Open Data in Practice

http://training.theodi.org/InPractice

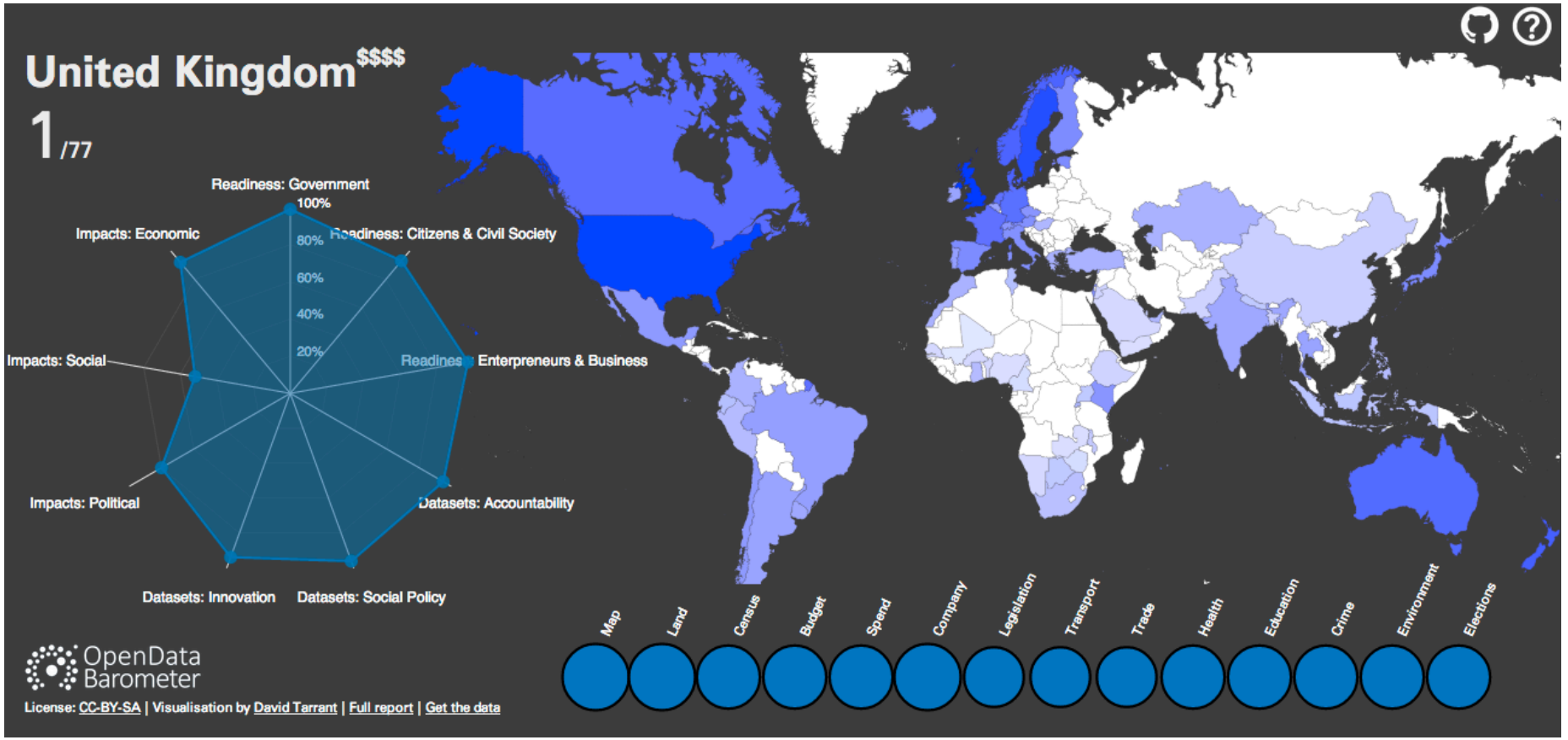
David Tarrant · @davetaz

Session 2
Data discovery patterns

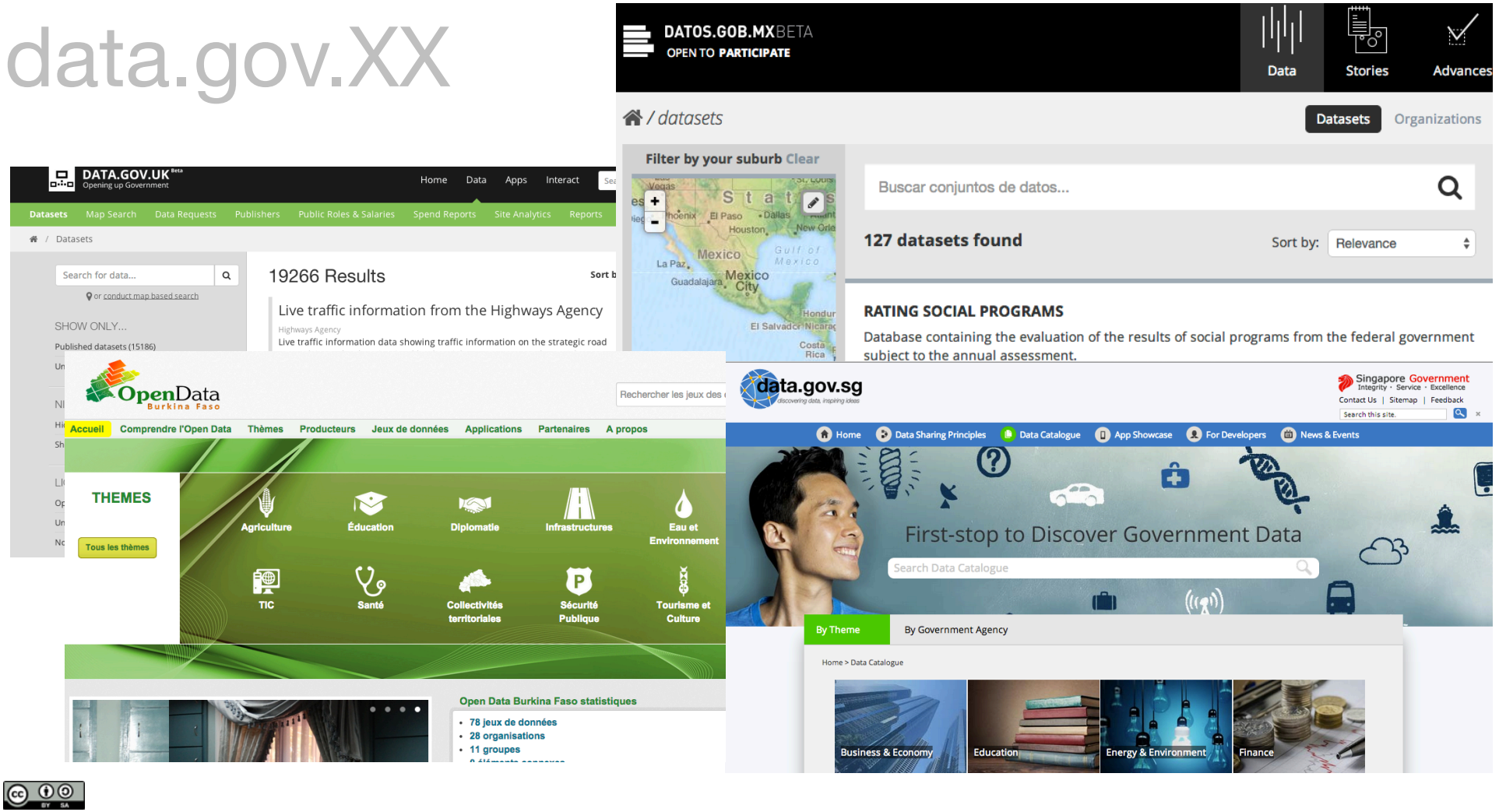
Finding data on the web **(of documents)**

- Government data
- Private sector data
- Google advanced
- Aggregators and portals
- Scraping

