

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 and transparent operation in order to provide the best in class services for its residents.



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
- 16:00** **END**



Introductions

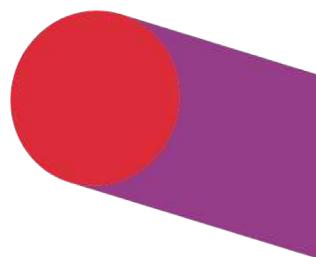
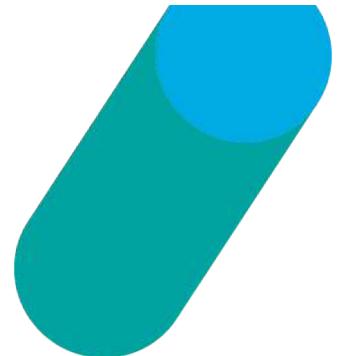
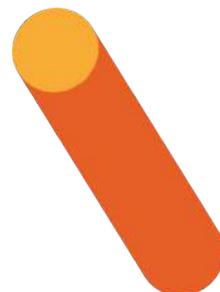
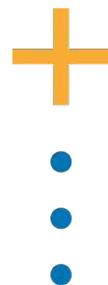
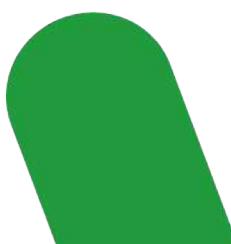
- Your name
- What do you do?
- What was your first experience of open data



theODI.org



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

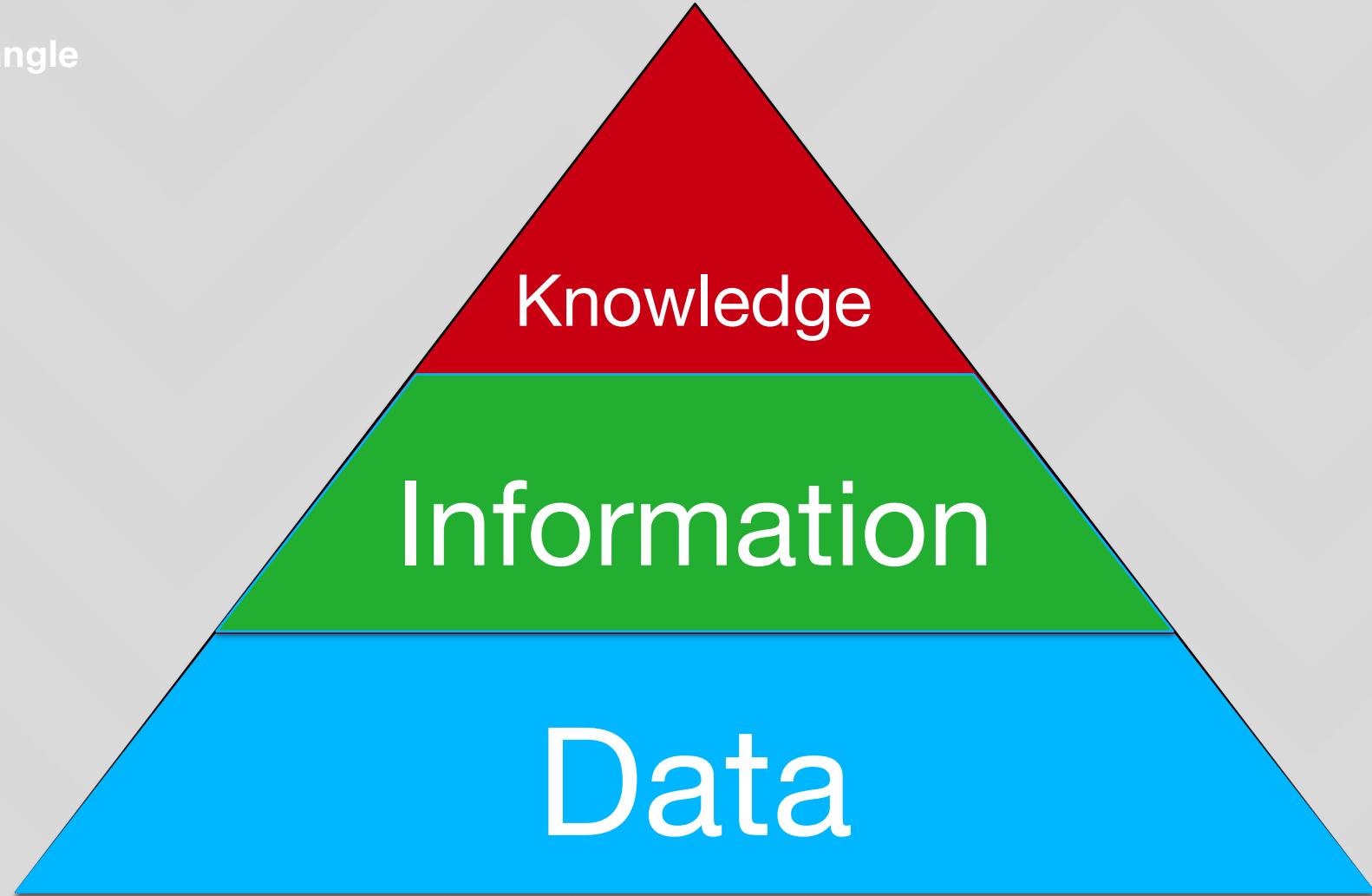


Definition of data

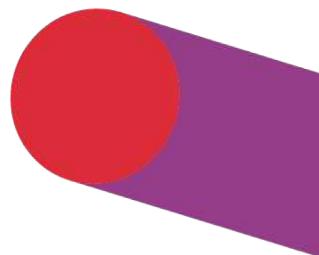
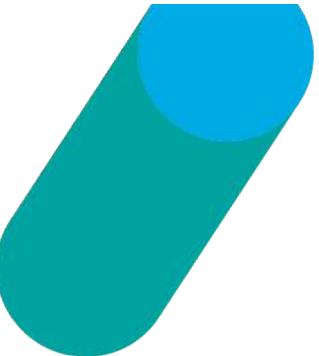
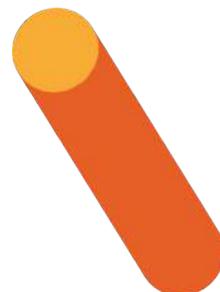
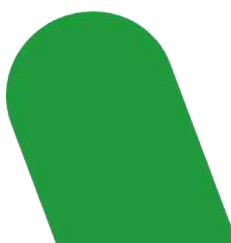
The lowest level of abstraction from which information and then knowledge are derived.



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



theODI.org



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)



A photograph of an oilfield at sunset, featuring numerous pumpjacks in various stages of their stroke. The sky is a warm orange and yellow. In the bottom right corner, there is a small white logo consisting of the letters "CDI" in a bold, sans-serif font, enclosed within a thin circular border.

