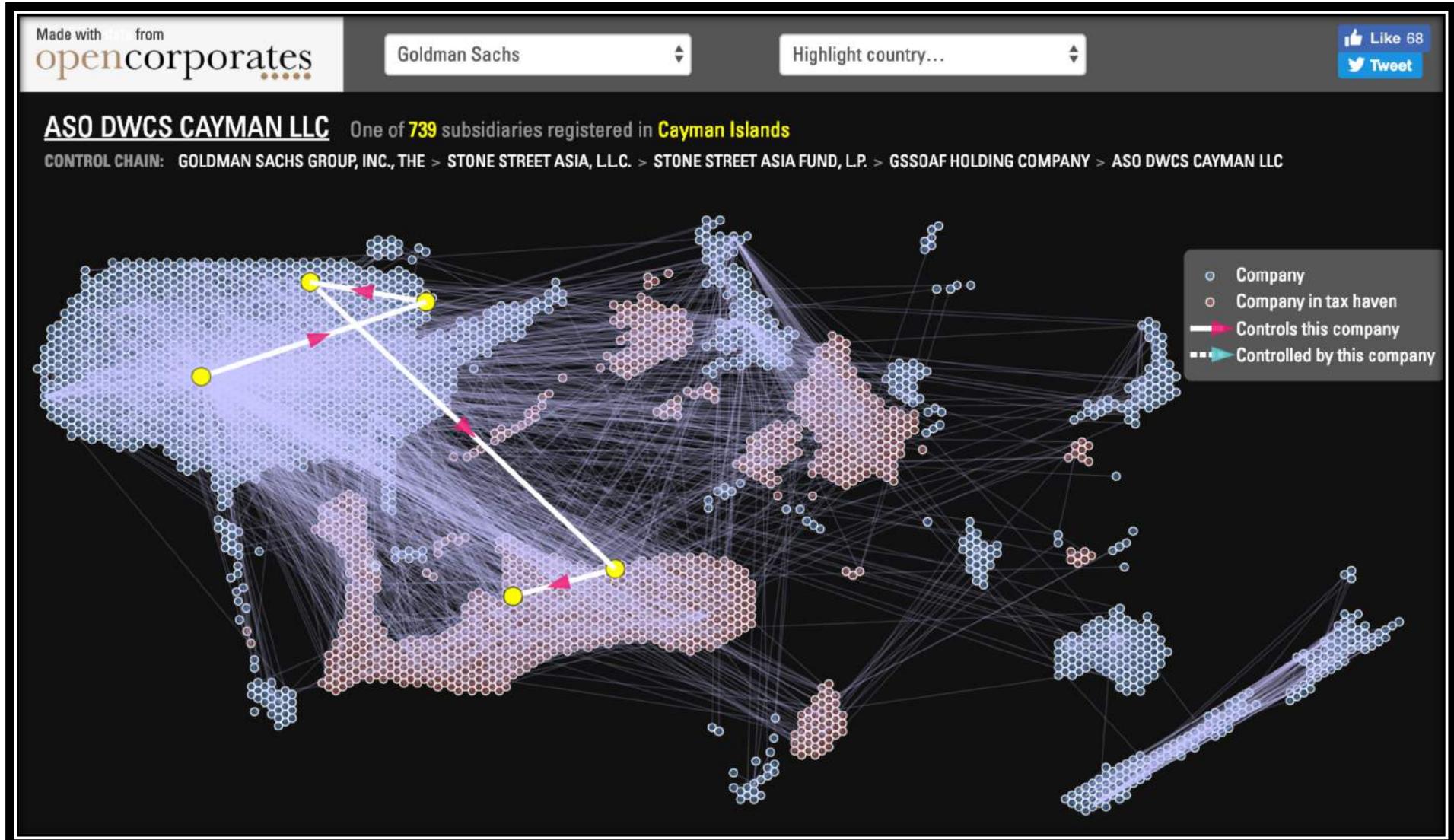
« < 1 2 3 4 5 6 7 > »

Planting the right urban forests – Mexico city

Open data revealed that habitats were being lost as the wrong trees were being planted.



Open Corporates – The Facebook network of companies



Open banking

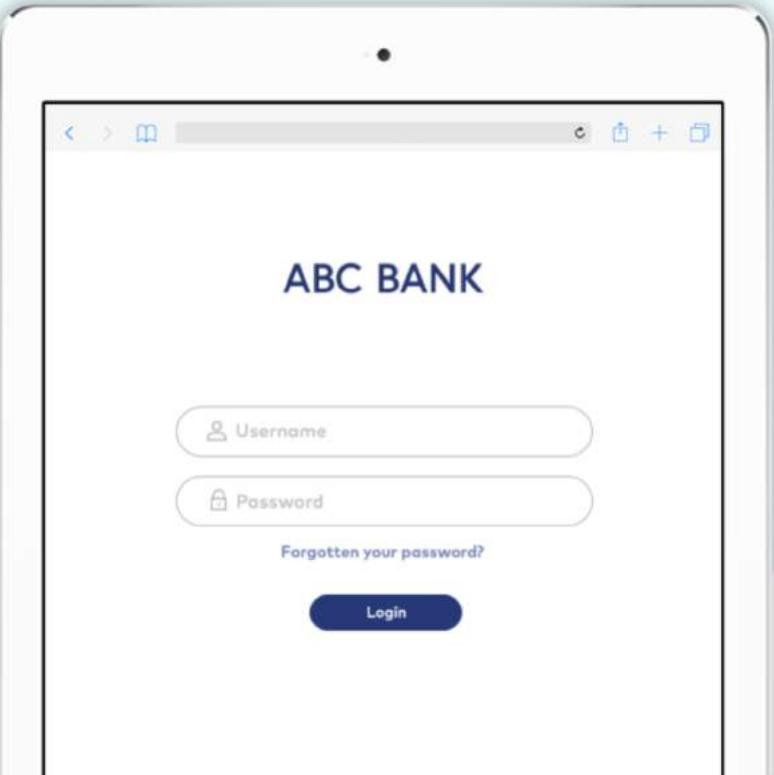
OPEN BANKING

What is Open Banking? FAQs Find a provider Insight

The future of money, where you're in control

Open Banking puts you in control of your data: an easier way to move, manage, and make more of your money.

Discover how



Saving money on FOI requests

The screenshot shows the Data.SA website homepage with a banner highlighting a \$3.2m saving. Below the banner is a navigation menu with links to Home, Datasets, Organisations, Groups, Strategies, Toolkit, Apps & Ideas, and a search bar. The main content area features a purple header "SA Police Reduce Costs Through Proactive Open Data" and a large blue title "SA Police Reduce Costs Through Proactive Open Data". The text discusses the estimated costs of administering the Freedom of Information Act 1991 (SA) in 2011-12 financial year, which was \$10.4 million. It notes that 85% of Freedom of Information requests are released in full or partially released. A sidebar on the right provides links to GovHack Website, Request a Dataset, Submit a Success Story, and Map Viewers and Portals.

Data.SA
South Australian Government Data Directory

\$3.2m saving

Government of South Australia

Home Datasets Organisations Groups Strategies Toolkit Apps & Ideas Search

House icon / SA Police Reduce Costs Through Proactive Open Data

SA Police Reduce Costs Through Proactive Open Data

Freedom of information access comes at a cost. In 2011-12 financial year the estimated costs of administering the *Freedom of Information Act 1991* (SA) was reported to be \$10.4 million (The Hon I.K.Hunter- Hansard 16/10/14). Records indicate 85% of Freedom of Information requests are released in full or partially released.

In 2014 SA Police released [Expiation Notice System Data](#) as open data on Data.SA.

Freedom of Information requests are received for expiation data from Members of Parliament and the media. In 2014 it became clear that the volume of requests were rising. These requests were often received at different times, requiring variations of the same data and significant time and resources expended to process individual requests.

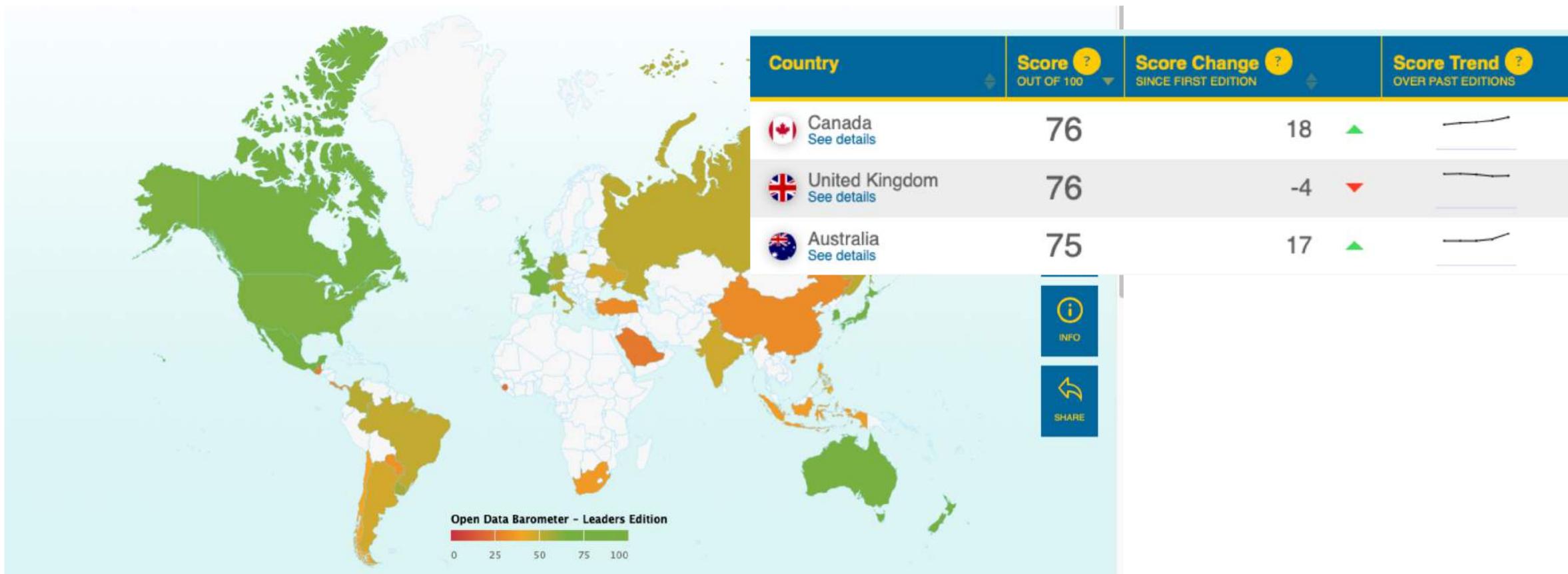
GovHack Website >

Request a Dataset >

Submit a Success Story >

Map Viewers and Portals >

The Open Data Barometer – Gamifying open data



Case study: LIDAR

 DATA.GOV.UK Beta
Opening up Government

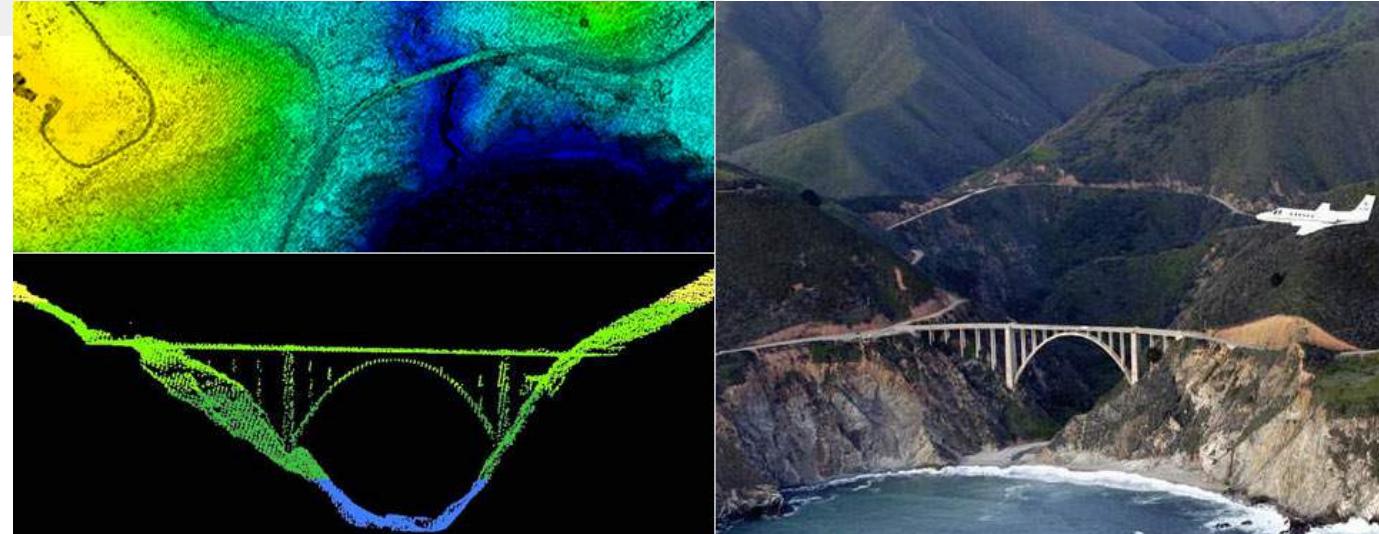
Home

Datasets Map Search Data Requests Publishers Data API Organograms Site Analytics

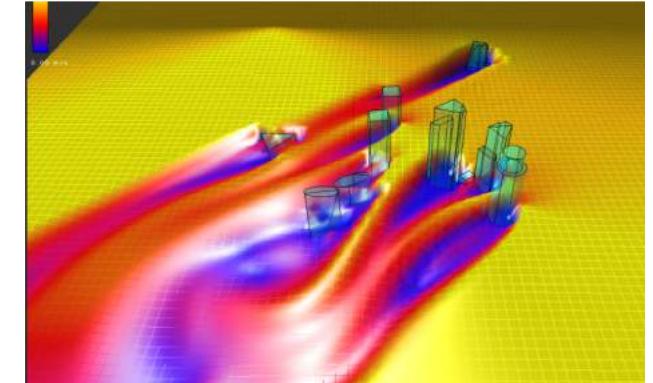
Home / Datasets / LIDAR Composite DTM - 50cm

