# Changing the World with Open Data

Dr David Tarrant
Dr Les Carr
University of Southampton

This work is licensed under a Creative Commons Attribution 3.0 Unported License.
Image attribution included when available.

# Data Rich Applications and **Open Data**

CC BY

# Course Topics

CC BY

# Open Data

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

CC BY

# Availability and Access

- The data must be available as a **whole**
- Available "**freely**", preferably to be downloaded
- Must be available in a convenient and modifiable **form**

CC BY

# Use

- Data must be available under an open **license**.
  - Open Government License
  - CC0, Creative Comments No Restrictions.

- Must be **reusable** and **redistributable**.

CC BY

# Universal Participation

- Must allow **everyone** to use the data.
- Must allow **unrestricted** use
- **Intermixing** and **Interoperability**.

Open Data Institute
Knowledge for everyone

*"The interesting thing about open data is the serendipity effect. It's about interconnecting all kinds of things."*

*"Open data is not just about transparency and accountability; there is potential for economic innovation and value to come from it." Wired 2012*

CC BY

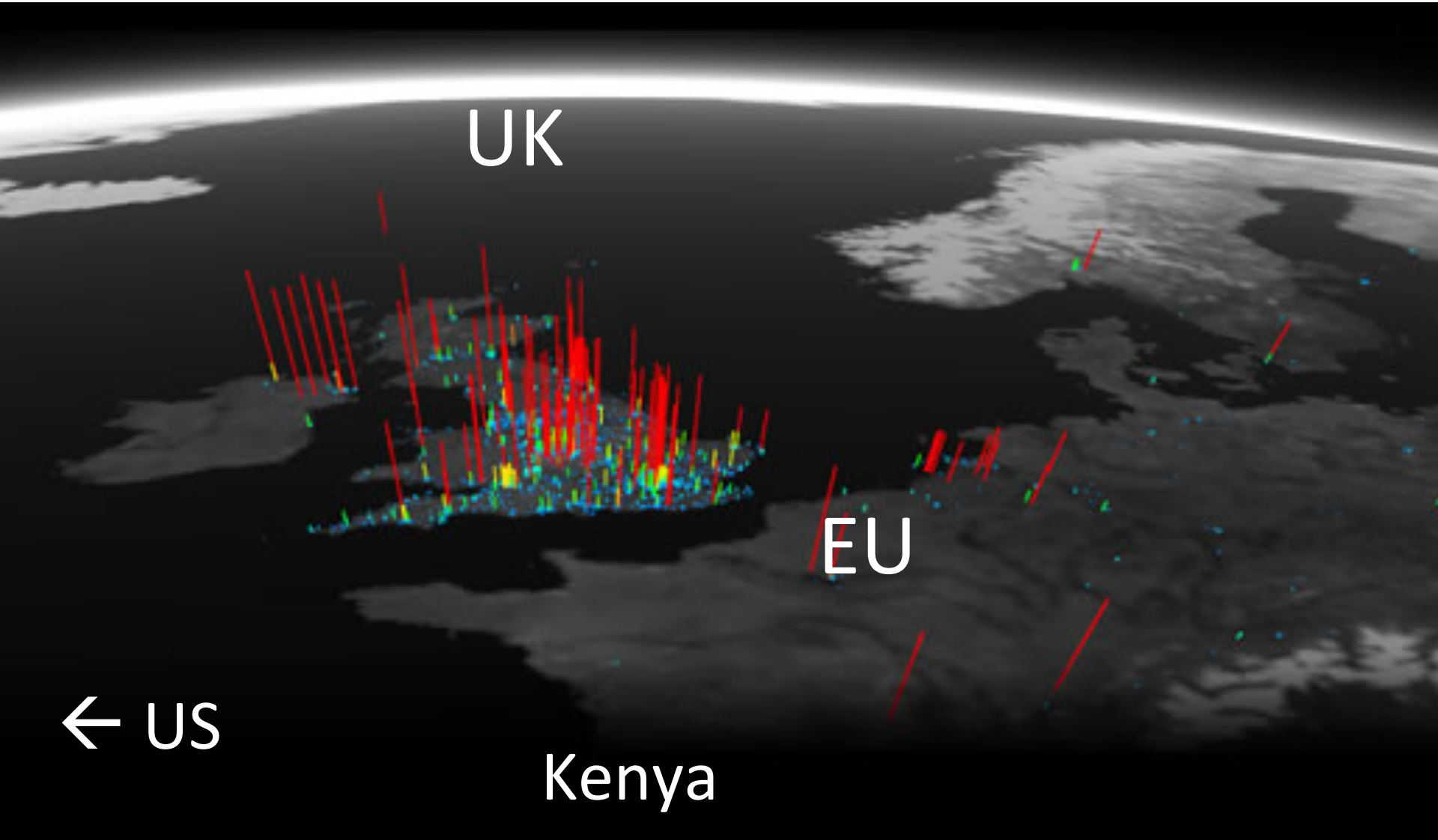
Open Data Institute
Knowledge for everyone