CDI

Our theory of change

**How value is
created from data**



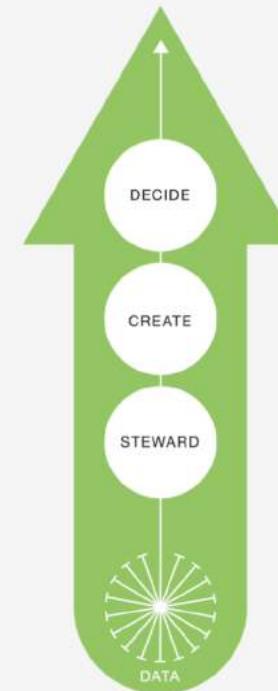
Our theory of change

Scenario:

What
happens
when we
hoard data –
the oil field



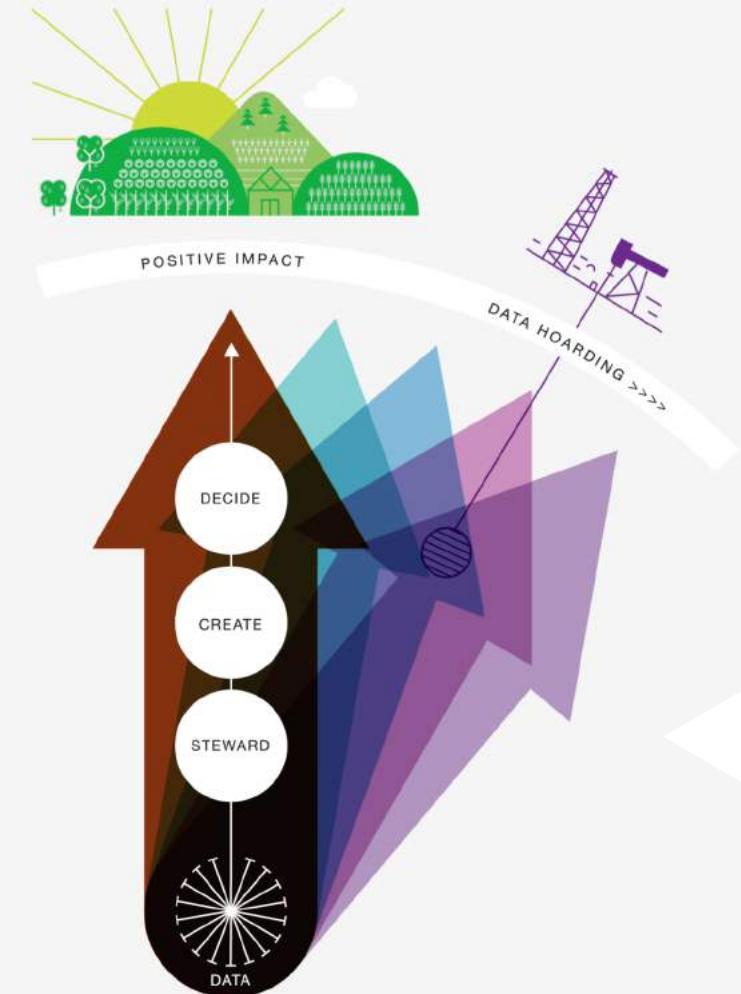
POSITIVE IMPACT



Our theory of change

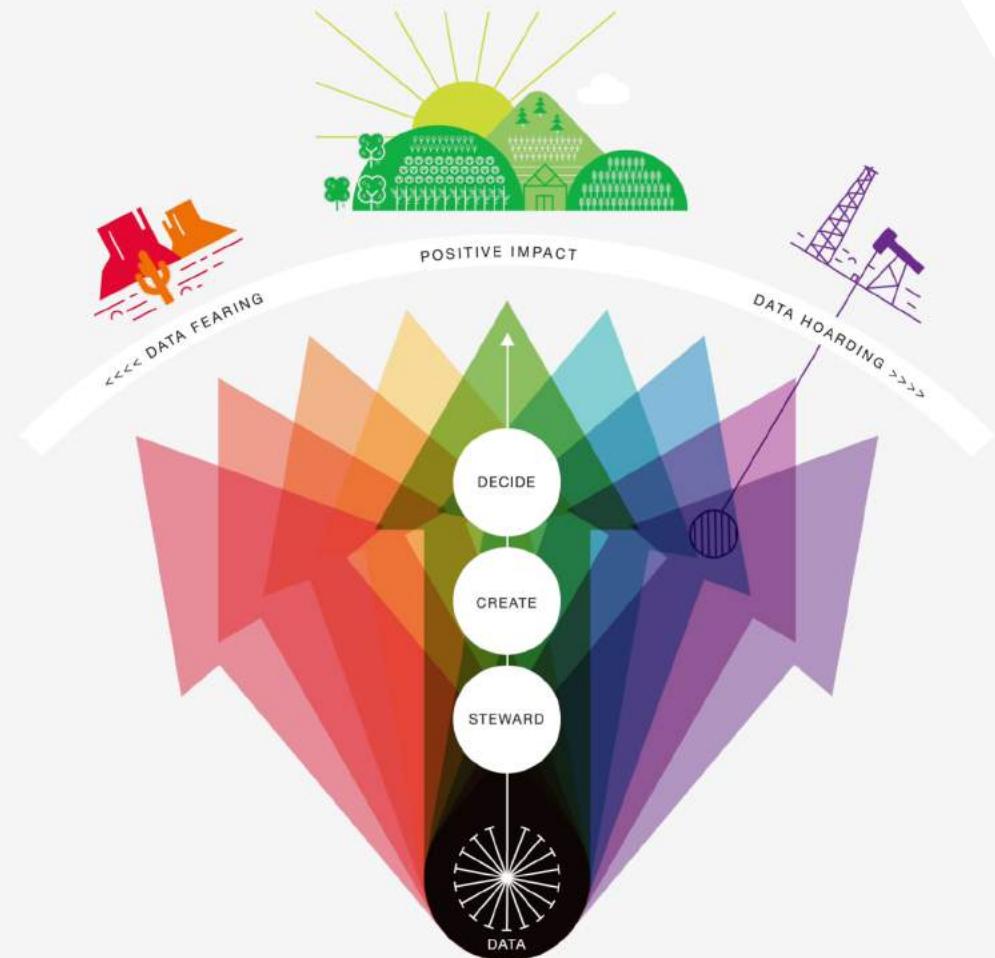
Scenario:

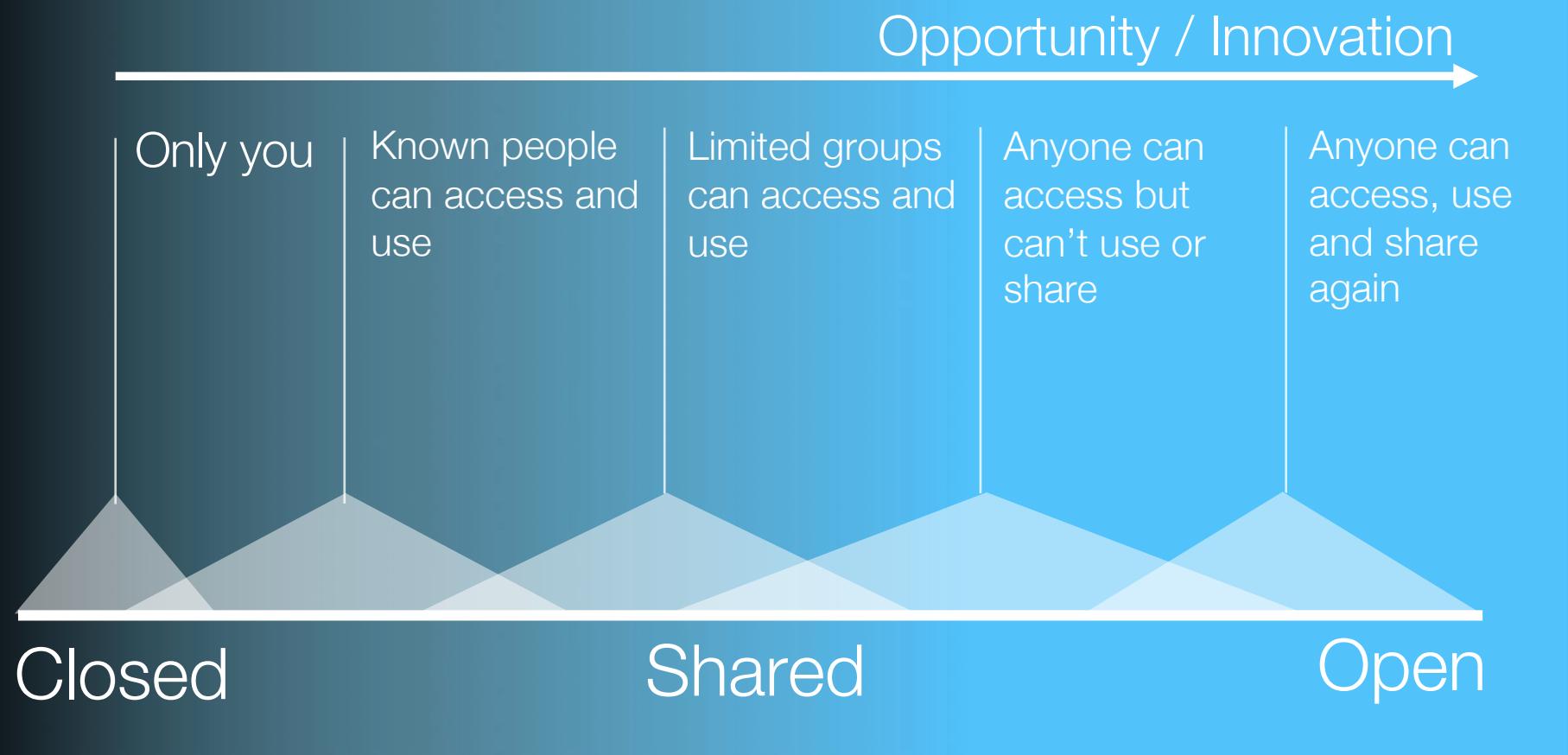
What
happens
when we
fear data –
the wasteland



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.