LIDAR Composite DTM - 50cm

Published by Environment Agency. Licensed under  Open Government Licence.
Openness rating: ★★★☆☆



Impact of LIDAR data



Wind modelling

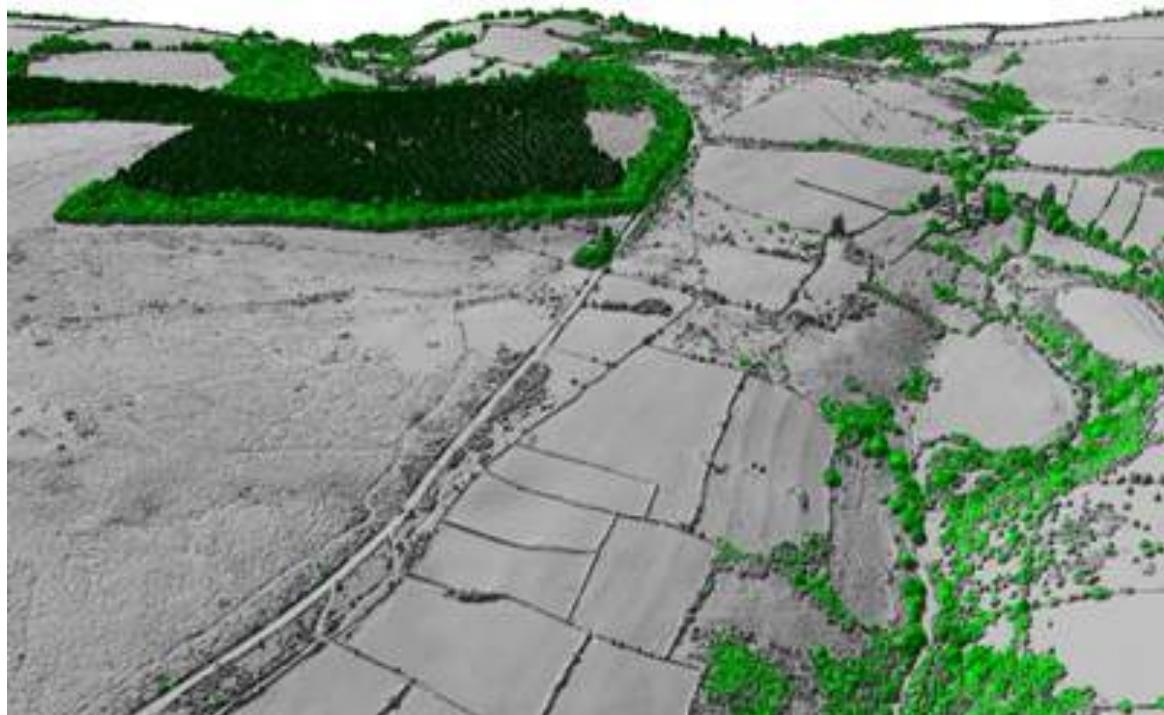


Archology



Educational games (Minecraft)

Cheers



The order

- Haiti
- Crime – London, Chicago
- Toilet map
- ~~London fire~~
- Public parking in Devon
- NASA – International space station data, LMAO
- NOAA – Farming and hurricanes, flooding
- ~~Plantwise~~
- ~~Wikidata~~
- Tanzania – Atlas
- ~~Tanzania – Education~~
- Schools in Burkina Faso
- FAO & World bank
- Hans Rosling
- Health – Dengue maps in Singapore
- Food – FSA data, Provenance, Food trade menu info from the NHS
- Sport – Open active, changing businesses
- ~~Transport~~ – same business model
- Emissions – ~~Heathrow~~, Car exhausts
- Mexico trees
- Open corporates – Talking of businesses
- Open banking – Not strictly open data but sort of in reverse.
- FOI – Talking of saving money
- ~~Socrata – Policy Govstat, statestat~~
- Canada – Barometer
- ~~Voting in Burkina Faso~~
- ~~Xalapa bins~~
- Bindicator
- ~~data.ac.uk facility sharing~~
- LIDAR – Flood mapping, wind, villas and wine.

equipment.data

Enabling access to UK HE research equipment

Search for equipment: e.g. Microscope



The city of Xalapa's waste management problem

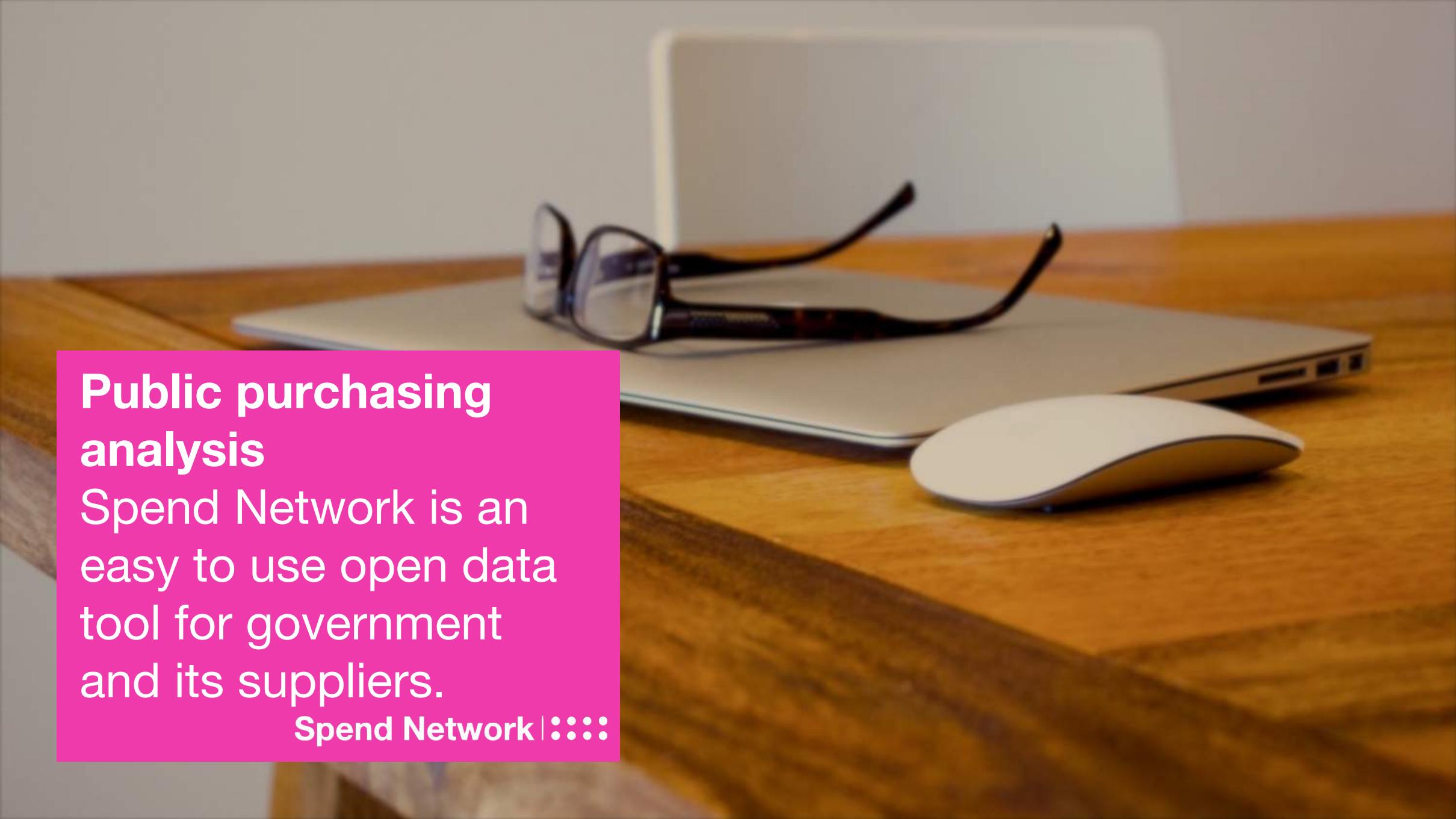
1. Put a GPS on every garbage truck
2. Collected the data
3. Opened the data
4. Engaged community in design
5. Analyzed the maps for new routes
6. Implemented new routes
7. Evaluated the impact, repeating 1-3
8. Civil society organizations organized a Hackathon
9. Developed an app (exact location, timetables, citizens reports)
10. Evaluated the solution

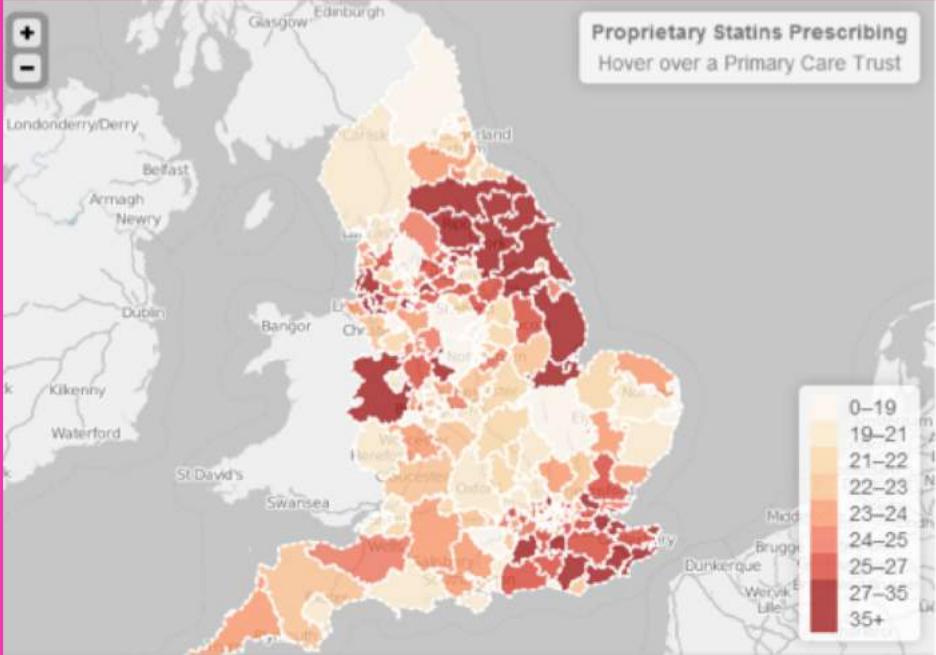


Public purchasing analysis

Spend Network is an
easy to use open data
tool for government
and its suppliers.

Spend Network | :::::





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

theodi.org/stories

Future impact

Opening up TfL bus data

Cost: £820,000

Time: 1 Year

Customer benefit: £8m

ROI: 10x

High Speed 2 (new railway)

Cost: £32.7bn

Time: 20 Years

Customer benefit: £105m

ROI: 0.003x

Source: <http://odimpact.org/case-united-kingdoms-transport-for-london.html>

Future impact

Opening up TfL bus data

Cost: £820,000

Time: 1 Year

Customer benefit: £8m

ROI: 10x

High Speed 2 (new railway)

Cost: £56bn

Time: 20 Years

Customer benefit: £105m

ROI: 0.0018x

Source: <http://odimpact.org/case-united-kingdoms-transport-for-london.html>



TfL Open Data Benefits

<http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf>

Passengers and other road users



Saved time for network passengers

- Passengers are able to *plan their journeys better with apps* that use TfL's open data to provide them real-time information and advice on how to adjust their routes.
- This provides *greater certainty* on when the next bus/tube will arrive *and saves time* – estimated at between *£70m and £90m pa*.



