Open Data: Where We Are, Where We're Going

European Data Forum, Copenhagen

Rufus Pollock
@rufuspollock[.org] - @okfn[.org]

Open Knowledge Foundation

SHUTTLEWORTH FELLOW

Open Knowledge Foundation

Community-based not-for-profit founded in 2004

We have **projects and collaborations around the world** - more than a **dozen local groups and chapters** and **15 working groups**

We build **tools, communities and applications** to **create, share and use** open data and knowledge - content and data that **everyone can use, share and build on**.

OKFestival = OGDcamp + OKCon. Helsinki, Finland. 17-22 Sept 2012.

We are delighted to invite you to the world's first Open Knowledge Festival: a week of participatory sessions, keynote lectures, workshops, hackathons and satellite events in Helsinki, organised by diverse communities from across the globe.

The 2012 theme of OKFestival is *Open Knowledge in Action*, looking at the *value* that can be generated by opening up knowledge, the ecosystems of organisations that can benefit from such sharing, and the impacts that transparency can have in our societies. What kinds of new professions, ideas and community initiatives can emerge within our governments, markets, networks and neighbourhoods as a result of these engagements?

The exploration of this theme will not only be visible in the festival's content, but also in its implementation as the first global event of its kind. This year, OKFestival will combine two popular annual events – the ***Open Government Data Camp*** and the ***Open Knowledge Conference***. This combination allows us to highlight a set of 13 diverse Topic Streams from open development to municipal data, all organised by global teams of Guest Programme Planners. With this collaborative format, we aim to highlight the diversity of open knowledge and data initiatives from around the world. We will bring together civil society representatives, programmers, data wranglers, designers, students, members of government, local communities and citizens for a week of building new things and sharing great ideas.

2 Stories

http://blog.okfn.org/2008/02/18/a-traffic-data-odyssey/

A Traffic Data Odyssey

February 18, 2008 in Exemplars, Open Government Data, Open/Closed Edit this entry

Recently, partly as an experiment regarding access to government data, partly out of genuine interest in the material itself, I looked into getting hold of some UK traffic count data — useful for, among other things, doing traffic analysis which is key to much road planning and policy (see e.g. this work by R J Gibbens and Y Saatchi at the University of Cambridge).

The results were rather disappointing and provide an interesting illustration of the kind of obstacles that can arise when trying to get access to Government data.

The Odyssey

From previous experience I knew count data was collected by UK's Department for Transport in the form of MIDAS (motorway incident detection and automatic signalling).

My journey then began with some simple searching which led me to here: . That page provided me with a clear link to "Traffic Count Data and Logs" (in nice bulk data form it appeared) but also informed me:

> The access of items marked with a padlock [the link to the data!] is restricted by username and password. If you don't have access to a username or password, contact the Mott MacDonald Helpdesk. Documents without a padlock icon are publicly available

The Request (and the Refusal)

Request for count data collected by UK's Department for Transport in the form of MIDAS (motorway incident detection and automatic signalling):

I'm a UK citizen interested in getting access to the Traffic Count Data and Logs dataset linked to from: http://www.midas-data.org.uk/

It appears that a username and password is required from yourselves in order to do this and so I wondered if you could therefore be kind enough to provide me with such a username and password.

The Conditions and the Refusal

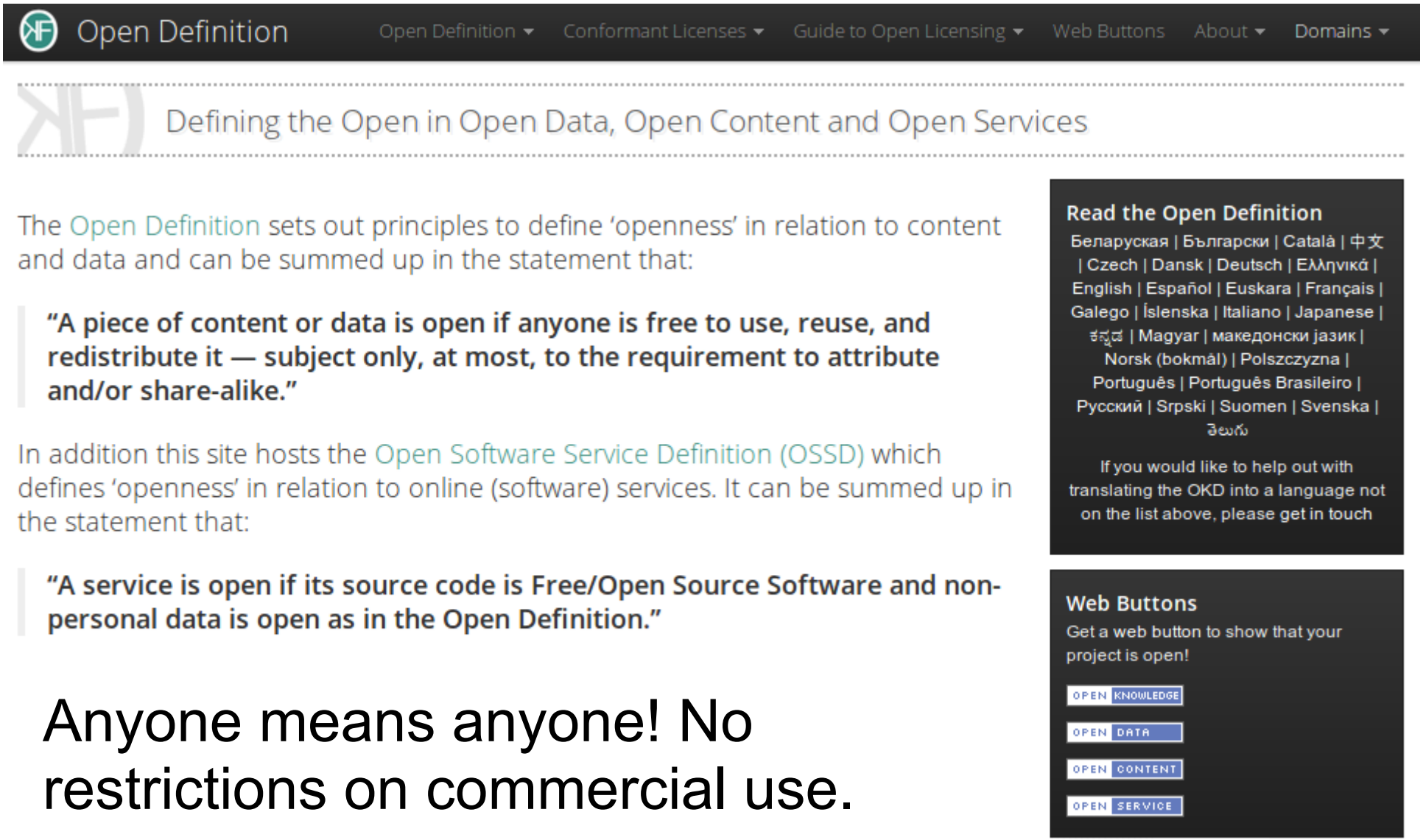
I need your **acceptance of the conditions stated below and some information regarding the research project you are undertaking before we allow you access to the data**. The conditions and information I have requested will allow the Group to **justify the costs associated with supplying this data** [what costs, it's already in a bzip file on a website?] and to **ensure the data is being used appropriately** [why is such paternalism needed?].