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

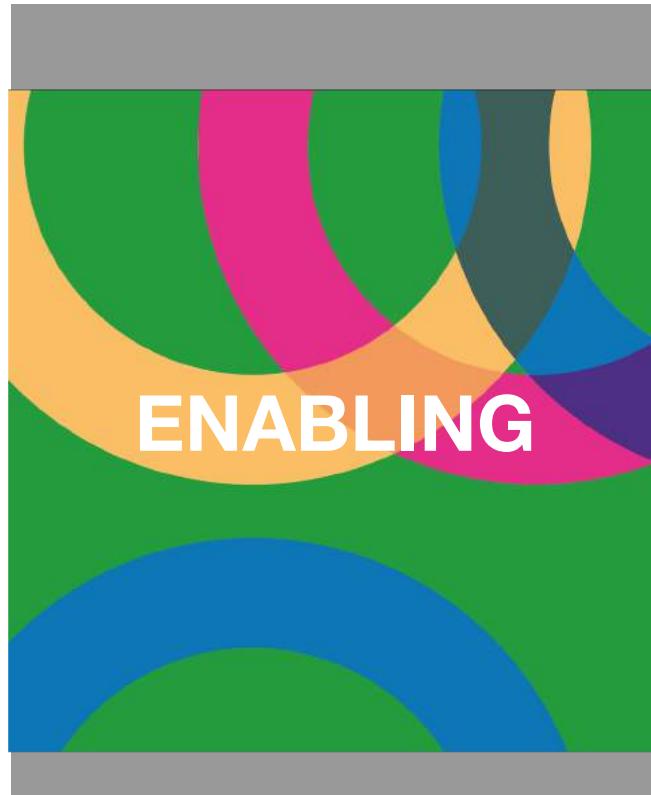
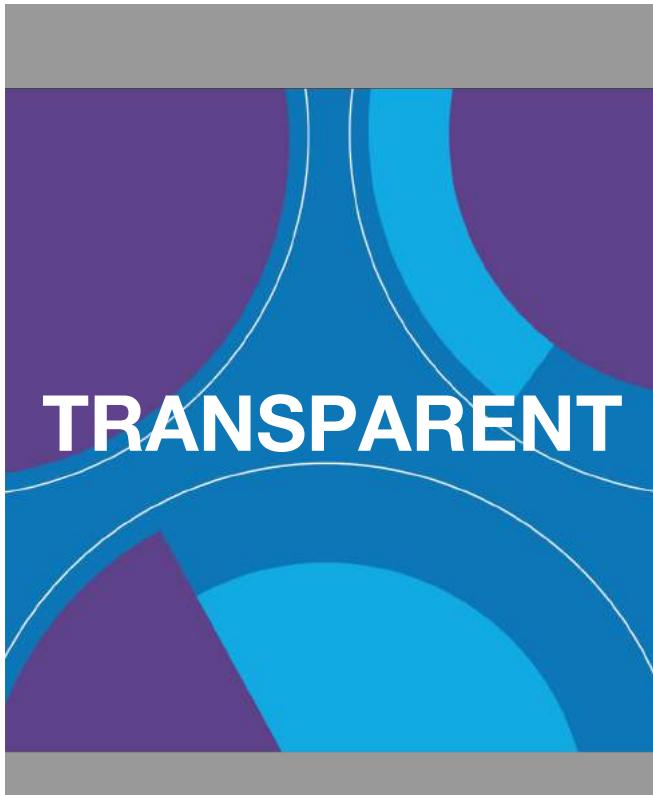
We work with
companies and
governments to
build an open,
trustworthy data
ecosystem.



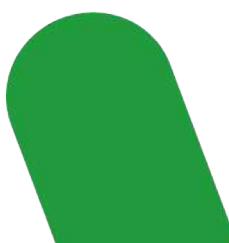
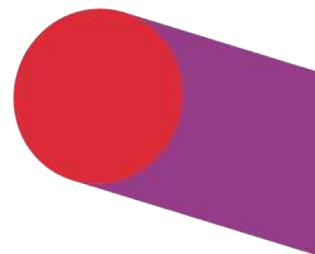
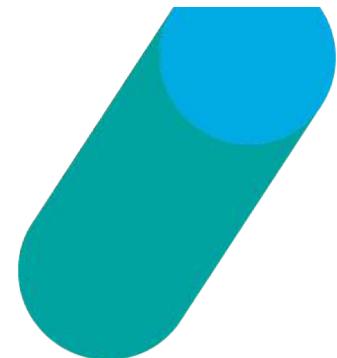
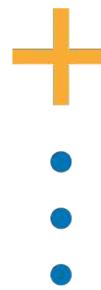
Our vision

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?



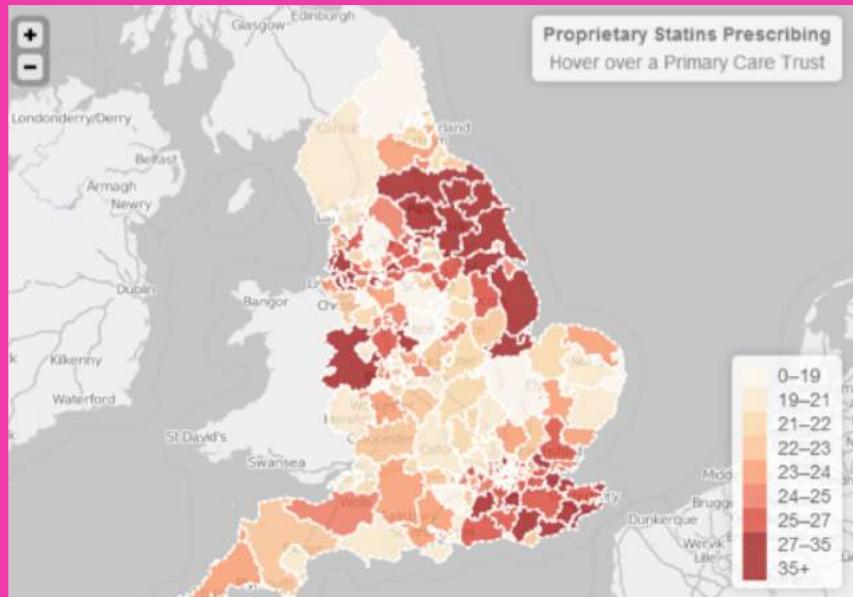
Public purchasing analysis

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

Spend Network | ::::



