

Open data in a day

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

Dr David Tarrant

@davetaz

theODI.org



Aim

**Enable you to unlock the value of open
data in 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:

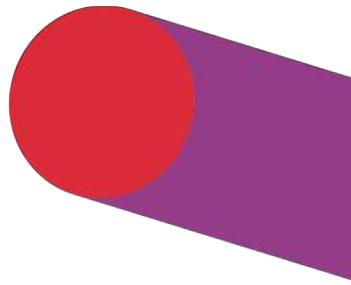
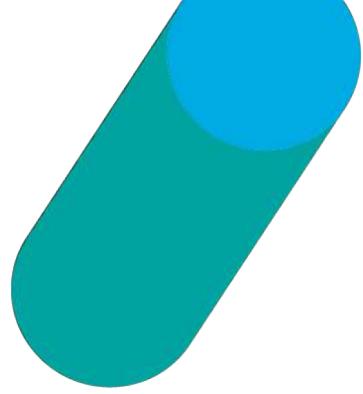
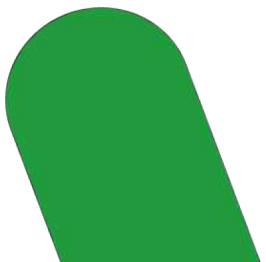
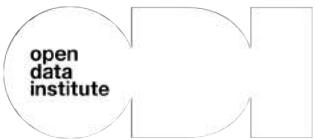
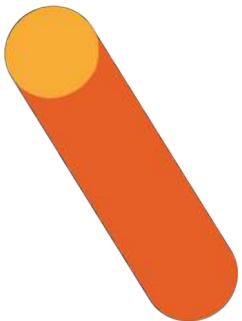
- 10:00** What open data is and isn't
- 11:30** Case studies in open and transparent ecosystems
- 12:45** **LUNCH**
- 13:30** Licenses and certificates
- 15:00** Accessing, cleaning and validating data
- 17:00** **END**

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

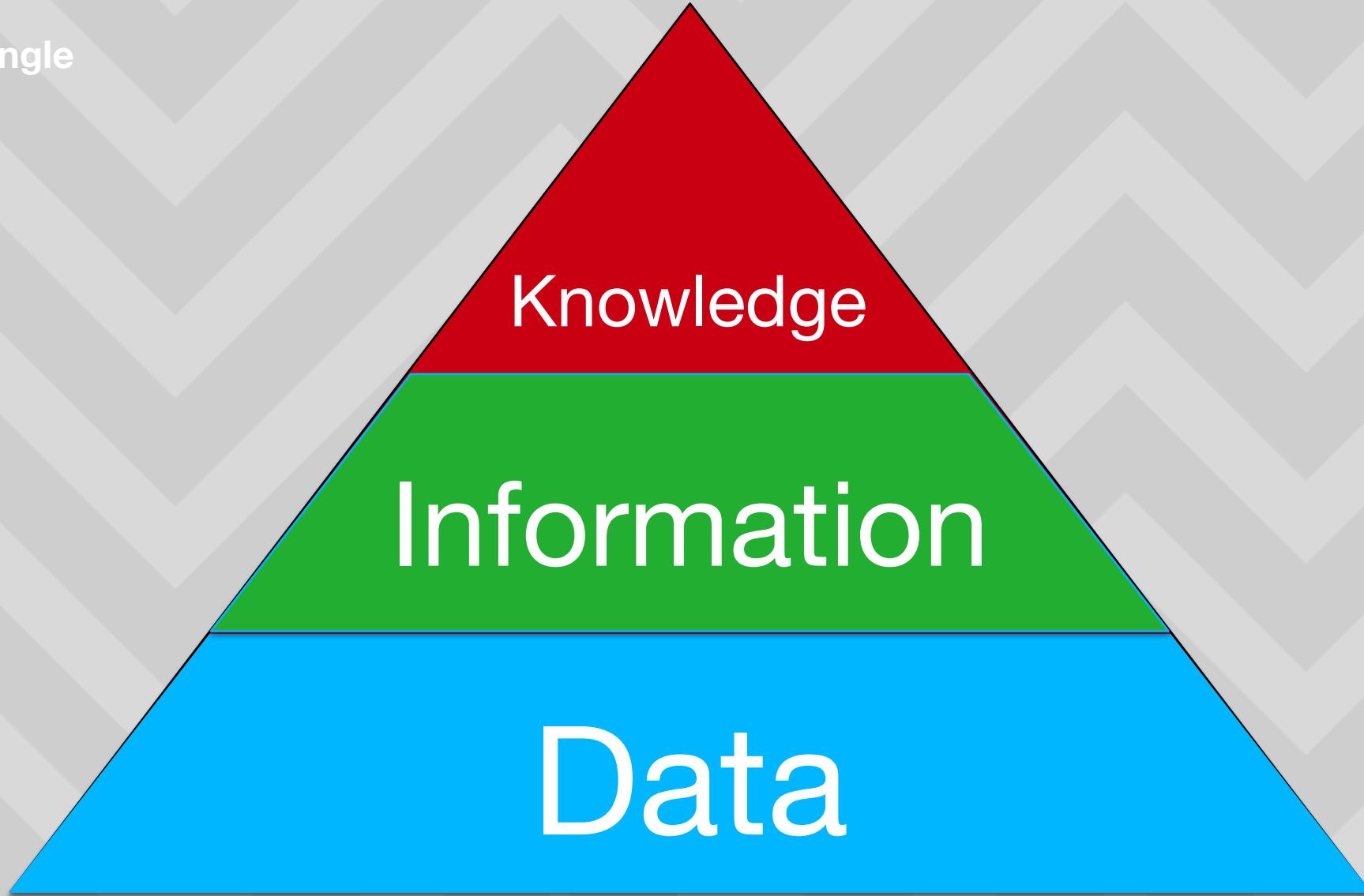


Definition of data

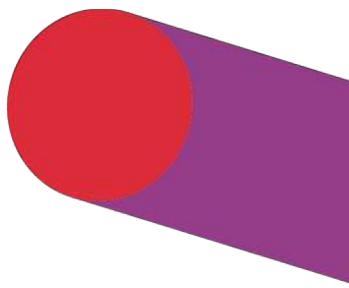
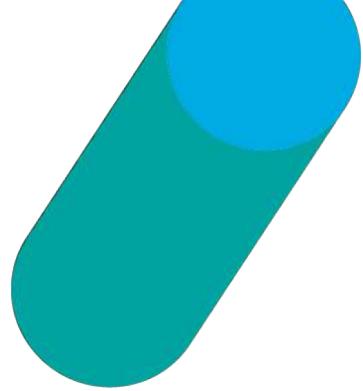
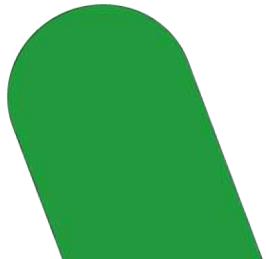
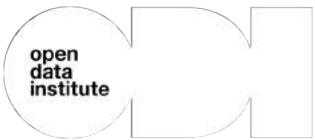
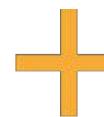
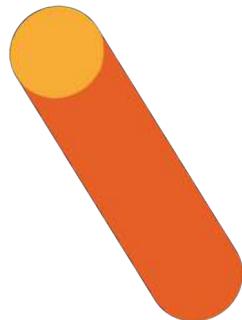
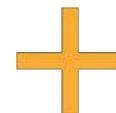
The lowest level of abstraction from which information and then knowledge are derived.



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)



CDN

Our theory of change

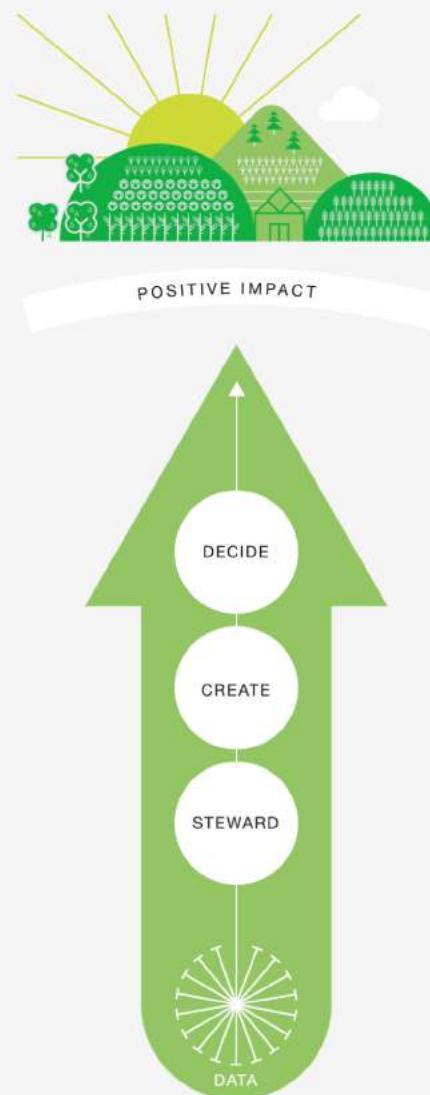
How value is
created from data



Our theory of change

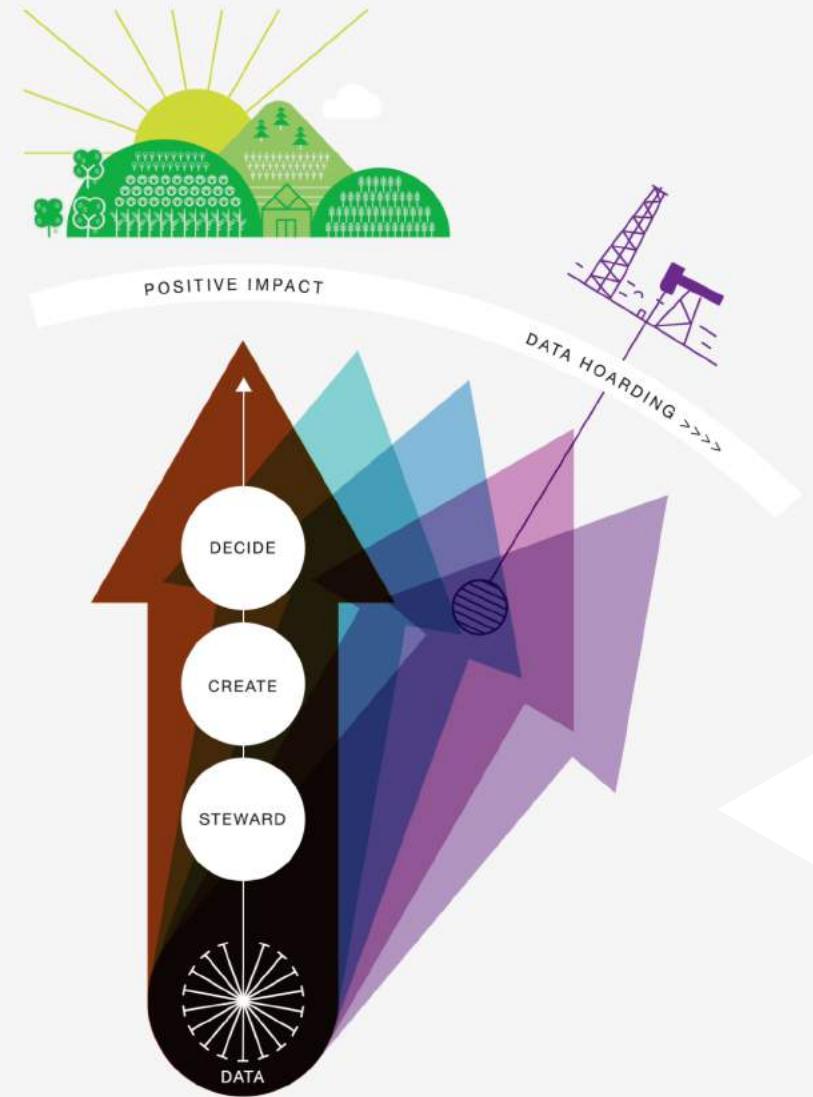
Scenario:

What
happens
when we
hoard data –
the oil field



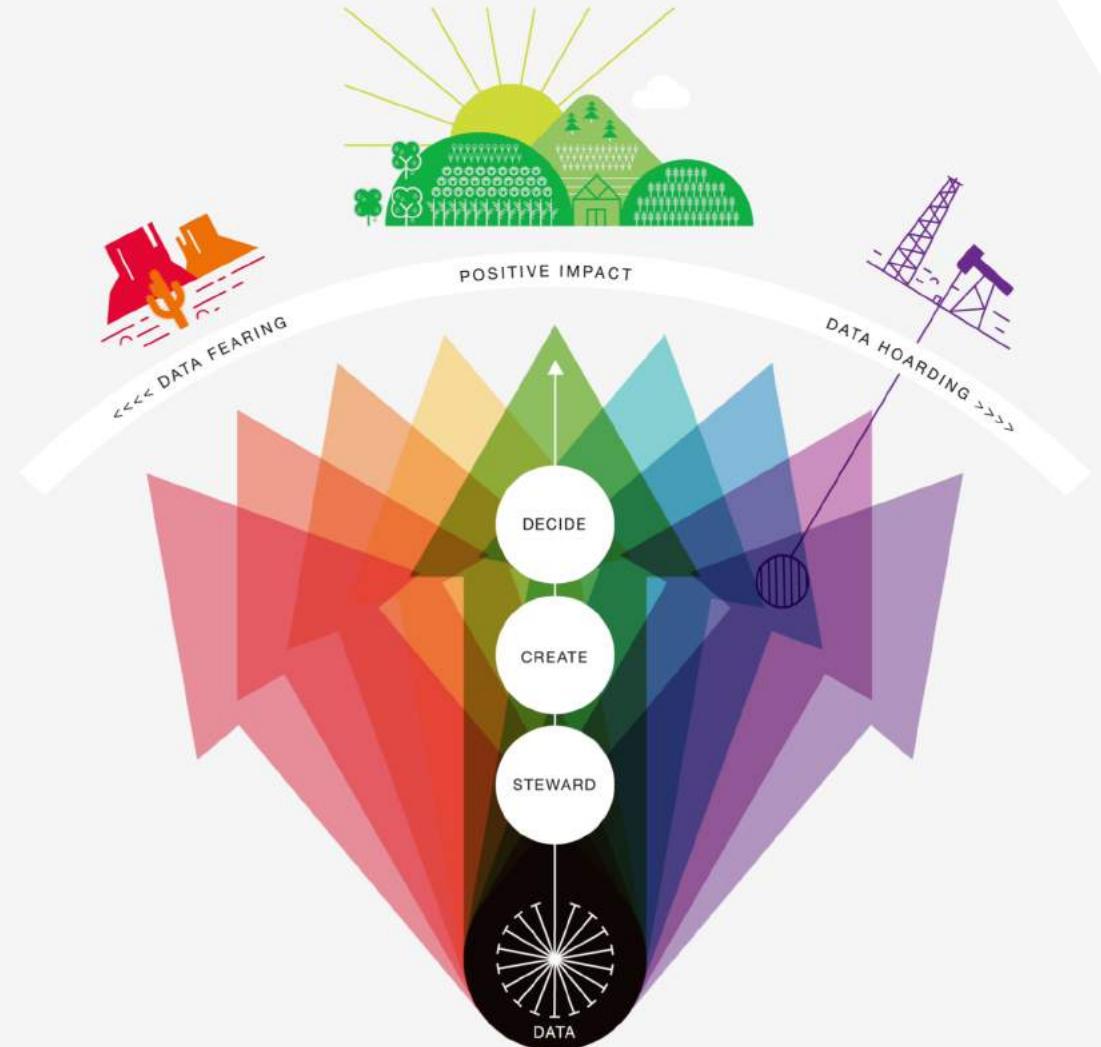
Scenario:

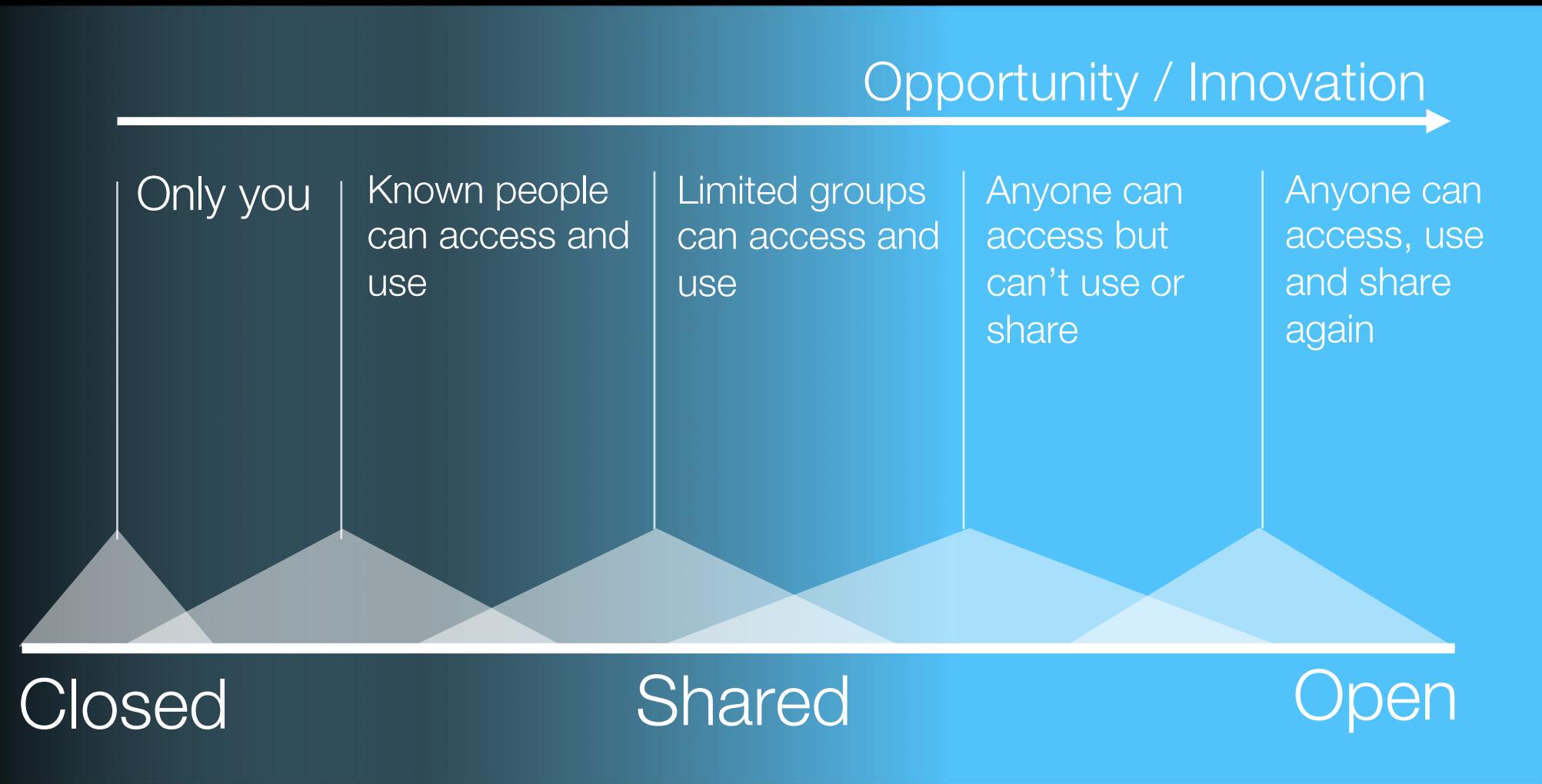
What happens when we fear data – the wasteland



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.