Saved time for other road users

- The availability of data on road works and traffic incidents can feed into Sat Navs, driving software and apps that can allow private and commercial drivers to adjust their routes to avoid congestion.
- This saves time and can reduce emissions as less time is spent waiting in traffic queues and journeys are shorter.

TfL Open Data Benefits

<http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf>

Passengers and other road users



Savings made from moving from SMS alerts

- Passengers are able to switch to using free apps or free web services for real-time data that use TfL's open data.
- This creates a cost saving for those who previously subscribed to fee-based SMS alerts, estimated to worth up to *£2m pa*. The *use value of new real time alert services* is estimated to be up to *£3m pa*.



Better information to plan journeys, travel more easily and take more journeys

- Passengers are now able to better plan journeys, enabling them to use TfL services more regularly and access other services.
- This can result in more journeys on the network. Conservatively the *value of these journeys* is estimated at up to *£20m pa*.

Plus improved customer satisfaction from having accurate and reliable information available instantly

TfL Open Data Benefits

<http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf>

London



Gross Value Added

- A number of companies use and re-use TfL data commercially, generating revenue, many of whom are based in London.
- We estimate that the total *Gross Value Add* from using TfL data by these companies directly and across the supply chain and wider economy is between £12m and £15m GVA pa.



High value Job creation

- TfL open data is estimated to directly support around 500 jobs that would not have existed otherwise.
- Many of these jobs are in sectors associated with high productivity.



Wider job creation in the supply chain

- A further 230 indirect jobs in the supply chain and wider economy have also been created.

Plus supporting the wider UK Digital Economy in London and other cities

TfL Open Data Benefits

<http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf>

Transport for London



Savings from not having to produce apps in-house

- With over *13,000 registered developers* currently, TfL is allowing the market to develop innovative new transport apps and services.
- This creates potential cost savings for TfL of not having to build apps itself or through co-developing with third party developers.



Savings from not having to invest in campaigns and systems

- The publication of open data gives passengers information directly, *reducing the pressure on the Contact Centre*.
- Undertaking an equivalent campaign to make available this information could cost £1m – open data allows TfL to make available the same data at a much reduced cost, *expanding customer reach and improving transparency*.
- The cost for TfL of publishing open data is estimated at around £1m annually, suggesting a *significant return on investment*.



Leveraging value and savings from partnerships

- Through partnerships with major data and software organisations, TfL *receives back significant data on areas it does not itself collect data* (e.g. crowdsourced traffic data).
- This allows TfL to undertake *new analyses and improve its operations*.

Plus new commercial opportunities arising from open data

TfL Open Data Benefits

<http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf>

Society



People are more likely to walk or cycle and lead healthier lifestyles

- Open data can help *integrate the first- and last-mile* encouraging alternative transport modes including cycling and walking, which has *health benefits*.
- This can support ambitions around *healthier streets*: since 2007, the number of people walking as their main mode of transport has increased by 13%.

Growth and productivity



Taking advantage of new opportunities

- The UK and London have already earnt a *reputation* as a leader in open data and the digital economy. Recent research by Tech City noted London's digital economy was worth £30bn in GVA and supported over 300,000 jobs.
- The provision of transport open data will be an *important foundational block for further development of new transport products*.



Encourages and facilitates increased innovation

- As TfL releases more data, it is able to encourage an increasing number of developers to innovate to create a range of new customer-facing services that can tackle social and economic issues (as demonstrated by the recent cyclist app challenge).

Environment

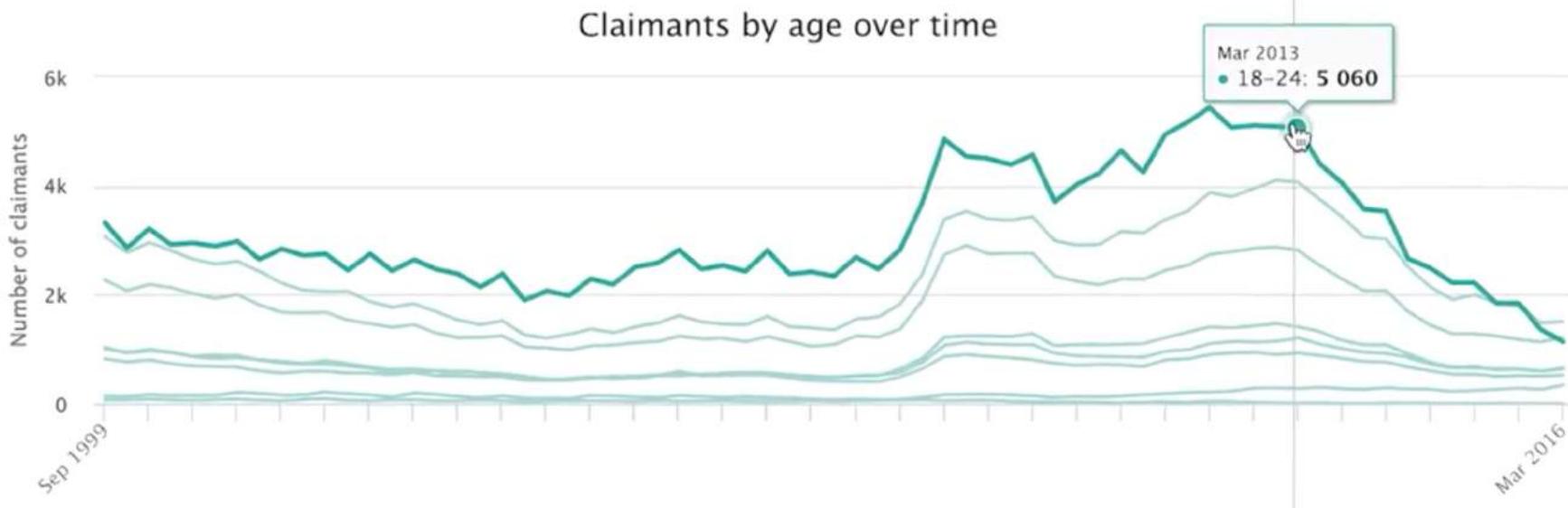
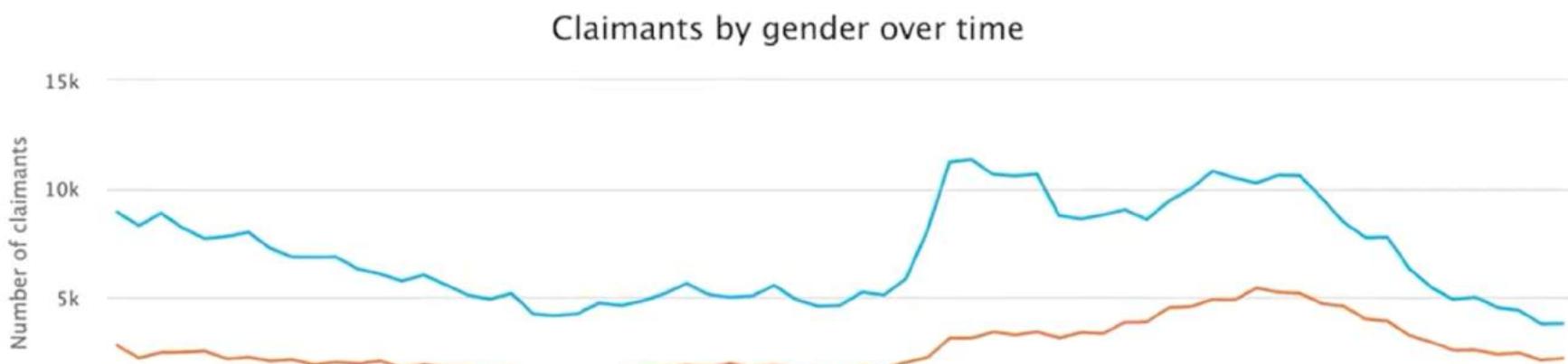
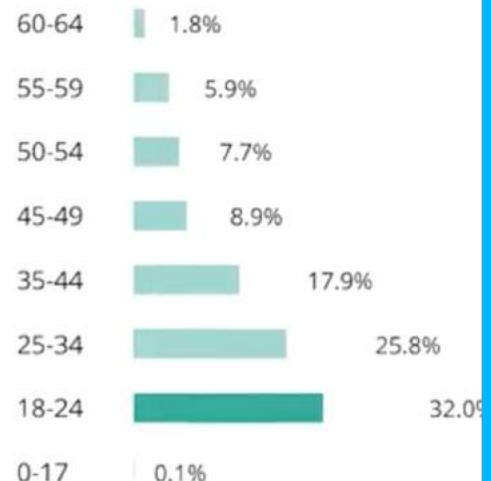


Contributing to improving air quality and reducing emissions

- As open data is used to develop new customer facing products that support modal shift from private and public vehicles, there will be greater numbers of pedestrians and cyclists.
- They will contribute to a lower carbon footprint in London and emit lower particulates being emitted, which can improve air quality.

People on Job Seekers Allowance in County Durham

Co

[Back](#)**15,790** total claimants

Public service delivery patterns

Efficiency

Transparency

Collaboration

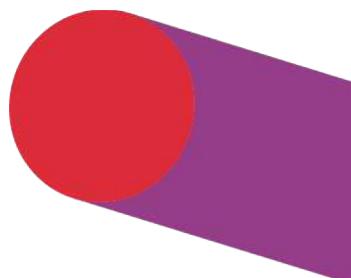
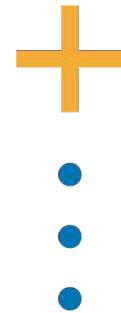
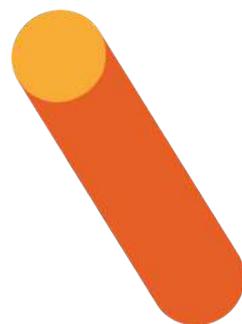


Using open
data to deliver
public services

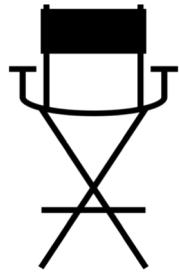
Open Data Institute



What other
opportunities could +
being more open
bring?

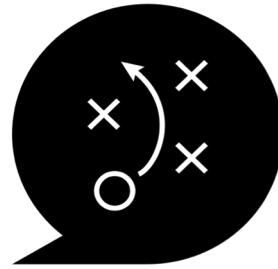


Benefits of open data



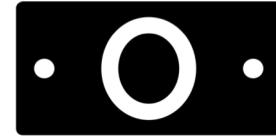
Leadership

Open data supports the wider data strategy of the organisation



Operations

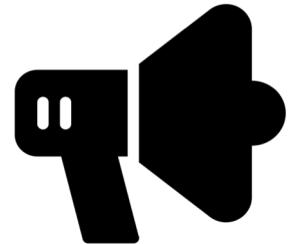
Open data creates significant data access efficiency savings



Finance/Ventures

Open data produces viable equity ventures & business lines

Benefits of open data



Communications

Open data provides free media and increased reach



CSR/Foundation

Open data improves outreach and forges social partnerships



Marketing/Product

Open data enhances product features and customer experience

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



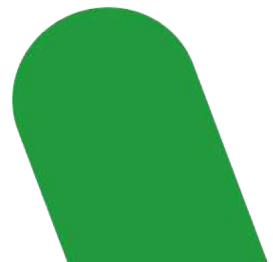
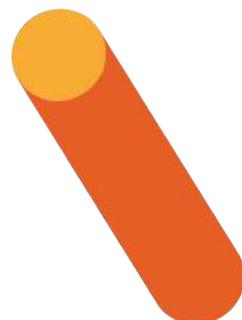
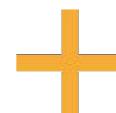
Joined up available transport data to enable choice. \$10m VC round and expanded to 29 cities



Emissions double reported figure. Evidence in parliament & planning debate



How would you
measure the success
and development of
an Open Data
Initiative?



