



# ODI Learning in Tanzania

Dr David Tarrant | @davetaz | The Open Data Institute

<http://tanzania.learndata.info>



# Course aim

Build a solid foundation and experience in  
publishing, consuming and building a  
business in Open Data



# Open data essentials

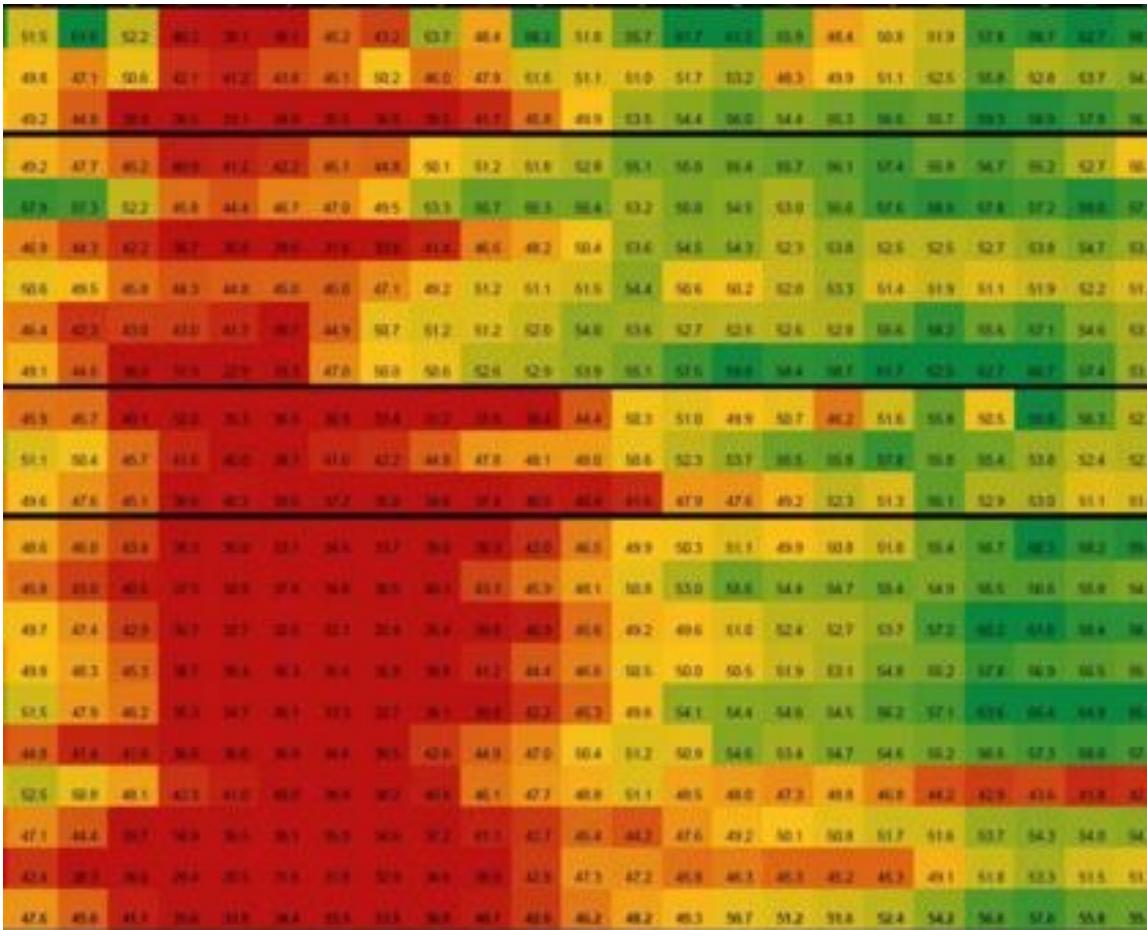
- ✓ Open data definitions
- ✓ Impact of open data
- ✓ Global impact
- ✓ Licensing
- ✓ Best practices
- ✓ Quality checking





## Practical publication

- ✓ Data types
- ✓ Standards
- ✓ Platforms
- ✓ Publishing hands-on



## Visulisation and communication

- ✓ Improving data quality
- ✓ Enriching data
- ✓ Filtering data
- ✓ Communication strategies
- ✓ Creating infographics

# Today

Discovering data

Best practice guidelines

Characteristics of data

Cleaning



# Discovering Open Data

The best way to apply your new experience of the open data field is to discover some open data on your own.

To start this process, there is an interactive exercise below.  
Start by access in the [Foreign and Commonwealth Office's Spend over £25000 dataset.](#)

Once you have dataset open, [explore the different aspects of the data with this exercise.](#)



# Discovering Open Data

- Correctly described?
- Access, use and share?
- Supported?
- Timely?
- Consistent?
- Clean?



# Exercise

What makes data usable?

(1 answer per post-it)