open
data
institute

/ Knowledge
for everyone

Case study: LIDAR

 DATA.GOV.UK Beta
Opening up Government

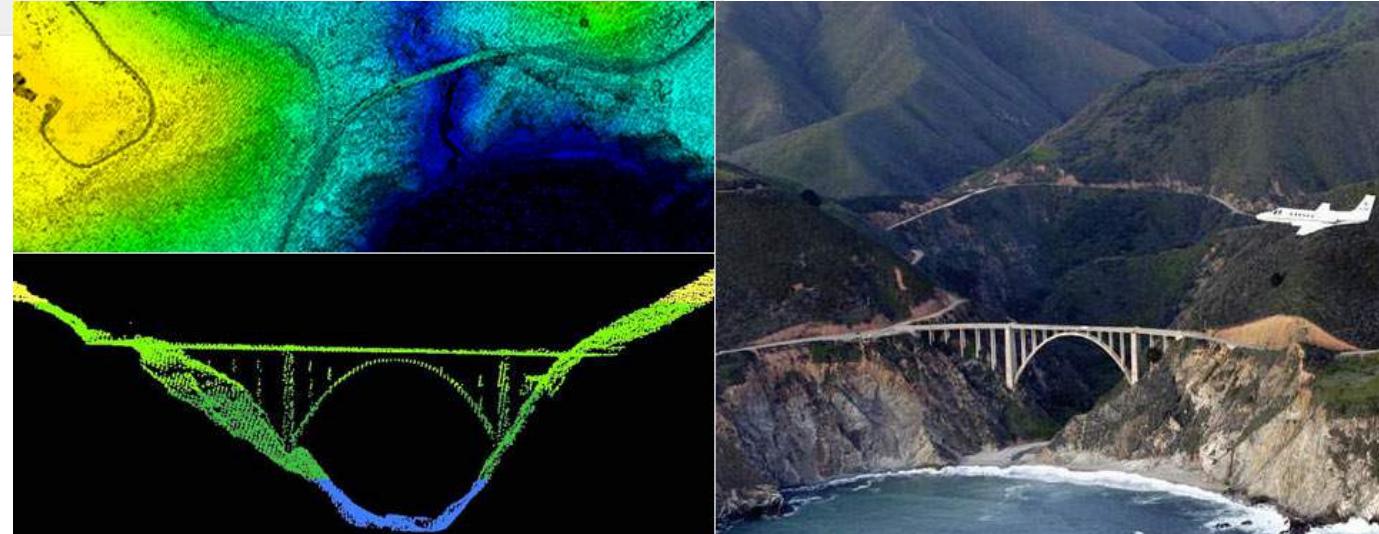
Home

Datasets Map Search Data Requests Publishers Data API Organograms Site Analytics

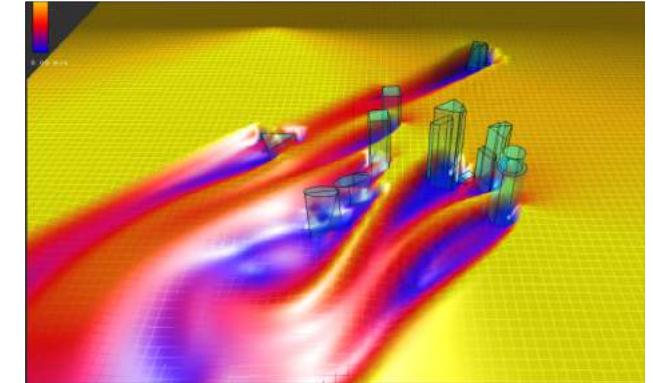
Home / Datasets / LIDAR Composite DTM - 50cm

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)



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**

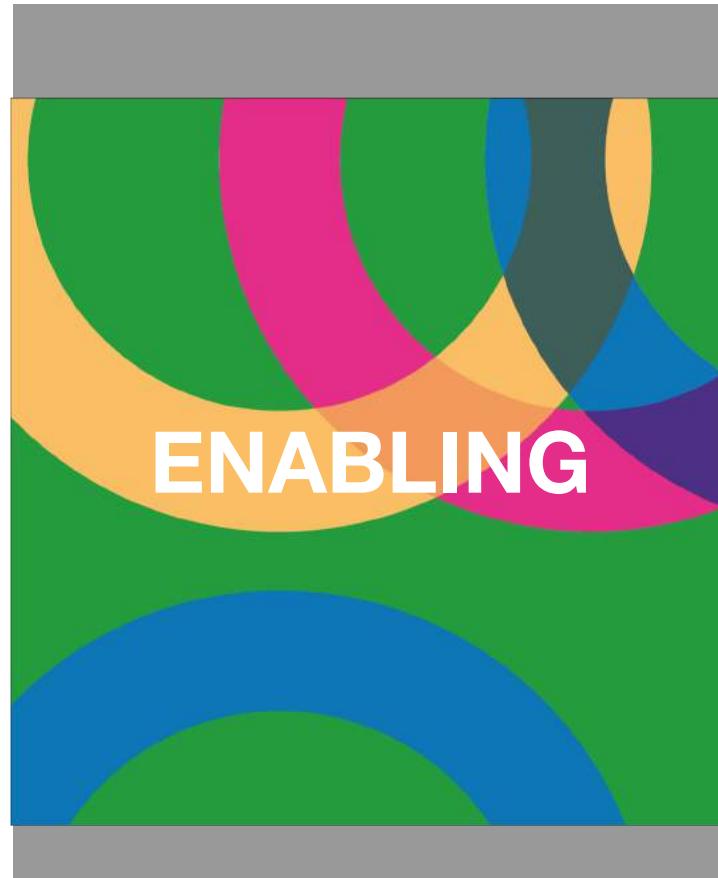
Our mission

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.

Justifications

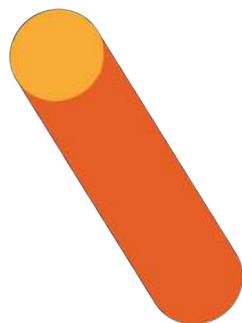


What are your favorite case studies involving open data?

+

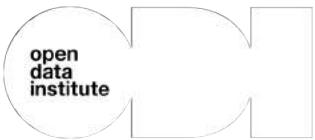
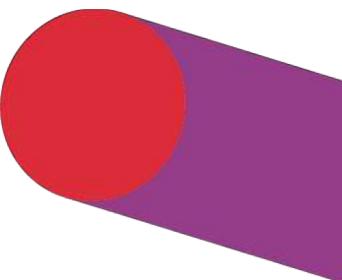
:

:



+

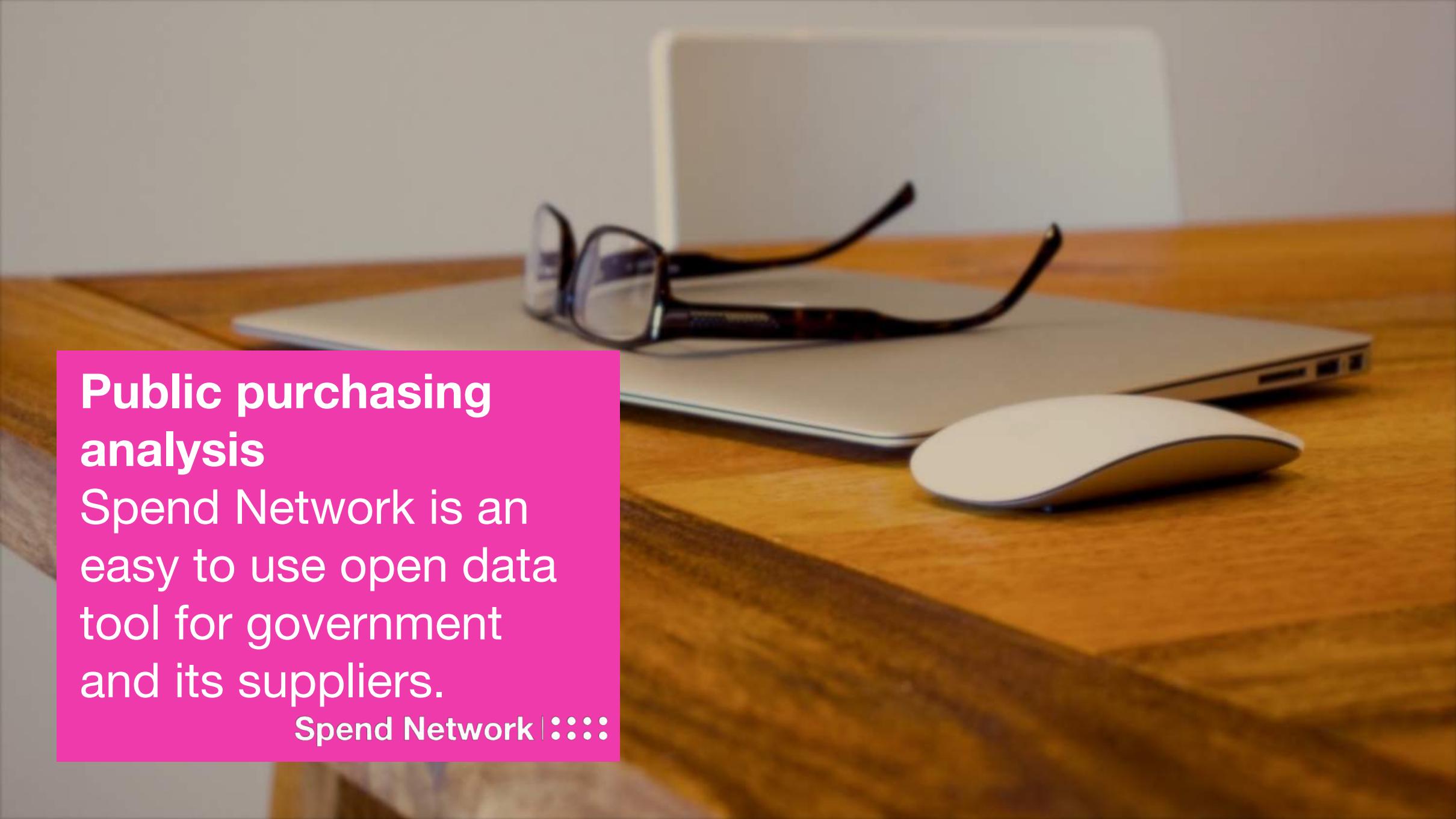
...