Note:- if the project is being undertaken jointly with **another organisation** then that organisation will **also be required to supply the information requested**. Please ensure **all grant and contract holders, staff and students** associated with the grant and project are **made aware of the conditions** contained within this letter.

Conditions

1. The data may not be copied to any other persons or organisations without the prior approval of the Highways Agency. The data may only be copied to another person or organisation after that person or organisation has confirmed with the HA the purpose for which the data is required and accepted the conditions laid down in this letter.
2. The data may not be used for any other purpose within your organisation without the prior written approval of the Highways agency.
3. **The data must not be sold or used for commercial gain.**
4. **The data will not be used to contradict or challenge any research project, works or statement made by the Government, the Department of Transport or the The Highways Agency as a result of analysis of the data by them or their agents.**
5. the Highways Agency will be provided, upon publication and free of charge, with: annual progress reports; any interim reports describing significant findings; a complete copy of the final report; and any technical papers resulting from the research.


OPEN DATA



Anyone means anyone! No restrictions on commercial use.

Open != Creative Commons. Many CC licenses NOT open (and most not appropriate for data).

HOUSEHOLD DATA
ANNUAL AVERAGES

1. Employment status of the civilian noninstitutional population, 1940 to date

(Numbers in thousands)

Year	Civilian noninstitutional population	Civilian labor force							
		Total	Percent of population	Employed				Unemployed	
				Total	Percent of population	Agriculture	Nonagricultural industries	Number	Percent of labor force
Persons 14 years of age and over									
1940	99,840	55,640	55.7	47,520	47.6	9,540	37,980	8,120	14.6
1941	99,900	55,910	56.0	50,350	50.4	9,100	41,250	5,560	9.9
1942	98,640	56,410	57.2	53,750	54.5	9,250	44,500	2,660	4.7
1943	94,640	55,540	58.7	54,470	57.6	9,080	45,390	1,070	1.9
1944	93,220	54,630	58.6	53,960	57.9	8,950	45,010	670	1.2
1945	94,090	53,860	57.2	52,820	56.1	8,580	44,240	1,040	1.9
1946	103,070	57,520	55.8	55,250	53.6	8,320	46,930	2,270	3.9
1947	106,018	60,168	56.8	57,812	54.5	8,256	49,557	2,356	3.9
Persons 16 years of age and over									
1947	101,827	59,350	58.3	57,038	56.0	7,890	49,148	2,311	3.9
1948	103,068	60,621	58.8	58,343	56.6	7,629	50,714	2,276	3.8
1949	103,994	61,286	58.9	57,651	55.4	7,658	49,993	3,637	5.9
1950	104,995	62,208	59.2	58,918	56.1	7,160	51,758	3,288	5.3
1951	104,621	62,017	59.2	59,961	57.3	6,726	53,235	2,055	3.3
1952	105,231	62,138	59.0	60,250	57.3	6,500	53,749	1,883	3.0
1953 (1)	107,056	63,015	58.9	61,179	57.1	6,260	54,919	1,834	2.9
1954	108,321	63,643	58.8	60,109	55.5	6,205	53,904	3,532	5.5
1955	109,683	65,023	59.3	62,170	56.7	6,450	55,722	2,852	4.4
1956	110,954	66,552	60.0	63,799	57.5	6,283	57,514	2,750	4.1
1957	112,265	66,929	59.6	64,071	57.1	5,947	58,123	2,859	4.3
1958	113,727	67,639	59.5	63,036	55.4	5,586	57,450	4,602	6.8
1959	115,329	68,369	59.3	64,630	56.0	5,565	59,065	3,740	5.5

```
def get_table_index():
    reader = econ.data.tabular.XlsReader()
    tabdata = reader.read(file(all_fn))
    data = [ row[0] for row in tabdata.data ]
    table_names = filter(lambda x: x.startswith('Table '), data)
    return table_names

class SheetParser(object):
    def get_sheet(self, index):
        reader = econ.data.tabular.XlsReader()
        tabdata = reader.read(file(all_fn), index)
        return tabdata.data

    def format_line(self, line):
        year = line[0]
        year = year.split('/')[0]
        year = int(year)
        def clean(value):
            if value == '--':
                return ''
            else:
                return econ.data.misc.floatify(value)
        out = [year] + [ clean(value) for value in line[1:] ]
        return out

    def extract_table_1(self):
        data = self.get_sheet(1)
        headings = ['Market Year', 'Planted acreage (millions)',
            'Harvested acreage (millions)', 'Production (millions of bushels)',
            'Yield (bushels per acre)', 'Weighted-average farm price ($ per bushel)'
            ]
        # remove headings and footnotes
        data = data[3:-3]
        # break into sections based on blank lines
        is_blank = lambda x: data[x][1] == ''
        blank_rows = filter(is_blank, range(len(data)))
        # put in start item
        blank_rows = [-1] + blank_rows
        sections = [ data[blank_rows[ii]+1:blank_rows[ii+1]] for ii in
```