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



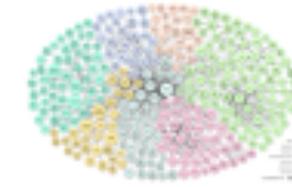
5 - S t a r s



# 5-Stars



<http://5stardata.info/>



<http://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

### 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

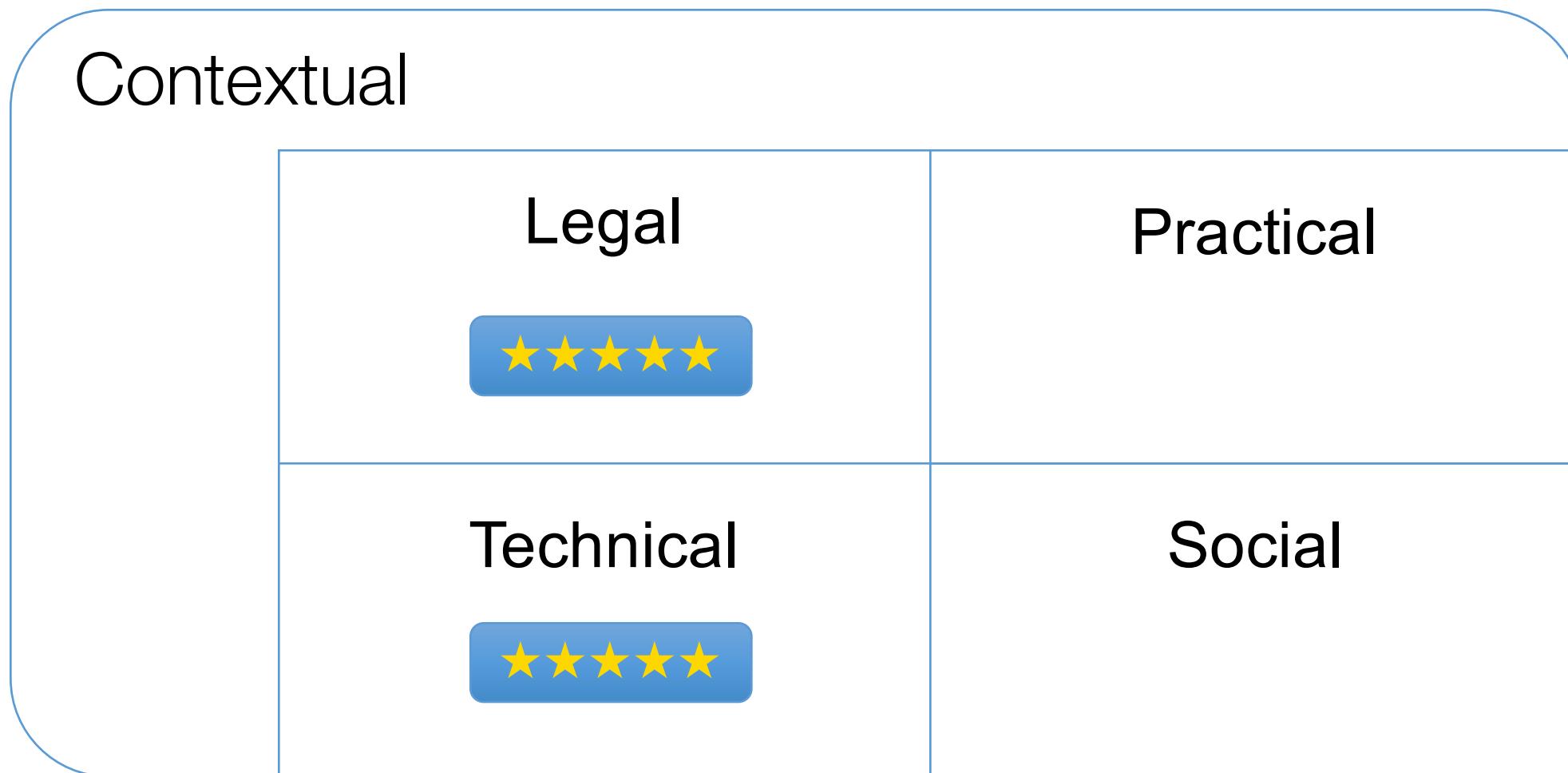


<http://certificates.theodi.org>



Content created by  
The Open Data Institute

# 5 Aspects



# 5 Aspects

Contextual



Legal



Practical



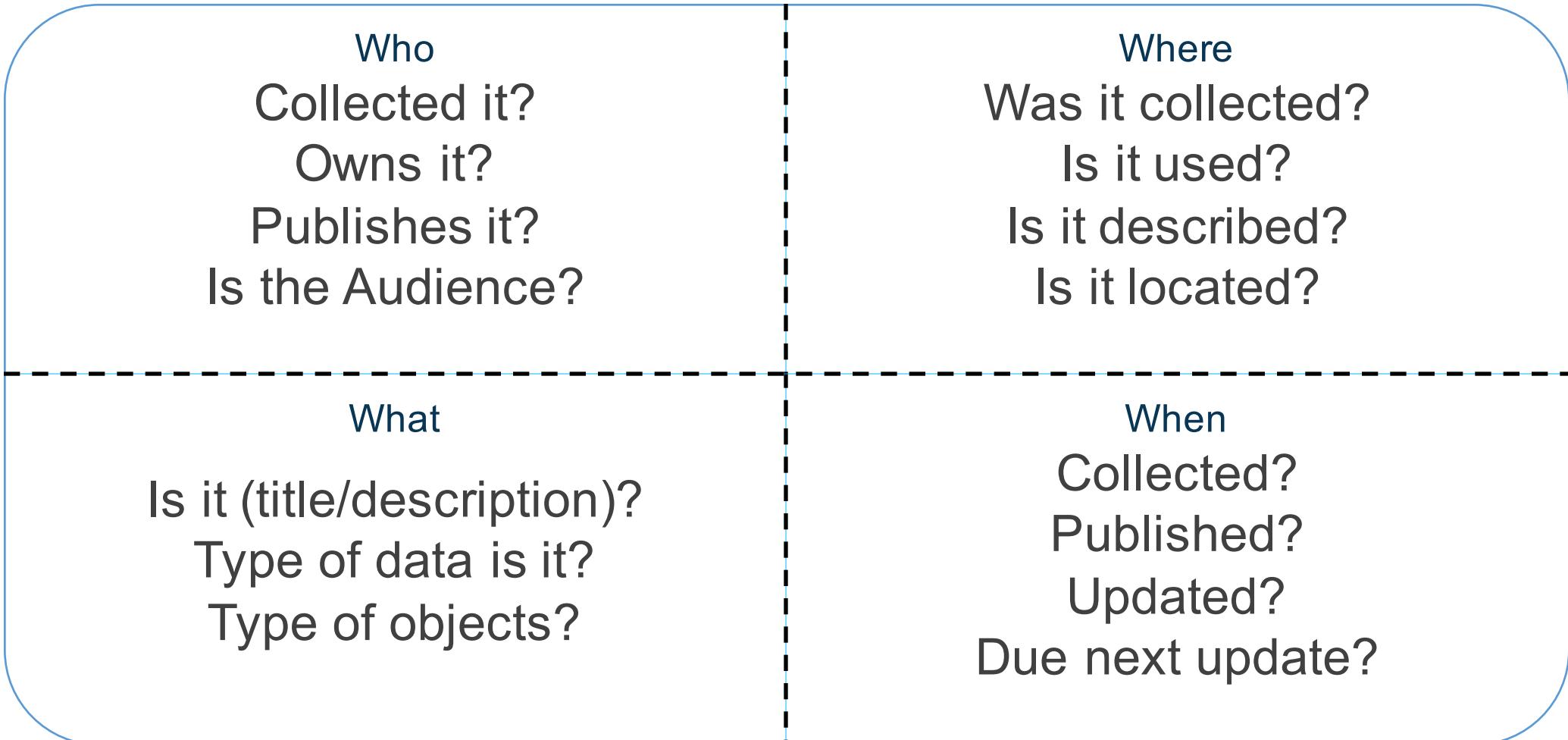
Technical



Social



# 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



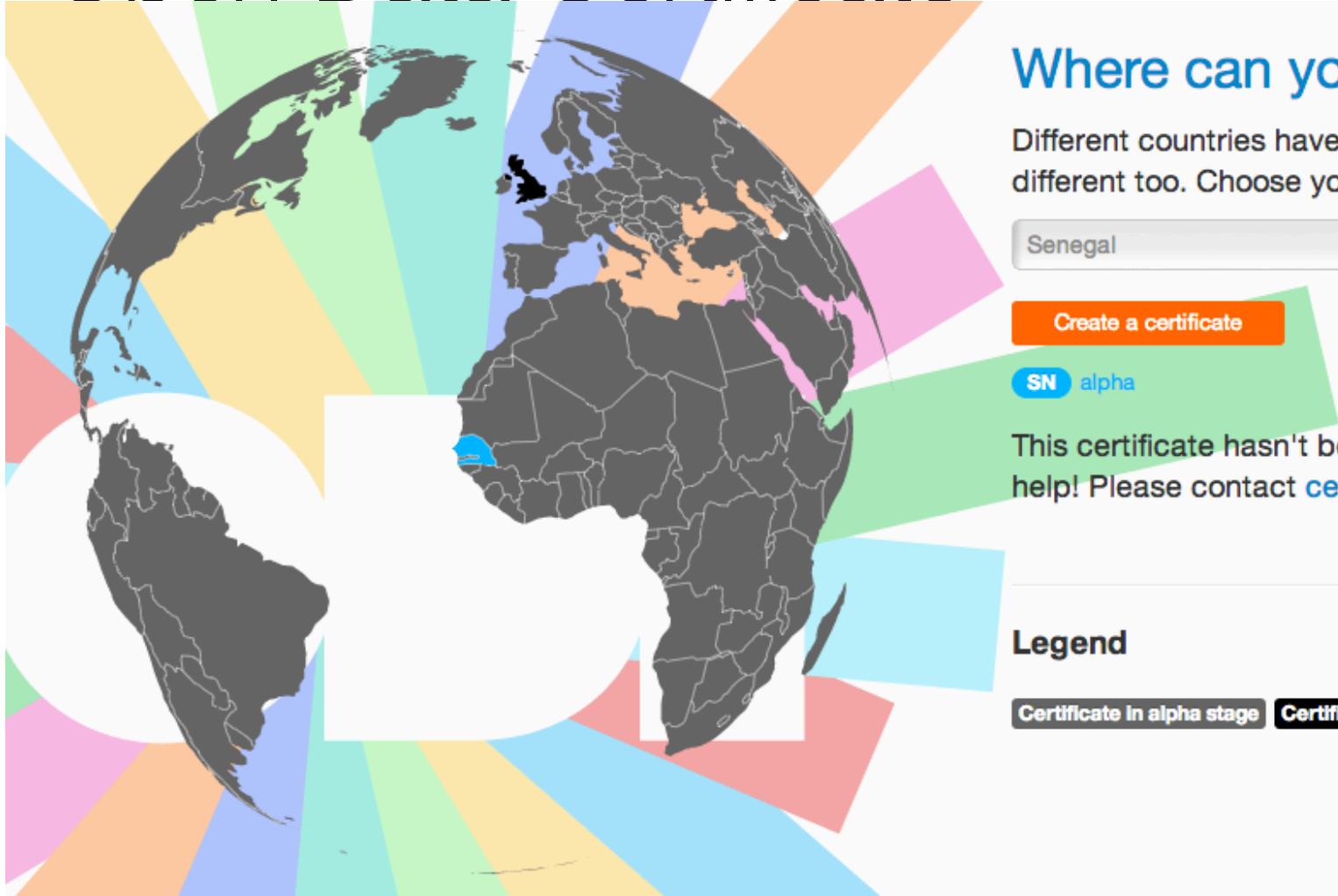
Supported community



Tools and guides available to work with data



# Open Data Certificate



## Where can you create certificates?

Different countries have different laws, so their certificates are slightly different too. Choose your country to create the right certificate.

Senegal

Create a certificate

SN alpha

This certificate hasn't been properly [localised](#) yet, and we need your help! Please contact [certificate@theodi.org](mailto:certificate@theodi.org)

### Legend

[Certificate in alpha stage](#) [Certificate in beta stage](#) [Certificate in final stage](#)



Content created by  
The Open Data Institute

# One Objective

Minimum Requirement



3-Star Data