*Open data is not just "nice to have" but critical for our future*

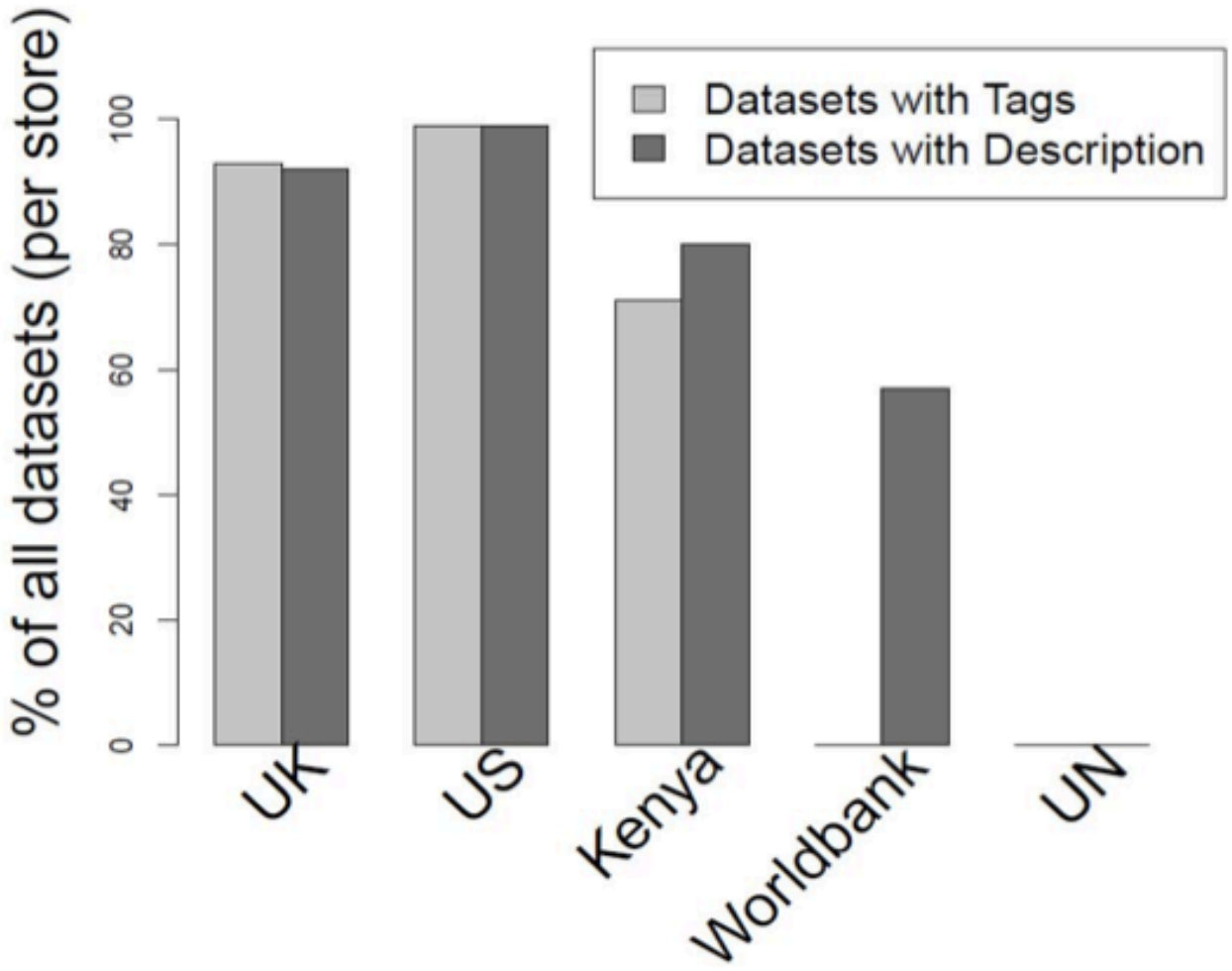
*Open data is an unstoppable tide, with transformative implications for both businesses and the public sector.*