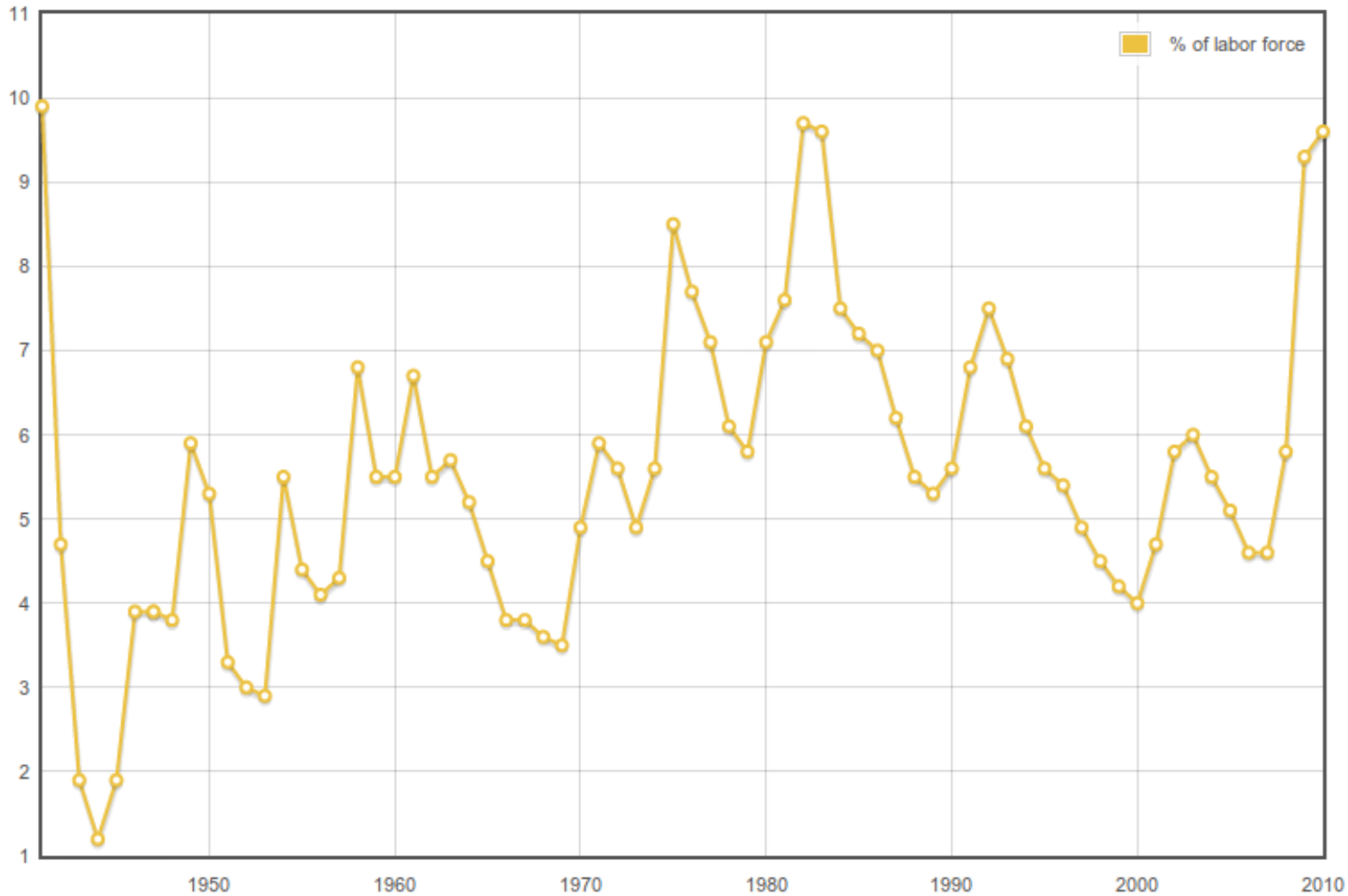
Recline Data Explorer

Load Data | Share and Embed | Settings

Grid | Graph | Map | Timeline

Results found 71 « 0 – 100 »

Search data ... Go » Filters Fields



Graph Type

Lines and Points

Group Column (x-axis)