Silver Open Data Certificate



# Today

Discovering data

Best practice guidelines

Characteristics of data

Cleaning



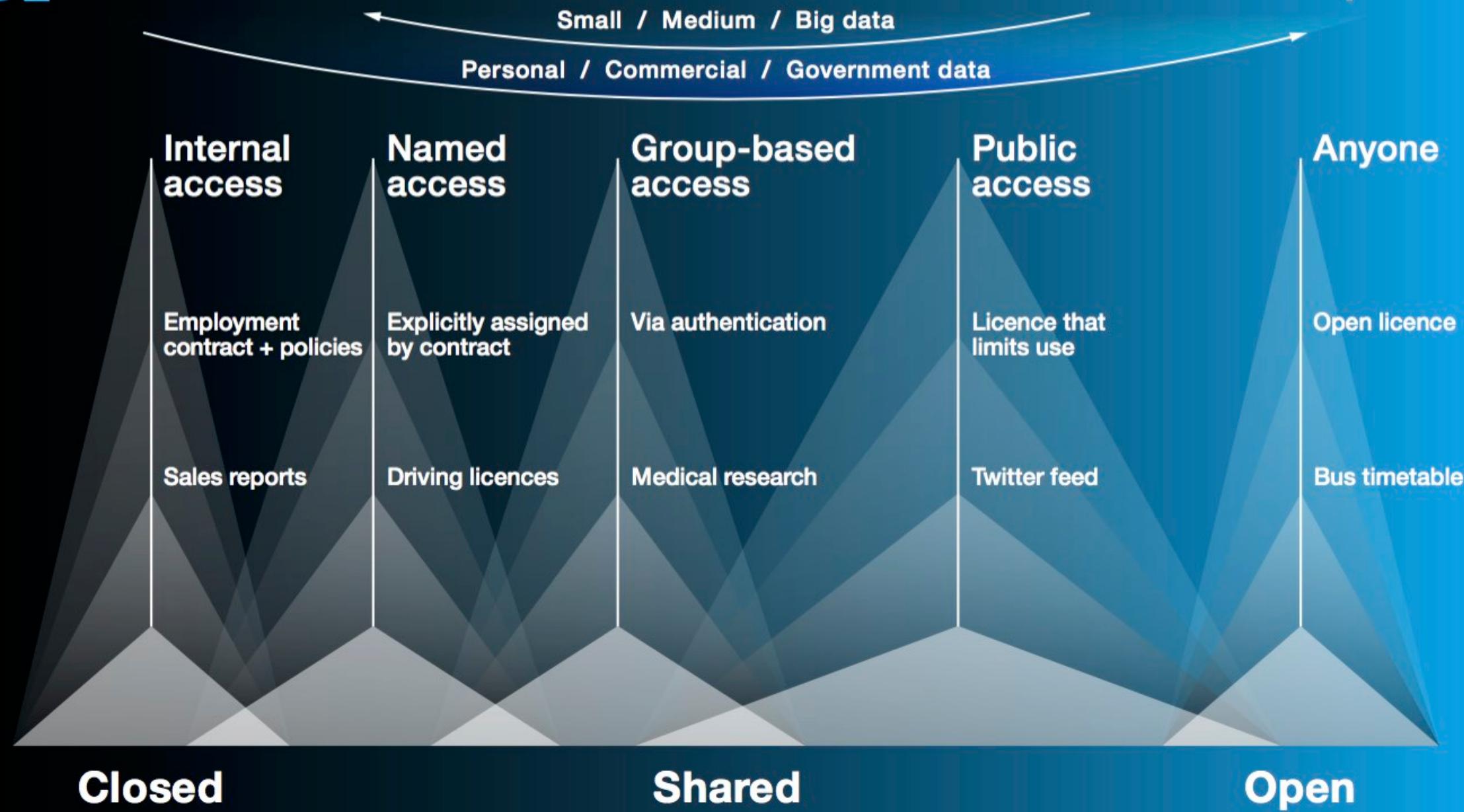
# The characteristics of open data



# Outcomes

- Identify a number of different characteristics of data
- Explain the justifications for publishing different types of data
- Evaluate the current open data ecosystem and future opportunities





# Exercise (part 1)

Pick three datasets, open, shared and closed from the spectrum.

Write the dataset title on a post-it

Write open, shared or closed on the post-it



# Types of Data



## Reference data

“things”

People Facilities Places  
Books Buildings

## Transaction data

“stats involving things”

Expenditure   Weather  
Consumption  
Observation



# Exercise

Categorize your data into reference and transactional data.

If they are all in one category you have 2 minutes to add some new datasets to the empty category.

When done, put a “T” or and “R” on each dataset post-it.



# Update frequency

Static

In frequent updates

Frequent updates

Live



# Exercise

Categorize your data into **frequency of updates**

If they are all in one category you have 2 minutes to add some new datasets to the empty category/ies

Put a number on your post-its representing the frequency of updates.