CC BY

# Global Supporters



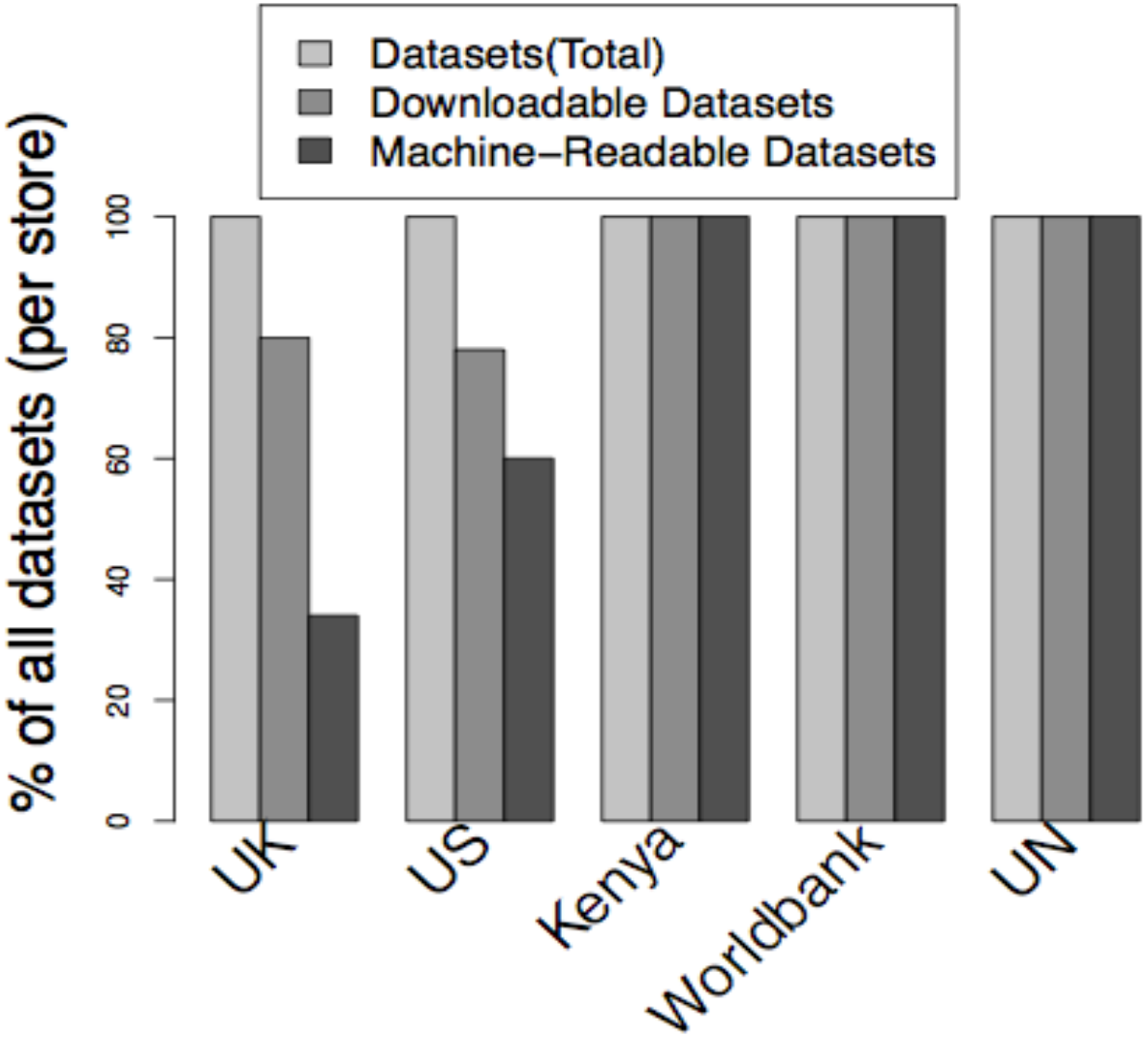
# Quality of Datasets




CC BY

Braunschweig, K., J. Eberius, M. Thiele, and W. Lehner. "The State of the Open Data-Limits of Current Open Data Platforms." In Proceedings of WWW. 2012.

# Dataset Availability




CC BY

Braunschweig, K., J. Eberius, M. Thiele, and W. Lehner. "The State of the Open Data-Limits of Current Open Data Platforms." In Proceedings of WWW. 2012.

# Question?

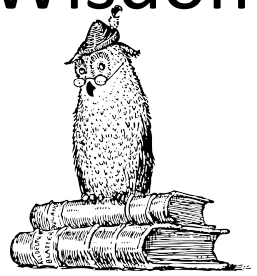
# Why expose data?