1. Infrastructure

Data is infrastructure

2. Equity

Data and rights

3. Engagement

Measuring success

4. Innovation

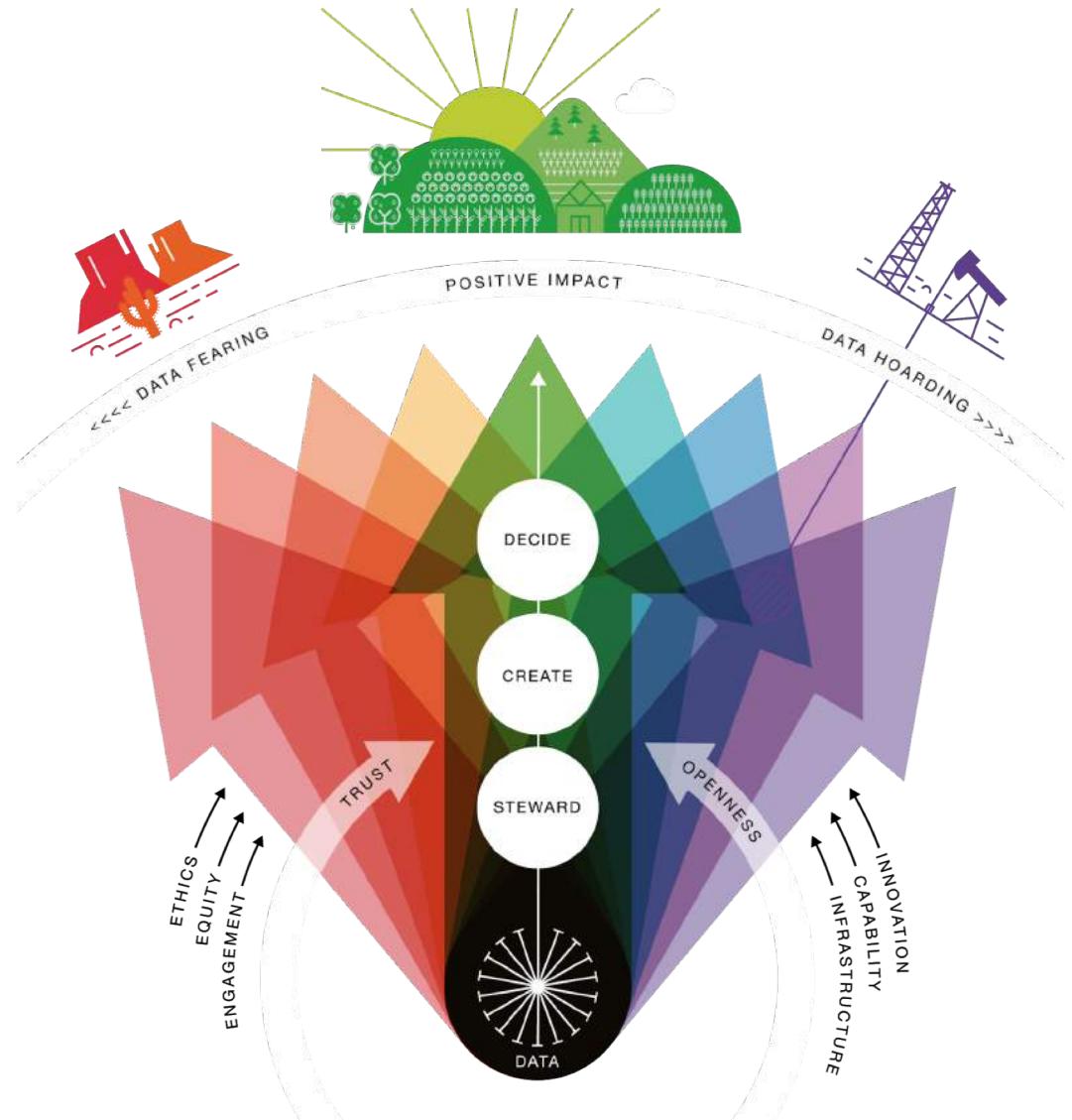
Growing economies

5. Capability

Making data usable

6. Ethics

Using data responsibly



1. Infrastructure

Data is infrastructure

Data infrastructure

Data is infrastructure. It underpins transparency, accountability, public services, business innovation and civil society.

<https://theodi.org/what-is-data-infrastructure>



freepik

3. Engagement

Measuring success

READINESS

Are governments ready to open their data?

IMPLEMENTATION

How well are governments fulfilling their open data promises?

IMPACT

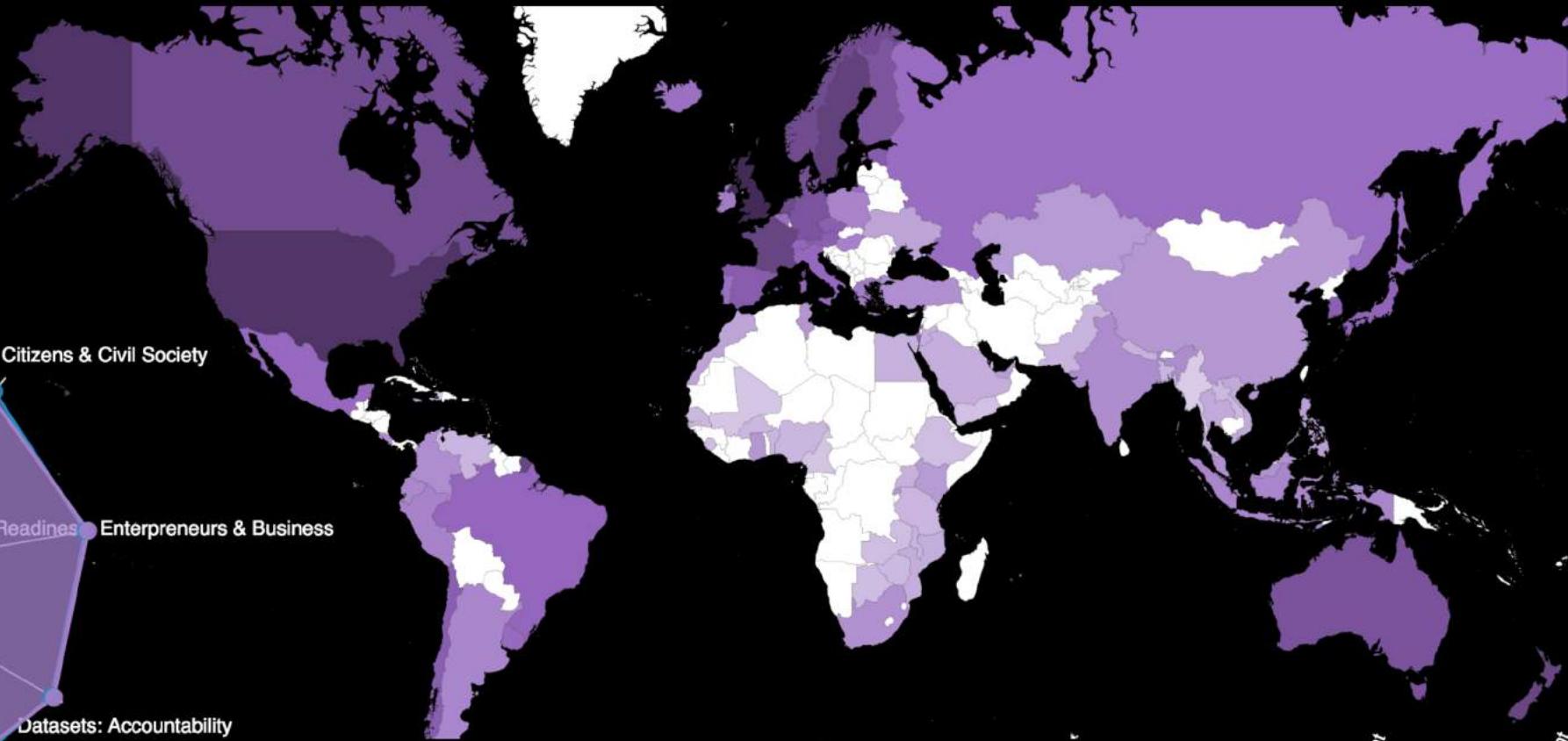
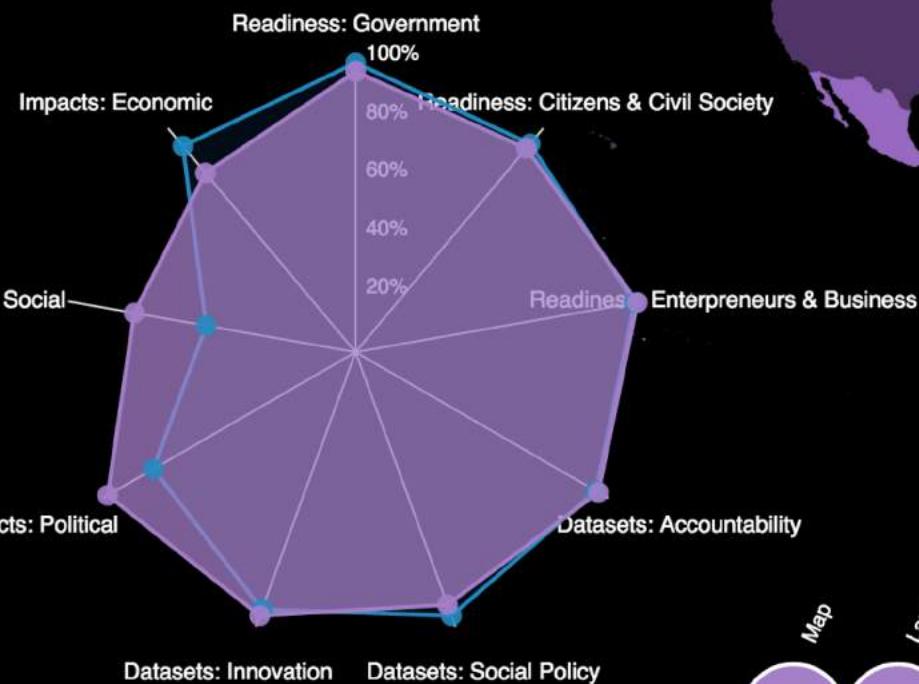
What impact is open data having on citizens' lives?

United Kingdom

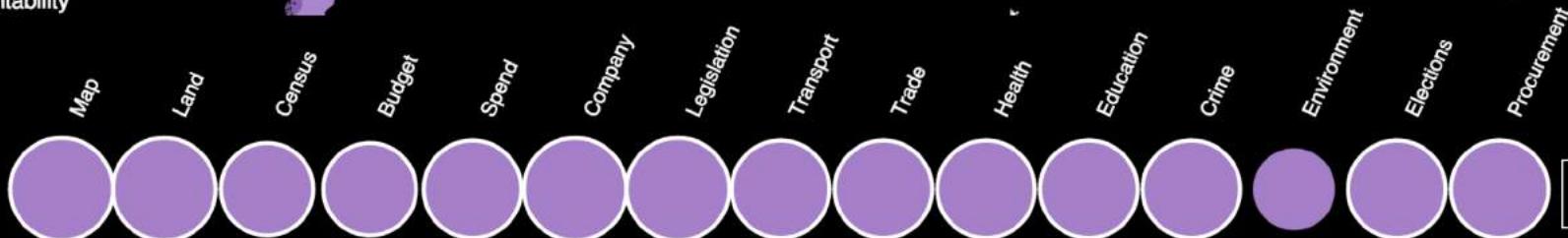
1 / 86

Score: 100.00

Filters: Off



License: CC-BY-SA | Visualisation by David Tarrant | Full report | Get the data



Agenda:

- 09:30** What open data is and isn't
- 11:15** Case studies in open and transparent ecosystems
- 12:45** **LUNCH**
- 13:30** Licenses and certificates
- 15:00** Accessing, cleaning and validating data
- 17:00** **END**