0 = static, 1= In frequent, 2 = Frequent, 3 = Live

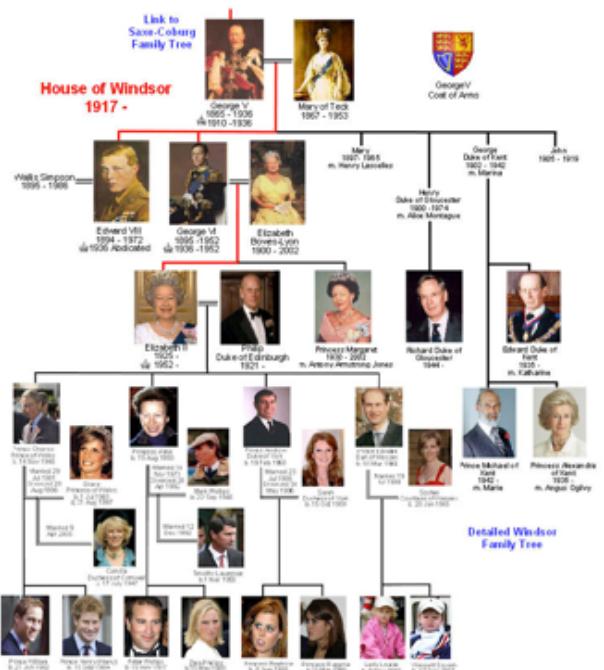


# Data Representations

Tabular

Region	Production			YTD Production (billion MT)
Country	Production (thousand MT)	Change from last year	Change from 5 year average	YTD Production (billion MT)
Brazil	57289	-4.05%	+2.00%	
Mato Grosso	18,008	0.90%	6.17%	
Parana	9,571	-19.55%	-9.08%	
Rio Grande do Sul	7,844	0.88%	9.35%	
Goias	6,820	4.23%	5.27%	
Mato Grosso do Sul	4,218	-7.60%	-1.97%	
Minas Gerais	2,667	5.12%	2.41%	
Bahia	2,512	-8.58%	4.84%	
Sao Paulo	1,392	-3.77%	-6.81%	
Maranhao	1,087	-13.93%	0.58%	
Santa Catarina	1,039	9.81%	13.35%	
Tocantins	902	-0.90%	7.05%	
Piaui	856	4.49%	23.75%	
Para	194	-3.13%	1.02%	
Distrito Federal	155	1.37%	-1.11%	
Roraima	22	-54.10%	-41.70%	

Hierarchical



Network/Graph



# Exercise

Categorize your data into **tabular, hierarchical (tree) and graph (network)**

If they are all in one category you have 2 minutes to add some new datasets to the empty category.

Add the word “**tab**”, “**tree**” or “**net**” to your post-its to represent the different structures.



# Justifications

Trust and  
Transparency

Enabling the  
economy



# One more

Categorize your data into **transparent** and **enabling**.



# Summing up

Do you have any obvious grouping of your datasets?

Is this reflective of the whole open data ecosystem?



# Outcomes

- Identify a number of different characteristics of data
- Explain the justifications for publishing different types of data
- Evaluate the current open data ecosystem and future opportunities



# Open data discovery patterns



# Outcomes

Identify a number of different sources of open data on the web.

Create search patterns that enable easy discovery of new sources of open data.

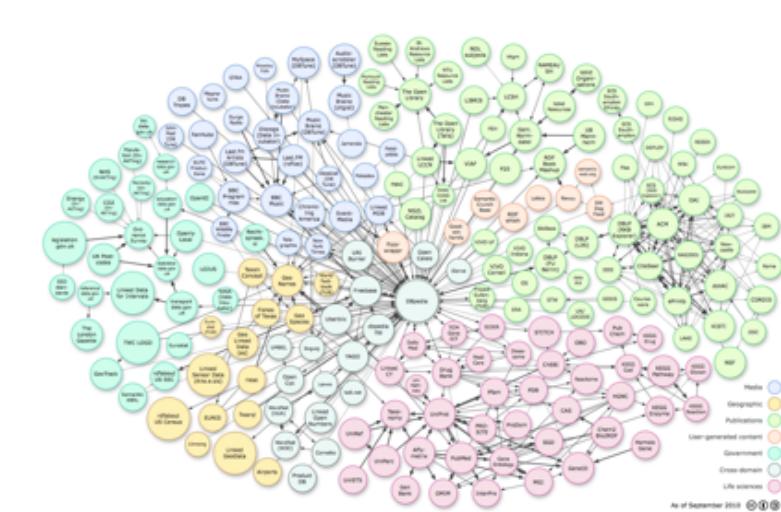


# Approaches to publishing data

ON the web



IN the web



# Finding data on the web (**of documents**)

Government data

Private sector data

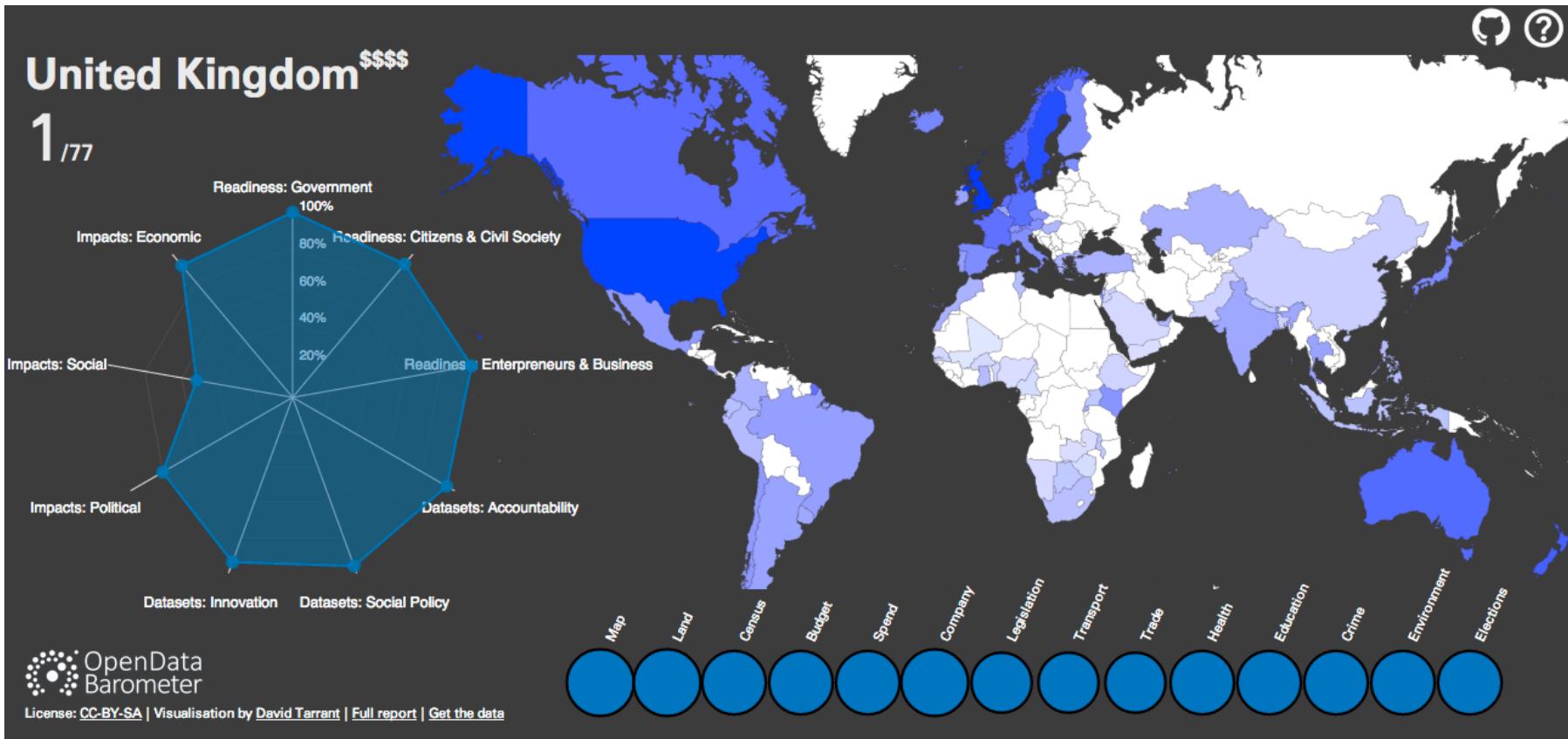
Google advanced

Aggregators and portals

Scraping



# Government data



<http://www.opendatabarometer.org/>



# data.gov.XX

The screenshot shows the DATA.GOV.UK homepage with a search bar and navigation menu. A large search result summary for "19266 Results" is displayed, with one item highlighted: "Live traffic information from the Highways Agency". Below this, there's a "SHOW ONLY..." section for published datasets and a "THEMES" section with various icons and categories like Agriculture, Education, Diplomatie, Infrastructures, Eau et Environnement, TIC, Santé, Collectivités territoriales, Sécurité Publique, and Tourisme Culture.