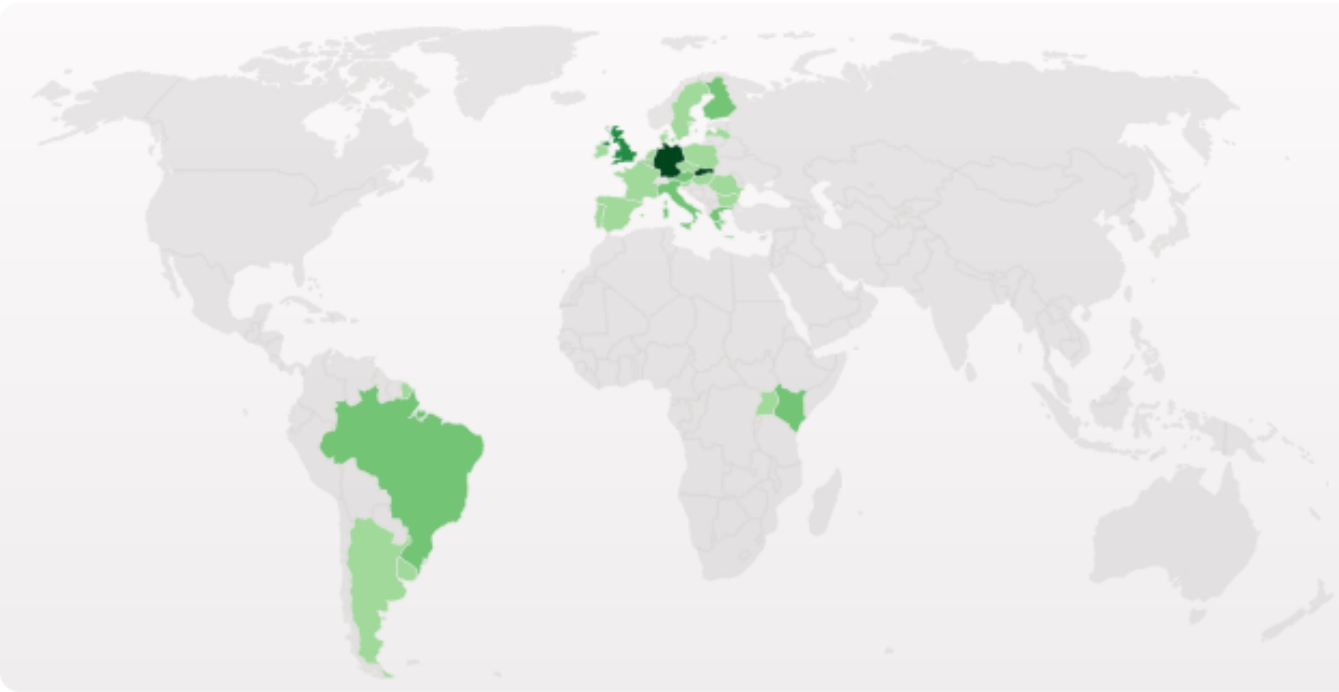
Year

Series A (y-axis) [Remove]

% of labor force

Add Series

Do you want to help map government finance from all around the globe? Get involved in Open Spending!



Our aim is to track every government and corporate financial transaction across the world and present it in useful and engaging forms for everyone from a school-child to a data geek.

Explore

OpenSpending presents financial information **from many different countries** and sources. Choose a dataset to explore or visit our Spending Blog, where we highlight spending stories and methods of analyzing the information.

Investigate

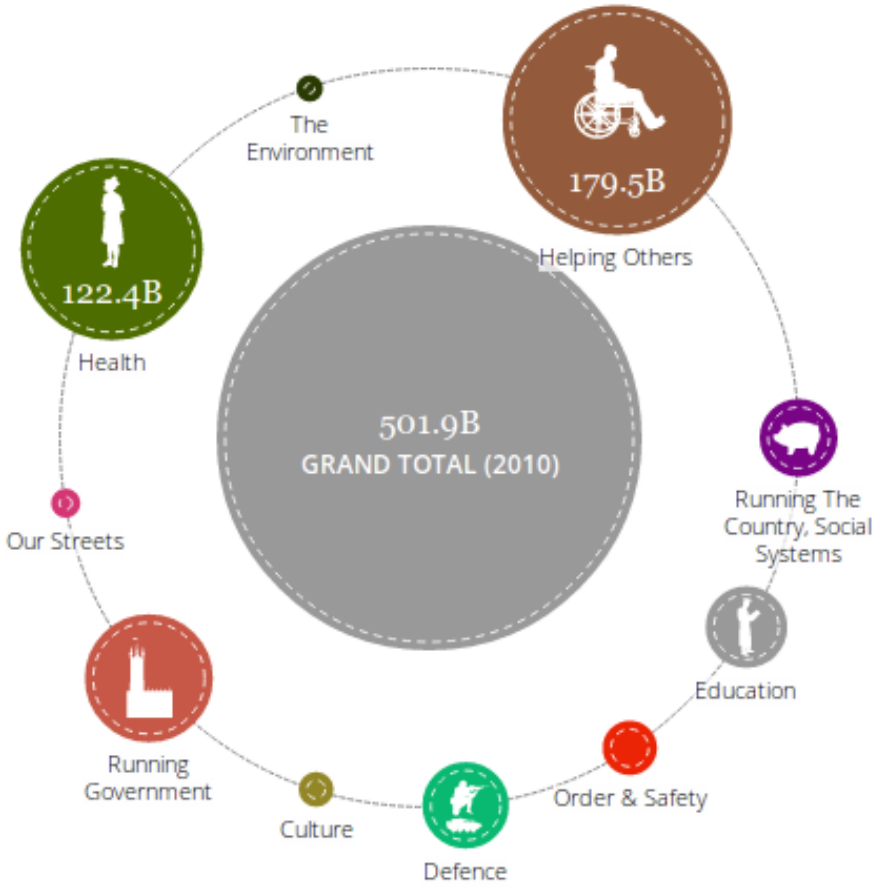
Do you have spending questions? Ask and help to answer them on our discussion list. When you've found interesting information in the data, become a contributor to the Spending Blog to share your findings.