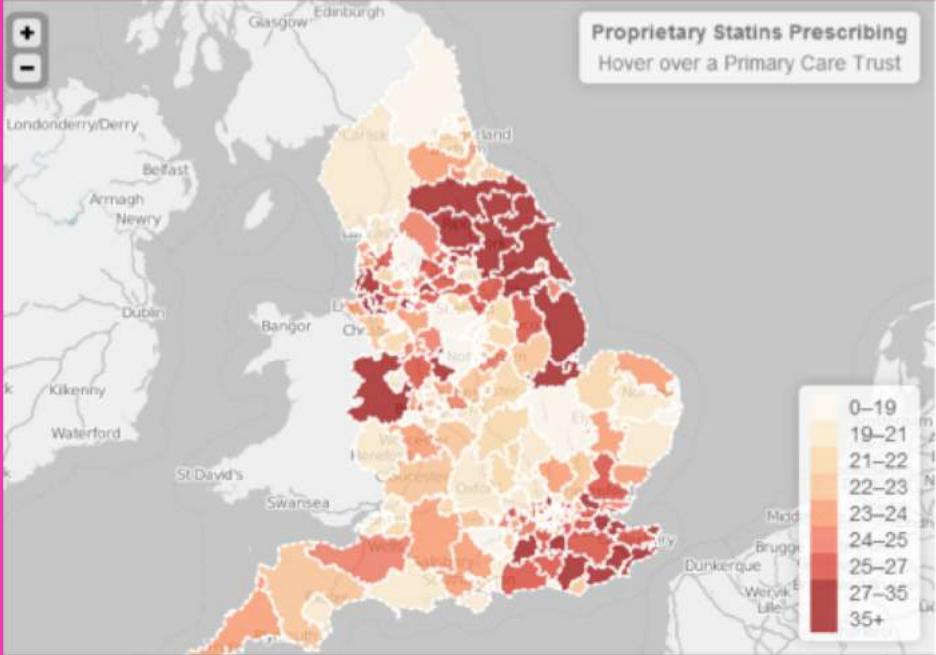


Public purchasing analysis

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

Section header: **subsection header**

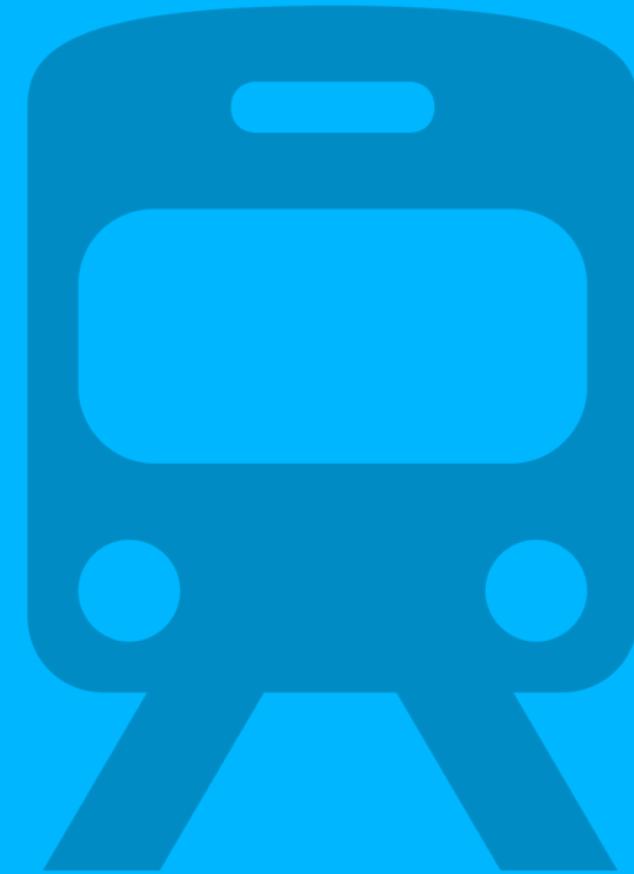
**Over 700 Apps now
rely on open
transport data:**

City Mapper (Millions of users)

Google Maps (Millions of users)

Real time trains (500,000 users)

Station master (10,000's users)

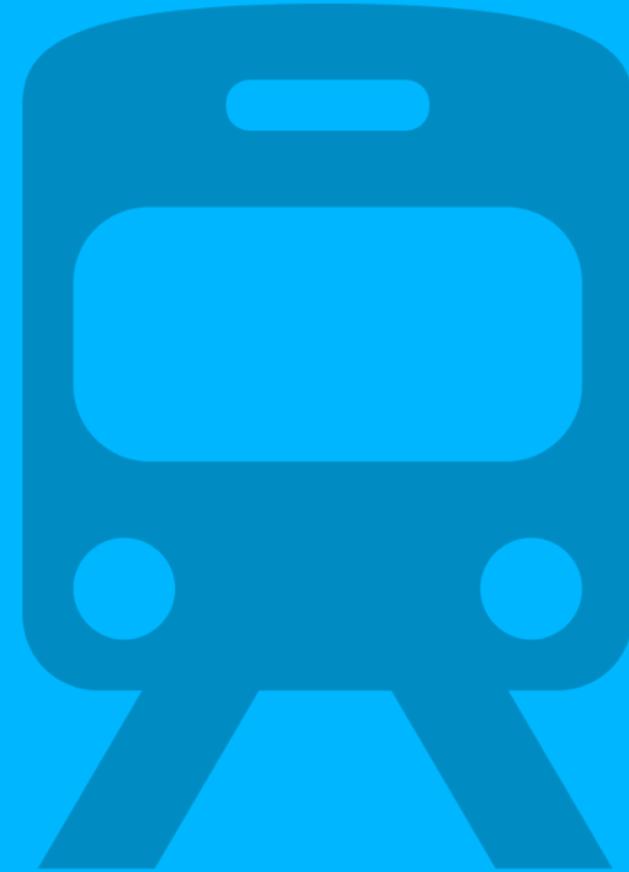


The London transport bubble

£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>

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>





*42% of Londoners
use an app powered
by TfL open data*

Transport for London

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).

TfL Open Data Benefits

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

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.

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.



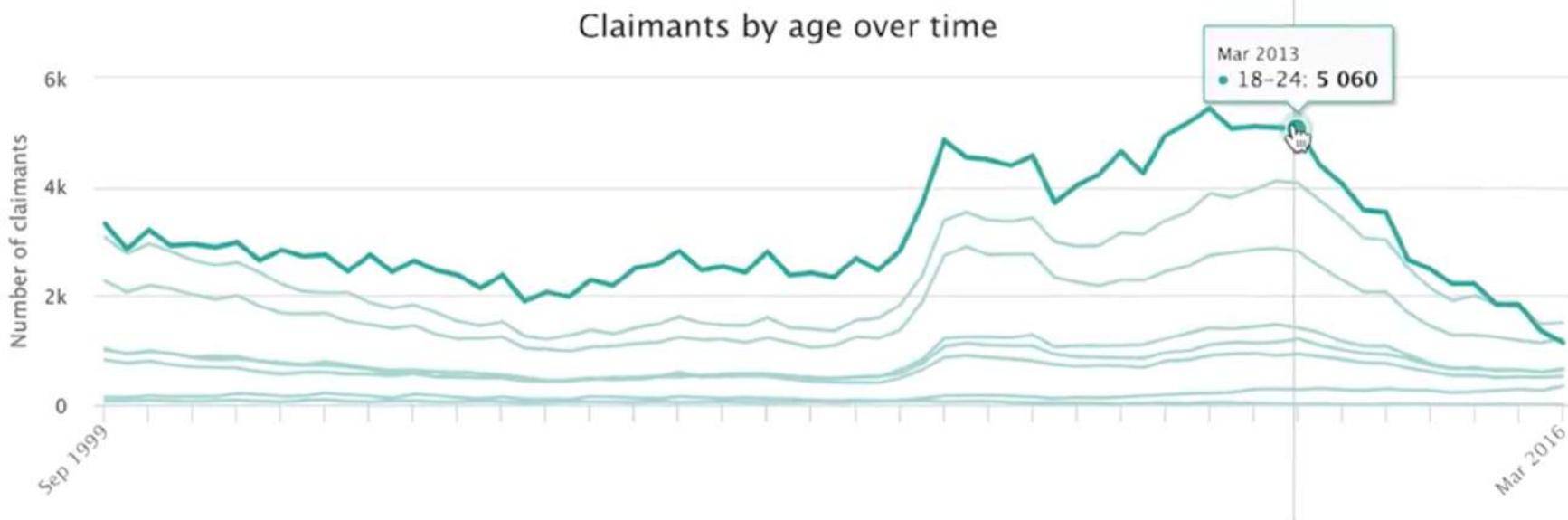
9:17 AM - 1 Jun 2018

6,002 Retweets 30,530 Likes

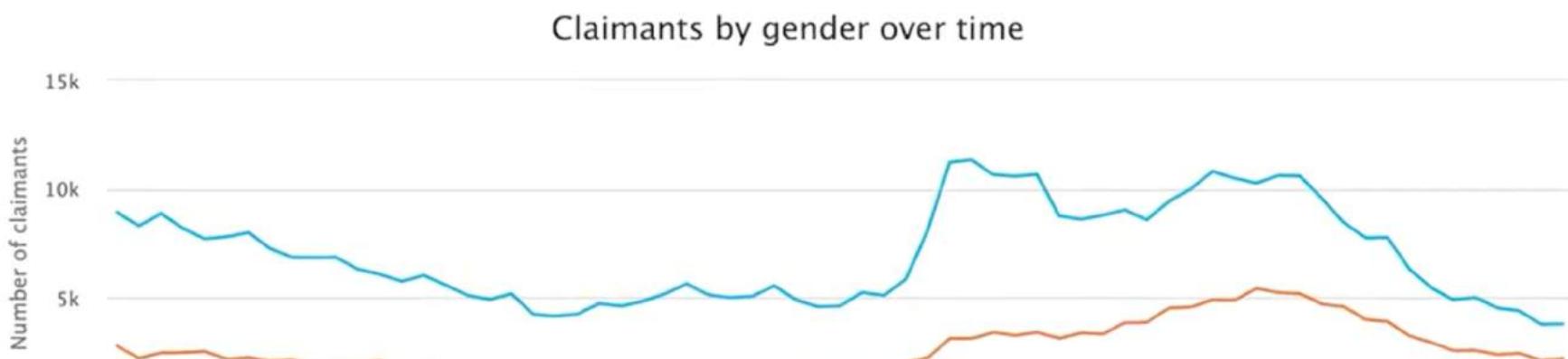
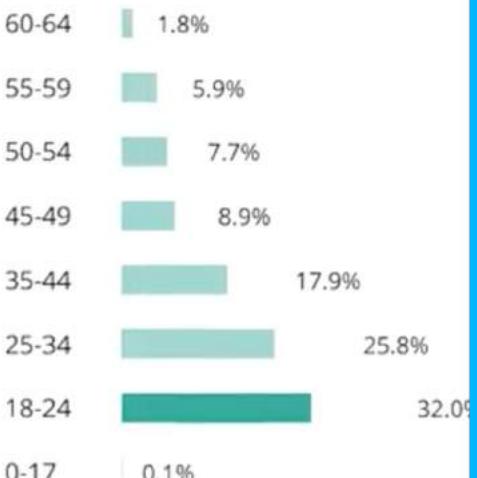


People on **Job Seekers Allowance** in **County Durham**[Back](#)

Co



15,790 total claimants



67.1%



32.9%

Public service delivery patterns

Efficiency

Transparency

Collaboration



Using open
data to deliver
public services

Open Data Institute



Data infrastructure

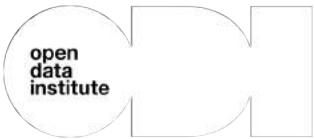
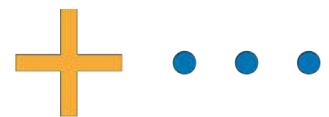
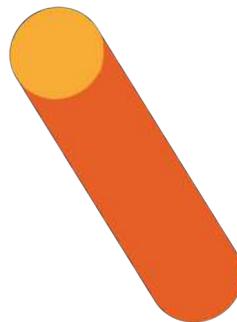
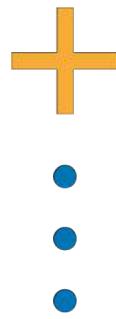
Data is infrastructure. It underpins transparency, accountability, public services, business innovation and civil society.

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

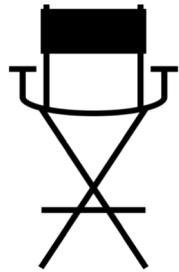


freepik

What other opportunities could being more open bring?