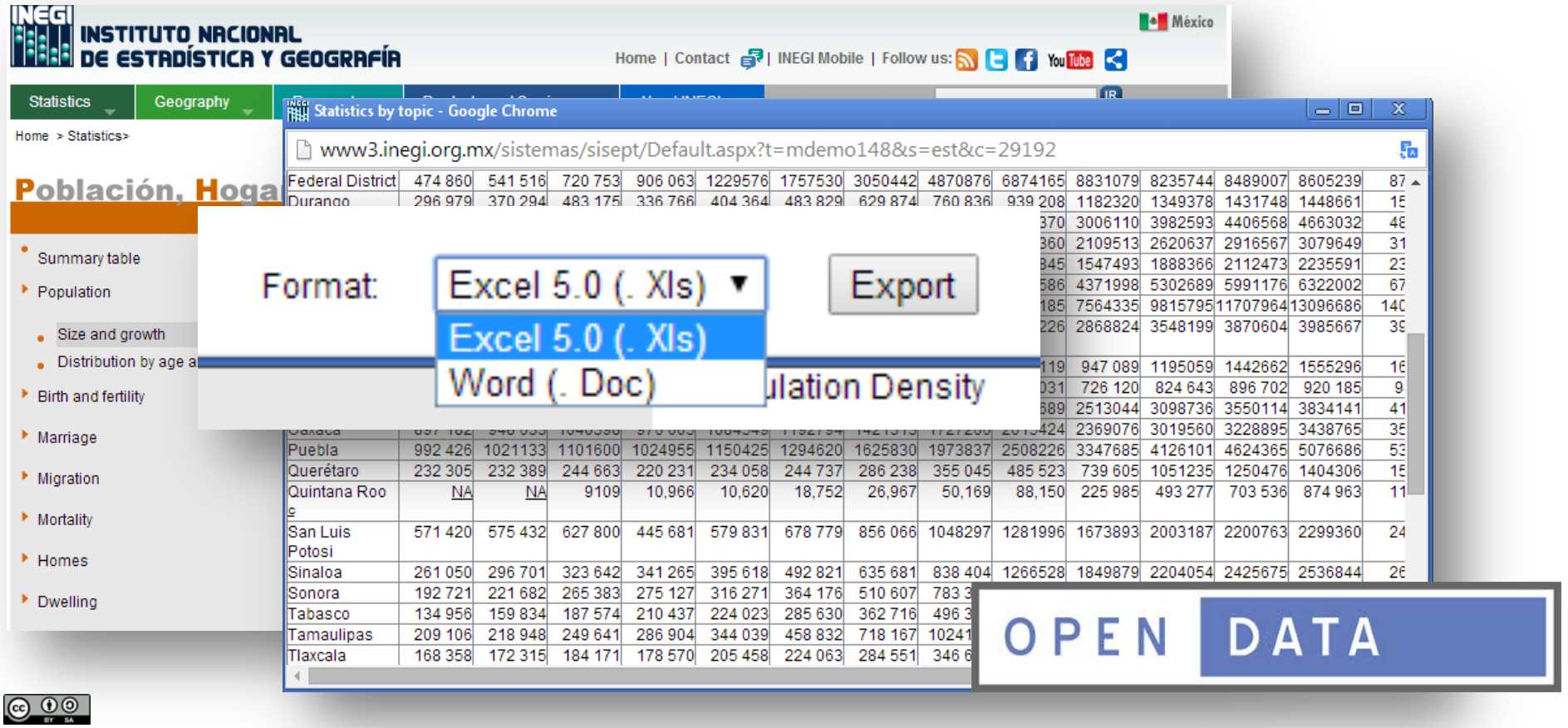
Government data



data.gov.XX

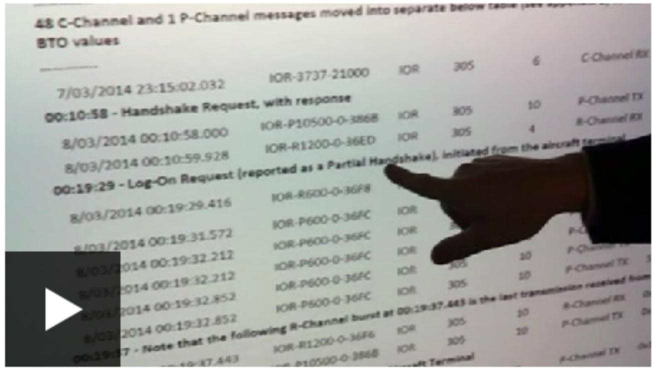

DATOS.GOB.MXBETA
OPEN TO PARTICIPATE
Data
Stories
Advances
/ datasets
Datasets
Organizations
Filter by your suburb Clear
Buscar conjuntos de datos...
127 datasets found
Sort by: Relevance
RATING SOCIAL PROGRAMS
Database containing the evaluation of the results of social programs from the federal government subject to the annual assessment.
DATA.GOV.UK
Opening up Government
Home
Data
Apps
Interact
Datasets
Map Search
Data Requests
Publishers
Public Roles & Salaries
Spend Reports
Site Analytics
Reports
Search for data...
19266 Results
Live traffic information from the Highways Agency
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 et Culture
Open Data Burkina Faso statistiques
78 jeux de données
28 organisations
11 groupes
data.gov.sg
Singapore Government
Integrity · Service · Excellence
Home
Data Sharing Principles
Data Catalogue
App Showcase
For Developers
News & Events
First-stop to Discover Government Data
Search Data Catalogue
By Theme
By Government Agency
Home > Data Catalogue
Business & Economy
Education
Energy & Environment
Finance

Government


INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA
México
Home | Contact | INEGI Mobile | Follow us:
Statistics
Geography
Home > Statistics>
Población, Hoga
Summary table
Population
Size and growth
Distribution by age a
