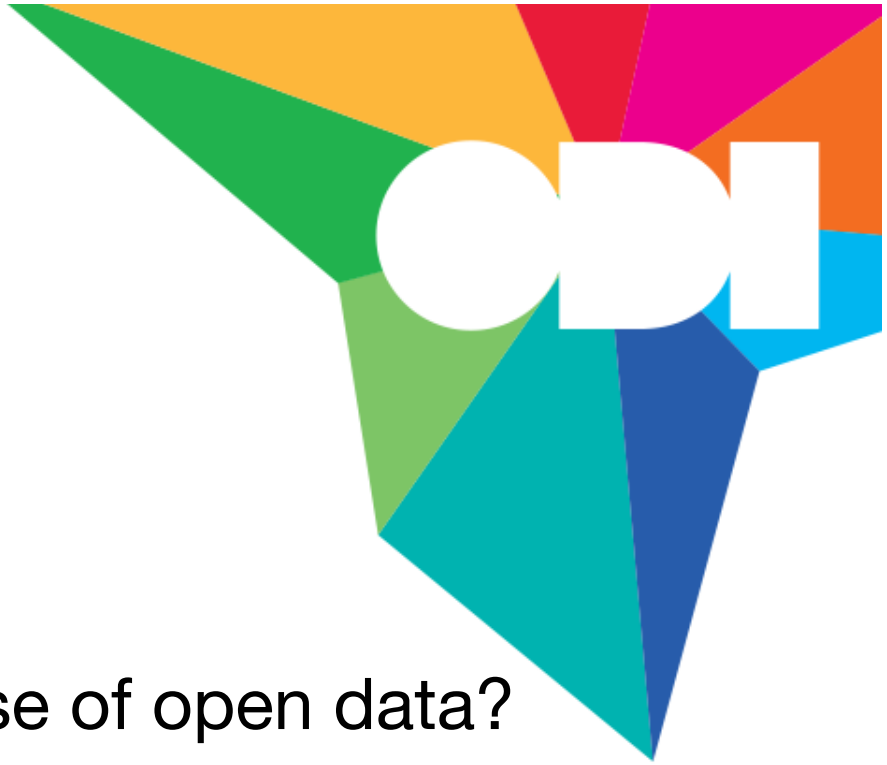

Open Data in Practice
http://training.theodi.org/InPractice
David Tarrant · @davetaz

Introductions



Your name

What is your favourite example/use of open data?

What do you want to do differently after the course?