1. Infrastructure

Data is infrastructure

2. Equity

Data and rights

3. Engagement

Measuring success

4. Innovation

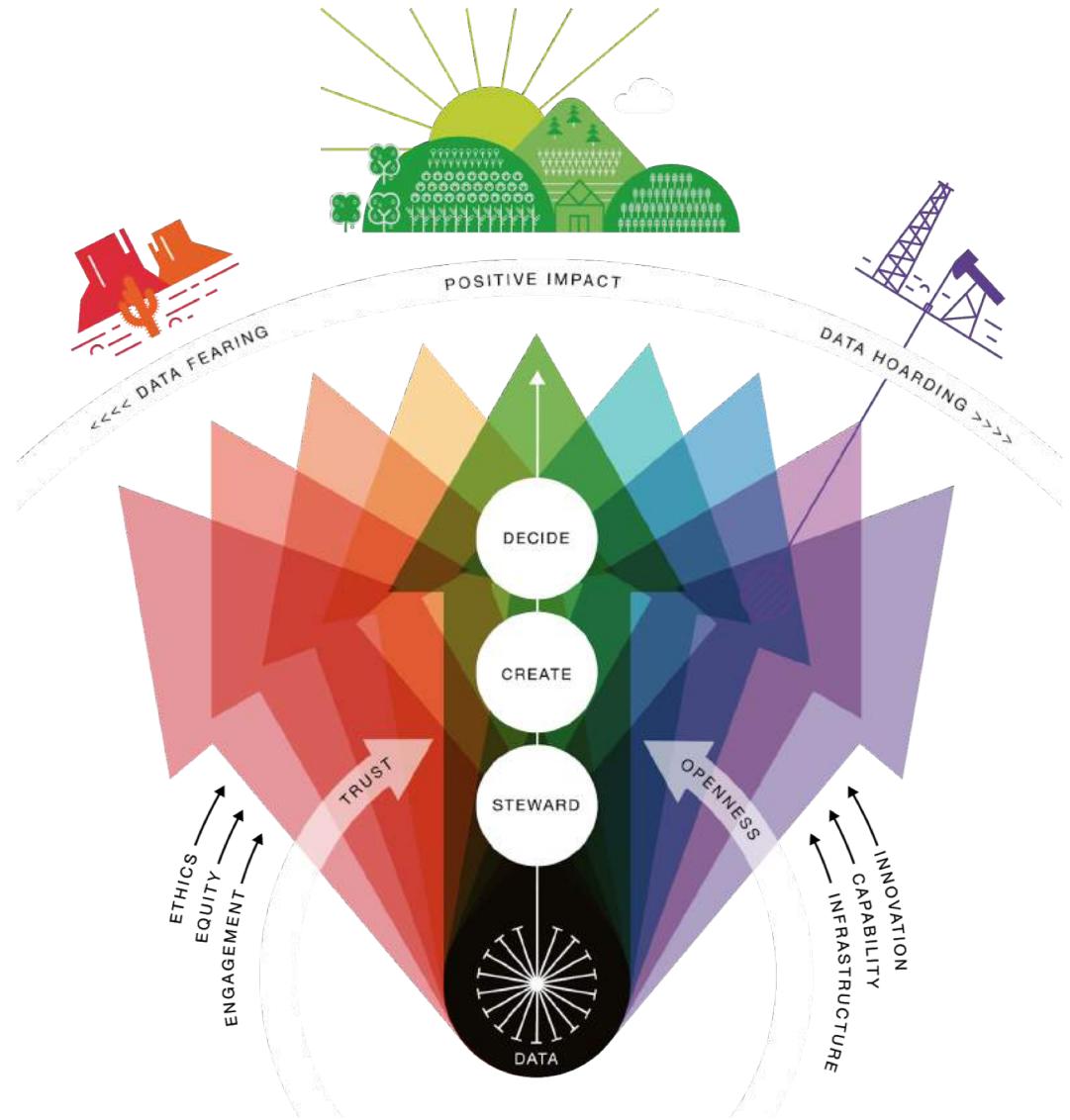
Growing economies

5. Capability

Making data usable

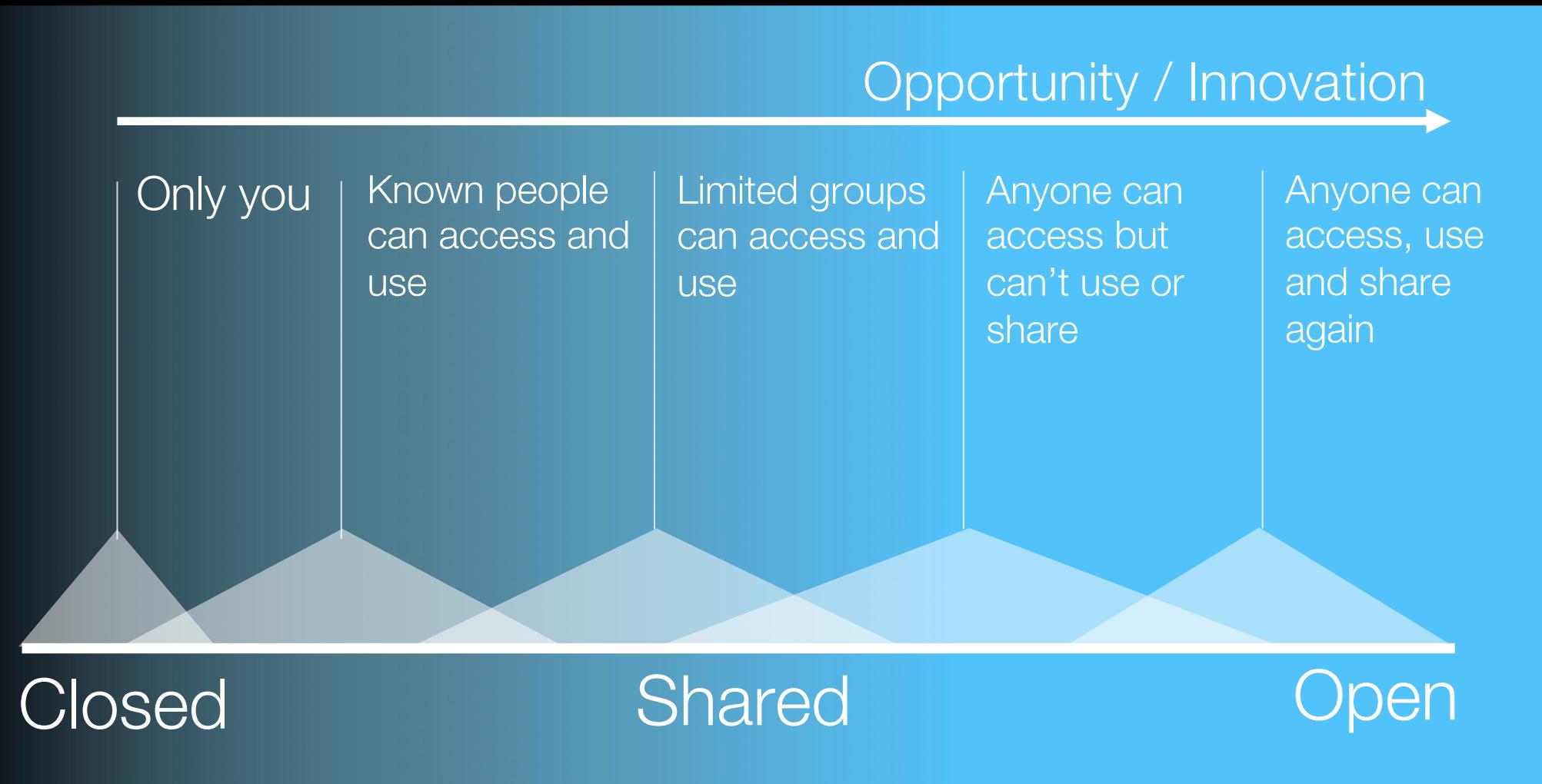
6. Ethics

Using data responsibly



2. Equity

Data and rights



open
data
institute

/ Knowledge
for everyone

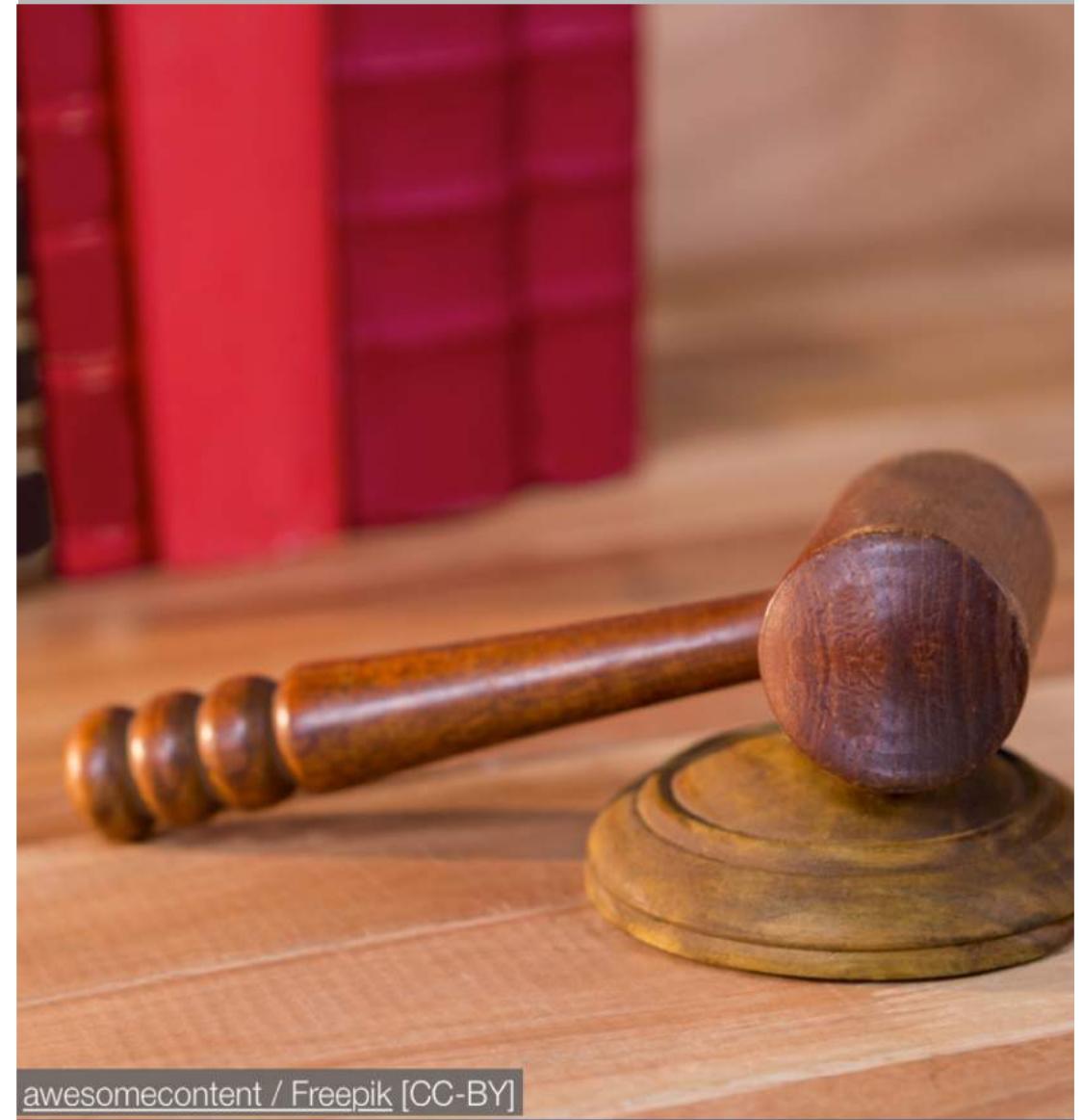
Law and Licensing

List the characteristics of an open licence

Identify instances of an open licence

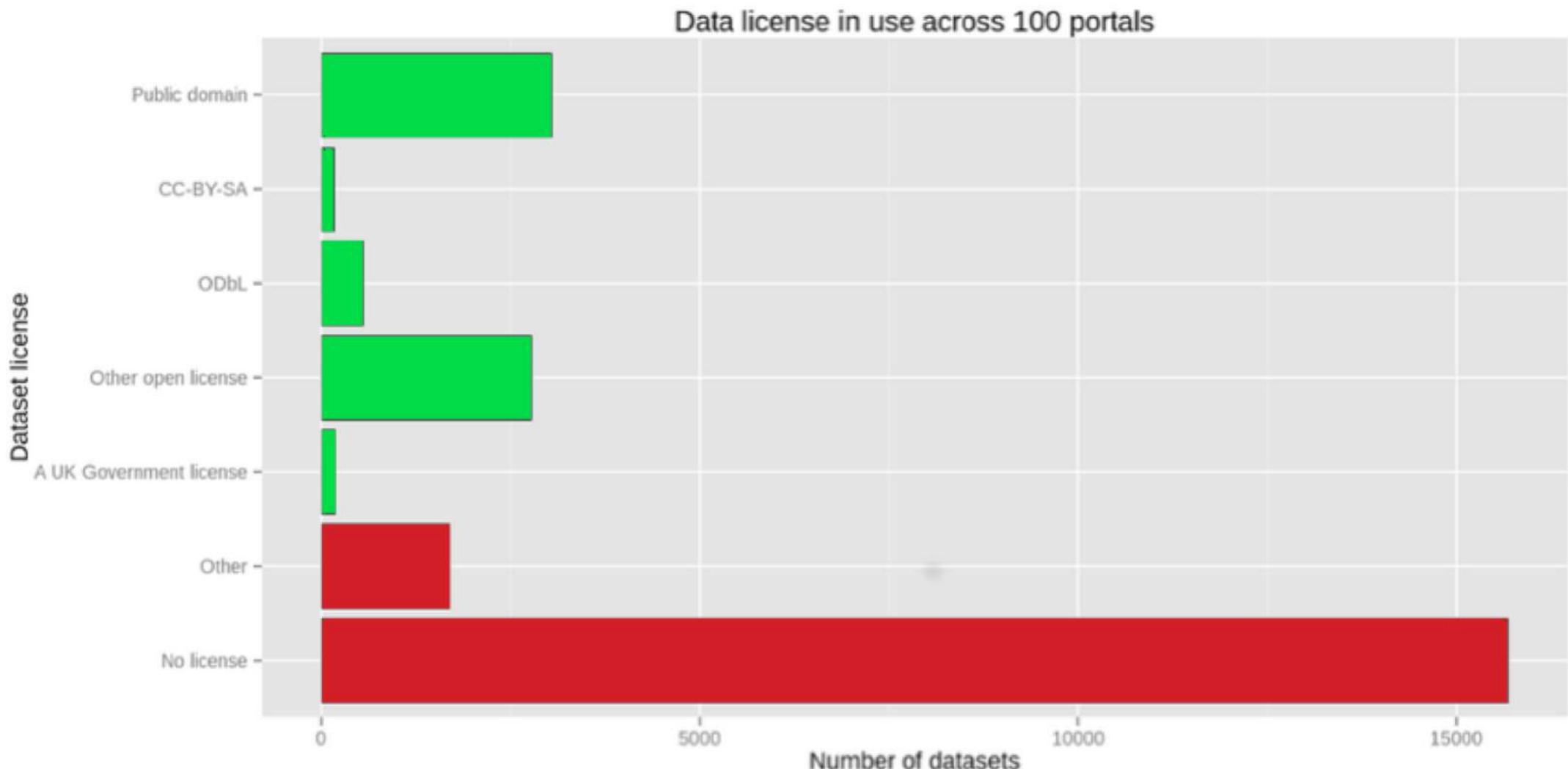
Please note:

I am not a lawyer and this should not be treated as legal advice.