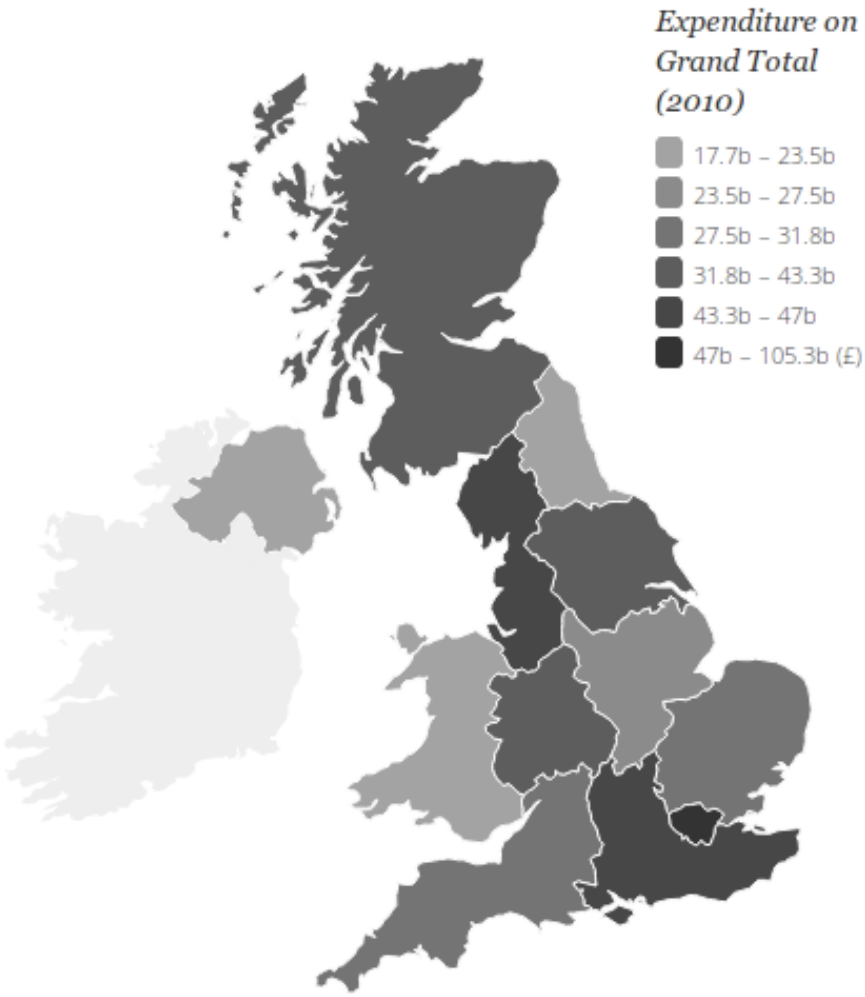
Contribute

Is your country, state or city not on OpenSpending yet? We're looking for spending data from all over the world. Import a dataset from your country or local community to visualize and explore government expenditure.

WHERE DOES MY MONEY GO?

Showing you where your taxes get spent





Where Does My Money Go? is based on OpenSpending, where you can also find the source data

Machine Readable Bulk Data

APIs are not enough!

CKAN, used by governments and user groups worldwide

Used to power both official and community data portals, CKAN was developed by the non-profit **Open Knowledge Foundation** to run **TheDatahub.org**. It now powers more than 40 data hubs around the world, including portals for local, national and international government, such as the UK's **data.gov.uk** and the European Union's **publicdata.eu**.

Take Tour | Pricing | Brochure

Feature Overview

- Complete catalog system with easy to use web interface and a powerful API
- Strong integration with third-party CMS's like Drupal and WordPress
- Data visualization and analytics
- Workflow support including moderated editing, full change history
- Fine-grained access control
- Integrated data storage
- Federated structure: easily set up new instances with common search

Read More

Support and Hosted Solutions

CKAN ensures that users have complete freedom both with regard to supplier and hosting but also customization and extension of their solution.

Managed by a full-time professional development team, CKAN has full support available, as well as hosted solutions with SLA.

Read More

Powered by CKAN

Ministerial gifts, hospitality, travel and meetings with external organisations in Cabinet Office / May-July 2010 - Rt Hon David Cameron MP meeting in CSV Forma...

May-July 2010 - Rt Hon David Cameron MP meeting in CSV Format

Part of dataset	Ministerial gifts, hospitality, travel and meetings with external organisations in Cabinet Office
Last updated	Unknown
Format	CSV
Licence	UK Open Government Licence (OGL) OPEN DATA