Course aim

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

Schedule

Day 1: Practical publication

Day 2: Business, the law and open data

Day 3: Enriching and visualising data

Agenda - Today

The characteristics of data

Data discovery patterns

*** Lunch ***

Data publication platforms

Quick big data break

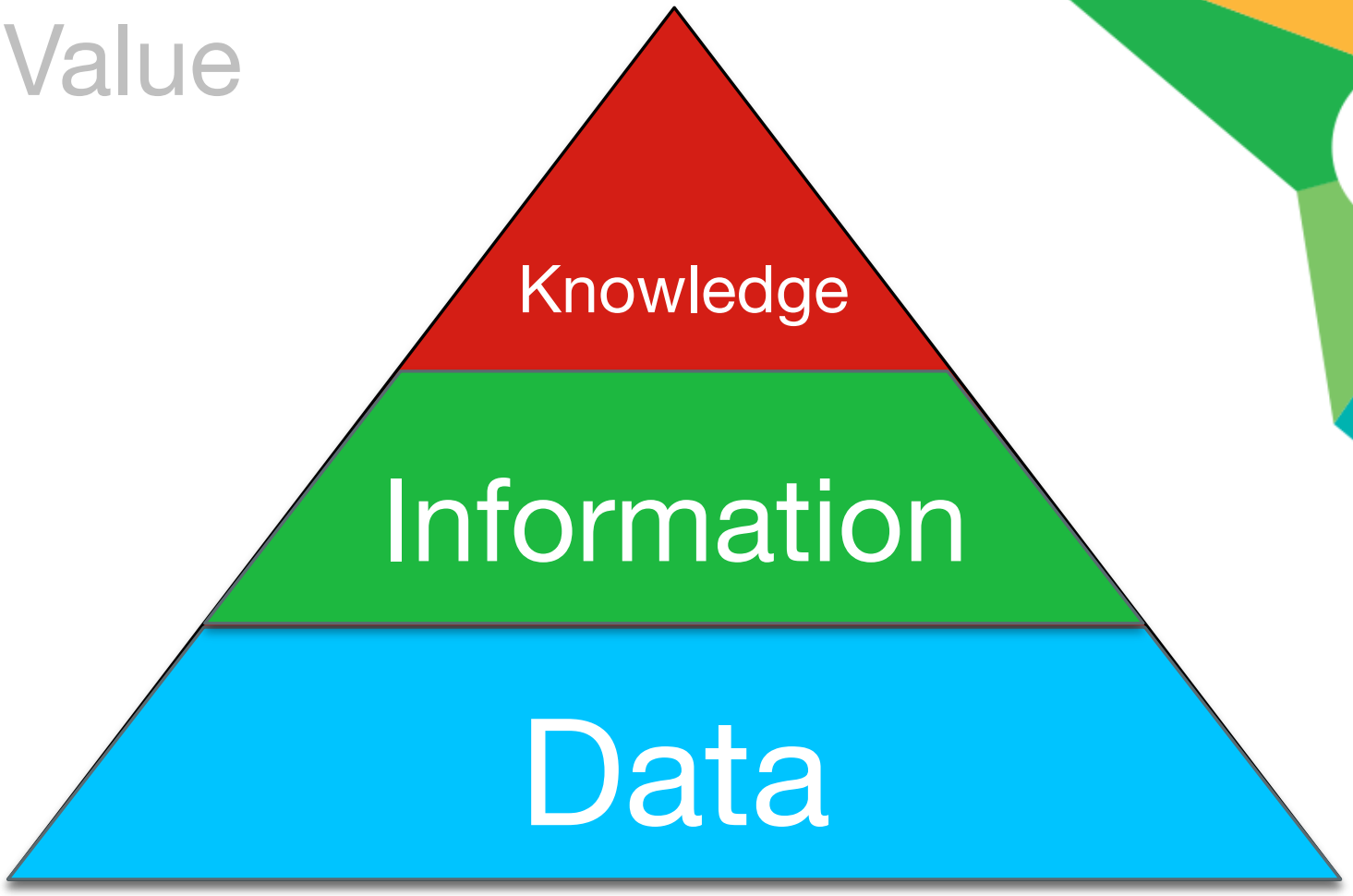
Practical publication hands-on

Recap session

Exercise

What is Data?

Value



Exercise

What is Open Data?

Option A

Open data is data that is made **available** by **organisations**, **businesses** and **individuals** for **anyone** to **access**, **use** and **share**.