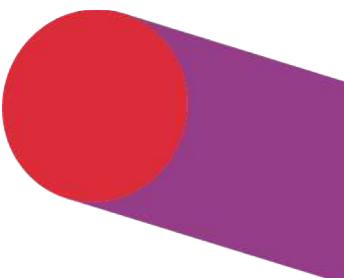
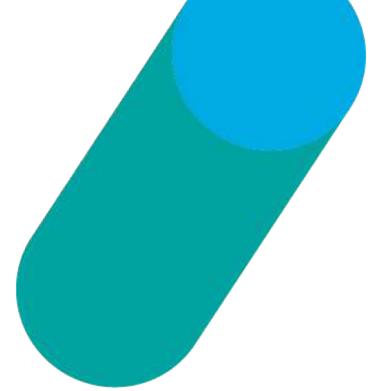
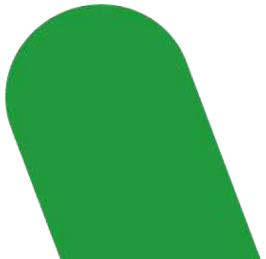
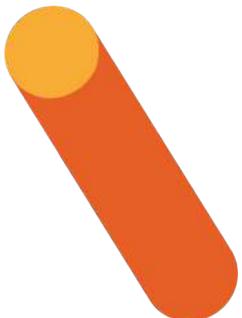
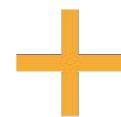


awesomecontent / Freepik [CC-BY]

Open data is hardly ever appropriately licensed.



Build your own open data licence



Build your own open data licence



can

You can...



must

You must...



cannot

You cannot...

Why licence?

A licence provides clarity

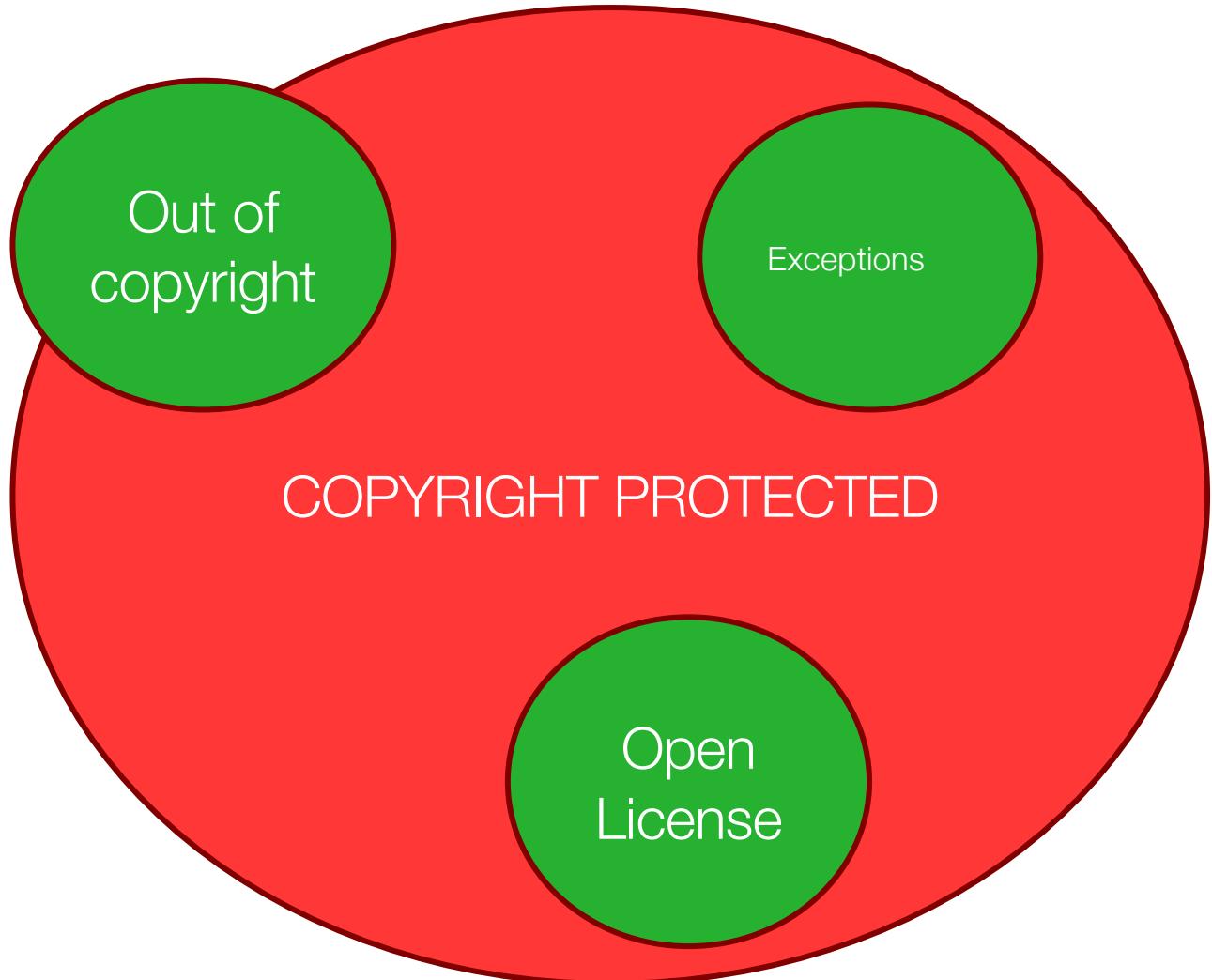
It sets out exactly what users and re-users are permitted to do with your dataset



Intellectual property

You can use it if [green]:

- It's out of copyright [public domain]
- It's openly licensed
- Your use is covered by a copyright exception



Copyright



Applies automatically to:

- literary works
- artistic works
- dramatic works
- musical works
- layouts of works
- sound and film works
- broadcasts

Copyright

Applies automatically to –

Anything where there is an intellectual, creative
and original process in creating a thing.

It does not apply when the thing created would be
the same whoever created it.

“it’s obvious”

Version 3.0 released in October 2014

[http://www.nationalarchives.gov.uk/doc/open-government-
licence/version/3/](http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/)



OGL - What does it permit?

Use of copyright and database right material expressly made available under this licence (the 'Information') indicates your acceptance of the terms and conditions below.

The Licensor grants you a worldwide, royalty-free, perpetual, non-exclusive licence to use the Information subject to the conditions below.

This licence does not affect your freedom under fair dealing or fair use or any other copyright or database right exceptions and limitations.

You are free to:

-  copy, publish, distribute and transmit the Information;
-  adapt the Information;
-  exploit the Information commercially and non-commercially for example, by combining it with other Information, or by including it in your own product or application.

Are there restrictions on its use?

You must (where you do any of the above):

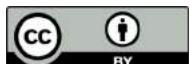
- ⚠ acknowledge the source of the Information in your product or application by including or linking to any attribution statement specified by the Information Provider(s) and, where possible, provide a link to this licence;

If the Information Provider does not provide a specific attribution statement, you must use the following:

Contains public sector information licensed under the Open Government Licence v3.0.

If you are using Information from several Information Providers and listing multiple attributions is not practical in your product or application, you may include a URI or hyperlink to a resource that contains the required attribution statements.

These are important conditions of this licence and if you fail to comply with them the rights granted to you under this licence, or any similar licence granted by the Licensor, will end automatically.



The main alternative



Open and community developed; not controlled by a central agency or body.



This image shows a screenshot of the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) license page. The header features the Creative Commons logo and the text "Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)". Below the header, there is a box containing the text "This is a human-readable summary of (and not a substitute for) the [license](#)". To the right of this box is a "Disclaimer" link. In the bottom right corner of the page, there is a green circular badge with the text "Free Cultural Works APPROVED FOR".

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:



Attribution — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



th

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.



Creative commons - restrictions



Attribution



No Derivative Works



Share Alike



Non-Commercial

| | PUBLIC DOMAIN | PUBLIC DOMAIN | BY | BY SA | BY NC | BY ND | BY NC SA | BY NC ND |
|-------------------|---------------|---------------|----|-------|-------|-------|----------|----------|
| (S) PUBLIC DOMAIN | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ |
| (Ø) PUBLIC DOMAIN | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ |
| CC BY | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ |
| CC BY SA | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ |
| CC BY NC | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ | ✓ | ✗ |
| CC BY ND | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| CC BY NC SA | ✓ | ✓ | ✓ | ✗ | ✗ | ✓ | ✓ | ✗ |
| CC BY NC ND | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |

Identifying open licenses



Recognising Open Licences

When discovering data online, it is important to be able to check the licence to understand whether the data is open and if so, under what conditions can you access, use and share it.

Take a look at each of the following licences and try to determine whether they are open licences

- [Twitter Terms of Use](#)
- [Open Corporates](#)
- [data.gov.sg](#)
- [Met Office](#)
- [Flickr](#)

1. Infrastructure

Data is infrastructure

2. Equity

Data and rights

3. Engagement

Measuring success

4. Innovation

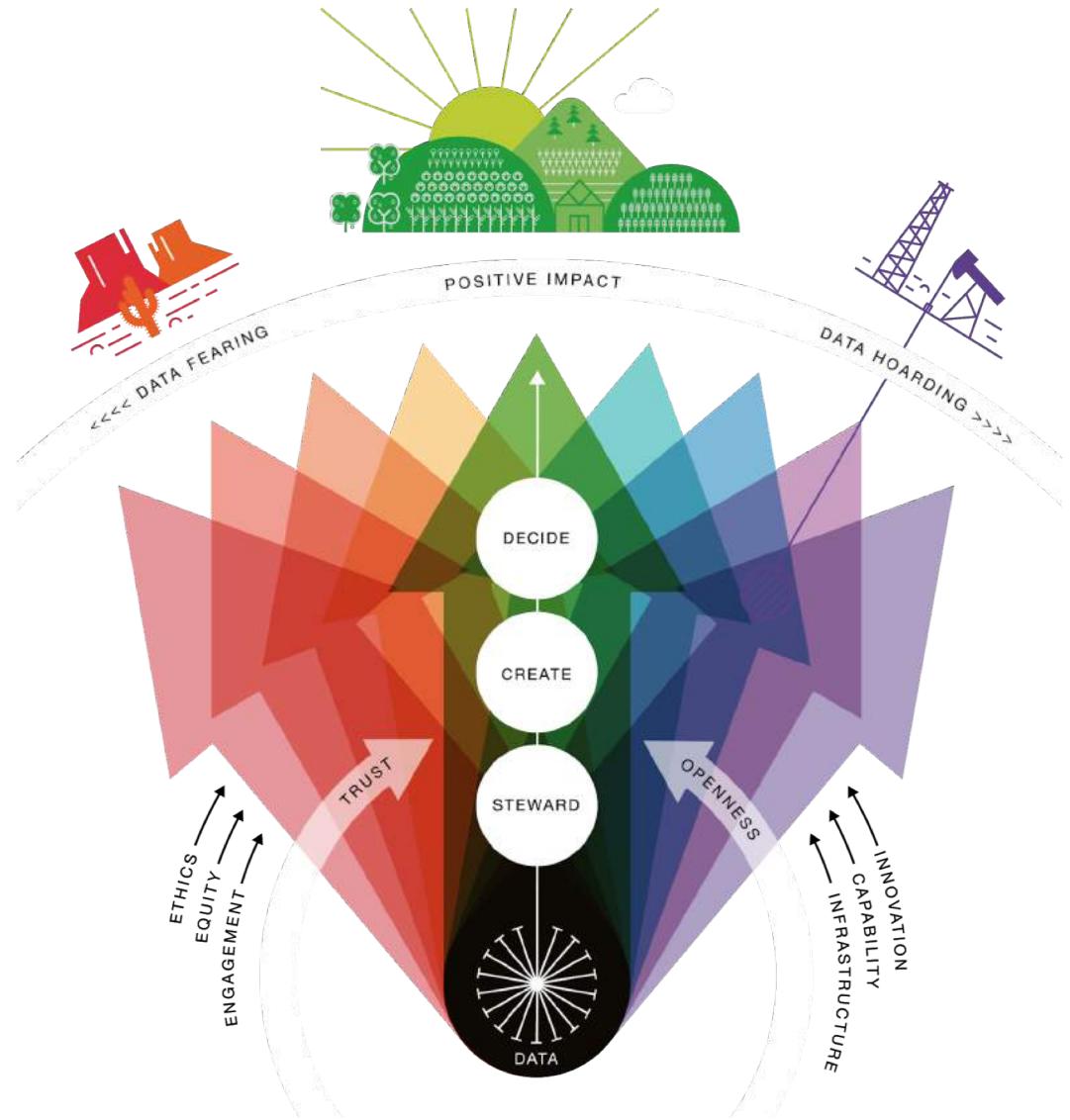
Growing economies

5. Capability

Making data usable

6. Ethics

Using data responsibly



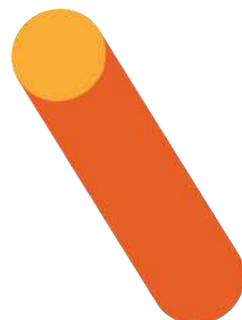
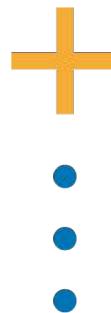
5. Capability

Making data usable

What makes data usable?