DATA.GOV.UK  
Opening up Government

Datasets Map Search Data Requests Publishers Public Roles & Salaries Spend Reports Site Analytics Reports

19266 Results

Live traffic information from the Highways Agency

Highways Agency  
Live traffic information data showing traffic information on the strategic road network.

Search for data...  or conduct map based search

SHOW ONLY...

Published datasets (15186)

OpenData Burkina Faso

Accueil Comprendre l'Open Data Thèmes Producteurs Jeux de données Applications Partenaires A propos

THEMES

Tous les thèmes

Agriculture Éducation Diplomatie Infrastructures Eau et Environnement

TIC Santé Collectivités territoriales Sécurité Publique Tourisme Culture

Open Data Burkina Faso statistiques

- 78 jeux de données
- 28 organisations
- 11 groupes

CC BY

The screenshot shows the DATOS.GOB.MX BETA website. It features a map-based filter for datasets across the United States and Mexico. A search bar at the top right shows "Buscar conjuntos de datos...". Below it, a summary states "127 datasets found". A section titled "RATING SOCIAL PROGRAMS" describes a database of social program evaluations. The page includes links for "Datasets" and "Organizations".

DATOS.GOB.MX BETA  
OPEN TO PARTICIPATE

/ datasets

Filter by your suburb Clear

StreetMap

Buscar conjuntos de datos...

127 datasets found

Sort by: Relevance

RATING SOCIAL PROGRAMS

Database containing the evaluation of the results of social programs from the federal government subject to the annual assessment.

Portal Rasmi Open Data Kerajaan Malaysia  
The Government of Malaysia's Open Data Official Portal

data.gov.my

HOME ABOUT US CATALOGUE INFOGRAPHICS MOBILE APPS CONTACT

DATASETS INFO

- 117 Datasets
- 11 Ministries
- 10 Sectors

RESOURCES

- Malaysia Directory
- Malaysian Open Source Centre
- Open Government Data

MOBILE APPS

- myHealth | myJakim | KPDKKK SPAD | SELAWATI | myMAHTAS
- 6 Android Mobile Apps 4 iOS Mobile Apps

QUICK LINKS

- Malaysia Informative Data Centre
- Malaysian Population Quick Info
- Tourism Malaysia - Facts & Figures

Latest Datasets Top Publishers Feedback

1 Aplikasi Pemetaan Belia Malaysia  
Info belia berdasarkan domain  
by : Ministry of Youth and Sports

Last Updated 2014-06-26 11:30:19

2 Aplikasi Pemetaan Belia Malaysia  
Statistik Sikap Belia Malaysia mengikut  
by : Ministry of Youth and Sports

Last Updated 2014-06-26 11:27:30

3 Aplikasi Pemetaan Belia Malaysia  
Statistik Pemetaan Belia Malaysia dalam bentuk aplikasi  
by : Ministry of Youth and Sports

Last Updated 2014-06-26 11:24:52

Latest News and Events

Currently, there are no latest event.

The Open Data Institute

# Government / Private



## Flight MH370: Malaysia releases raw satellite data



The BBC's Richard Westcott visited Inmarsat's headquarters to find out what the data tells us about MH370's fate

**The Malaysian government has released the raw data used to determine that the missing Malaysia Airlines flight MH370 crashed into the southern Indian Ocean.**

The data was first released to relatives of passengers, who have been asking for greater transparency, before copies were also provided to media.

The document released on Tuesday comprises 47 pages of data, plus notes, from British firm Inmarsat.

### MH370 mystery

[Deep sea challenge](#)

[Ocean maps problem](#)

[Costs of the search](#)

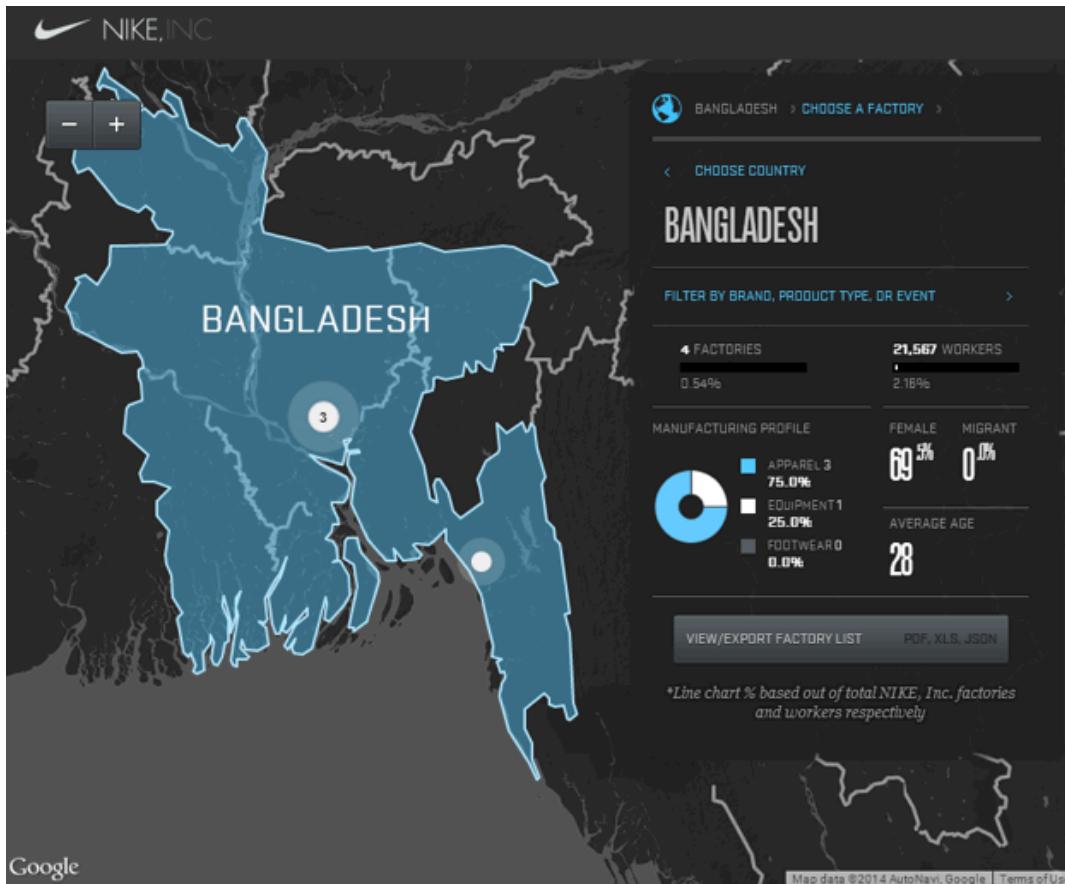
[What we know](#)