CC BY

# Data: The New Raw Material?

- Data:
  - Single piece
    of information
      of every nature

Data --> Information --> Knowledge --> Wisdom



CC BY

# Why expose data?

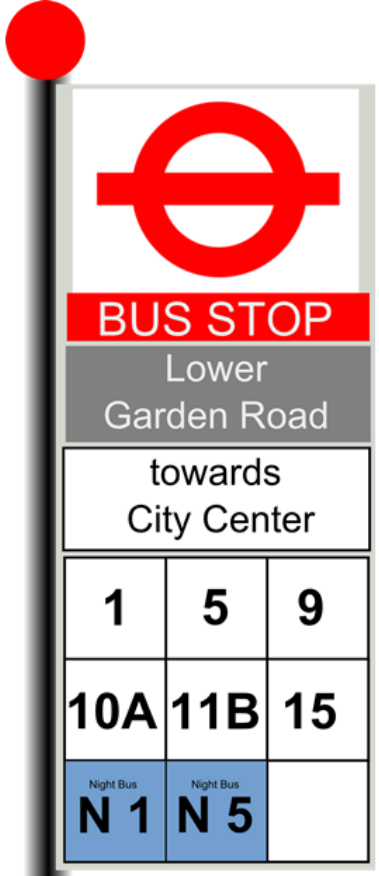
- Transparency
- Trust
- Marketing
- Domain-Leadership

- Crowd Sourcing
- Streamlining
- Knowledge Building
- Money Saving

CC BY

Savings
Advocacy
Collaboration
Transparency
Planning
Marketing
Activities
Evaluation
Efficiency
Community
Integration
Trust
Campaigning
Policy

CC BY




# Public Service Improvement

- UK released data on 300,000 bus stops.
- 18,000 corrected by Open Street Map community.

# Health

- Prescription savings worth millions identified by using open data.
  - Combined usage data with drug effectiveness data.

http://www.theodi.org/news/prescription-savings-worth-millions-identified-odi-incubated-company

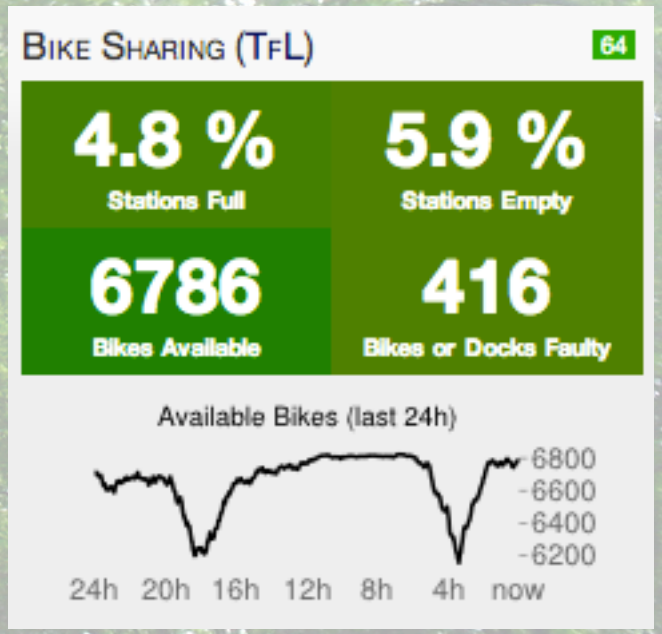
- 95% reduction in MRSA following open release of hospital performance data.

https://www.mja.com.au/journal/2008/188/11/significant-reductions-methicillin-resistant-staphylococcus-aureus-bacteraemia

http://www.desmogblog.com/2012/11/15/why-climate-deniers-have-no-credibility-science-one-pie-chart