You are a data startup. You want to use government open data. You are going to invest time and money in your startup. What guarantees do you need that you can rely on the open data?

What makes it usable? e.g. Consistant

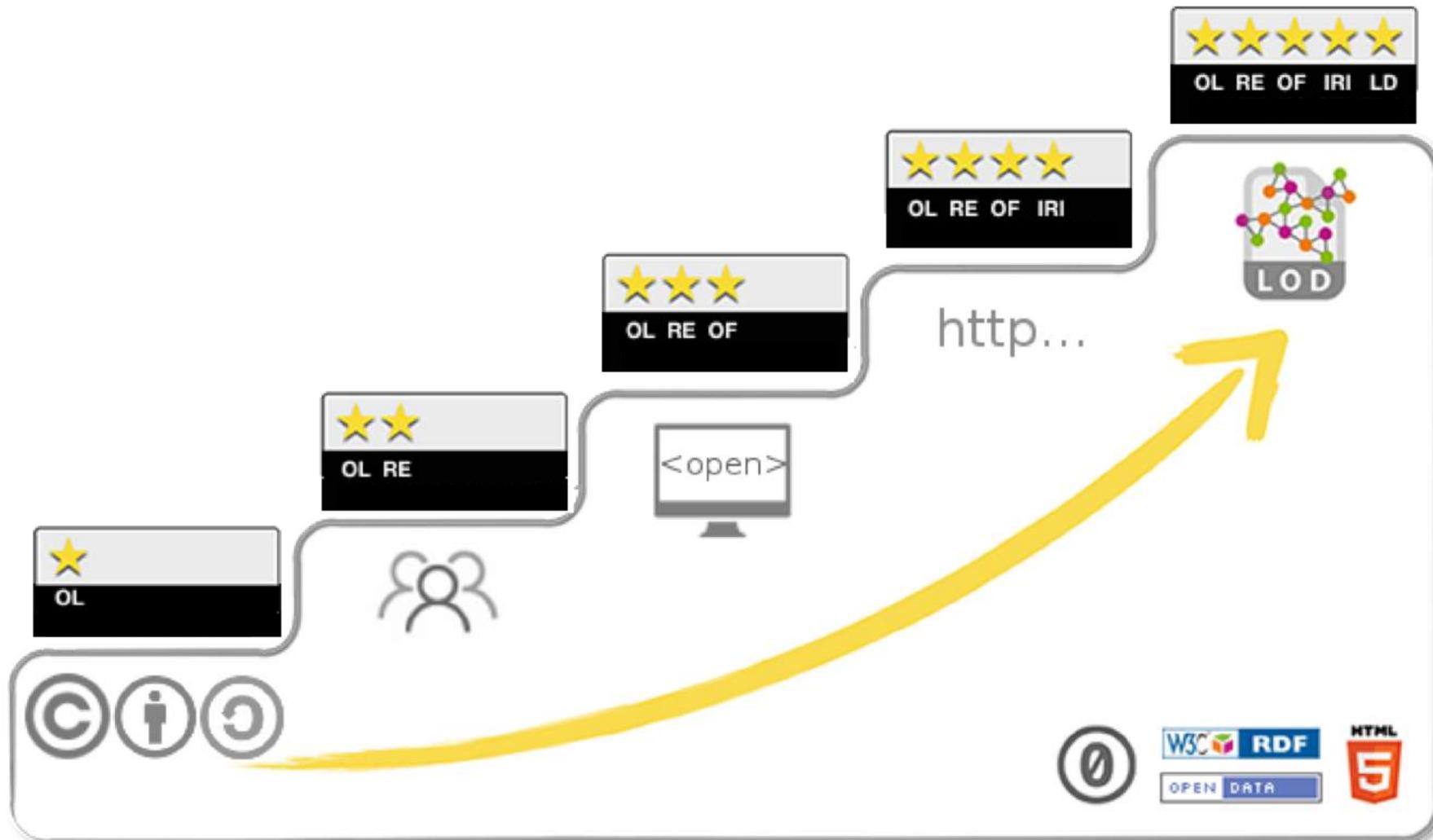


Guidelines



5 - S t a r s

The 5-Stars of Linked Open Data



Open Data Certificate

 **Bronze level**
self certified 

This data has achieved [Bronze level](#) on 29 March 2016 which means this data makes a great start at the basics of publishing open data.