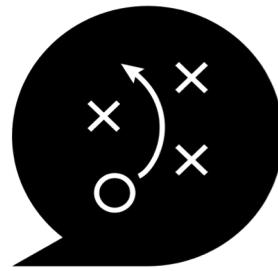


Benefits of open data



Leadership

Open data supports the wider data strategy of the organisation



Operations

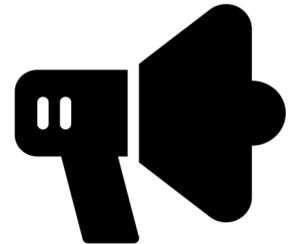
Open data creates significant data access efficiency savings



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



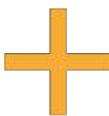
Joined up available transport data to enable choice. \$10m VC round and expanded to 29 cities



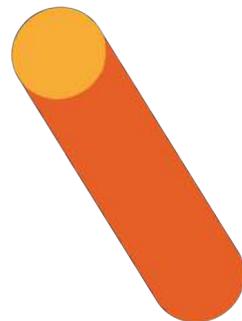
Emissions double reported figure. Evidence in parliament & planning debate



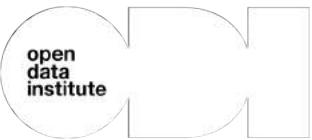
How would you
measure the success +
and development of
an Open Data
Initiative?



⋮



⋮



READINESS

Are governments ready to open their data?

IMPLEMENTATION

How well are governments fulfilling their open data promises?

IMPACT

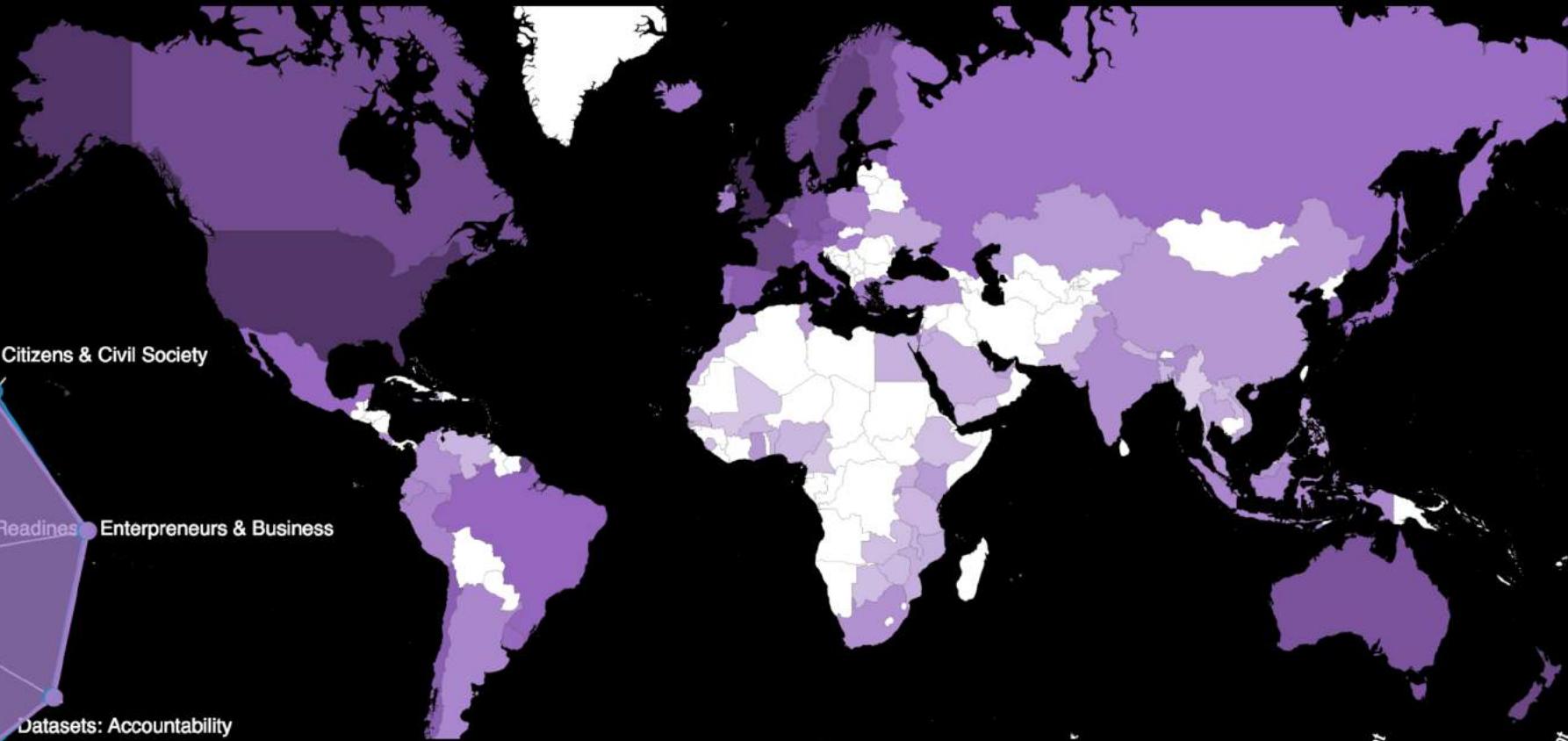
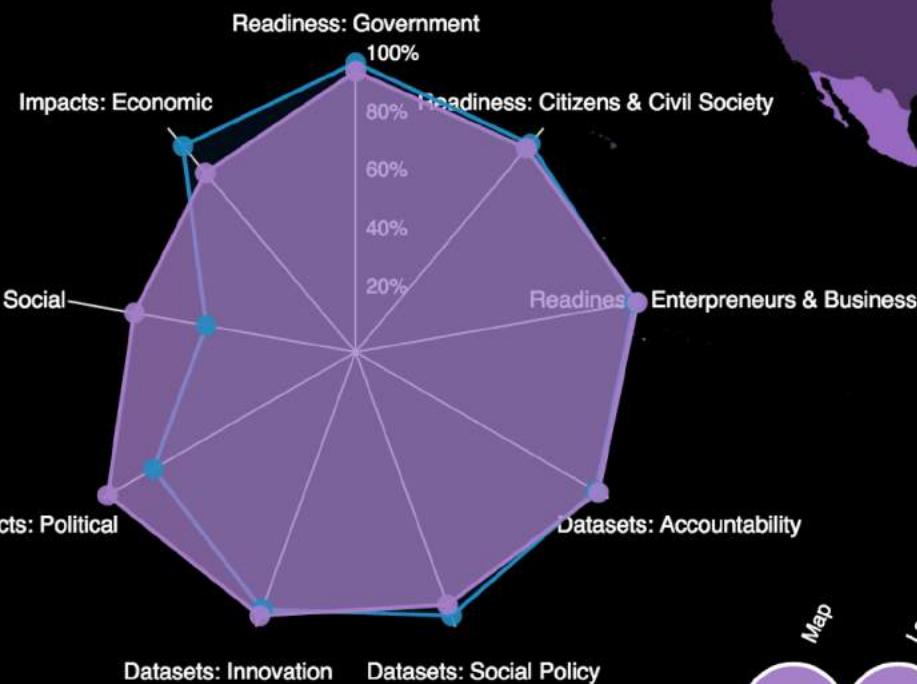
What impact is open data having on citizens' lives?