# The Environment

## When open access becomes open data




## Bike sharing

http://citydashboard.org/london/

CC BY

# Finances

MPs jailed after expenses exposed

http://www.telegraph.co.uk/news/newstopics/mps-expenses/

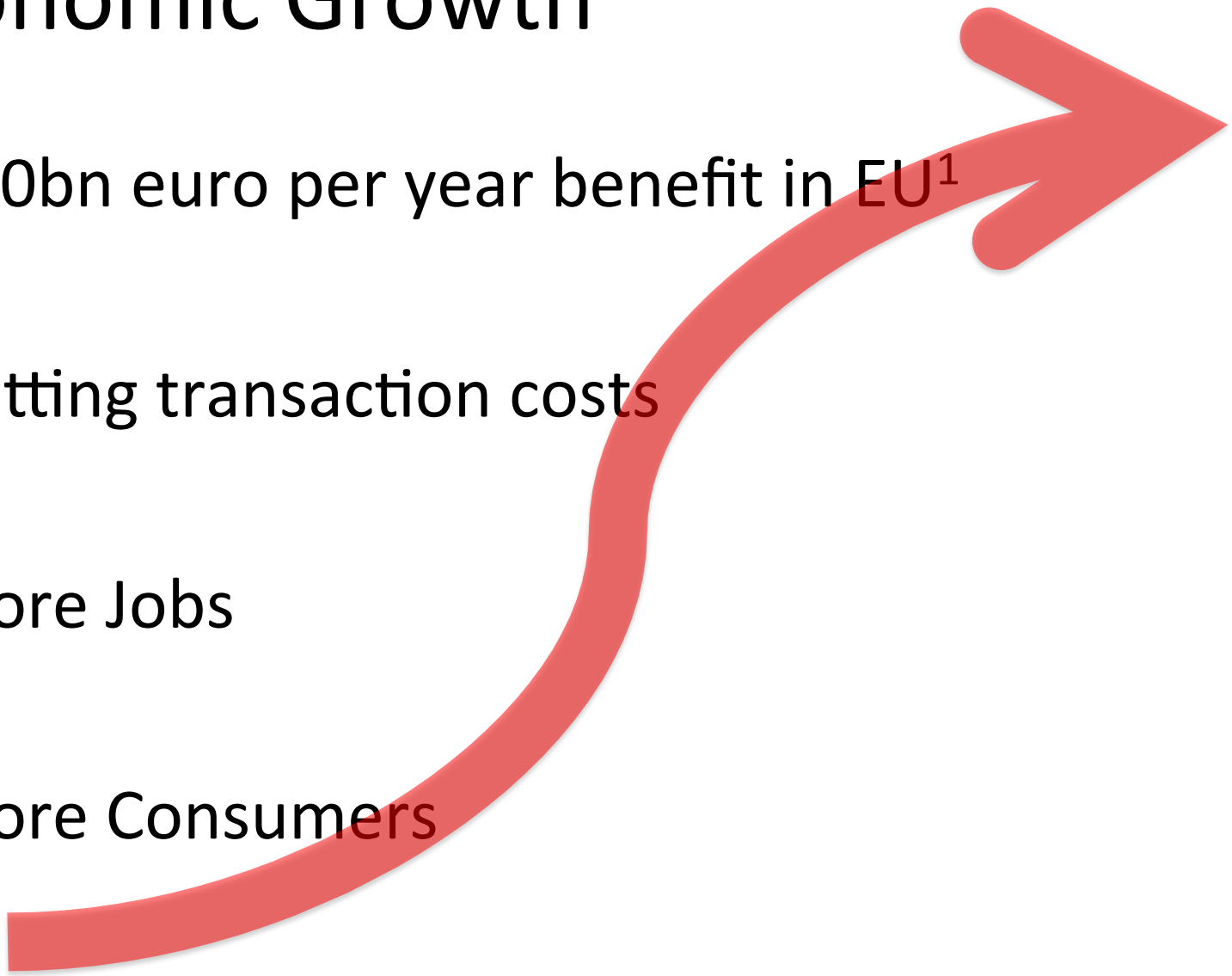
CAN$3.2 in money laundered via misuse of charity tax codes

http://eaves.ca/2010/04/14/case-study-open-data-and-the-public-purse/

CC BY

# Economic Growth

- 140bn euro per year benefit in EU[1]
- Cutting transaction costs
- More Jobs
- More Consumers