Prescribing analytics



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

theodi.org/stories



theODI.org



Section header: **subsection header**

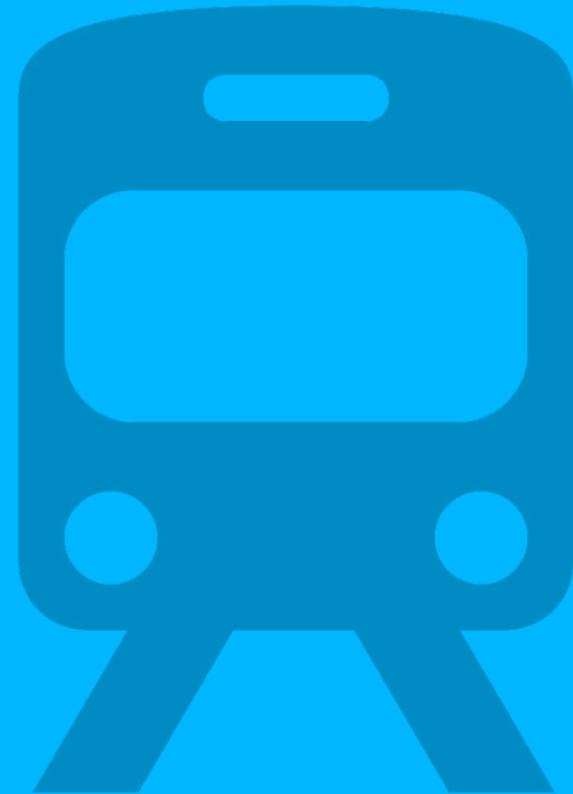
Over 700 Apps now rely on open transport data:

City Mapper (Millions of users)

Google Maps (Millions of users)

Real time trains (500,000 users)

Station master (10,000's users)



The release of open data by TfL is generating annual economic benefits and savings of up to £130m for travellers, London and TfL itself

42% of Londoners use an app powered by TfL data and 83% use its website with similar data: this benefits all transport users in the Capital, TfL itself, and supports London's economic agenda

TfL Passengers and Other Road Users	London	Transport for London
 Saved time for network passengers <ul style="list-style-type: none"> Passengers are able to <i>plan their journeys better with apps</i> that use TfL's open data to provide them real-time information and advice on how to adjust their routes. This provides <i>greater certainty</i> on when the next bus/tube will arrive and <i>saves time</i> - estimated at between <i>£70m and £90m pa</i>. 	 Gross Value Added <ul style="list-style-type: none"> A number of companies use and re-use <i>TfL data commercially, generating revenue</i>, many of whom are based in London. We estimate that the total <i>Gross Value Add</i> from using TfL data by these companies directly and across the supply chain and wider economy is between <i>£12m and £15m GVA pa</i>. 	 Savings from not having to produce apps in-house <ul style="list-style-type: none"> With over <i>13,000 registered developers</i> 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.
 Saved time for other road users <ul style="list-style-type: none"> 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. 	 High value job creation <ul style="list-style-type: none"> TfL open data is estimated to <i>directly support around 500 jobs</i> that would not have existed otherwise. Many of these jobs are in sectors associated with <i>high productivity</i>. 	 Savings from not having to invest in campaigns and systems <ul style="list-style-type: none"> The publication of open data gives passengers information directly, <i>reducing the pressure on the Contact Centre</i>. 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, <i>expanding customer reach and improving transparency</i>. The cost for TfL of publishing open data is estimated at around £1m annually, suggesting a <i>significant return on investment</i>.
 Savings made from moving from SMS alerts <ul style="list-style-type: none"> 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 <i>£2m pa</i>. The <i>use value of new real time alert services</i> is estimated to be up to <i>£3m pa</i>. 	 Wider job creation in the supply chain <ul style="list-style-type: none"> A further <i>230 indirect jobs in the supply chain and wider economy</i> have also been created. 	 Leveraging value and savings from partnerships <ul style="list-style-type: none"> 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 <i>new analyses and improve its operations</i>.
 Better information to plan journeys, travel more easily and take more journeys <ul style="list-style-type: none"> 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 <i>value of these journeys</i> is estimated at up to <i>£20m pa</i>. 		
Plus improved customer satisfaction from having accurate and reliable information available instantly	Plus supporting the wider UK Digital Economy in London and other cities	Plus new commercial opportunities arising from open data

TfL's open data can also contribute to improving societal outcomes, encourage innovation and the wider environment.

The use of open data can also change behaviours and position London and the UK to take advantage of new commercial opportunities

Society	Growth and productivity	Environment
<p> People are more likely to walk or cycle and lead healthier lifestyles</p> <ul style="list-style-type: none">Open data can help <i>integrate the first- and last-mile</i> encouraging alternative transport modes including cycling and walking, which has <i>health benefits</i>.This can support ambitions around <i>healthier streets</i>: since 2007, the number of people walking as their main mode of transport has increased by 13%.	<p> Taking advantage of new opportunities</p> <ul style="list-style-type: none">The UK and London have already earned a <i>reputation</i> 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 <i>important foundational block for further development of new transport products</i>.	<p> Contributing to improving air quality and reducing emissions</p> <ul style="list-style-type: none">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.
	<p> Encourages and facilitates increased innovation</p> <ul style="list-style-type: none">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).	<p>http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf</p>



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.