United Kingdom

1 / 86

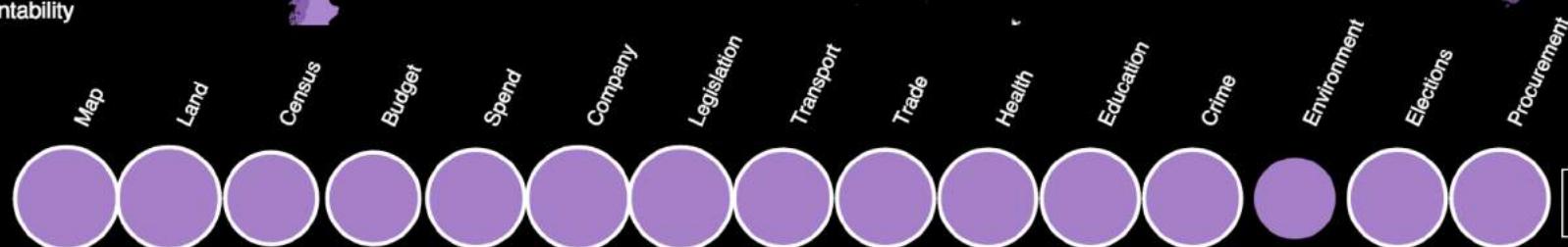
Score: 100.00

Filters: Off



 OpenData
Barometer

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



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
- 17:00** **END**

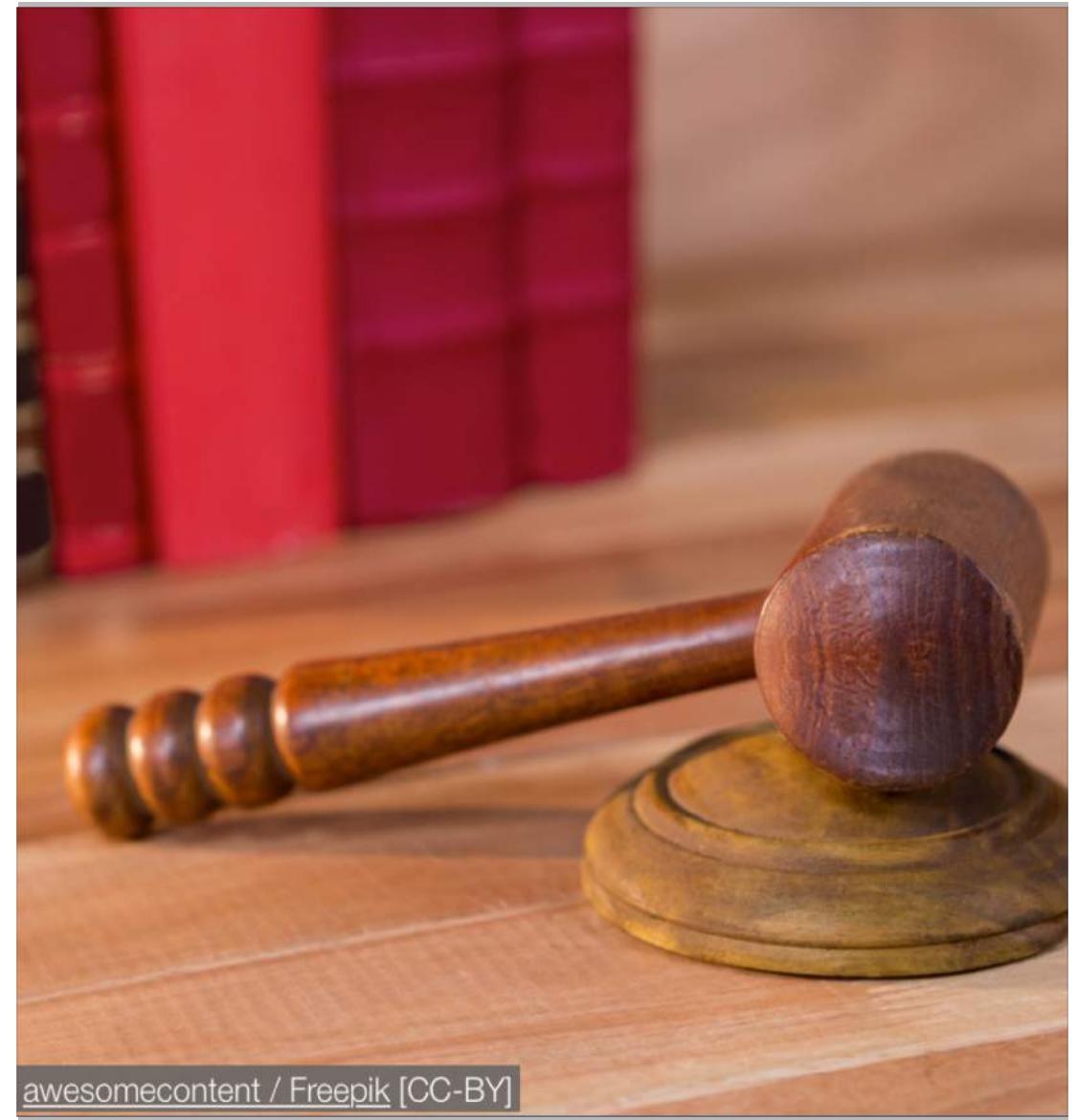
Law and Licensing

List the characteristics of an open licence

Identify instances of an open licence

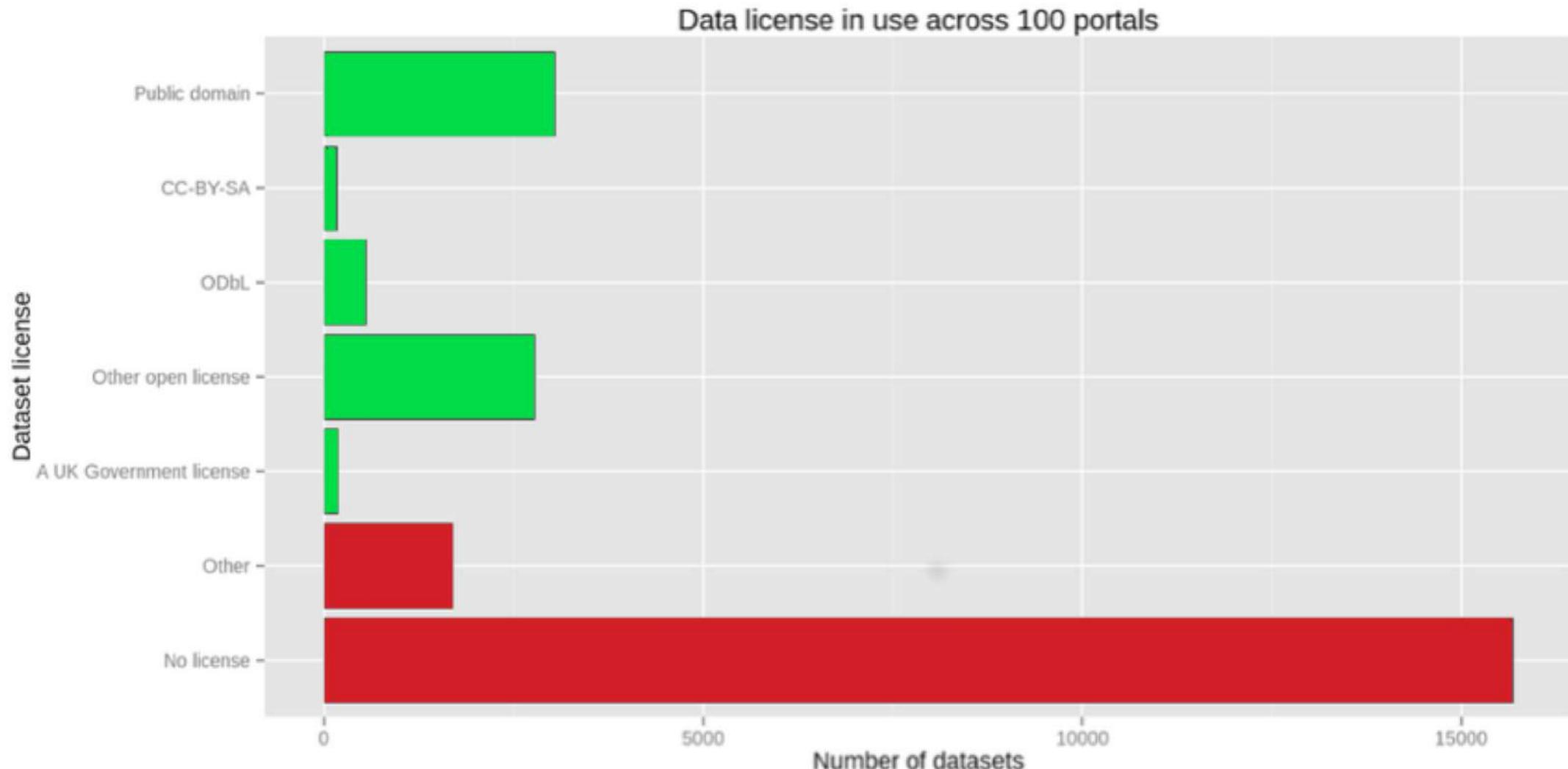
Please note:

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

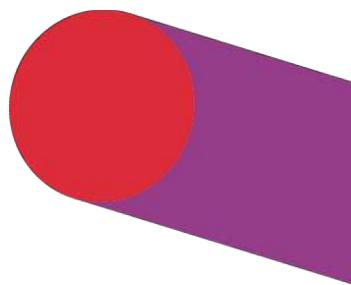
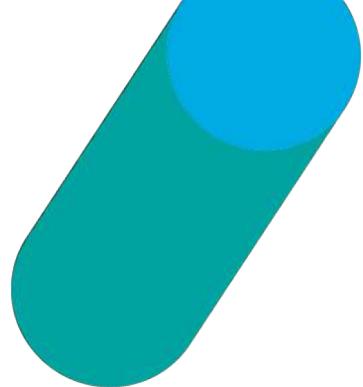
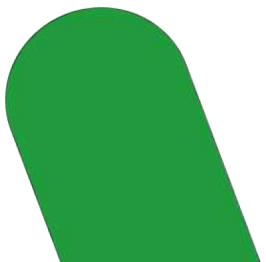
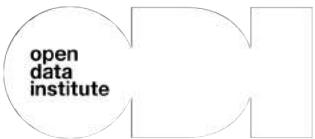
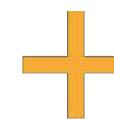
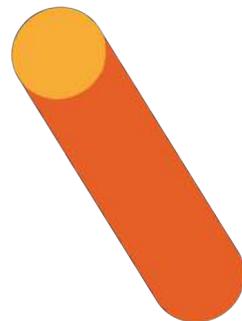
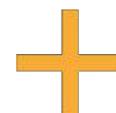


[awesomecontent / Freepik \[CC-BY\]](#)

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

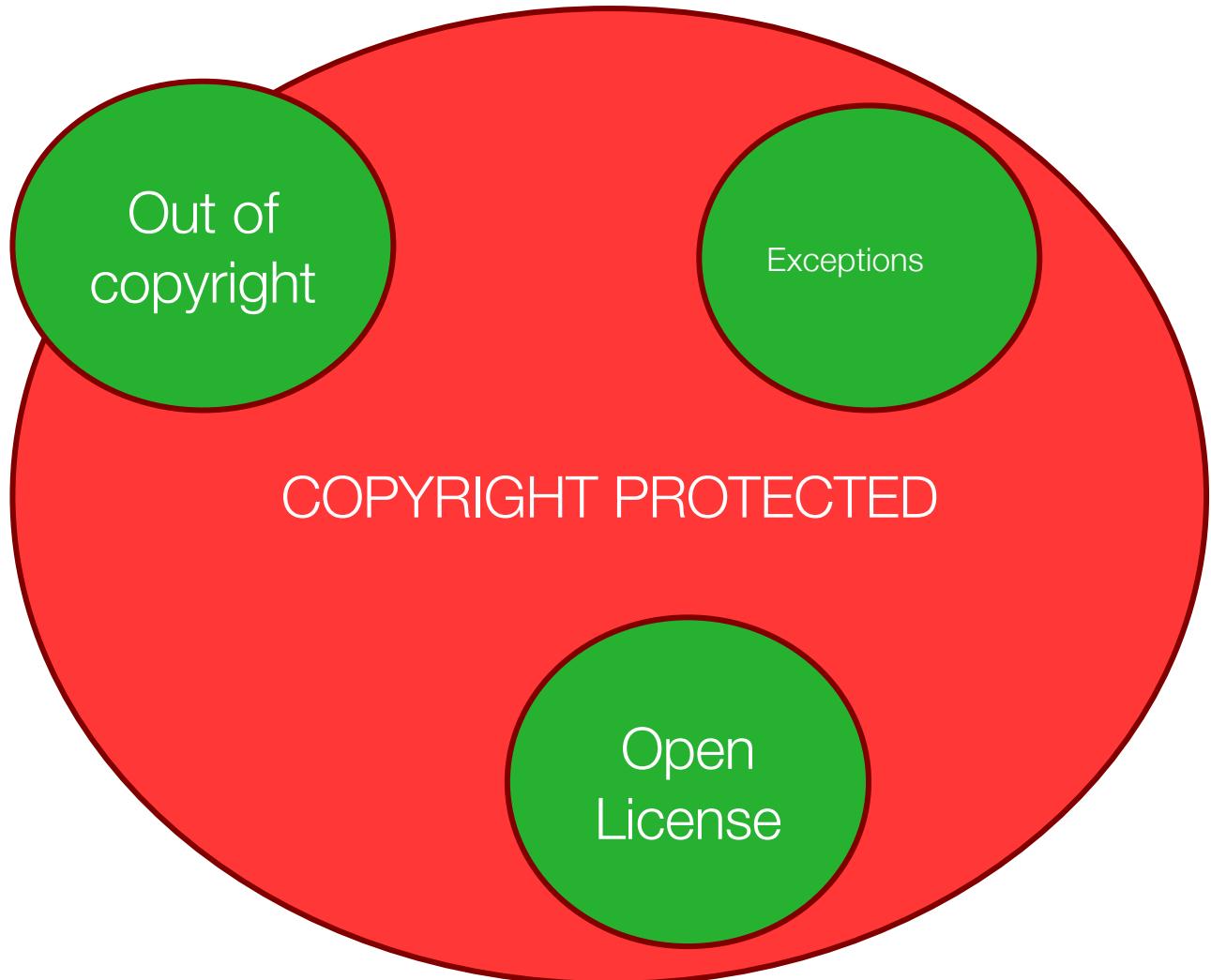
It sets out exactly what users and re-users are permitted to do with your dataset



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.