Download

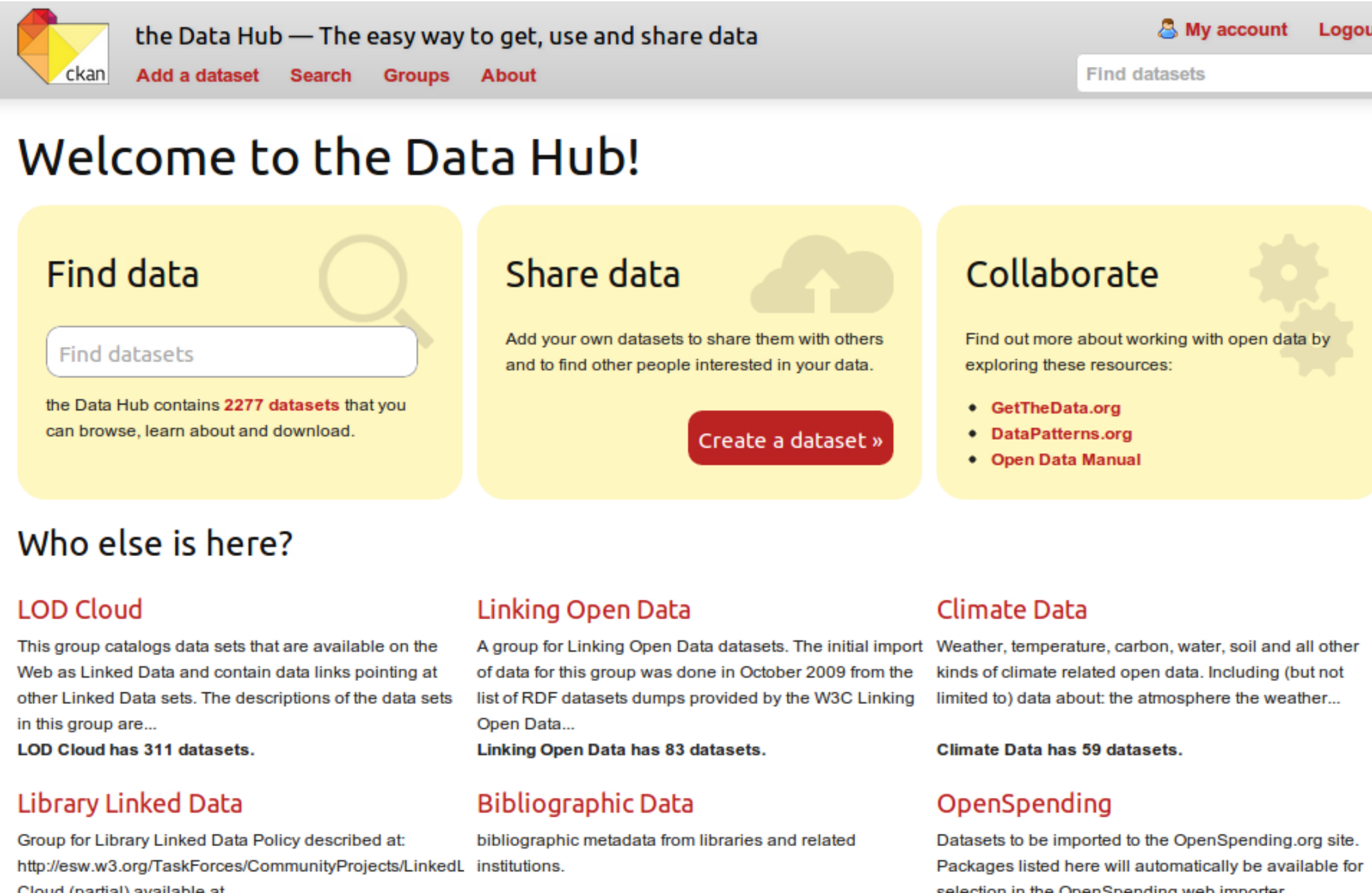
Preview

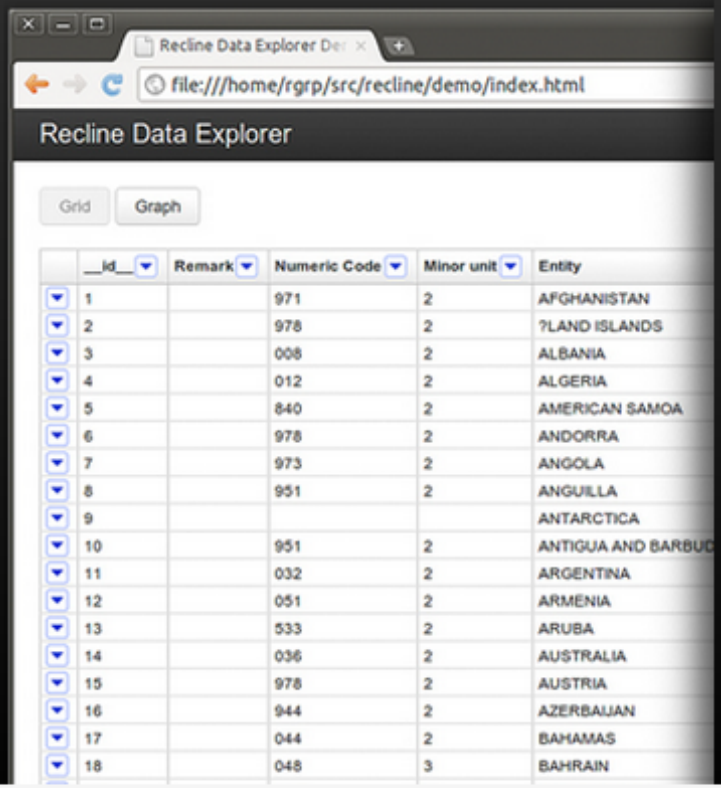
Grid Graph

Showing 0 to 50 of Unknown

Minister	Date	Name of External Organisation	Purpose of meeting
Prime Minister, The Rt Hon David Cameron MP	May, 2010	Young Foundation, Community Links, Antigone, Big Society Network, Balsall Health Forum, London Citizens, Participle, Talk About Local, CAN Breakthrough, Mayor of Middlesborough, Business in the Community, Esmee Fairbairn, Greener Leith, St Giles Trust, Big Issue Invest, Kids Company	To discuss Big Society
	May, 2010	Rupert Murdoch	General meeting
	June, 2010	World Bank	To discuss business issues
	June, 2010	Bob Geldof	To discuss development issues
	June, 2010	Ratan Tata	To discuss business issues