<u>Pre Take-Off</u>								
Time	Channel Name	Ocean Region	GES ID (octal)	Channel Unit ID	Channel Type	SU Type	Burst Frequency Offset (Hz) BFO	Burst Timing Offset (microseconds) BTO
7/03/2014 16:00:13.406	IOR-R1200-0-3603	IOR	305	8	R-Channel RX	0x15 - Log-on/Log-off Acknowledge		
7/03/2014 16:00:13.906	IOR-P10500-0-3859	IOR	305	10	P-Channel TX	0x15 - Log-on/Log-off Acknowledge	103	14820
7/03/2014 16:00:17.430	IOR-R1200-0-3603	IOR	305	8	R-Channel RX	Eleven Octet User Data		
7/03/2014 16:00:17.906	IOR-R1200-0-3603	IOR	305	8	R-Channel RX	Eleven Octet User Data	103	14740
7/03/2014 16:00:18.406	IOR-R1200-0-3603	IOR	305	8	R-Channel RX	Eight Octet User Data	103	14780
7/03/2014 16:00:18.905	IOR-P10500-0-3859	IOR	305	10	P-Channel TX	0x62 - Acknowledge User Data	103	14820
7/03/2014 16:00:20.906	IOR-P10500-0-3859	IOR	305	10	P-Channel TX	0x71 - User Data (ISU) + RLS		
7/03/2014 16:00:20.906	IOR-P10500-0-3859	IOR	305	10	P-Channel TX	Subsequent Signalling Unit		
7/03/2014 16:00:22.906	IOR-R1200-0-3603	IOR	305	10	P-Channel TX	Subsequent Signalling Unit		
7/03/2014 16:00:23.407	IOR-R1200-0-3603	IOR	305	8	P-Channel TX	Subsequent Signalling Unit		
7/03/2014 16:00:23.905	IOR-P10500-0-3859	IOR	305	8	R-Channel RX	0x62 - Acknowledge User Data		
7/03/2014 16:00:27.741	IOR-T1200-0-3607	IOR	305	8	R-Channel RX	Subsequent Signalling Unit		
7/03/2014 16:00:27.901	IOR-T1200-0-3607	IOR	305	8	R-Channel RX	Subsequent Signalling Unit		
7/03/2014 16:00:28.051	IOR-T1200-0-3607	IOR	305	8	R-Channel RX	Subsequent Signalling Unit		
7/03/2014 16:00:28.221	IOR-T1200-0-3607	IOR	305	8	R-Channel RX	Subsequent Signalling Unit		
7/03/2014 16:00:28.405	IOR-T1200-0-3607	IOR	305	8	R-Channel RX	Subsequent Signalling Unit		
7/03/2014 16:00:28.541	IOR-T1200-0-3607	IOR	305	8	R-Channel RX	Subsequent Signalling Unit		

**OPEN DATA?**



# Suppliers



You agree not to change or delete any ownership notices from materials downloaded or printed from the Platform. You agree not to modify, copy, translate, broadcast, perform, display, distribute, frame, reproduce, republish, download, display, post, transmit or sell any Intellectual Property or Content appearing on the Platform



<http://manufacturingmap.nikeinc.com/#>



# Google advanced

The screenshot shows a Google search results page with the query "site:gov filetype:xls" entered into the search bar. The results are categorized under "Web". The first result is a link to "Code List or Concept (Acronym)" from "www.acquisition.gov/short\_codelistsTS.xls", described as a Microsoft Excel file. The second result is a link to "Approps - Foreign Assistance.gov" from "www.foreignassistance.gov/Full\_ForeignAssistanceData.xls", also described as a Microsoft Excel file. The third result is a link to "TSB Monthly Cash Flow Projection" from "www.dia.iowa.gov/tech/cashflow.xls", described as a Microsoft Excel file.

site:gov filetype:xls

Web Images Maps Shopping More ▾ Search tools

About 4,150,000 results (0.22 seconds)

[XLS] [Code List or Concept \(Acronym\)](#) ↗  
www.acquisition.gov/short\_codelistsTS.xls Share  
File Format: Microsoft Excel - [View as HTML](#)  
A, B, C, D, F, G, H, 1, Code List or Concept (Acronym), Definition, Authoritative or Maintenance Agency, Current Version, Update Frequency and/or ...

[XLS] [Approps - Foreign Assistance.gov](#) ↗  
www.foreignassistance.gov/Full\_ForeignAssistanceData.xls  
File Format: Microsoft Excel  
A, B, C, D, E, F, G, H, 1, Planning Data, 2, 3, 4, Fiscal Year, Fiscal Year Type Account Name, Agency Name, Operating Unit, Category, Sector, Amount ...

[XLS] [TSB Monthly Cash Flow Projection](#) ↗  
www.dia.iowa.gov/tech/cashflow.xls

**site:** Get results only from certain sites or domains

**link:** Find pages that link to a certain page

**related:** Find sites similar to one you already know

**filetype:** Find certain file types only

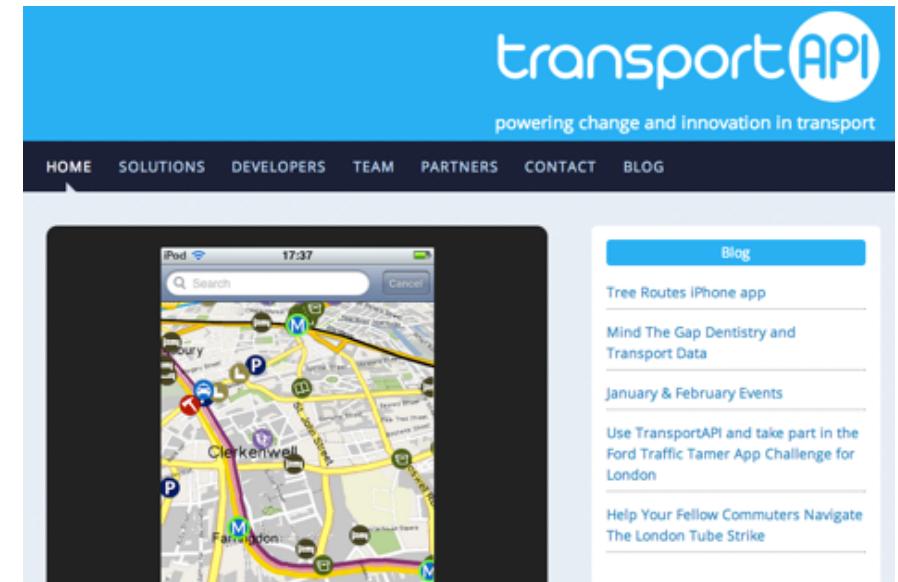


# Aggregators and portals

Collect together data from across the web into one place.



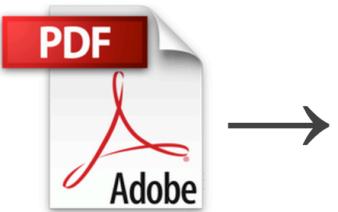
enigma.io



transportAPI

# Scraping

If you can't obtain usable data (csv, xls) then you may have to resort to scraping.