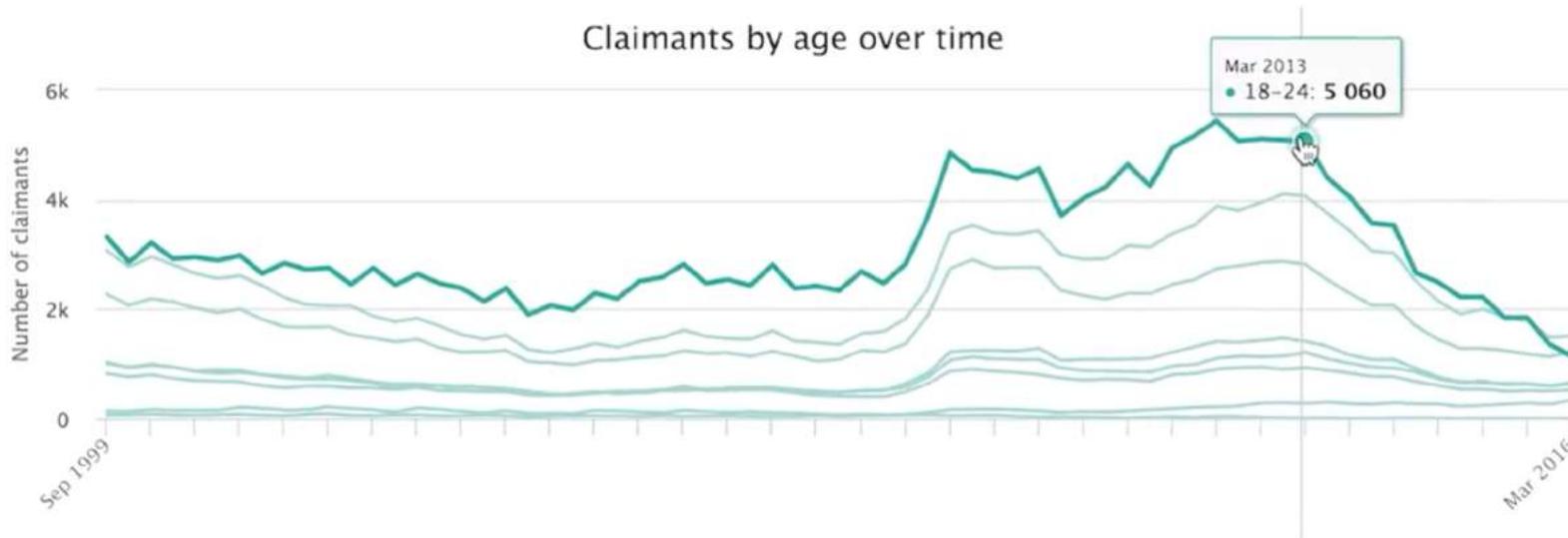
9:17 AM - 1 Jun 2018

6,002 Retweets 30,530 Likes

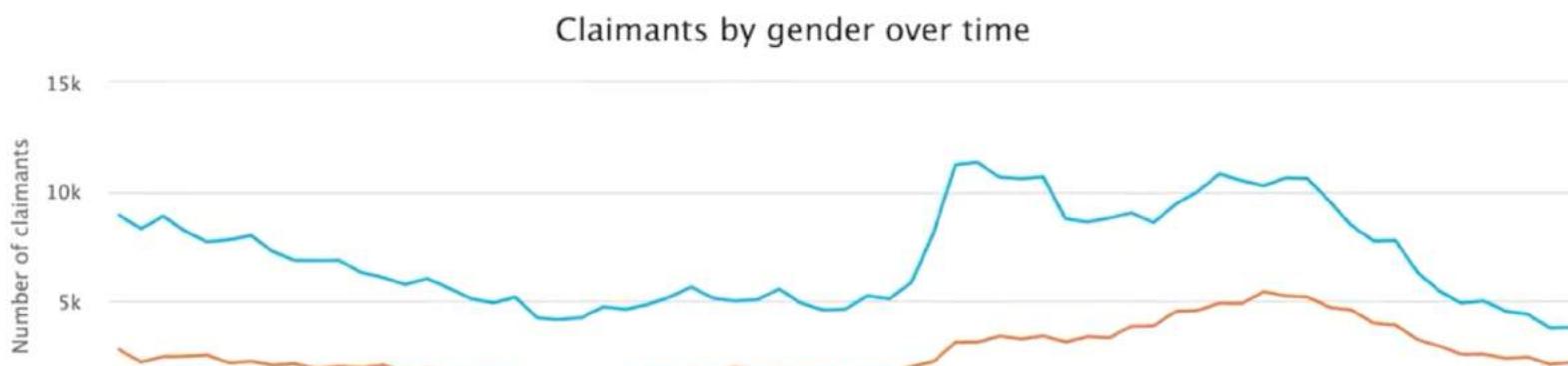


People on **Job Seekers Allowance** in County Durham[Back](#)

Co

**15,790** total claimants

60-64	1.8%
55-59	5.9%
50-54	7.7%
45-49	8.9%
35-44	17.9%
25-34	25.8%
18-24	32.0%
0-17	0.1%



Public service delivery patterns

Efficiency

Transparency

Collaboration



**Using open
data to deliver
public services**

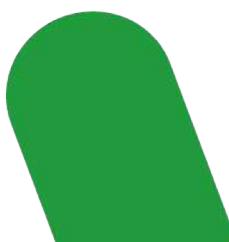
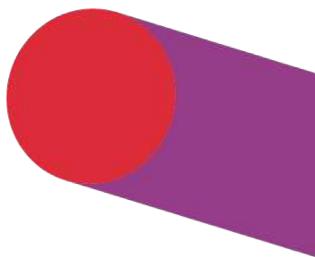
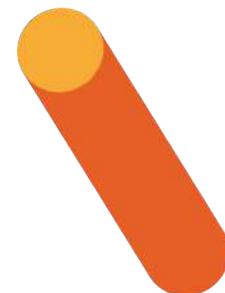
Open Data Institute



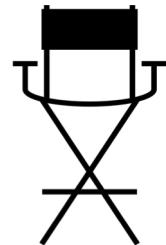
theODI.org



What other opportunities could being more open bring?



Benefits of open data



Leadership

Open data supports the wider data strategy of the organisation



Operations

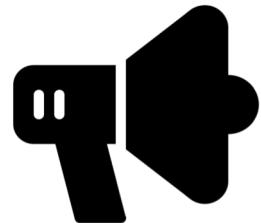
Open data creates significant data access efficiency savings



Finance/Ventures

Open data produces viable equity ventures & business lines

Benefits of open data



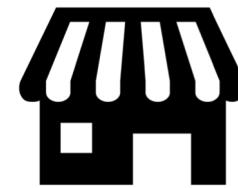
Communications

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 choice. statins and better delivery routes



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

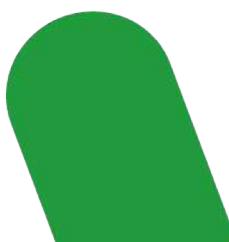
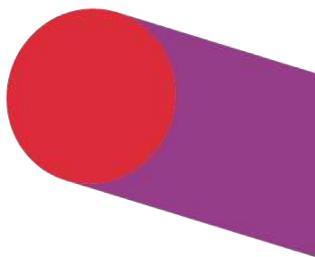
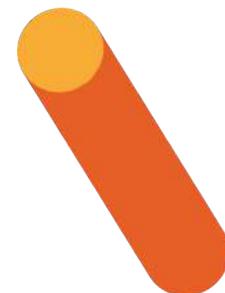


Emissions double reported figure. Evidence in parliament & planning debate



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

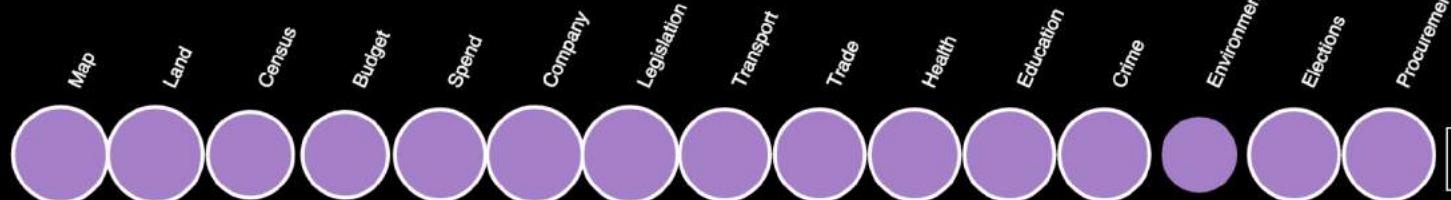
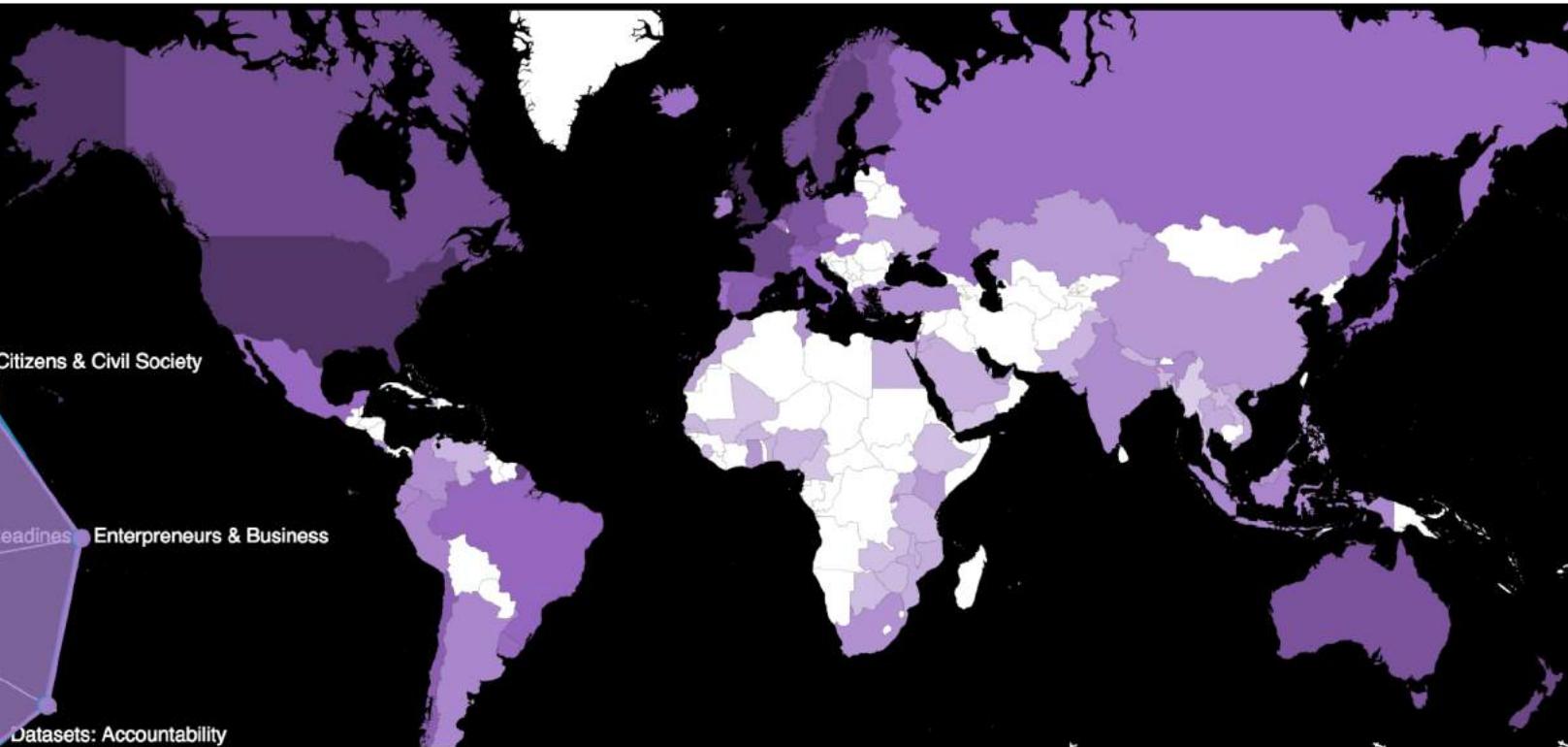
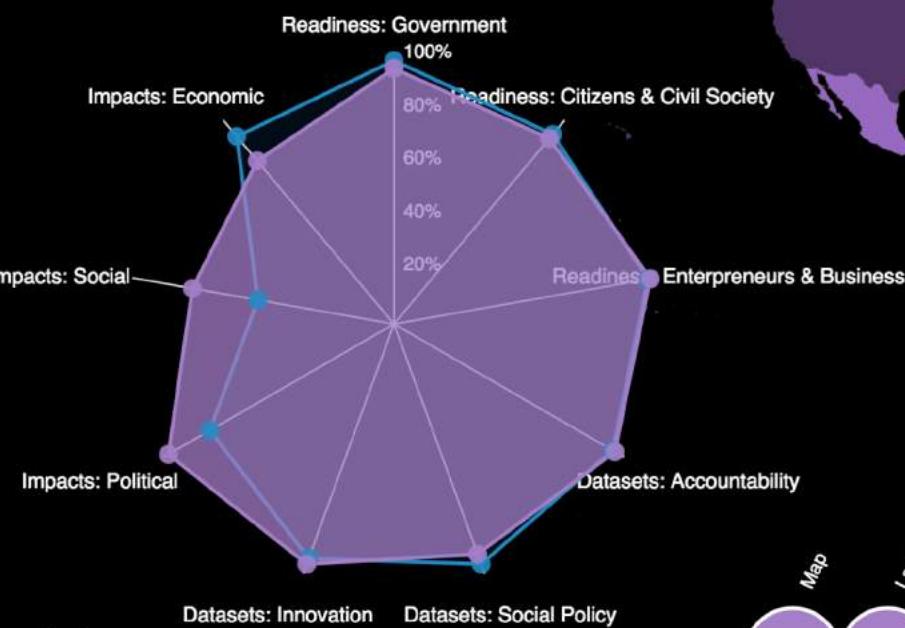
What impact is open data having on citizens' lives?