Birth and fertility
Marriage
Migration
Mortality
Homes
Dwelling
Statistics by topic - Google Chrome
www3.inegi.org.mx/sistemas/sisept/Default.aspx?t=mdemo148&s=est&c=29192
Format:
Excel 5.0 (. Xls)
Word (. Doc)
Export
ulation Density
Federal District 474 860 541 516 720 753 906 063 1229576 1757530 3050442 4870876 6874165 8831079 8235744 8489007 8605239
Puebla 992 426 1021133 1101600 1024955 1150425 1294620 1625830 1973837 2508226 3347685 4126101 4624365 5076686
Querétaro 232 305 232 389 244 663 220 231 234 058 244 737 286 238 355 045 485 523 739 605 1051235 1250476 1404306
Quintana Roo 9109 10,966 10,620 18,752 26,967 50,169 88,150 225 985 493 277 703 536 874 963
San Luis Potosi 571 420 575 432 627 800 445 681 579 831 678 779 856 066 1048297 1281996 1673893 2003187 2200763 2299360
Sinaloa 261 050 296 701 323 642 341 265 395 618 492 821 635 681 838 404 1266528 1849879 2204054 2425675 2536844
Sonora 192 721 221 682 265 383 275 127 316 271 364 176 510 607
Tabasco 134 956 159 834 187 574 210 437 224 023 285 630 362 716
Tamaulipas 209 106 218 948 249 641 286 904 344 039 458 832 718 167
Tlaxcala 168 358 172 315 184 171 178 570 205 458 224 063 284 551
OPEN DATA

Government / Private

Flight MH370: Malaysia releases raw satellite data



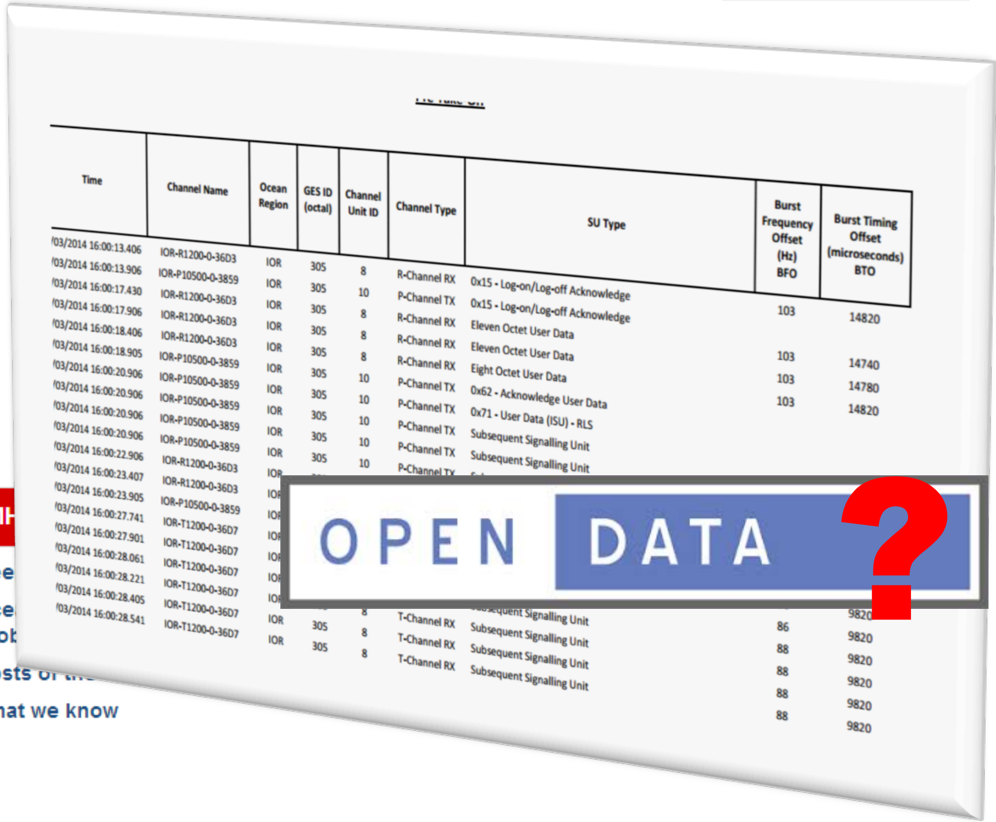
The BBC's Richard Westcott visited Inmarsat's headquarters to find out what the data tells us about MH370's fate