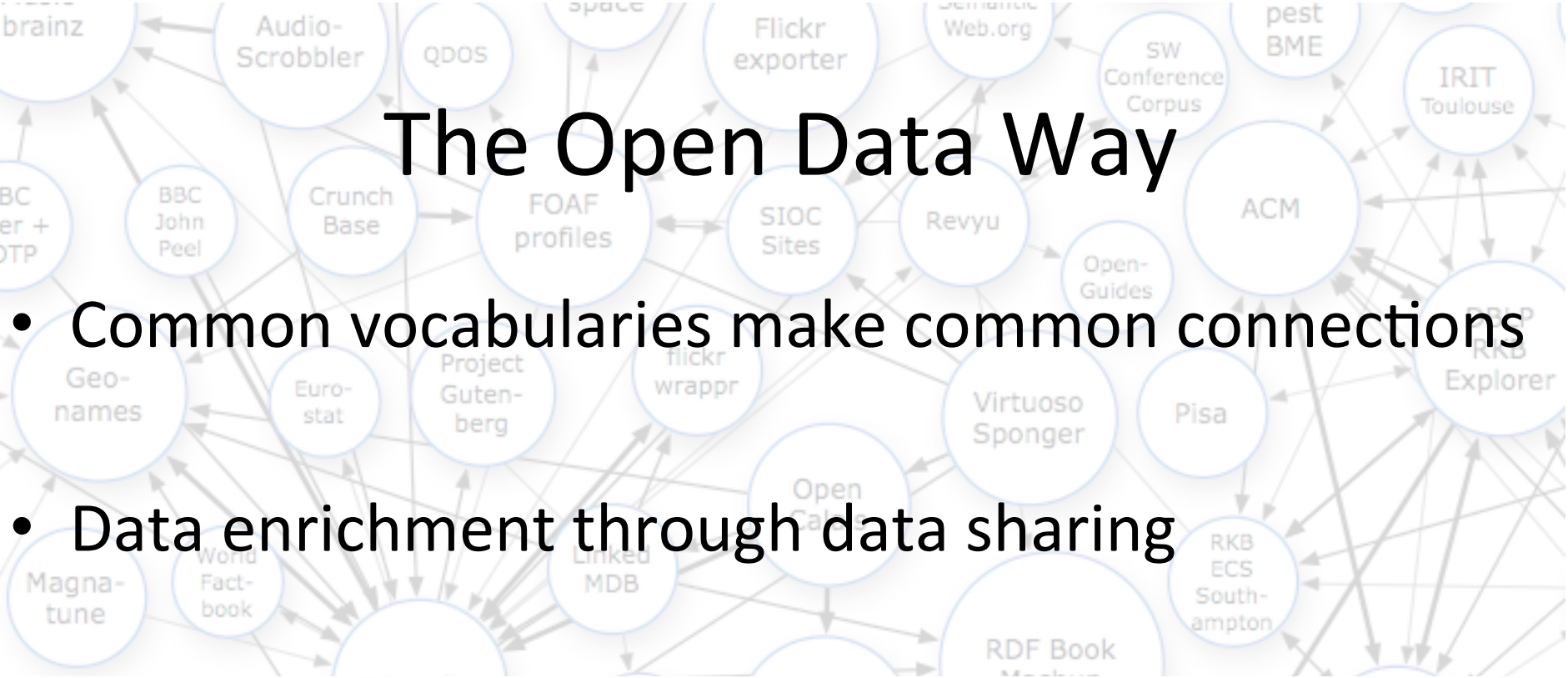


CC BY

[1] - http://epsiplatform.eu/content/review-psi-re-use-studies-published

# The Open Data Way

- Common vocabularies make common connections
- Data enrichment through data sharing




Browse CPV Codes Homepage

## Aircraft and spacecraft

**ID:** 34700000
**ID with checksum:** 34700000-4
**URI:** http://purl.org/cpv/2008/code-34700000
**Scheme:** CPV 2008

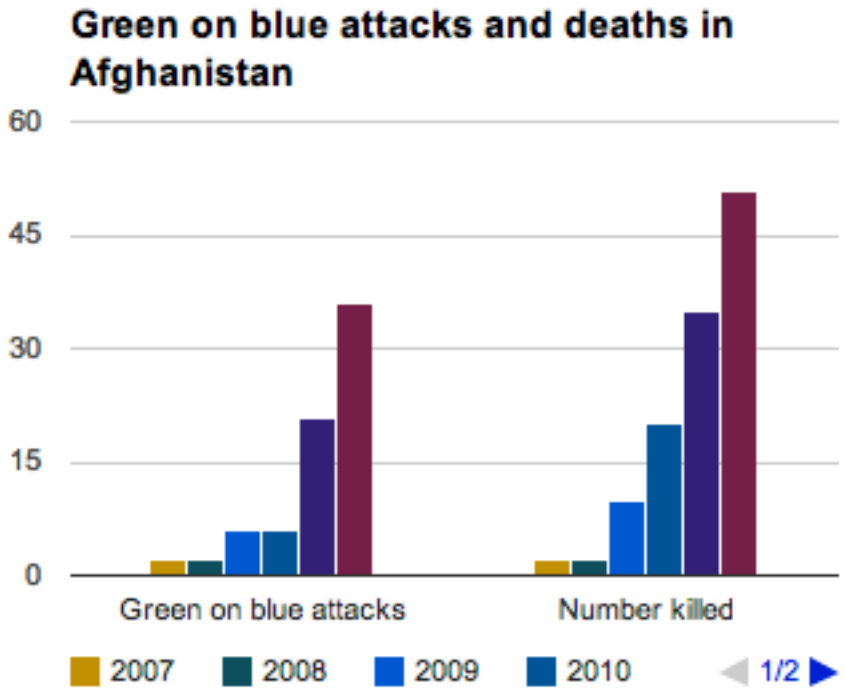
### Data

**Aircraft and spacecraft** - RDF Turtle

CPV2008

# New Jobs

# Data Journalism




**Data summary**

**Total green on blue attacks by year**

Click heading to sort table. Download this data

| Year | Green on blue attacks | Number ki |
|---|---|---|
| 2007 | 2 | 2 |
| 2008 | 2 | 2 |
| 2009 | 6 | 10 |
| 2010 | 6 | 20 |
| 2011 | 21 | 35 |
| 2012 | 36 | 51 |

Source: Isaf

CC BY

# New Jobs

# Open Data Managers

## Building: EEE Building

http://id.southampton.ac.uk/building/32 ← This is the URI

**Site:** Highfield Campus

**Date of Construction:** 2006

**Architect:** John McAslan & Partners

**Occupants:**
- Electronics & Computer Science
- Agents, Interactions & Complexity
- Web & Internet Science

**Features:** Building 32 is non-residential

### Disability Information

- View Disability Report for this Building

### Email

Bulk email lists may only be emailed by members of the university *and in addition* are moderated.