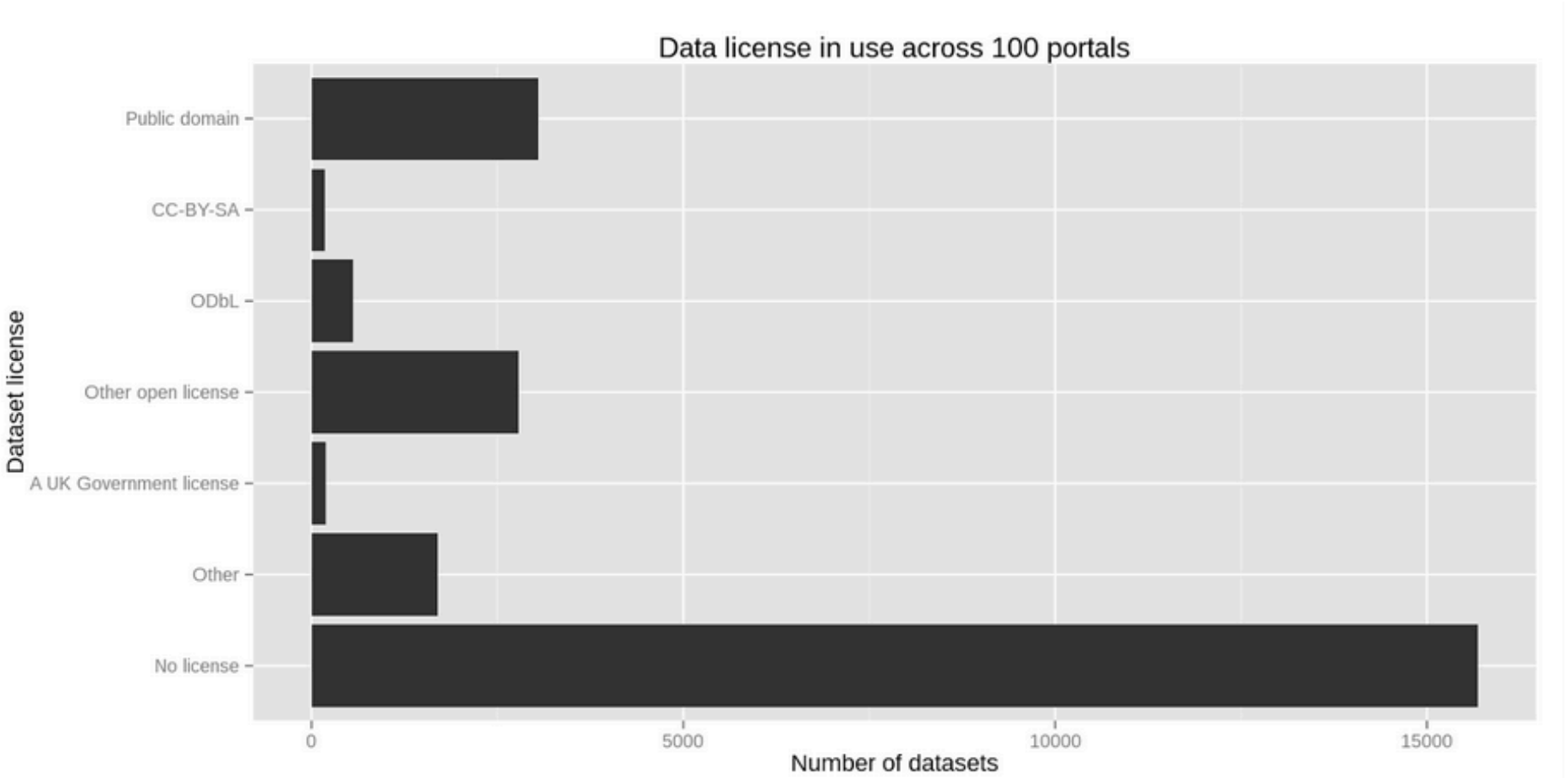
- Open Data Institute

Introduced November 2014


Open Definition Open Knowledge Foundation
OMB Memo, 2013 The White House Sylvia Burwell et al.
Data.Gov.UK Antonio Acuña
"DBpedia: A Nucleus for a Web of Open Data" Sören Auer et al.
Open Data Institute (ODI) Open Data Institute
LinkedGov LinkedGov
McKinsey James Manyika et al.
Open Data Now Joel Gurin
Open Data Barometer Tim Davies
The World Bank The World Bank
Free
Negligible Cost
Publicly Available
Re-usable
Can be Redistributed
Non-exclusive (No Restrictions from copyright, patents, etc.)
Structured for Usability
Requires "Open" License
Non Personally Identifiable
Produced during business operation
Belongs to the Taxpayer (when not in violation of laws/privacy)
Accessible in Bulk