United Kingdom

1 /86 Score: 100.00

Filters: Off



 Open Data 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
- 16:00** **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.



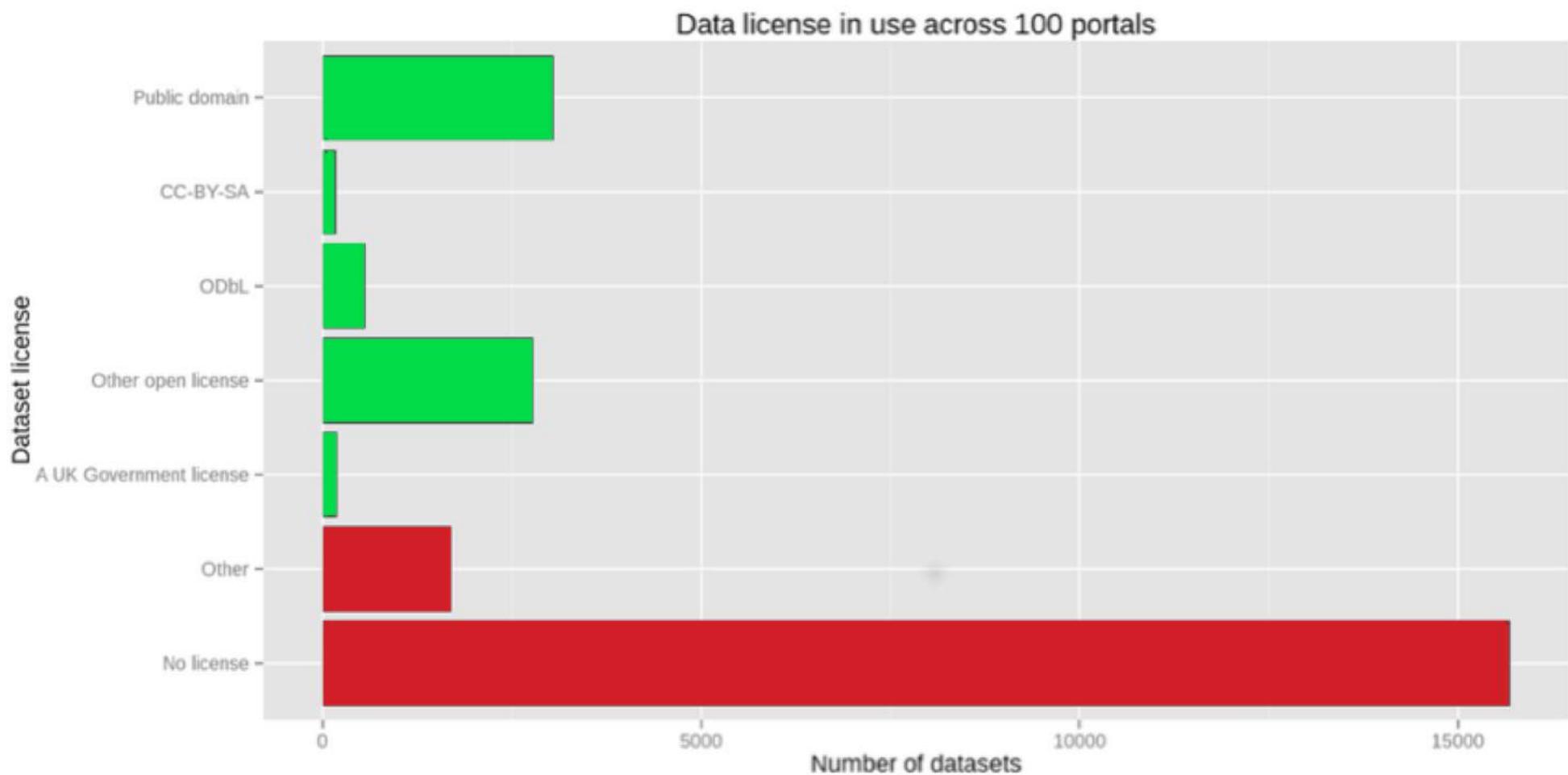
[awesomecontent / Freepik \[CC-BY\]](#)



theODI.org



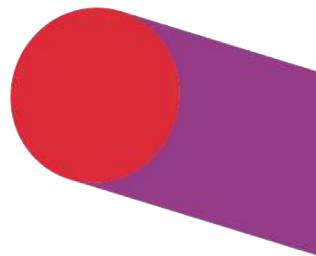
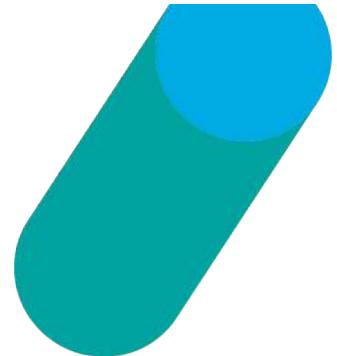
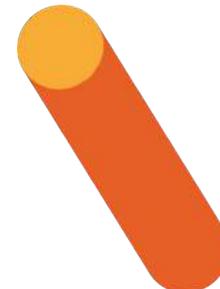
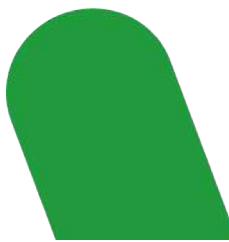
Open data is hardly ever appropriately licensed.