- building-32@all.soton.ac.uk

### Points of Sale & Service

**Bottle Vending Machine (Coke)**

**Offers:** Caffeine, Soft Drinks.

**Physical and Applied Science Faculty Deanery**

**Snack Vending Machine (Snakky Max)**

**Offers:** Snacks.

**Social and Human Sciences Faculty Deanery**

### Upcoming Events

**Tuesday 25th September**

14:00 - The use of Vibration in Microfluidics for Particle Handling and Sorting

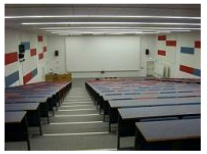
### Teaching Locations

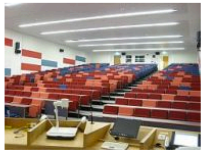
Generally these are lecture theatres, but this also includes any location used in the syllabus database. The feature codes are a little esoteric due to them having only been intended for internal office use.

**32 / 1015 (Lecture Theatre)**

**Capacity:** 426

- RSC- BENCH PC
- RSC- DATA/VIDEO PROJECTION
- RSC- DISABLED ACCESS - AUDIENCE (A) Ground Floor Room Fully Accessible
- RSC- DISABLED ACCESS - AUDIENCE (B) Upper Floor Room Egress Standard Lift & Evac Management Plan
- RSC- DISABLED ACCESS - INSTRUCTOR (B) Upper Floor Room Egress Standard Lift & Evac Management Plan
- RSC- DVD
- RSC- INDUCTION LOOP (HEARING)
- RSC- LAPTOP/NETWORK CONNECTION
- RSC- MICROPHONE (VOICE ENHANCEMENT)
- RSC- PC (NETWORKED; REQUIRES UNIVERSITY LOGIN)
- RSC- SEATS FIXED TIERED
- RSC- VIDEO CONFERENCING
- RSC- VIZUALISER (VIDEO OHP)
- RSC- WIRELESS NETWORK





**32 / 3077 (Seminar Room)**

**Capacity:** 40

- RSC- BENCH PC
- RSC- DATA/VIDEO PROJECTION
- RSC- DISABLED ACCESS - AUDIENCE (B) Upper Floor Room Egress Standard Lift & Evac Management Plan
- RSC- DISABLED ACCESS - INSTRUCTOR (B) Upper Floor Room Egress Standard Lift & Evac Management Plan
- RSC- DVD
- RSC- INDUCTION LOOP (HEARING)
- RSC- LAPTOP/NETWORK CONNECTION
- RSC- SEATS MOVEABLE
- RSC- WHITEBOARD
- RSC- WIRELESS NETWORK