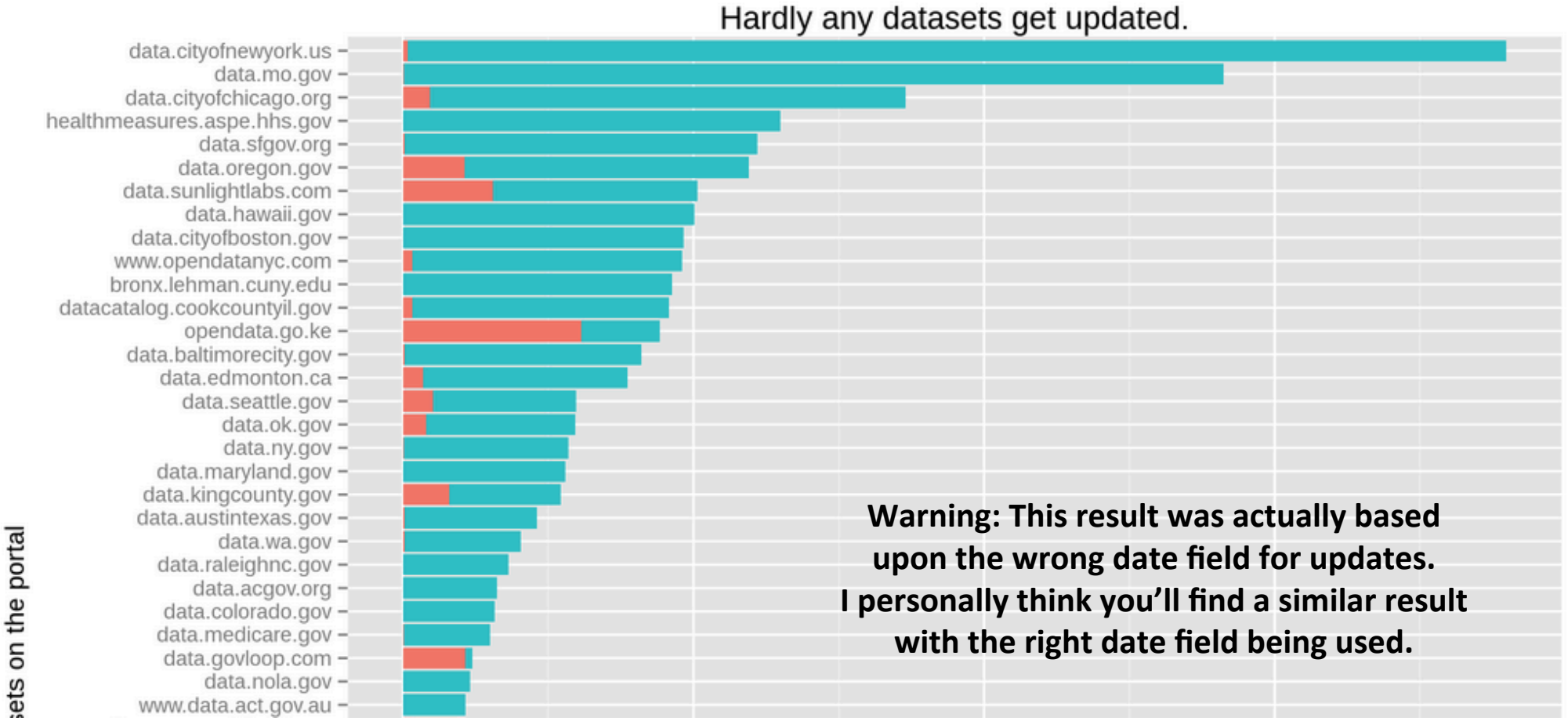
GOVLAB

Open data is hardly ever appropriately licensed.