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

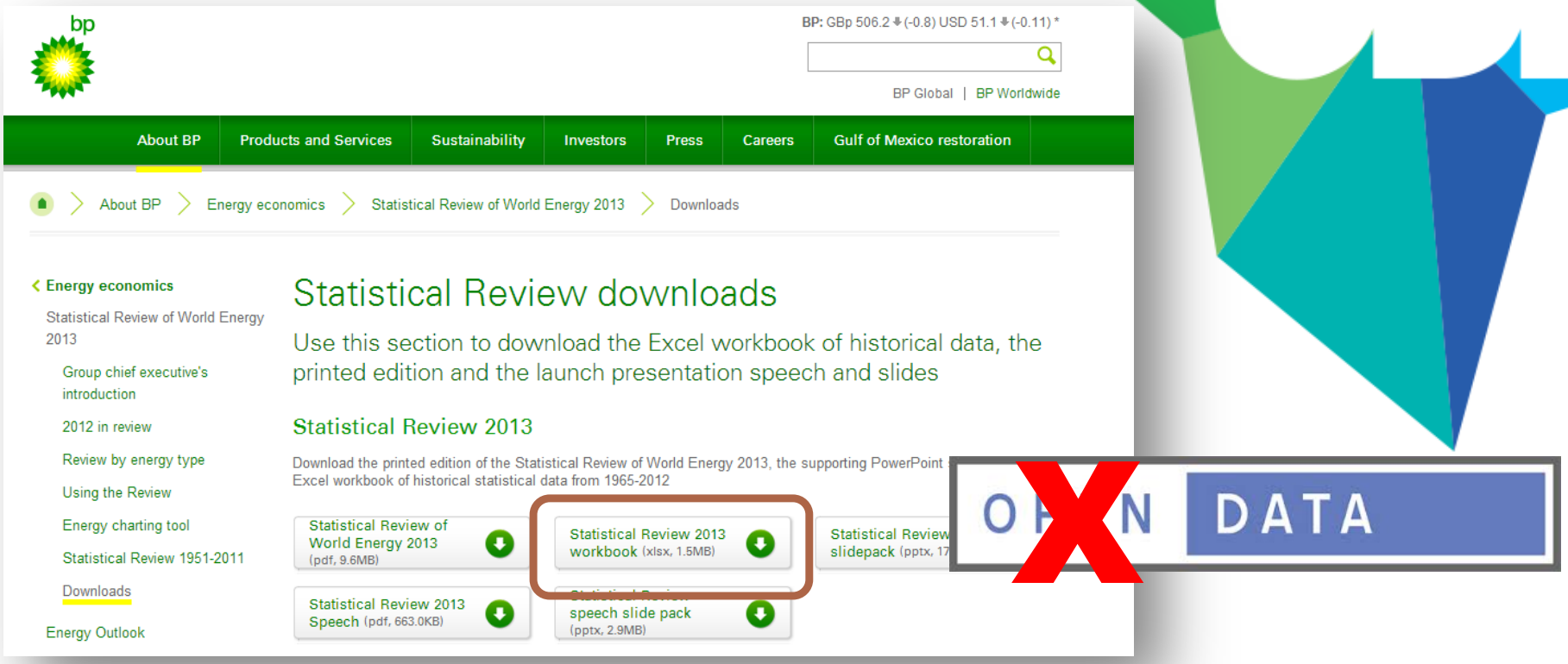
The data was first released to relatives of passengers, who have been asking for greater transparency, before copies were also provided to media.