DataHub.io Community DataHub powered by CKAN





Recline.js

relax with your data

A. Powerful data explorer built in pure javascript and html
B. A library of data components - grid, graphing and data connectors

— All built with Backbone

Use the Explorer Use the Library

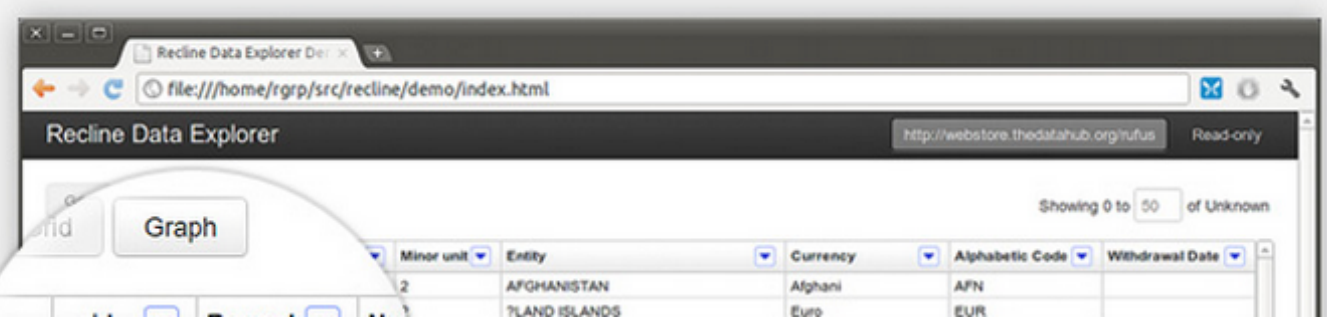
Recline is Two Things

- A Data Explorer combining a data grid, Google Refine-style data transforms and visualizations all in lightweight javascript and html.
- A simple but powerful library of extensible of data components - data grid, graphing, and data connectors - which you can selectively use and build on.

The Explorer can be used standalone (just download and use) or can be embedded into your own site. Recline builds on the powerful but lightweight Backbone framework making it extremely easy to extend and adapt. The library's modular design mean means you only have to take what you need.

Main Features

- View and edit your data in a clean grid / table interface
- Bulk update/clean your data using an easy scripting UI
- Easily extensible with new Backends so you can connect to your database or storage layer
- Visualize data



rgrp / ISO 3166 2 Letter Country Codes

python • public scraper

Fork this scraper | Download SQLite3 database | Explore with ScraperWiki API

Download spreadsheet (CSV)

swdata (249 rows)

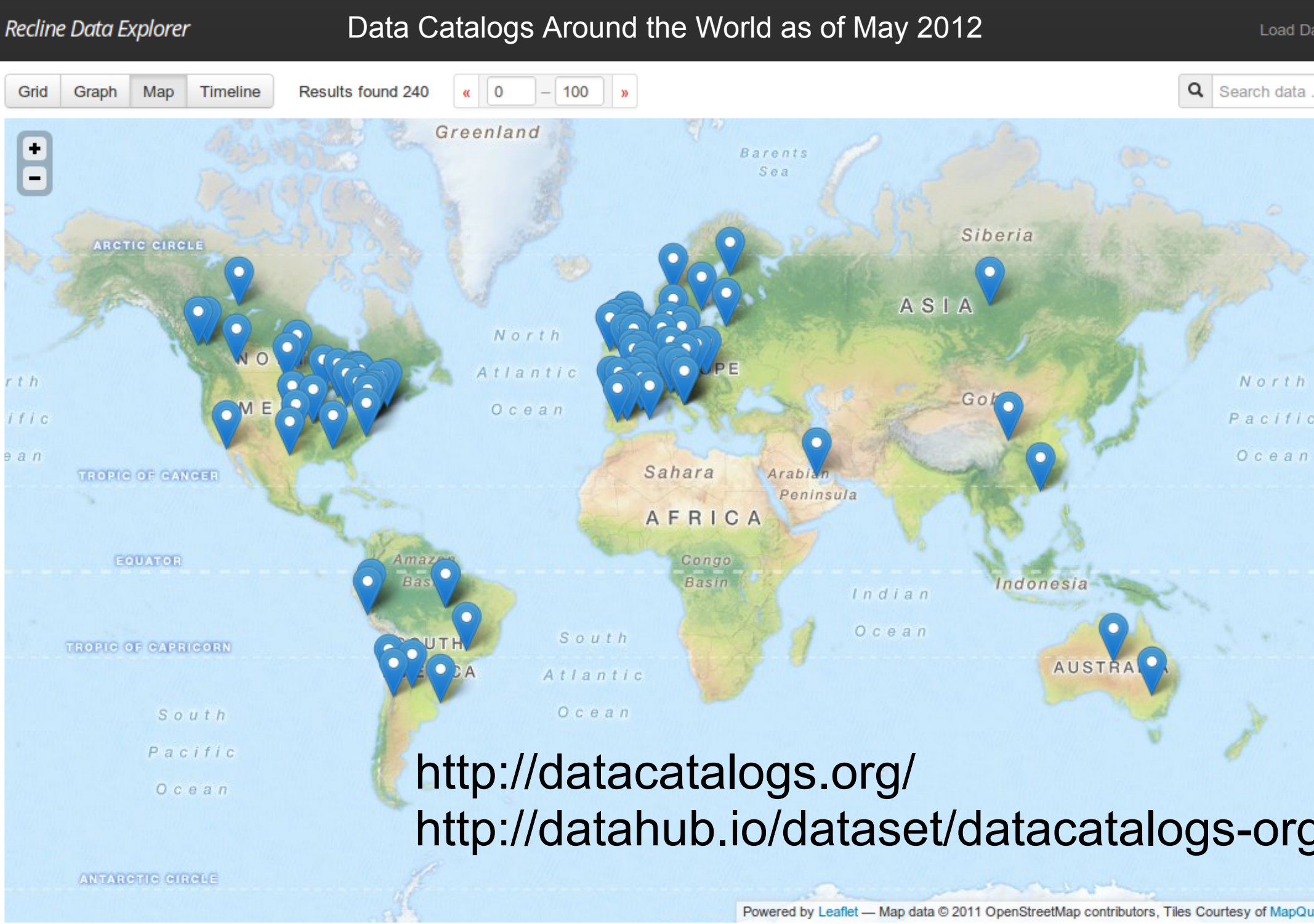
date_scraped	code	name
	ZW	Zimbabwe
	ZM	Zambia
	YE	Yemen
	EH	Western Sahara
	WF	Wallis And Futuna
	VI	Virgin Islands, U.S.
	VG	Virgin Islands, British
	VN	Viet Nam
	VE	Venezuela, Bolivarian Republic Of
	see	Vatican City State