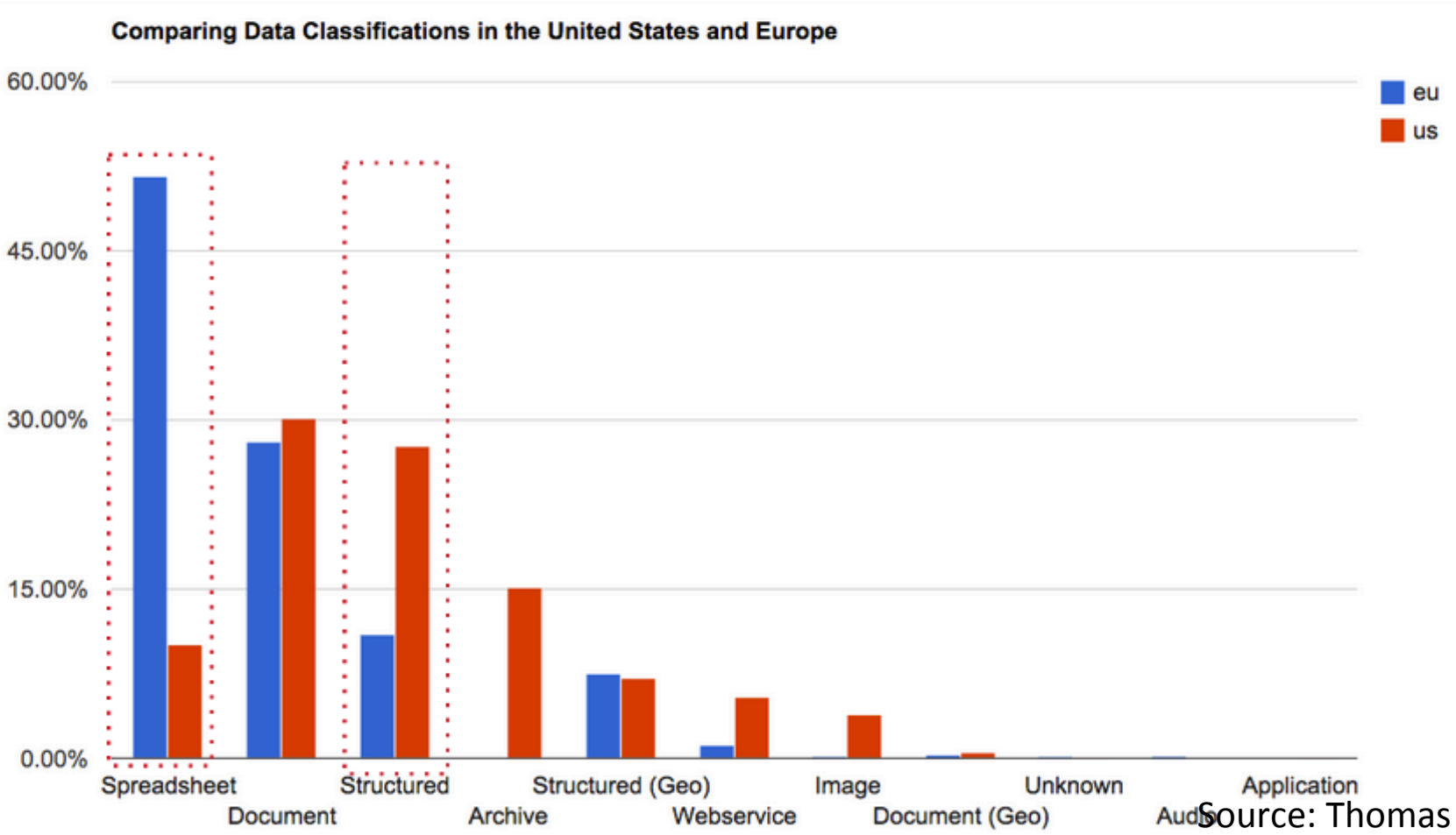
Source: Thomas Levine

Source: Thomas Levine



Source: Thomas Levine

Open data is rarely structured.



Source: Thomas Levine

Publication phases

Phase 1: Get the data online, in some form. This will help with the trust and transparency and community building.

Phase 2: Increase the usability of the data by potentially publishing differently and keeping it up to date.

Today's mission

To move to phase 2 of publishing open data and solve some of the phase 1 problems.

What best practice guidelines and tools will help us achieve this phase 2 goal?