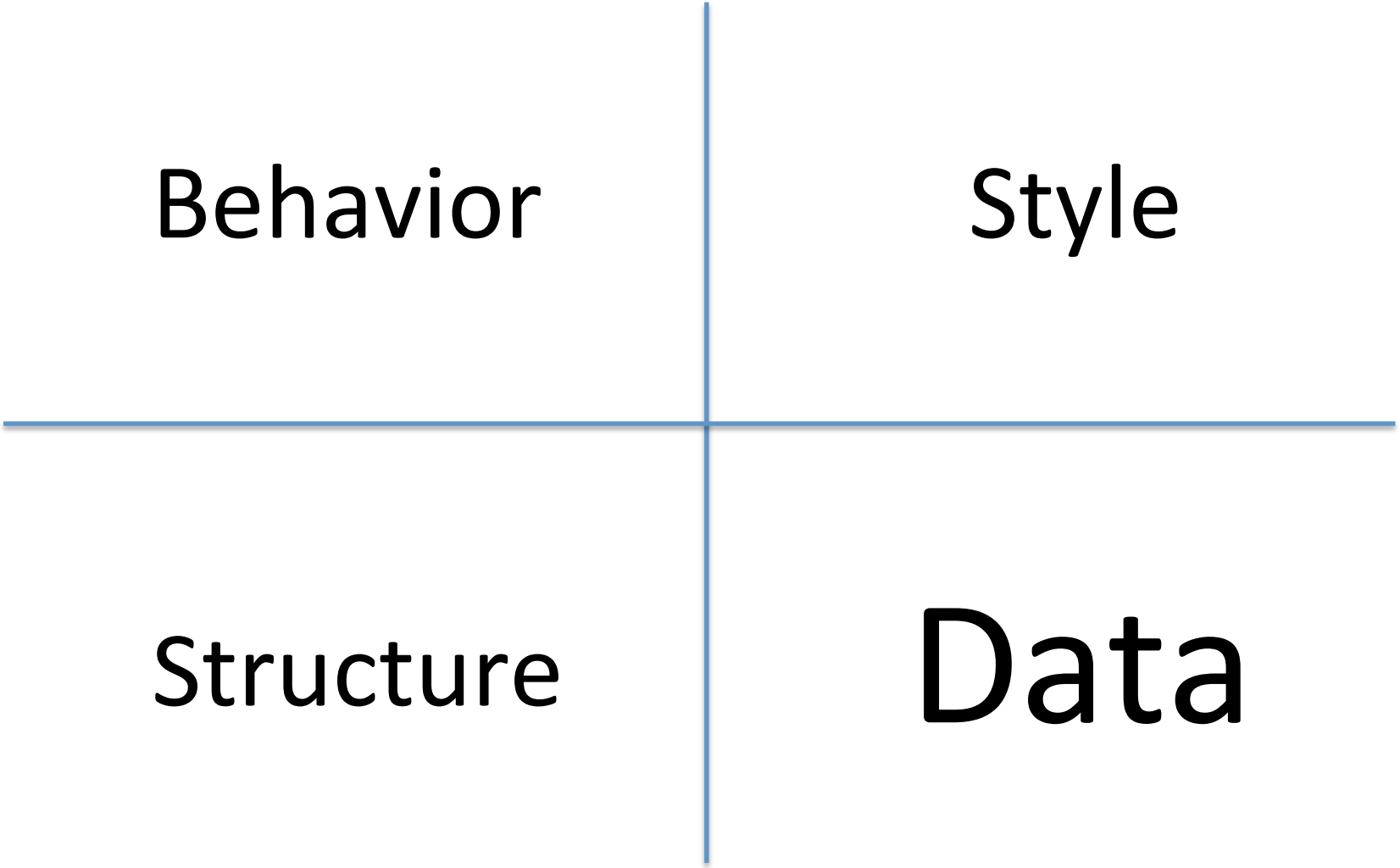
CC BY

# New Companies

- Placr
- MastadonC

CC BY

# Data and the Web




CC BY

★ Available on the web (whatever format) *but with an open licence, to be Open Data*

★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)

★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★ All the above, plus: Link your data to other people's data to provide context

CC BY

★ Available on the web (whatever format) *but with an open licence, to be Open Data*

★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)

★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★ All the above, plus: Link your data to other people's data to provide context

CC BY

★ Available on the web (whatever format) *but with an open licence, to be Open Data*

★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)

★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★ All the above, plus: Link your data to other people's data to provide context

CC BY

★ Available on the web (whatever format) *but with an open licence, to be Open Data*

★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)

★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★ All the above, plus: Link your data to other people's data to provide context

CC BY

★ Available on the web (whatever format) *but with an open licence, to be Open Data*

★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)

★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★ All the above, plus: Link your data to other people's data to provide context

CC BY

data.ac.uk
data.gov.uk
data.redbridge.gov.uk
ordnancesurvey.co.uk
data.southampton.ac.uk
opendata.go.ke
data.gov

CC BY

# Open Data

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

CC BY

# Exercise 1

- Locate a number of different sources of open datasets and pick one dataset from each.
- Identify which star rating the dataset is?
- What license is the dataset available under?
- How is this license referenced?
- What features make this data set usable (or not)?

CC BY

# Exercise 2

- Over the weekend, update your ECS profile.
- www.ecs.soton.ac.uk/people/username
- Then look at the data form.
- P.S. Do not change your name to “Professor”, it’s only funny the first time!

DATA

CC BY

# Further Reading


- Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic web." Scientific american 284.5 (2001): 28-37.The Semantic Web Revisited – Berners-Lee
- Shadbolt, Nigel, Wendy Hall, and Tim Berners-Lee. "The semantic web revisited." Intelligent Systems, IEEE 21.3 (2006): 96-101.
- Braunschweig, K., J. Eberius, M. Thiele, and W. Lehner. "The State of the Open Data-Limits of Current Open Data Platforms." In Proceedings of WWW. 2012.
- Manola, Frank, Eric Miller, and Brian McBride. "RDF primer." W3C recommendation, http://www.w3.org/TR/rdf-primer/
- http://www.theodi.org/faq
- http://opendatahandbook.org/


CC BY