Source: Thomas Levine

<http://thomaslevine.com/l/open-data-licensing/>

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



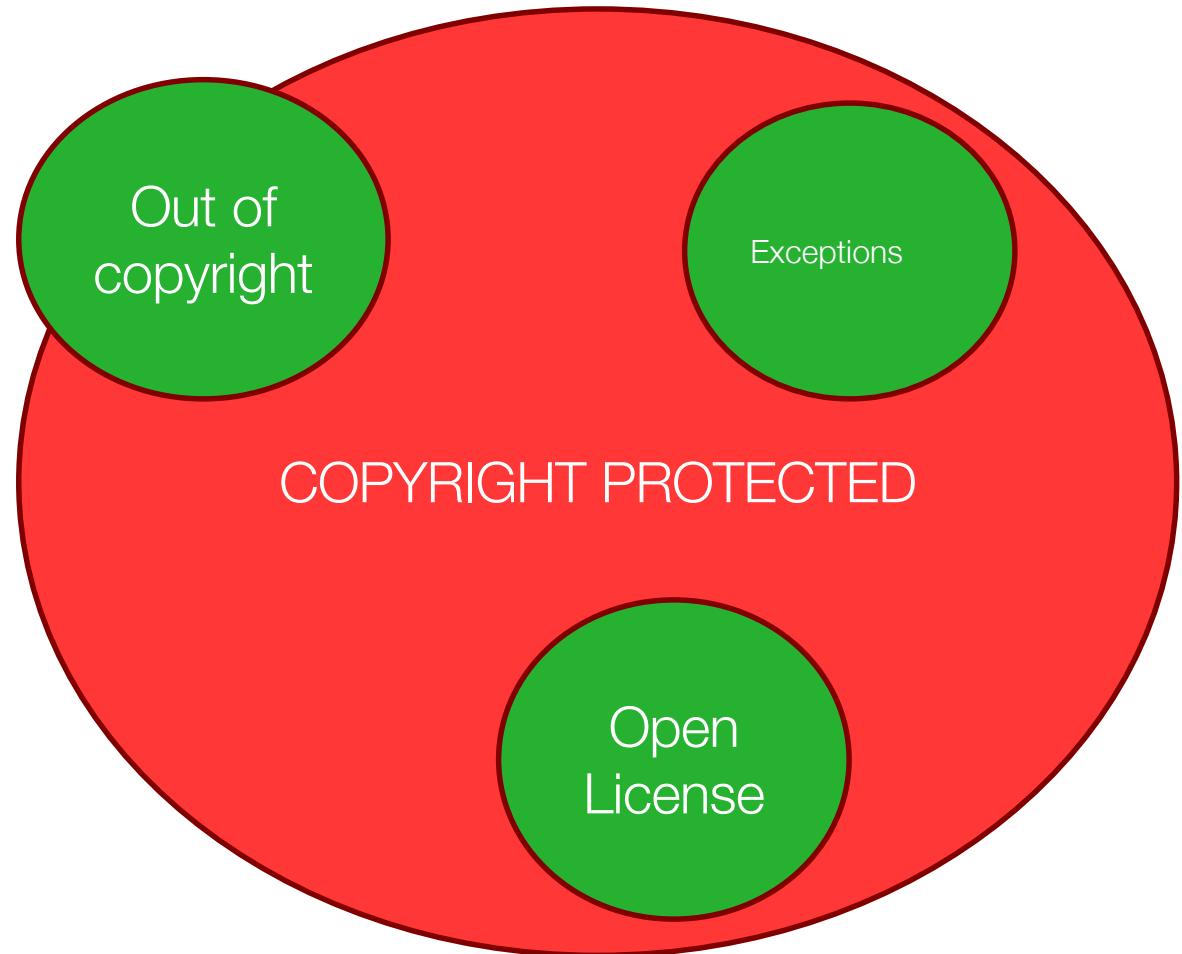
theODI.org



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”



The UK Open Government Licence

Version 3.0 released in October 2014

<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.



creative commons
Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

This is a human-readable summary of (and not a substitute for) the [license](#).

[Disclaimer](#)



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.

 **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



theODI.org



	(S) PUBLIC DOMAIN	(0) PUBLIC DOMAIN	CC BY	CC BY SA	CC BY NC	CC BY ND	CC BY NC SA	CC BY NC ND
(S) PUBLIC DOMAIN	✓	✓	✓	✓	✓	✗	✓	✗
(0) 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.

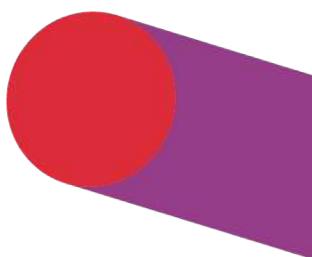
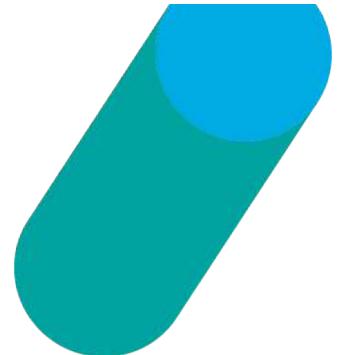
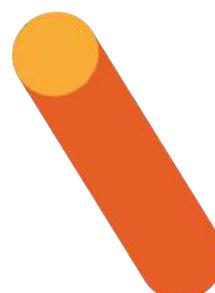
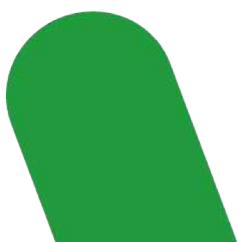
Take a look at each of the following licences and try to determine whether they are open licences