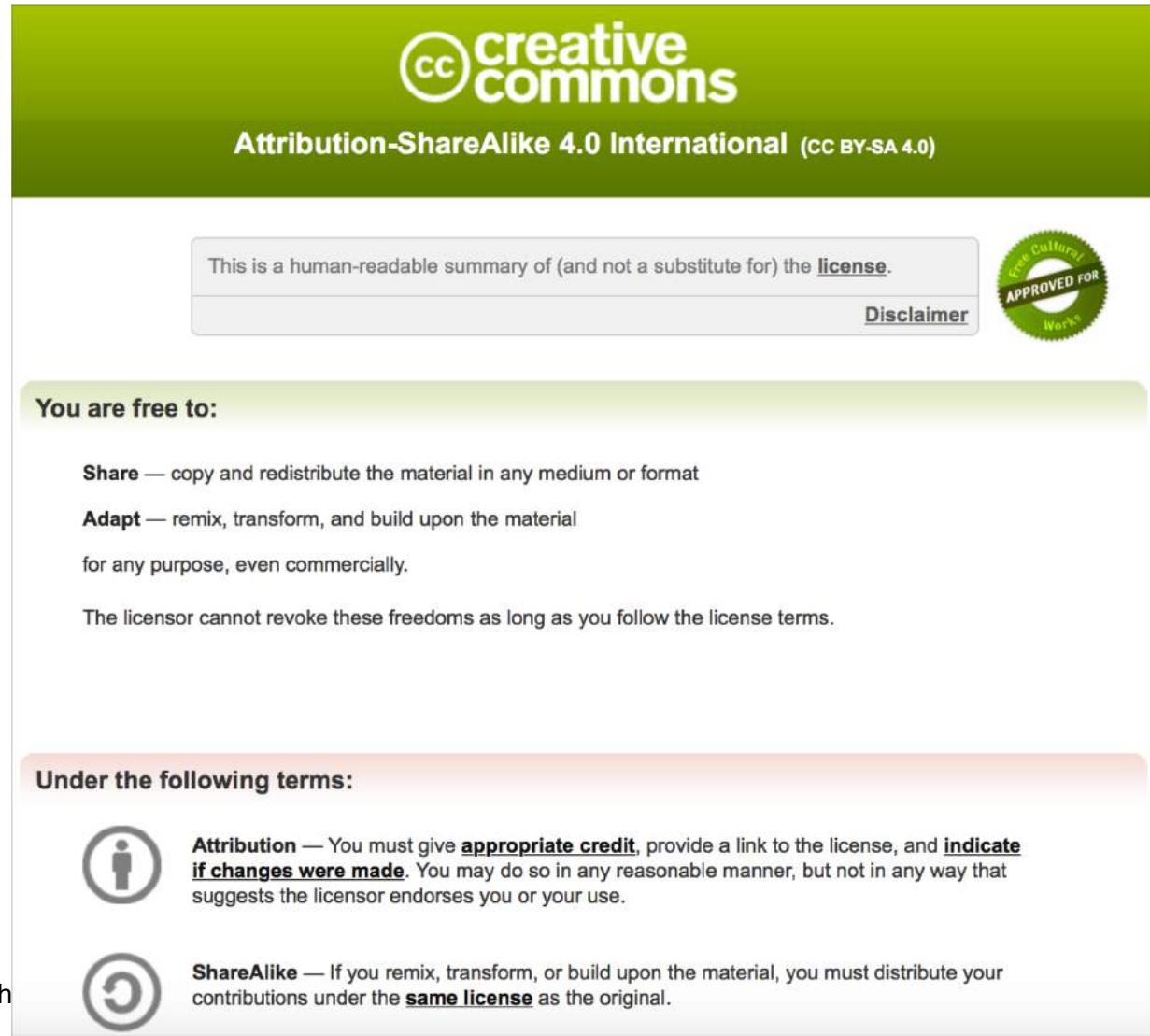


The main alternative



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

the



This image shows a screenshot of the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) license page. The top header is green with the CC logo and the text "creative commons". Below it, the license name "Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)" is displayed. A summary box contains the text: "This is a human-readable summary of (and not a substitute for) the [license](#)". Below the summary is a "Disclaimer" link. To the right is a green circular badge with the text "Free Cultural Works APPROVED FOR". The main content area has a green-to-white gradient background. It starts with "You are free to:" followed by two bullet points: "Share — copy and redistribute the material in any medium or format" and "Adapt — remix, transform, and build upon the material for any purpose, even commercially." A note below states: "The licensor cannot revoke these freedoms as long as you follow the license terms." At the bottom, there's a section titled "Under the following terms:" with two items: "Attribution" (represented by a person icon) and "ShareAlike" (represented by a circular arrow icon). Both sections describe the requirements for attribution and the distribution of derivative works under the same license.

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.

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



Share Alike



No Derivative Works



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](#)

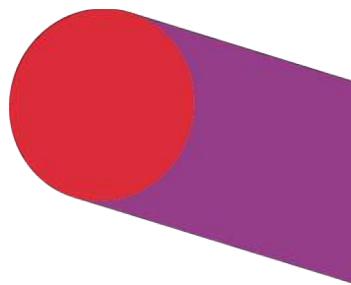
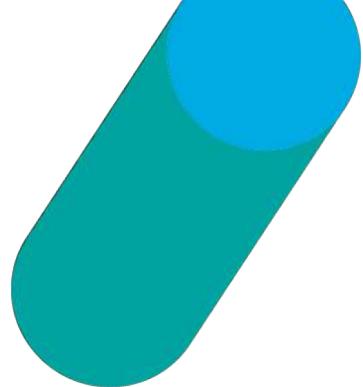
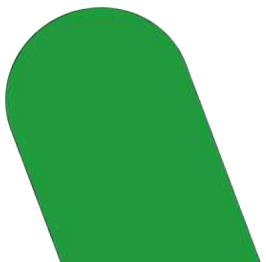
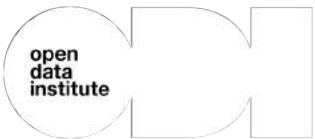
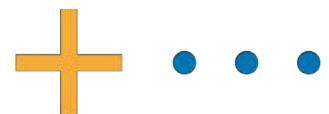
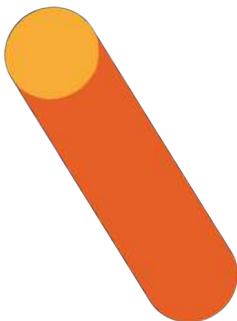
What makes data usable?