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

What we know

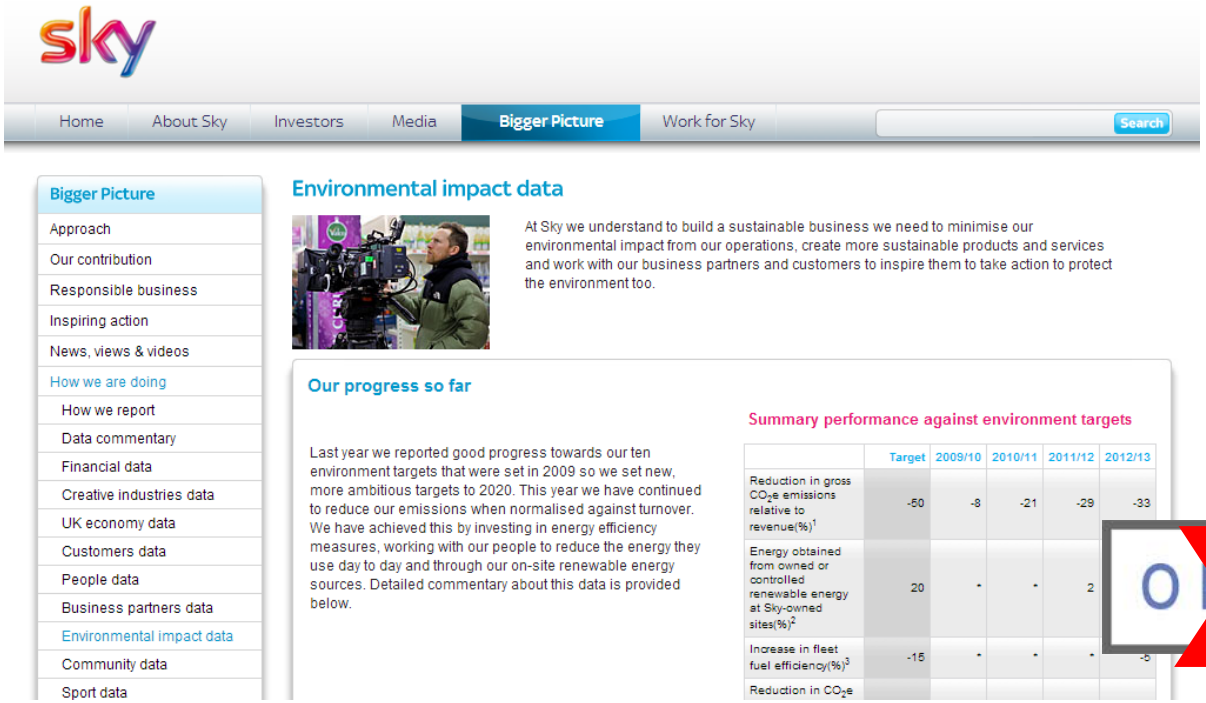


Suppliers



To find this Google "**bp statistical review**"

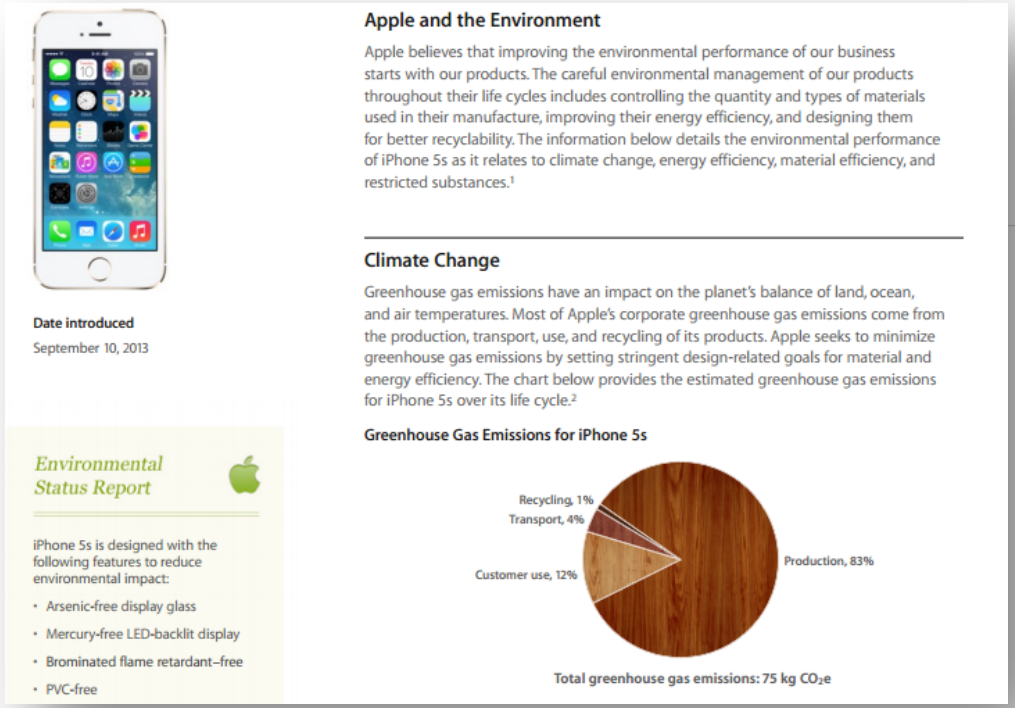
Suppliers (ish)



http://corporate.sky.com/the_bigger_picture/how_we_are_doing/environmental_impact_data

Suppliers (ish)

Product Environment Reports



Date introduced

September 10, 2013

Environmental Status Report

iPhone 5s is designed with the following features to reduce environmental impact:

- Arsenic-free display glass
- Mercury-free LED-backlit display
- Brominated flame retardant–free
- PVC-free

Apple and the Environment

Apple believes that improving the environmental performance of our business starts with our products. The careful environmental management of our products throughout their life cycles includes controlling the quantity and types of materials used in their manufacture, improving their energy efficiency, and designing them for better recyclability. The information below details the environmental performance of iPhone 5s as it relates to climate change, energy efficiency, material efficiency, and restricted substances.[1]