Variety Name	Total Receipts	Total Receipts	Total Inedibles	Receipts
Aldrich	49,043,476	48,900,432	560,403	2.61%
Aviation	8,999,200	8,932,000	159,796	7.79%
Butte	144,865,828	147,182,233	1,186,402	7.70%
Butte/Padre	215,647,498	213,129,006	1,118,151	11.47%
Carmel	171,682,007	170,460,781	1,571,020	9.13%
Canyon	138,180,000	137,500,000	1,680,000	1.21%
Fritz	110,175,472	110,246,890	1,854,841	5.86%
Harvey	36,907	36,907	1,181	0.06%
Hashem	347,618	348,680	3,812	0.02%
LeGrand	9,797	9,557	261	0.26%
Livingston	7,394,450	7,500,200	159,210	0.39%
Marchini	418,442	418,442	9,096	0.02%
Mered	57,800	57,768	846	0.06%
Mission	17,265,270	17,174,884	78,967	0.92%

The image shows the import.io platform's user interface, which includes a search bar at the top labeled "Enter a URL for a list page" and a pink "Extract Data" button. Below the search bar are four examples of scraped data:

- Reseller Ratings:** A screenshot of a web page showing a list of products with their ratings and reviews.
- Zoopla:** A screenshot of a real estate listing page from Zoopla.
- 500px:** A screenshot of a photo sharing website.

Below these are three more examples:

- Growth Hackers:** A screenshot of a blog or community website.
- Udemy:** A screenshot of a course marketplace.
- Stack Exchange:** A screenshot of a Q&A site.

“excellent, so excited beyond description”  
George Ofosu, Doctoral Student, UCLA

pdftables.com

magic.import.io

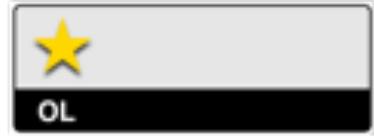


# 5-Stars



<http://5stardata.info/>

## ON THE WEB



## IN THE WEB

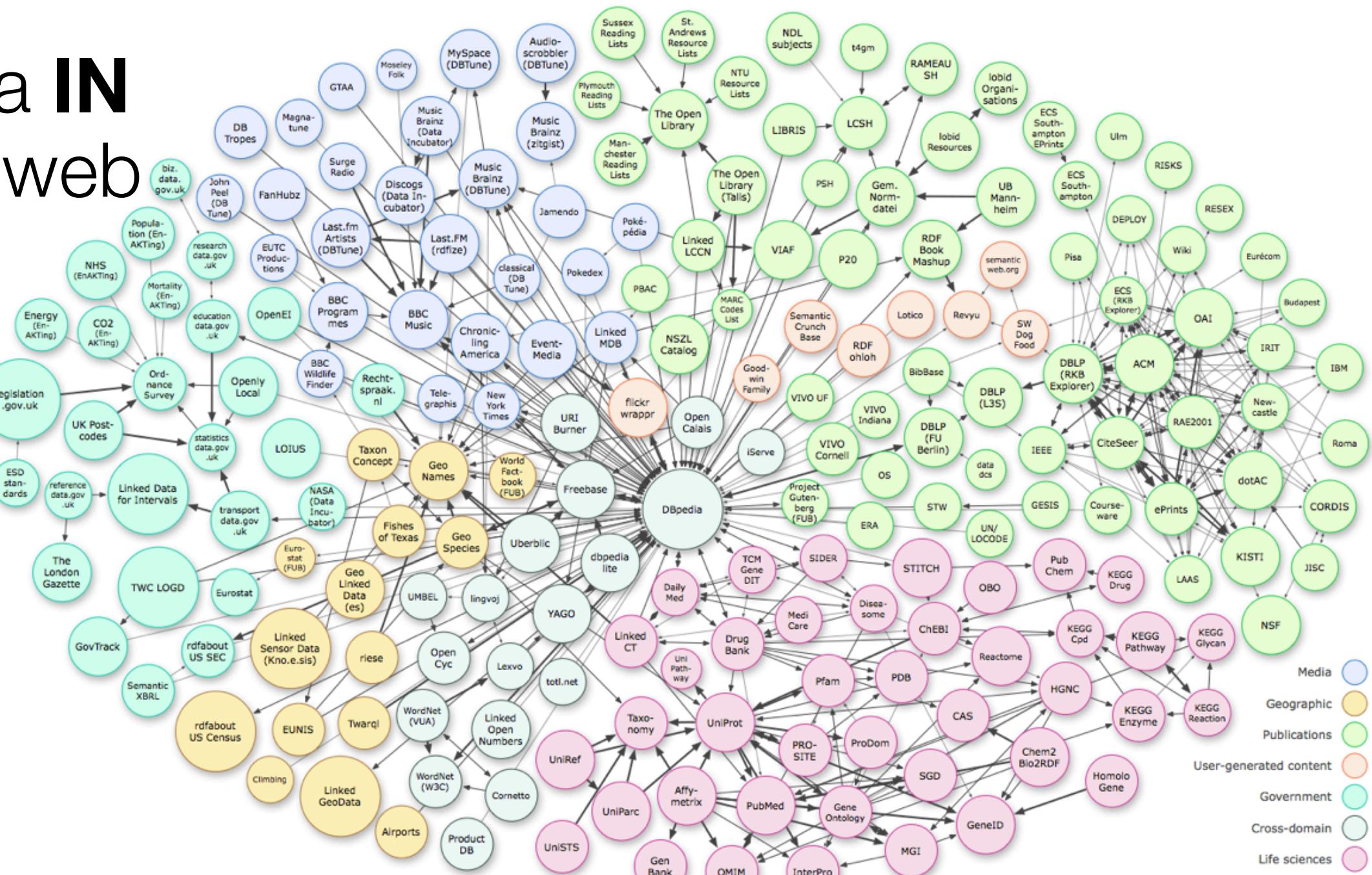


OPEN DATA



Content created by  
The Open Data Institute

# Data IN the web



# Linked data

## Amazing but challenging to publish and use

### EEE Building

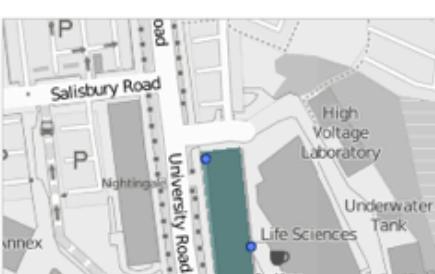
[http://id.southampton.ac.uk/building/32 ← This is the URI](http://id.southampton.ac.uk/building/32)

Detail	Facilities	Services	Energy
<p>Site: Highfield Campus</p> <p>Construction: 2006</p> <p>Architect: John McAslan &amp; Partners</p> <p>Features: Building 32 is non-residential</p> <p><a href="#">View Disability Report for this Building</a></p>			

 ©2010 Francois-Xavier Beckers (CC-BY)

**Occupants**

- Electronics & Computer Science
- Southampton Education School
- Agents, Interactions & Complexity
- Web & Internet Science
- Leadership School Improve &Effectiveness
- Lifelong & Work-Related Learning
- Mathematics & Science Education
- Social Justice & Inclusive Education
- Teaching Only Staff
- Deanery