Grants to voluntary community and social enterprise organisations

Summary

Type of release
ongoing release of a series of related datasets

Data Licence
UK Open Government Licence

Content Licence
Not Applicable

Verification
self certified

Community verification
Other people can verify whether the answers on this

General Information

This data is described at <http://data.hounslow.gov.uk/View/loc...> 

This data is published by London Borough of Hounslow 

The data is published on <http://data.hounslow.gov.uk/> 

Legal Information

This data was originally created or generated by its curator 

The rights statement is at <http://www.hounslow.gov.uk/index/c...> 

This data is available under UK Open Government Licence 

There are no rights in the content of the data 

<http://certificates.theodi.org>



5 Aspects

Contextual

Legal



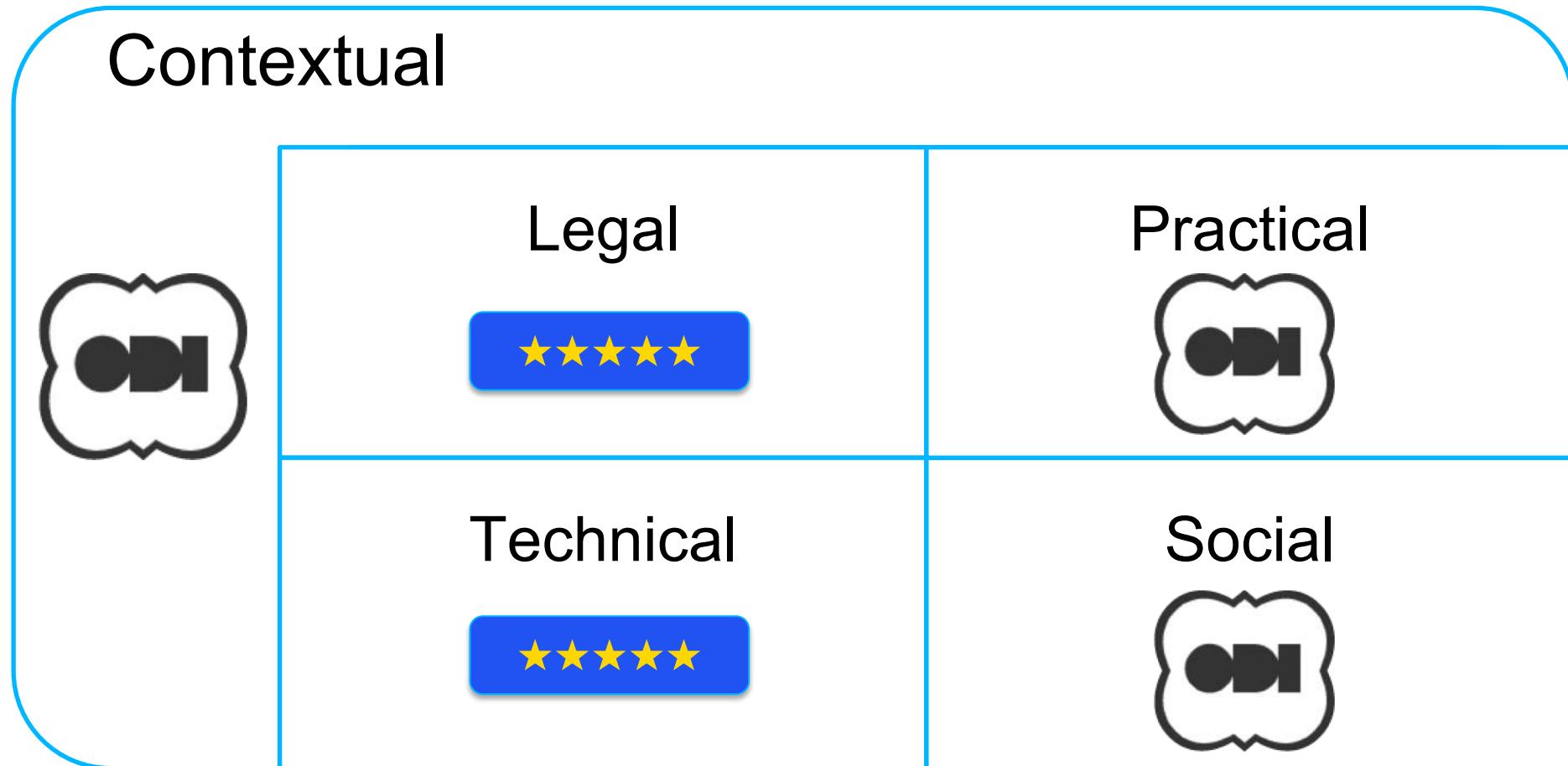
Practical

Technical

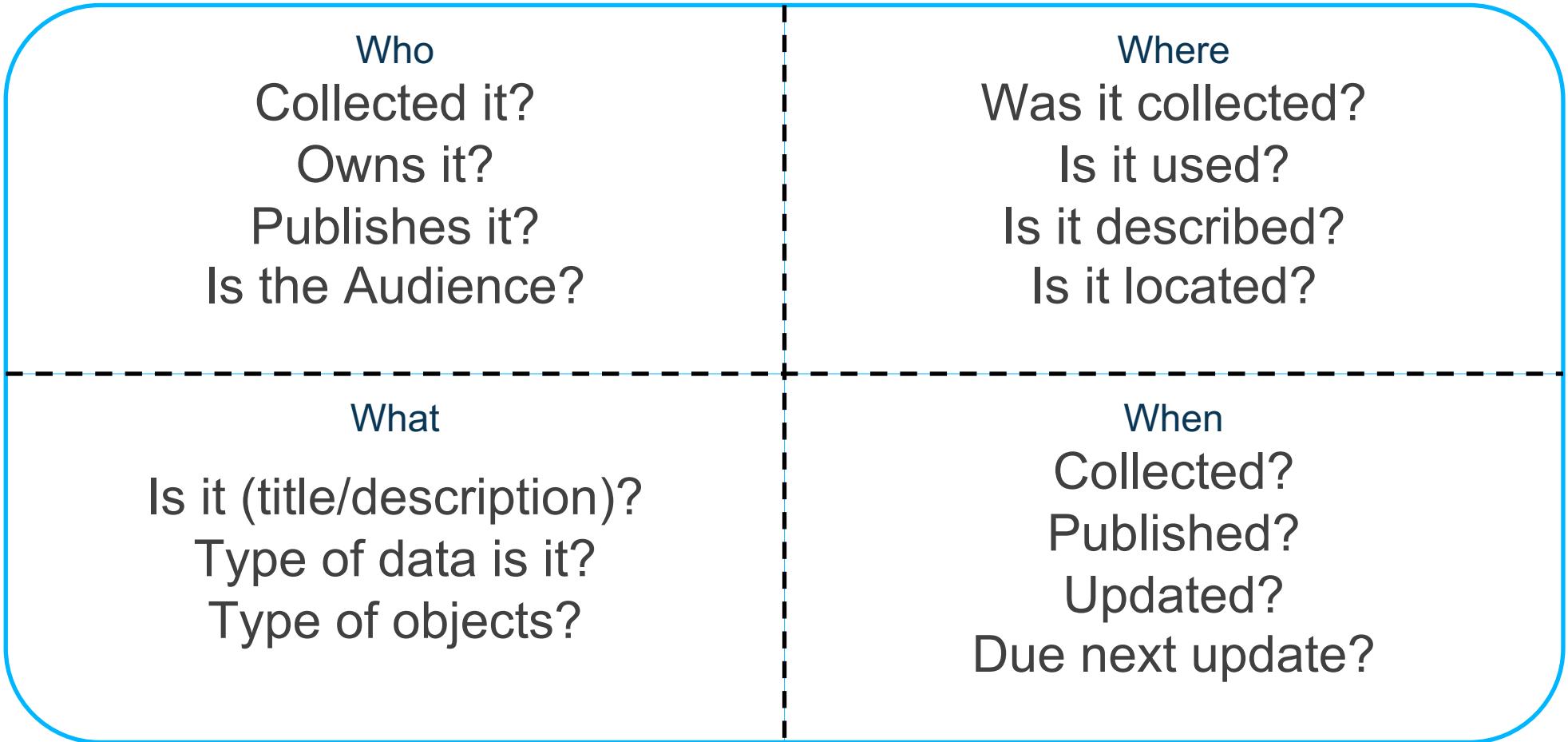


Social

5 Aspects



Adding Context



General

Title/Description/Publisher/URL



Release Type



Legal

Rights to publish



Data licensed



Content licensed



Clear privacy statement



Sources of data documented



Audited anonymisation



Practical

Usable period described



Availability period described



Discoverable from home page



Listed in collection



Referenced from publication/application



Quality problems listed



Quality control process described



Technical

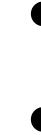
| |  |  |  |  |
|--------------------------------------|---|---|---|---|
| Data hosted online | • | • | • | • |
| Type of data defined | • | • | • | • |
| Machine readable metadata | | • | • | • |
| Clear technical documentation | | • | • | • |
| Persistent & common identifiers used | | | • | • |
| 5-Star Linked Data | | | | • |
| Machine readable provenance | | | | • |
| Data can be verified | | | | • |

Social

Support for improving/fixing



Email support



Discussion groups/forums



Social media channels



Supported community



Tools and guides available to work with data



Agenda:

- 09:30** What open data is and isn't
- 11:00** Case studies in open and transparent ecosystems
- 12:45** **LUNCH**
- 13:30** Licenses and certificates
- 15:00** Accessing, cleaning and validating data
- 17:00** **END**

Open Refine

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

Mass edit 2380 cells in column Type of Contract Undo

Facet / Filter Undo / Redo Refresh Reset All Remove All

5200 rows Show as: rows records Show: 5 10 25 50 100 rows

Extensions: Freebase

Type of Contract

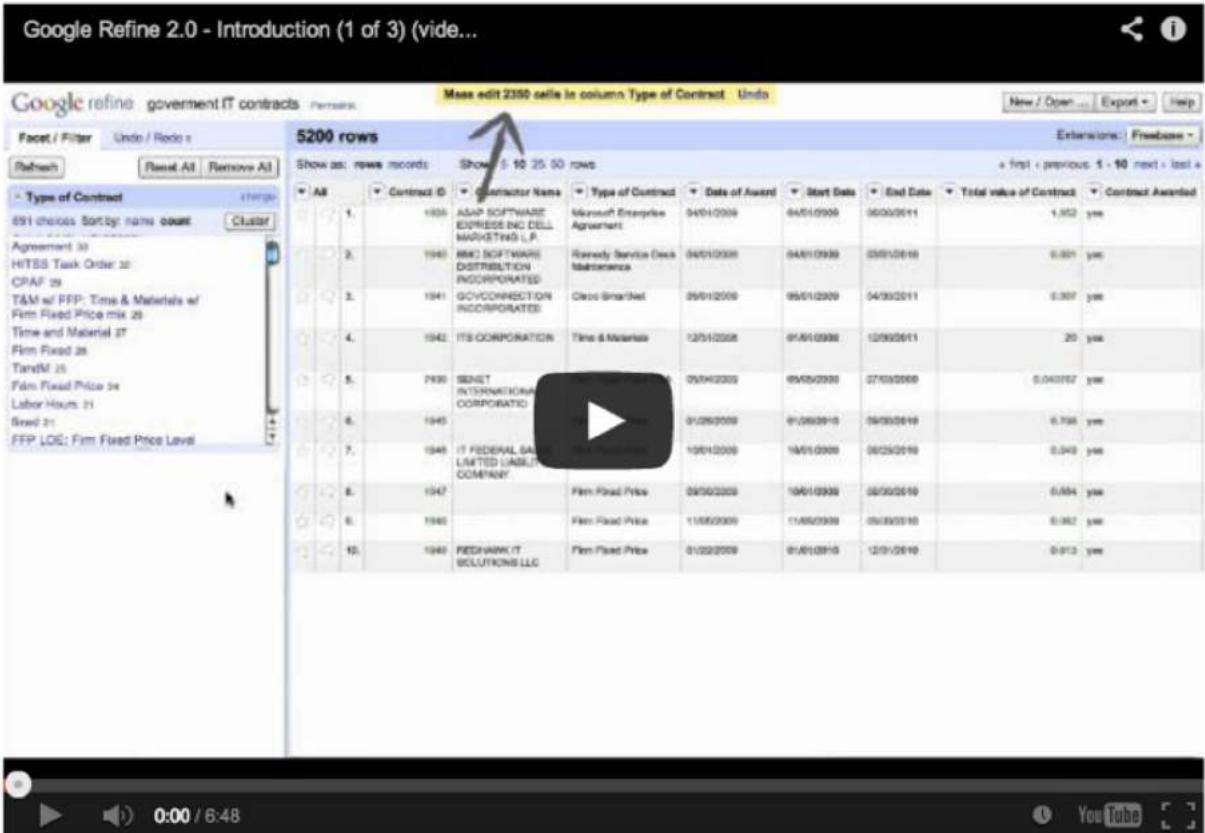
691 choices Sort by: name count Cluster

Agreement 30
HTSS Task Order 22
CPAF 29
T&M w/ FFP: Time & Materials w/
Firm Fixed Price mix 29
Time and Material 27
Firm Fixed 26
TandM 25
Firm Fixed Price 24
Labor Hours 21
Revol 21
FFP LOE: Firm Fixed Price Level

| # | 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/CELL MARKETING L.P. | Microsoft Enterprise Agreement | 04/04/2009 | 04/05/2009 | 06/03/2011 | 1,352 | yes |
| 2. | 1940 | AMC SOFTWARE DISTRIBUTION INCORPORATED | Ramsey Service Desk Maintenance | 04/01/2009 | 04/01/2009 | 03/01/2010 | 0.001 | yes |
| 3. | 1941 | GOVCONNECTION INCORPORATED | Cisco Standard | 05/01/2009 | 05/01/2009 | 04/03/2011 | 0.307 | yes |
| 4. | 1942 | IT'S CORPORATION | Time & Materials | 12/01/2008 | 01/01/2009 | 12/03/2011 | 20 | yes |
| 5. | 2130 | ISBET INTERNATIONAL CORPORATION | Standard | 05/04/2009 | 05/05/2009 | 07/03/2009 | 0.000172 | yes |
| 6. | 1945 | | Standard | 01/06/2009 | 01/06/2009 | 06/03/2010 | 0.708 | yes |
| 7. | 1946 | IT FEDERAL SALES LIMITED LIABILITY COMPANY | Standard | 10/01/2009 | 10/01/2009 | 08/26/2010 | 0.049 | yes |
| 8. | 1947 | | Firm Fixed Price | 05/05/2009 | 10/01/2009 | 08/03/2010 | 0.004 | yes |
| 9. | 1948 | | Firm Fixed Price | 11/05/2009 | 11/05/2009 | 01/03/2010 | 0.002 | yes |
| 10. | 1949 | PEDDLEWELL IT SOLUTIONS LLC | Firm Fixed Price | 01/03/2009 | 01/01/2010 | 12/01/2010 | 0.012 | yes |

0:00 / 6:48

YouTube



<http://openrefine.org>

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

1. Infrastructure

Data is infrastructure

2. Equity

Data and rights

3. Engagement

Measuring success

4. Innovation

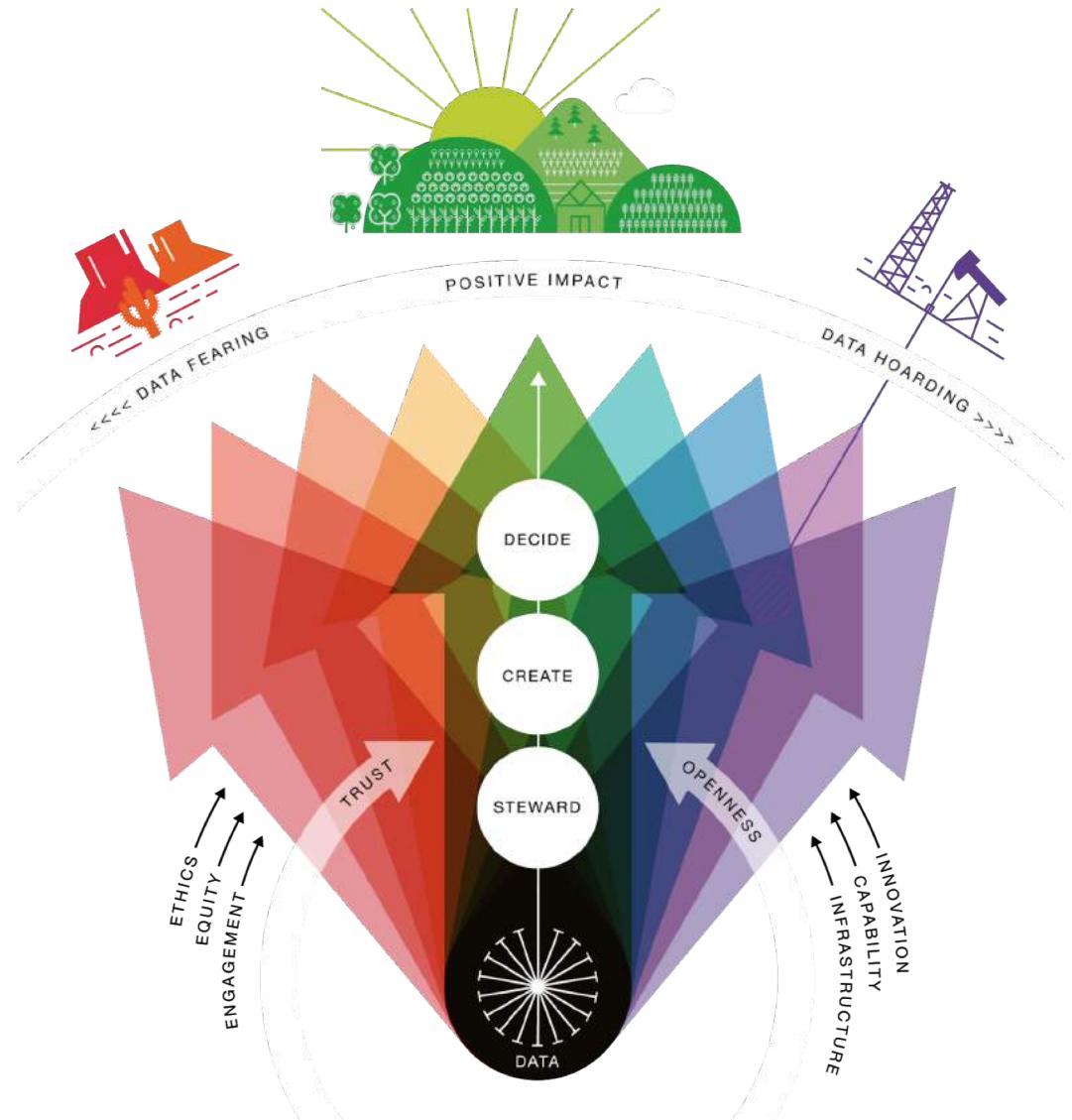
Growing economies

5. Capability

Making data usable

6. Ethics

Using data responsibly



What's next?

theodi.org/courses

theodi.org/eLearning

eLearning

Open data fundamentals

Finding stories in data

Blended courses

Better data, better decisions.

Face to face courses

Change or grow your business with Open Data

Machine learning AI and Ethics

Introduction to the data ethics canvas



Open data essentials

- ✓ Open data definitions
- ✓ Licensing
- ✓ Impact of open data
- ✓ Discovering
- ✓ Best practices
- ✓ Quality checking





THIS IS FOR EVERYONE



Dr David Tarrant
@davetaz
theODI.org



Any questions?

Get in touch

If you would like to talk to us about collaborating, partnering, supporting our work, or anything else, we'd love you to get in touch.

info@theodi.org

+44 (0)20 3598 9395

@ODIHQ