Climate Change

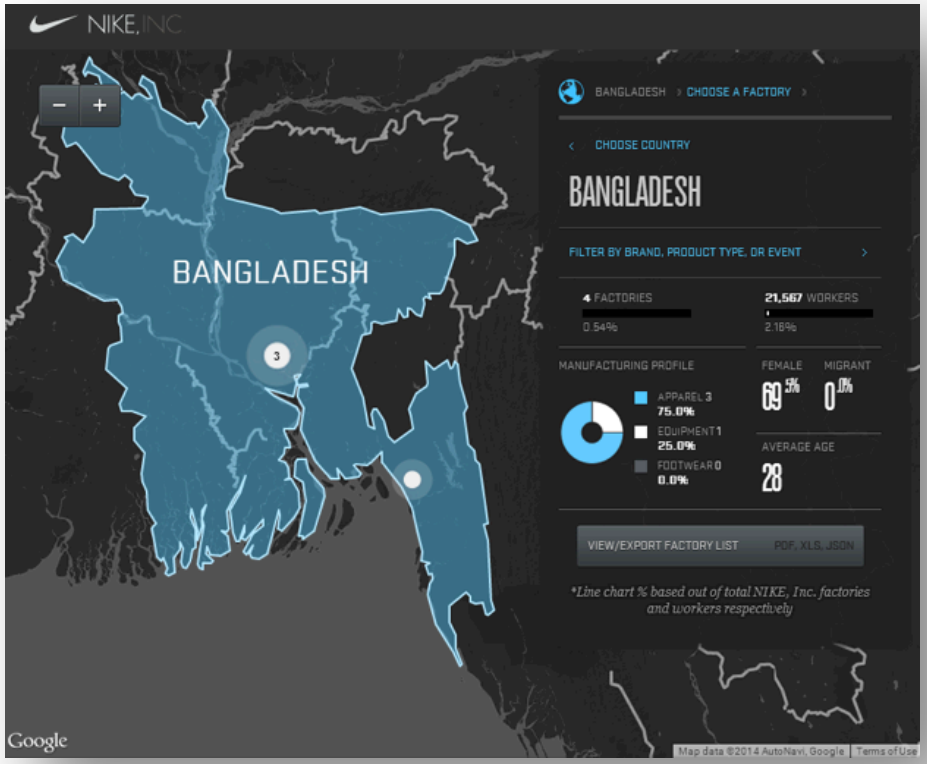
Greenhouse gas emissions have an impact on the planet's balance of land, ocean, and air temperatures. Most of Apple's corporate greenhouse gas emissions come from the production, transport, use, and recycling of its products. Apple seeks to minimize greenhouse gas emissions by setting stringent design-related goals for material and energy efficiency. The chart below provides the estimated greenhouse gas emissions for iPhone 5s over its life cycle.[2]

Greenhouse Gas Emissions for iPhone 5s

Total greenhouse gas emissions: 75 kg CO_2e

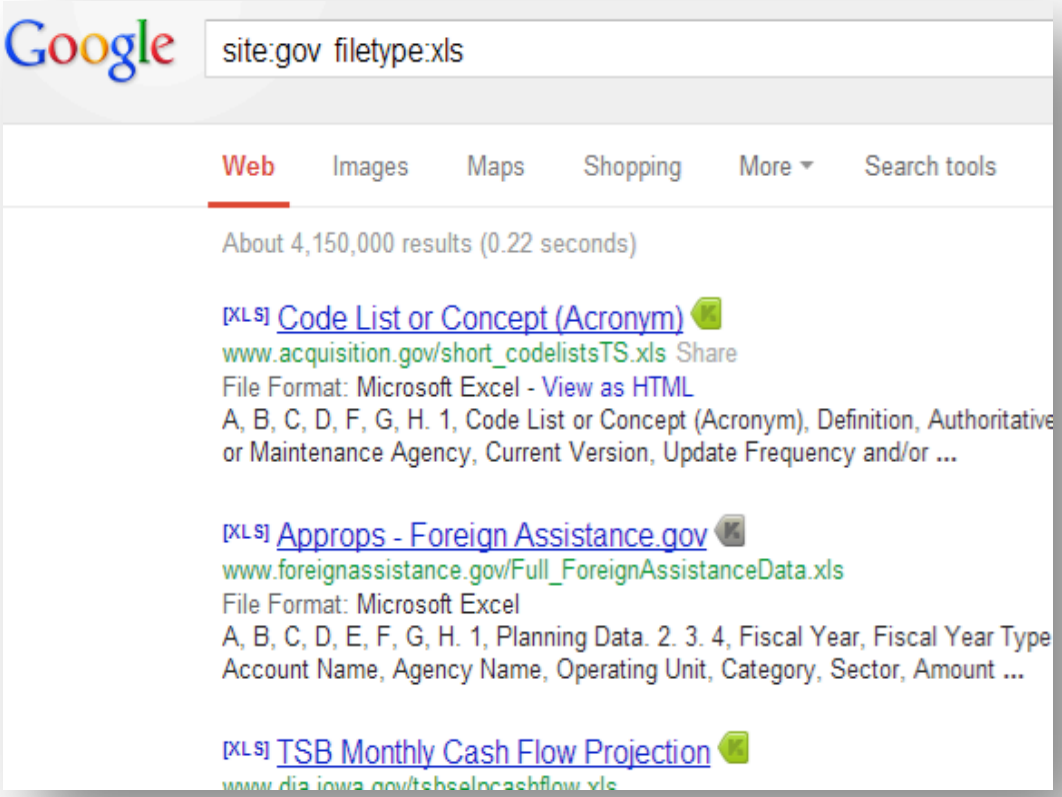
http://www.apple.com/uk/environment/reports/