**Properties and Values**

- soton.ac.uk/building/32 → rooms:Building, http://id.southampton.ac.uk/ns/UoSBuilding → "EEE Building"
- occupant → Electronics & Computer Science, Southampton Education School, Agents, Interactions Complexity, Web & Internet Science, ...show 8 more...
- construction → "32"^^http://id.southampton.ac.uk/ns/building-code-scheme
- locations:within → Highfield Campus
- <http://www.soton.ac.uk/estates/ourestate/buildings/highfield/32.html>
- "50.9364157"^^xsd:float
- "-1.395905"^^xsd:float
- southampton.ac.uk/ns/disabledGoPage → <http://www.disabledgo.com/en/access-guide/building-32>
- locations:easting → "442544"^^xsd:integer
- locations:northing → "115392"^^xsd:integer
- Organization → University of Southampton
- southampton.ac.uk/ns/ombielName → "Bldg 32 (EEE)"
- feature → Building 32 is non-residential
- southampton.ac.uk/ns/buildingDate → "2006"
- southampton.ac.uk/ns/buildingArchitect → John McAslan & Partners
- spatial → "POLYGON((-1.3961073411331264 50.93683868764933,-1.3958347895092957 50.9368567227702,-1.3956958407975968 50.936065737417,-1.3959558923017397 50.93603859197583,-1.3961073411331264 50.93683868764933))"
- southampton.ac.uk/ns/electricityTimeSeries → "elec/b32/ekw"
- ← is spatialrelations:within of ← 32 / 3077, 32 / 1015, Physical and Applied Science Faculty Deanery, Social and Human Sciences Faculty Deanery, ...show 54 more...
- ← is foaf:depicts of ← <http://data.southampton.ac.uk/image-archive/buildings/raw/32.jpg>, <http://data.southampton.ac.uk/image-archive/buildings/1000/32.jpg>, <http://data.southampton.ac.uk/image-archive/buildings/800/32.jpg>, ...show 5 more...
- ← is event:place of ← RaeS Solent Branch Christmas Special Lecture - The Red Arrows

 a Building  
Google Images Ltd & Bluesky, The GeoInformation Group



The Open Data Institute

# Finding data on the web (**of data**)

1. Add random extensions (.xml, .json, .csv etc)
2. Look for alternative links (rss feeds etc)
3. Look for embedded data
4. Do some content negotiation
5. Spot the API
6. Scrape (or search google again)

**IN THE WEB**

How the web should work, but people forgot that Tim put this in when he invented it!



Content created by  
The Open Data Institute

# Duck typed data

If it looks like a duck  
and quacks like a duck,  
then it's probably a duck.

Basically, keep an eye out for tables,  
lists and other stuff that looks like data.



# 1. Adding random extensions

The screenshot shows the GOV.UK Trade Tariff homepage. At the top, there's a navigation bar with the GOV.UK logo and a search bar. Below it, a breadcrumb trail reads "Home > Business and self-employed > Imports and exports". The main title "Trade Tariff" is displayed. A search form allows users to "Search the tariff" by name or code and includes a "Search" button. Below the search form, a note says "This tariff is for 6 August 2014" with a link to "change date". There are also links to "View all sections" and "A-Z Index". A table lists tariff sections from I to IX, each with a title:

Section	Chapters	Title
I	1 to 5	Live animals; animal products
II	6 to 14	Vegetable products
III	15	Animal or vegetable fats and oils and their cleavage products; prepared edible fats; animal or vegetable waxes
IV	16 to 24	Prepared foodstuffs; beverages, spirits and vinegar; tobacco and manufactured tobacco substitutes
V	25 to 27	Mineral products
VI	28 to 38	Products of the chemical or allied industries
VII	39 to 40	Plastics and articles thereof; rubber and articles thereof
VIII	41 to 43	Raw hides and skins, leather, furskins and articles thereof; saddlery and harness; travel goods, handbags and similar containers; articles of animal gut (other than silkworm gut)
IX	44 to 46	Wood and articles of wood; wood charcoal; cork and articles of cork; manufactures of straw, of esparto or of other plaiting materials; basket-ware and wickerwork

UK Trade Tariff

Try using the following: .csv .json .xml .rss .rdf

The screenshot shows the BBC Doctor Who website. At the top, there's a navigation bar with the "one" BBC logo and a search bar. The main title "DOCTOR WHO" is prominently displayed with the TARDIS logo. Below the title, a menu bar offers links to "Home", "Episodes", "Clips", "Galleries", "Latest News", "Characters", "Monsters", "Fun and Games", and "More". A central image features the Twelfth Doctor and his companion. To the left, a box promotes the "Launch" event with a thumbnail of the Doctor and a woman. To the right, two boxes show "On iPlayer" (with a note about availability) and "On TV" (listing the "Day of the Doctor" broadcast details).

BBC Music and Programmes





## 2. Look for alternative links

The screenshot shows the NewsAsia website homepage from Wednesday, August 06, 2014. The top navigation bar includes 'Business Insight' (highlighted in red), 'NEWS', 'TV', 'WATCH LIVE', and various news categories like 'ASIA PACIFIC', 'SINGAPORE', 'WORLD', 'BUSINESS', 'SPORT', 'ENTERTAINMENT', 'TECHNOLOGY', 'HEALTH', 'LIFESTYLE', 'VIDEOS', 'WEATHER', and 'MORE'. A large blue box highlights the 'NEWS' section, and a large black arrow points downwards from it towards the main content area. The main content features a large image of two men shaking hands at a ceremony. Headlines include: '12% for Home Team officers, bonuses of up to S\$30,000', 'Pay rise, special bonus for about 23,000 nurses', '50,000 openings on Jobs Bank for Singaporeans, PRs', and 'NUS University Town Identified as a high-risk dengue cluster'. The sidebar on the left lists categories: 'SIMPLY SINGAPORE', 'WORLD', 'BUSINESS', 'SPORT', 'LIFESTYLE', 'HEADLINES', 'VIDEOS', and 'MEDIACORP'.

Scroll down!

12% for Home Team officers, bonuses of up to S\$30,000

Officers from the Home Team, both past and present were recognised at the Home Team National Day Observance Ceremony on Wednesday.

9 hours ago

Pay rise, special bonus for about 23,000 nurses

10 hours ago

50,000 openings on Jobs Bank for Singaporeans, PRs

1 hour ago

NUS University Town Identified as a high-risk dengue cluster

10 hours ago





# 2. Look for alternative links

 <b>CHANNEL NEWSASIA</b>  MediaCorp News Group. © 2014 MediaCorp Pte Ltd. All Rights Reserved.  Terms and Conditions Privacy Policy About MediaCorp Pte Ltd	<b>NEWS</b> Asia Pacific Singapore World Business Sport Entertainment Technology Health Lifestyle Videos Photos Special Reports Archives	<b>TV</b> Live TV TV Videos TV Schedule  <b>SERVICES</b> Weather  <b>ADVERTISE WITH US</b> Online Advertising Mobile Advertising TV Advertising Contact Sales	<b>ABOUT US</b> About Channel NewsAsia Our Logo Our Coverage Our Tagline Presenters and Correspondents Contact Us  <b>GET OUR NEWS</b>  
---	---	---	---

RSS

# 3. Look for embedded data

The screenshot shows the 'Hidden data extractor' tool. At the top, there's a dark header bar with the text 'Hidden data extractor' in blue on the left and 'ODI Experiment' in orange on the right. To the right of the header is the ODI logo. Below the header is a light gray main area. In the center, it says 'Hidden data extractor' and 'Enter the URL of any webpage to see what JSON data is hidden within it.' Below this is a large input field for URLs. Underneath the input field is a blue 'Submit' button. Further down, there's a section titled 'Try these' with two links: 'Products from Marks and Spencer UK' and 'Products from ASOS'.

<http://odinprac.theodi.org/hidden-data-extractor/>



# Finding data on the web (**of data**)

1. Add random extensions (.xml, .json, .csv etc)
2. Look for alternative links (rss feeds etc)
3. Look for embedded data
4. Do some content negotiation
5. Spot the API
6. Scrape (or search google again)

**IN THE WEB**

How the web should work, but people forgot that Tim put this in when he invented it!



Content created by  
The Open Data Institute

# Exercise

Find a data set using one of the routes we've just looked at.....

Ask yourself – (and discuss in groups)

Is it usable?

What makes it usable?

What more do you need to know?



# Outcomes

Identify a number of different sources of open data on the web.

Create search patterns that enable easy discovery of new sources of open data.



# Outcomes

Identify a number of different sources of open data on the web.

Create search patterns that enable easy discovery of new sources of open data.

Understand the difference between “data on the web” and the “web of data”.