- [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



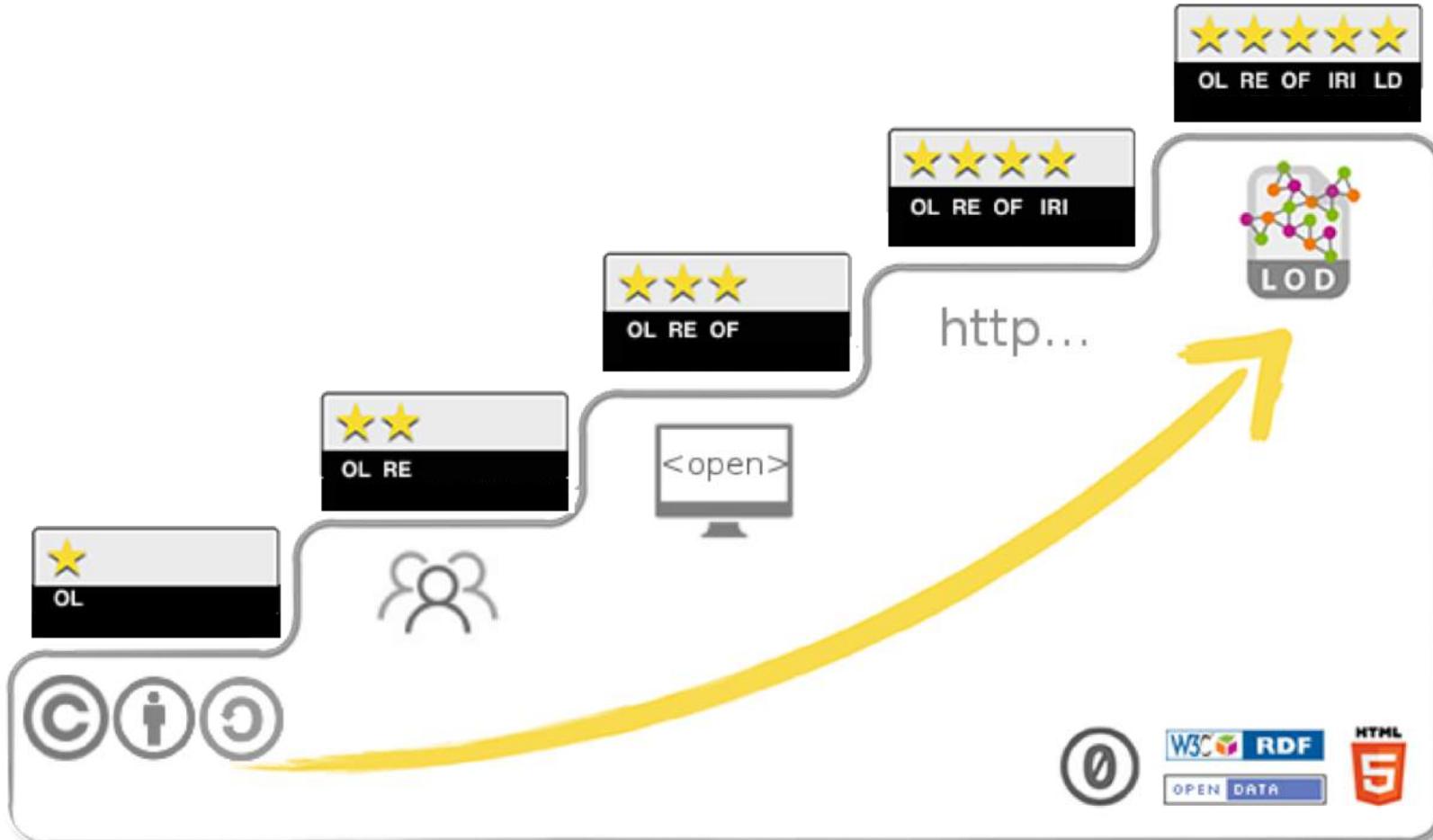
5 - S t a r s



theODI.org



The 5-Stars of Linked Open Data



Open Data Certificate



Bronze level

self certified

GB final

Embed this
on your site

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

Contextual



Legal



Practical



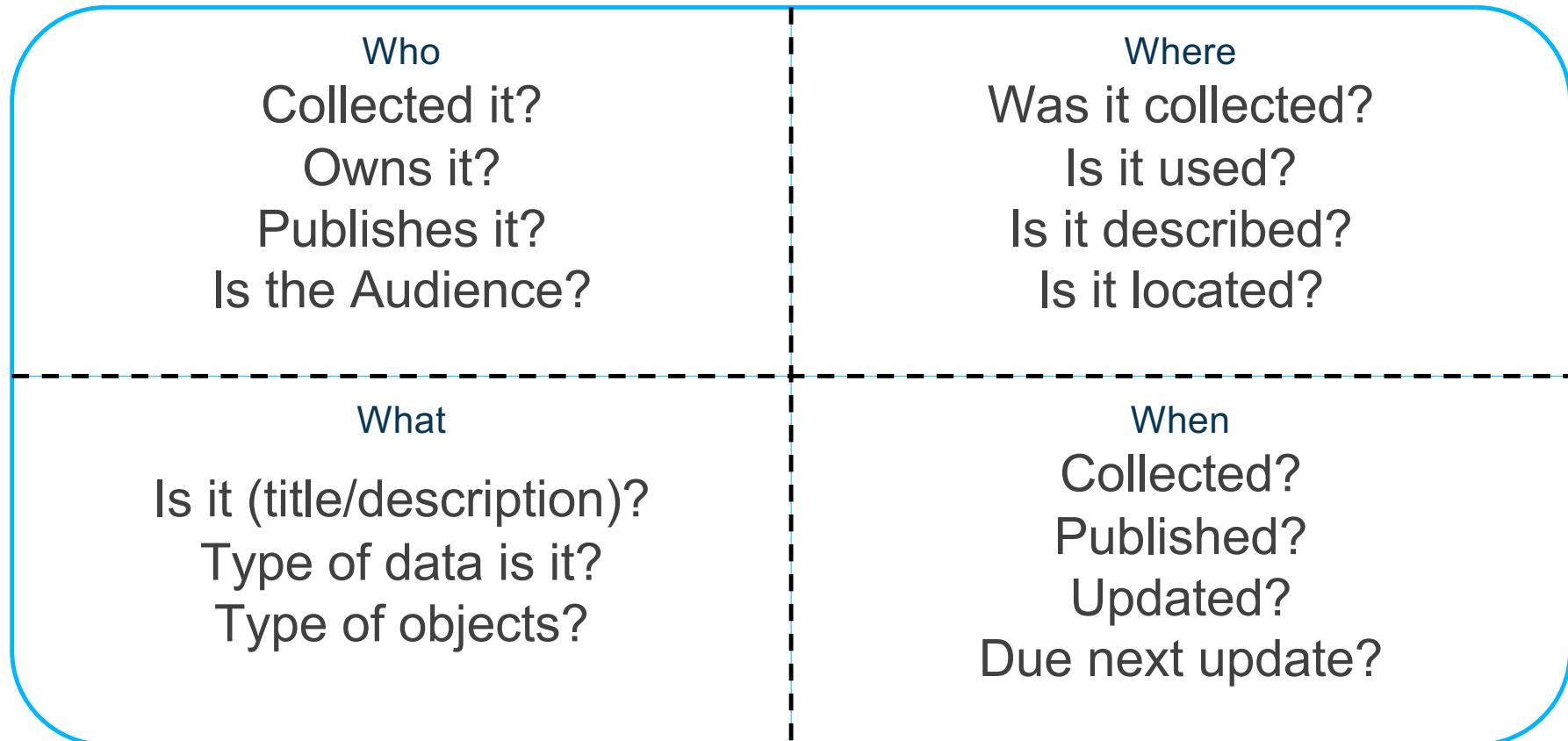
Technical



Social



Adding Context



General



Title/Description/Publisher/URL

-
-
-
-

Release Type

-
-
-
-



theODI.org



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
- 16:00** **END**



Open Refine

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

Mass edit 2300 cells in column Type of Contract Undo

Google refine government IT contracts

Facet / Filter Units / Record

Refresh Reset All Remove All

Type of Contract

891 choices Sort by name asc

Agreement 33

HITECS Task Order 22

CPAF 28

TAM w/ FFP: Time & Materials w/
Firm Fixed Price mix 28

Time and Material 37

Firm Fixed 26

TandM 23

Firm Fixed Price 34

Labor Hours 21

Reel 21

FFP LOG: Firm Fixed Price Level

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

Contract ID	Contractor Name	Type of Contract	Date of Award	Start Date	End Date	Total value of Contract	Contract Awarded
1039	ASAP SOFTWARE EXPRESS INC DELL MARKETING L.P.	Microsoft Enterprise Agreement	04/01/2009	04/01/2009	05/03/2011	1,352	year
1040	MICROSOFT DISTRIBUTION INCORPORATED	Remote Service Desk Maintenance	04/01/2009	04/01/2009	03/01/2010	0.301	year
1041	SYNCHRONIZATION INCORPORATED	Class Branches	05/01/2009	05/01/2009	04/03/2011	0.307	year
1042	ITC CORPORATION	Time & Materials	12/01/2008	01/01/2009	12/03/2011	20	year
1043	SERENET INTERNATIONAL CORPORATION		05/04/2009	05/05/2009	07/03/2009	0.040107	year
1045			01/05/2009	01/06/2009	01/06/2009	0.708	year
1046	IT FEDERAL SA LTD LIABILITY COMPANY		10/01/2009	10/02/2009	08/05/2010	0.049	year
1047		Firm Fixed Price	08/02/2009	08/01/2009	08/02/2010	0.004	year
1048		Firm Fixed Price	11/05/2009	11/05/2009	05/03/2010	0.062	year
1049	PREDICTION IT SOLUTIONS LLC	Firm Fixed Price	01/02/2009	01/01/2009	12/01/2010	0.012	year

0:00 / 6:48

<http://openrefine.org>

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

Open data essentials

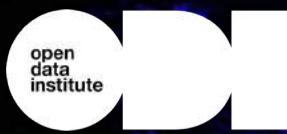
- ✓ Open data definitions
- ✓ Licensing
- ✓ Impact of open data
- ✓ Discovering
- ✓ Best practices
- ✓ Quality checking





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