Suppliers



http://manufacturingmap.nikeinc.com/#

Google advanced



site: Get results only from certain sites or domains

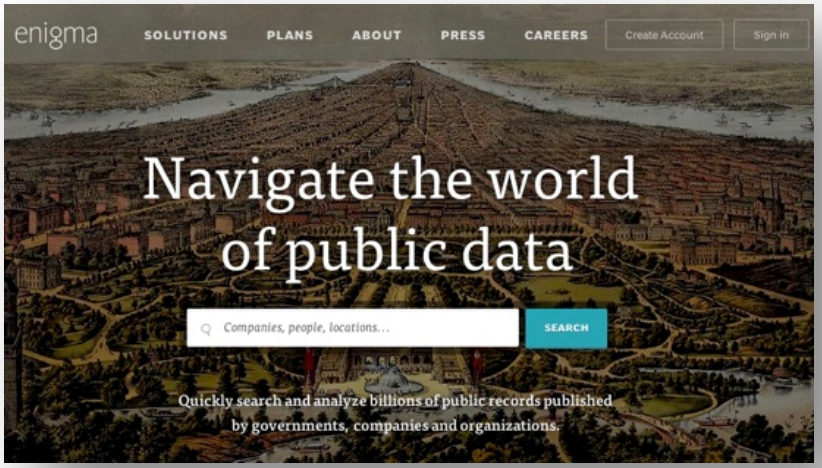
link: Find pages that link to a certain page

related: Find sites similar to one you already know

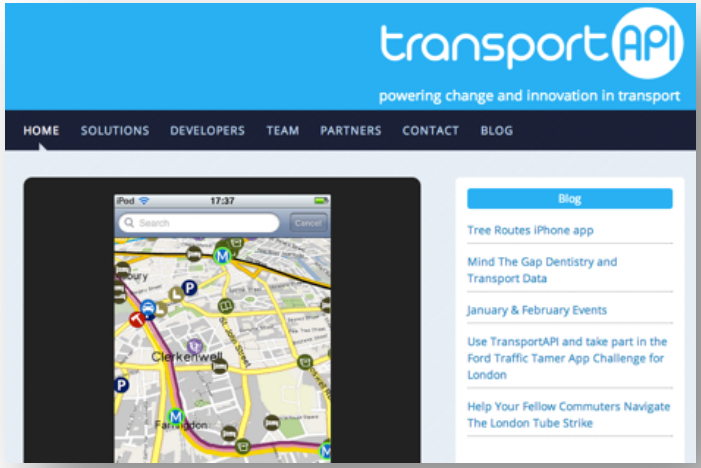
filetype: Find certain file types only

Aggregators and portals

Collect together data from across the web into one place.



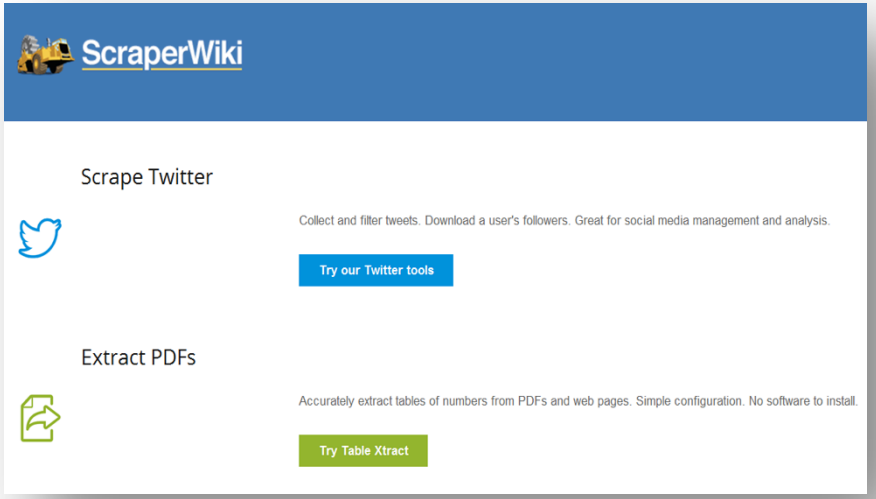
enigma.io



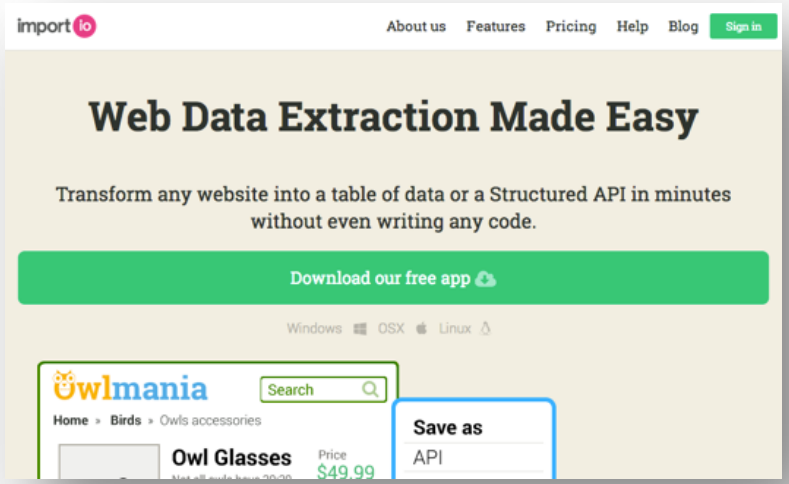
transportAPI

Scraping

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



scraperwiki.com



import.io

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)