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



-
-
-

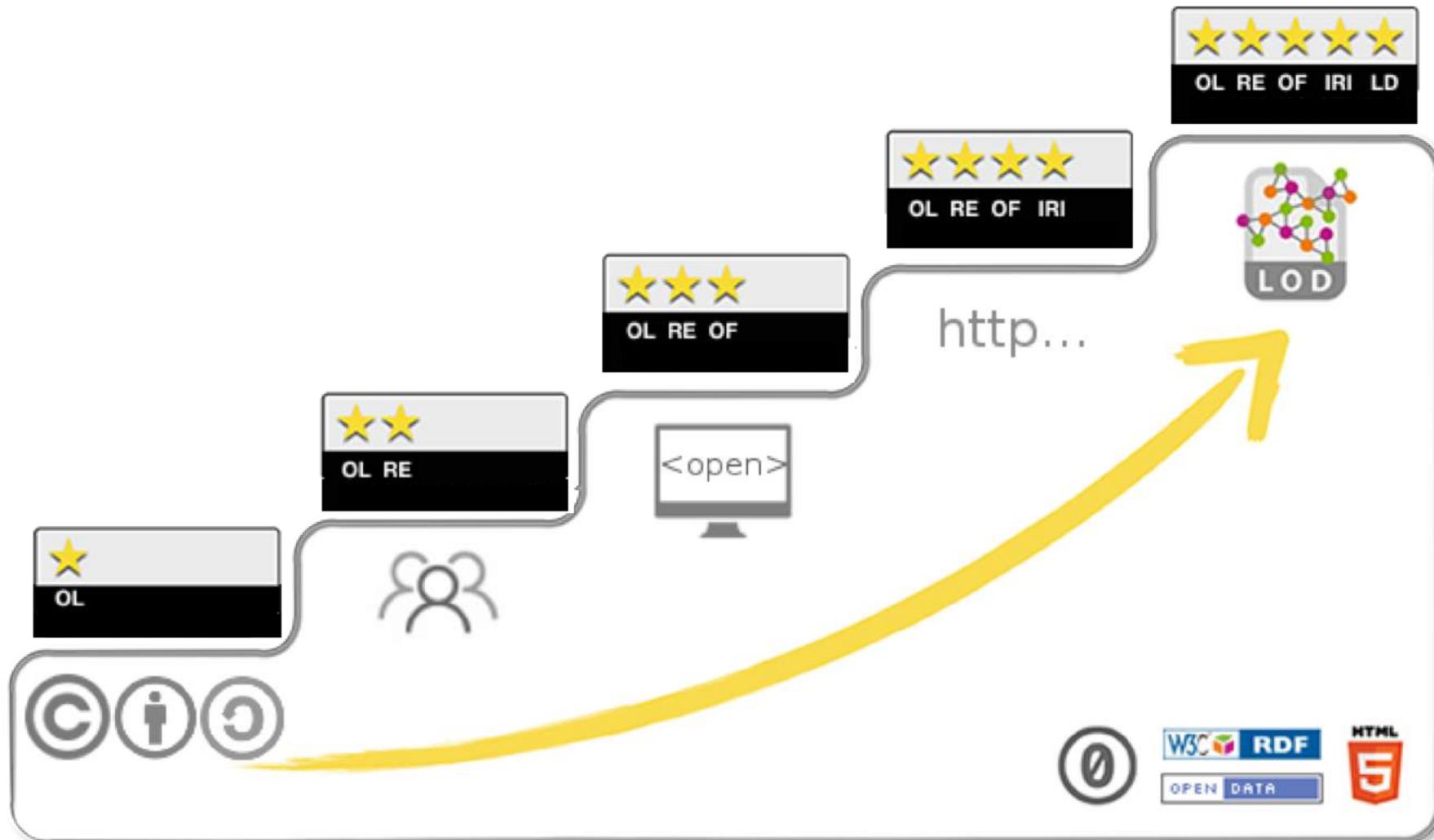


Guidelines



5 - S t a r s

The 5-Stars of Linked Open Data



Open Data Certificate

 **Bronze level**
self certified
GB final

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

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...>

Community verification

Other people can verify whether the answers on this

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



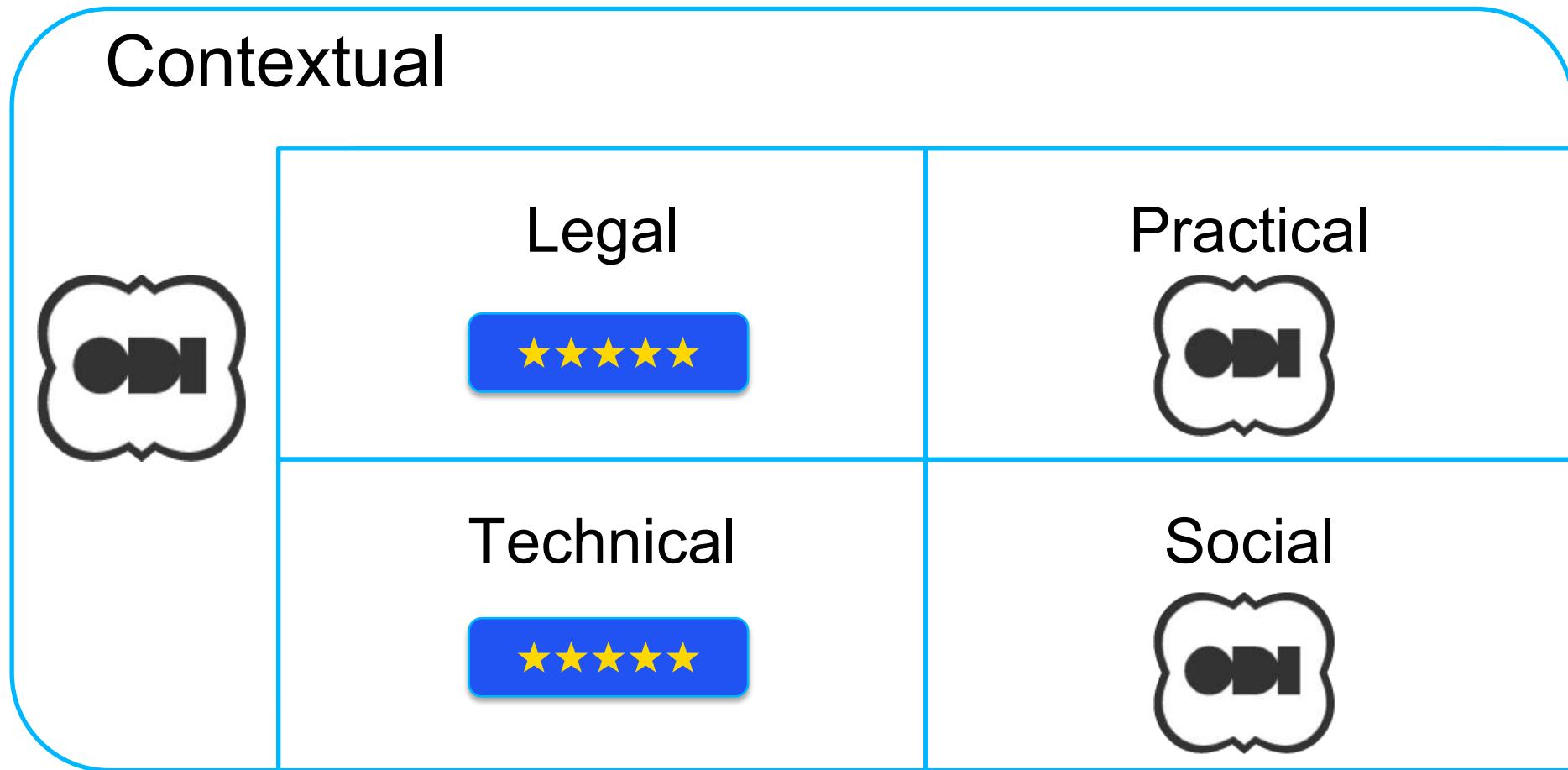
Practical

Technical

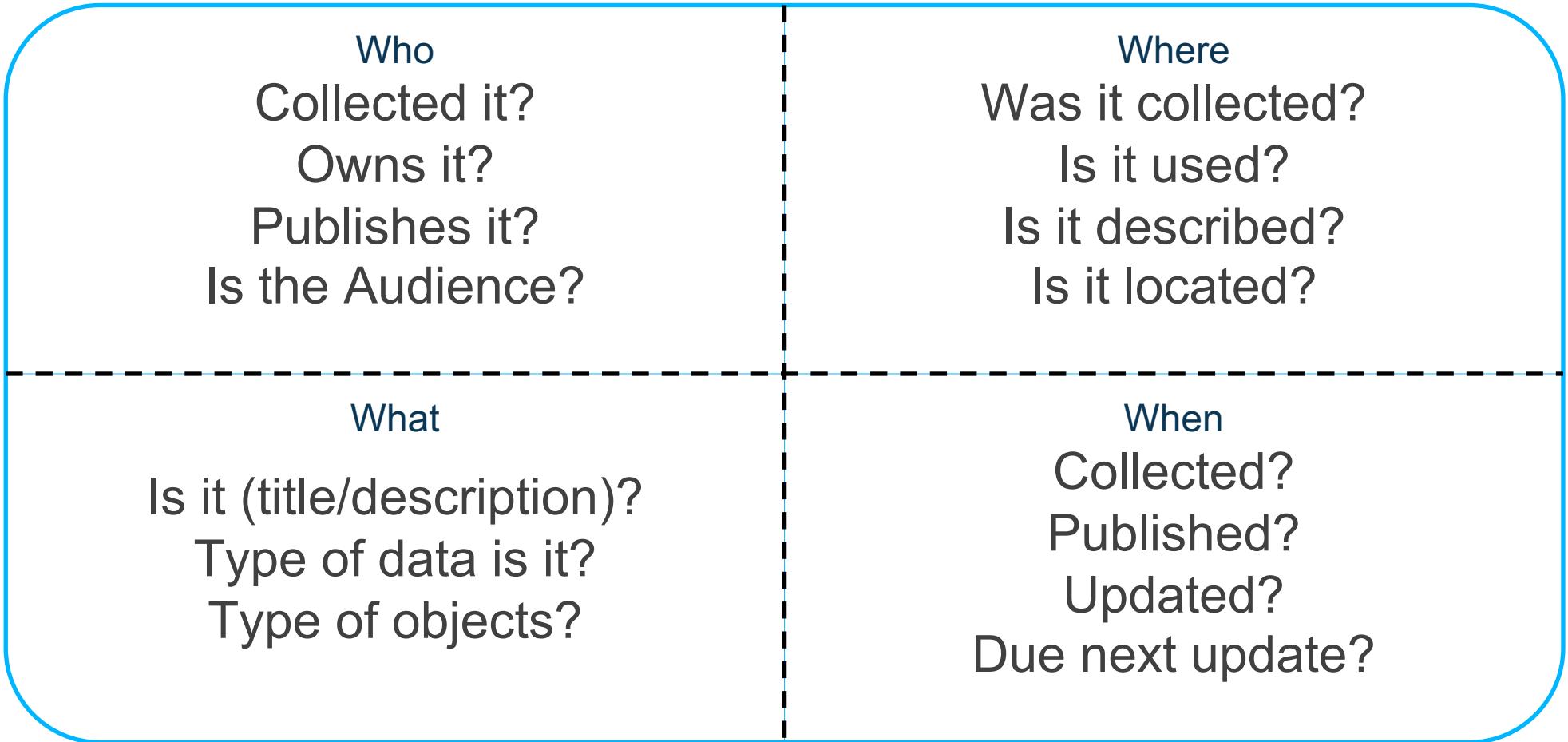


Social

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

-

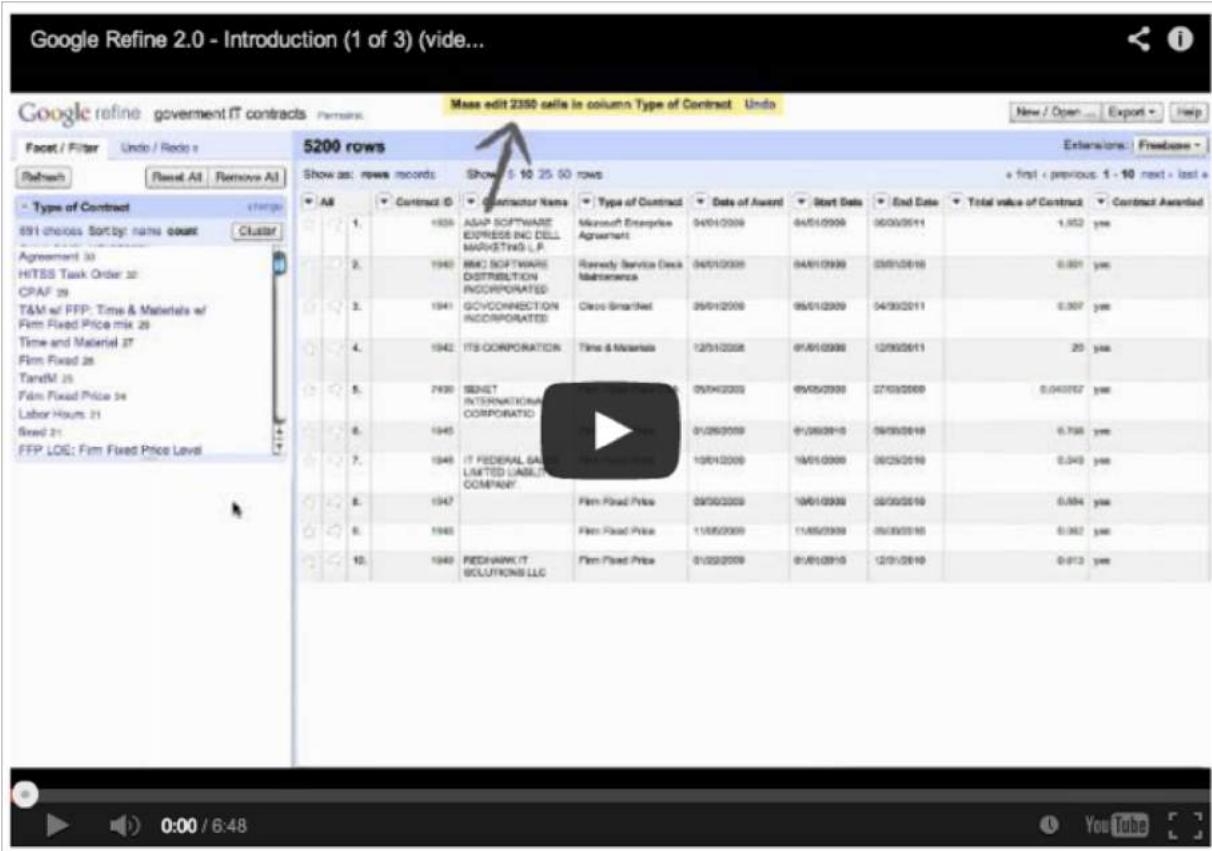
Tools and guides available to work with data

-

Agenda:

- 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
- 17:00** **END**

Open Refine



<http://openrefine.org>

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

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