Guidelines

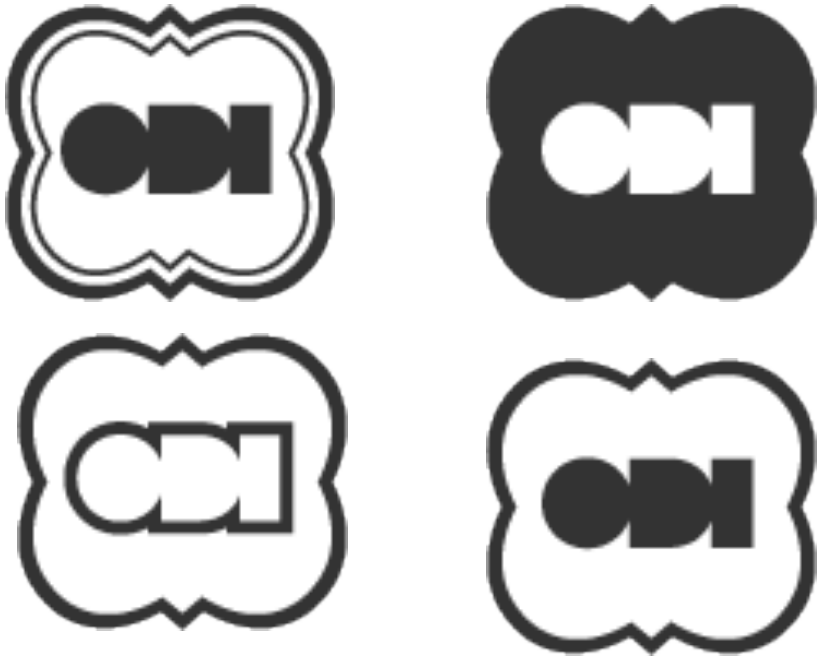
open data certificate™

5-Stars

5-Stars

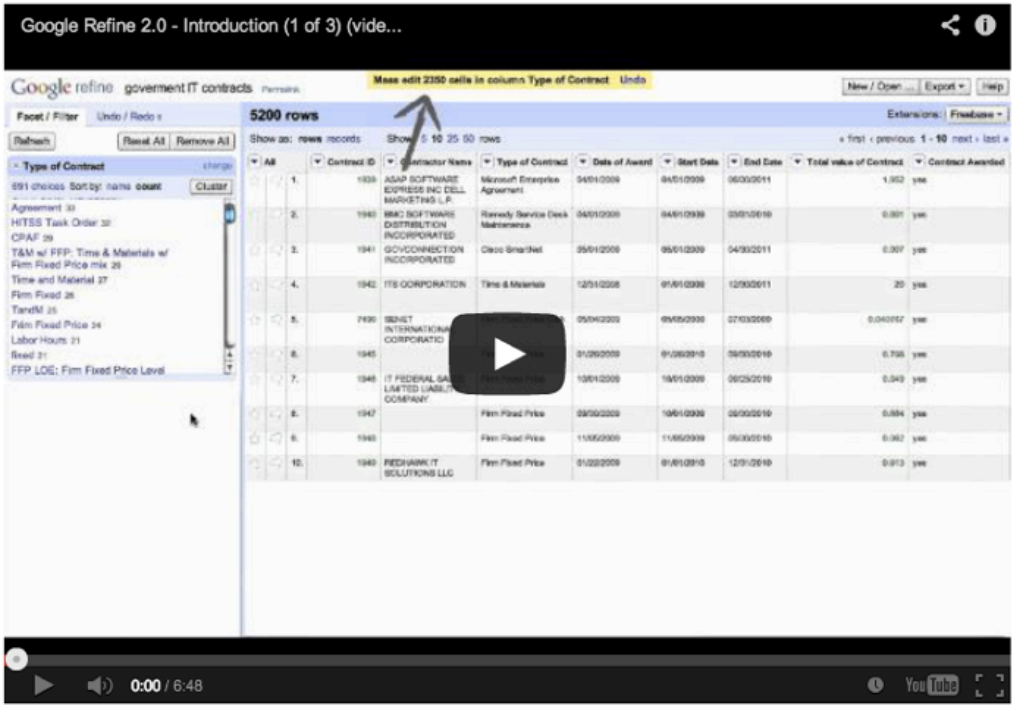
http://5stardata.info/

Open Data Certificate



http://certificates.theodi.org

Introducing Open Refine



http://openrefine.org

Session 1

The characteristics of 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)

In your pre-training exercise, you were all asked to identify a dataset.

In your groups briefly discuss each others datasets and write down some key characteristics of each.

Also write the dataset title on a post-it, one per post-it.

Types of Data

Reference data

“things”

Transaction data

“stats involving things”

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.

Types of Data

Reference data

“things”

People Facilities Places
Books Buildings

Transaction data

“stats involving things”

Expenditure
Weather Consumption
Observation

Update frequency

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.

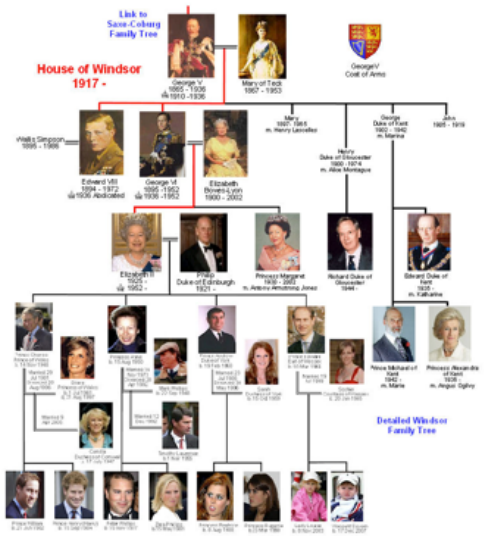
0 = static, 1= In frequent, 2 = Frequent, 3 = Live

Data Representations

Tabular

Region	Production			Y
Country Level 2	Production (thousand MT)	Change from last year	Change from 5 year average	Y (b
Brazil	57289	-4.05%	+2.68%	
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	-7.69%	-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?

Policy paper

G8 Open Data Charter and Technical Annex

Published 18 June 2013

Contents

Exercise

Pick one “group” of datasets that share similar colours and come up with a data publication strategy for getting these datasets online and usable.

What are the publication requirements on the human publisher?

What are the requirements on potential users?

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


Thank-you