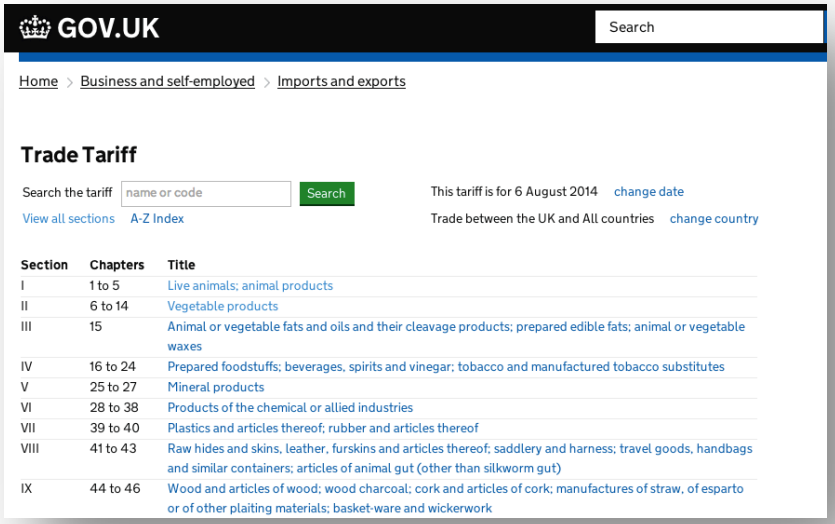
How the web should work, but people forgot that Tim put this in when he invented it!

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

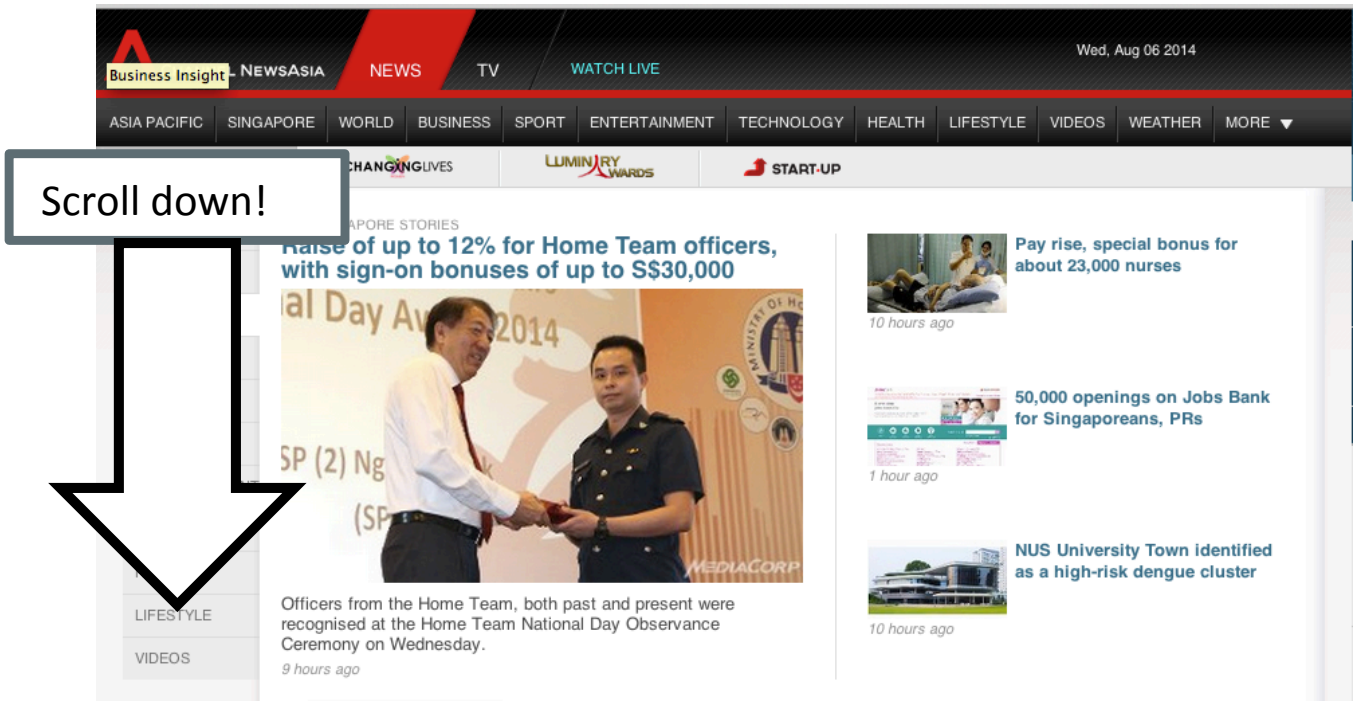


UK Trade Tariff

BBC Music and Programmes

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

2. Look for alternative links



2. Look for alternative links

CHANNEL NEWSASIA

MediaCorp News Group.
© 2014 MediaCorp Pte Ltd.
All Rights Reserved.

Terms and Conditions
Privacy Policy
About MediaCorp Pte Ltd

NEWS
Asia Pacific
Singapore
World
Business
Sport
Entertainment
Technology
Health
Lifestyle
Videos
Photos
Special Reports
Archives

TV
Live TV
TV Videos
TV Schedule

SERVICES
Weather

ADVERTISE WITH US
Online Advertising
Mobile Advertising
TV Advertising
Contact Sales

ABOUT US
About Channel NewsAsia
Our Logo
Our Coverage
Our Tagline
Presenters and Correspondents
Contact Us

GET OUR NEWS

RSS

Finding data on the web (c

Techniques 3-5 are not covered in this session. Please ask your trainer for more information if there is time.

1. json, .csv etc)
2. eeds etc)
3. Look for embedded data
4. Do some content negotiation
5. Spot the API
6. Scrape (or search google again)

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?


Thank-you