This dataset has a total of 249 records in 1 table (show schema).

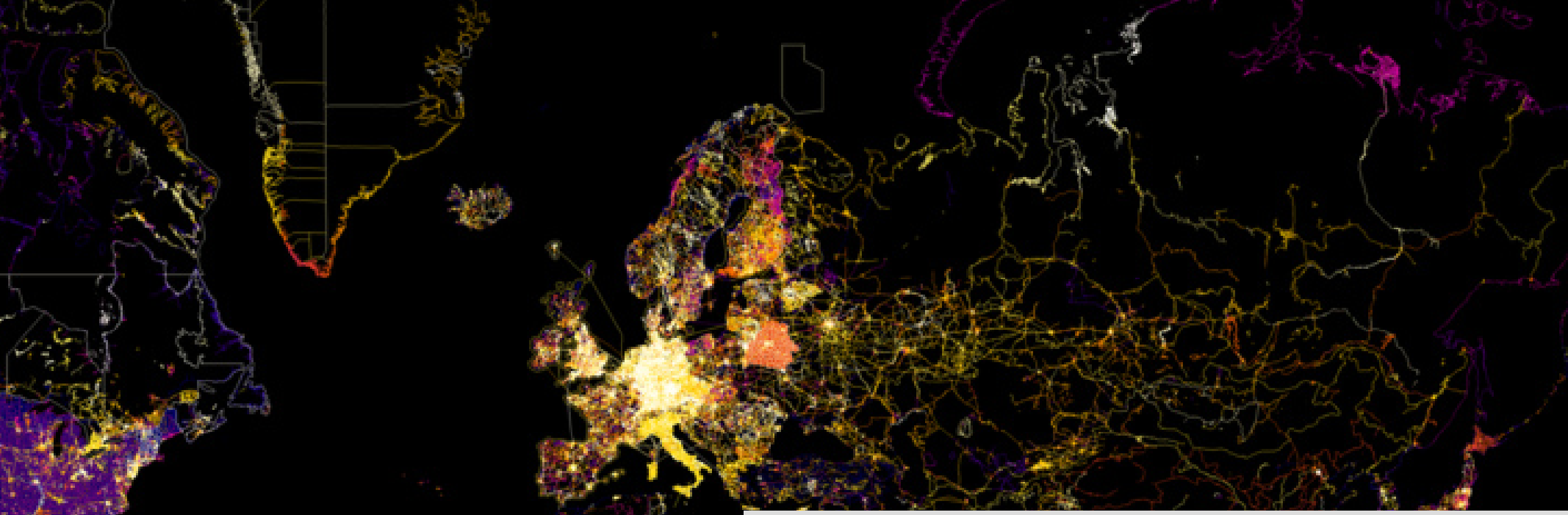
Where We Are

Huge Growth in Last Few Years

Especially for Government Data


Recline Data Explorer
Data Catalogs Around the World as of May 2012
Load Da
Grid
Graph
Map
Timeline
Results found 240
«
0
–
100
»
Search data
Greenland
Barents Sea
Siberia
ARCTIC CIRCLE
ASIA
North Atlantic Ocean
North Pacific Ocean
TROPIC OF CANCER
Sahara
Arabian Peninsula
AFRICA
EQUATOR
Congo Basin
Indian Ocean
Indonesia
TROPIC OF CAPRICORN
South Atlantic Ocean
South Pacific Ocean
ANTARCTIC CIRCLE
http://datacatalogs.org/
http://datahub.io/dataset/datacatalogs-org
Powered by Leaflet — Map data © 2011 OpenStreetMap contributors, Tiles Courtesy of MapQu


Controladoria-Geral da União
BRASIL
Open Government Partnership
Parceria para Governo Aberto
Conferência Anual
Annual Meeting



foursquare BLOG

Home Eng Blog Support Jobs

Feb 29th

foursquare is joining the OpenStreetMap movement! Say hi to pretty new maps!

We usually use this blog for big product announcements, but, as a startup also often think about how we can make life easier for other startups. Tod we're doing both – a little announcement, and hopefully some help for oth startups that are thinking about the same things.

So, the announcement:

Starting today, we're embracing the OpenStreetMap movement, so all the you see when you go to foursquare.com will look a tiny bit different (we th new ones are really pretty). Other than slightly different colors and buttons though, foursquare is still the same site you know and love.

Where Next?

Toy vs Core Datasets

Location of park benches vs National Map

Open Data Census - Response Summary

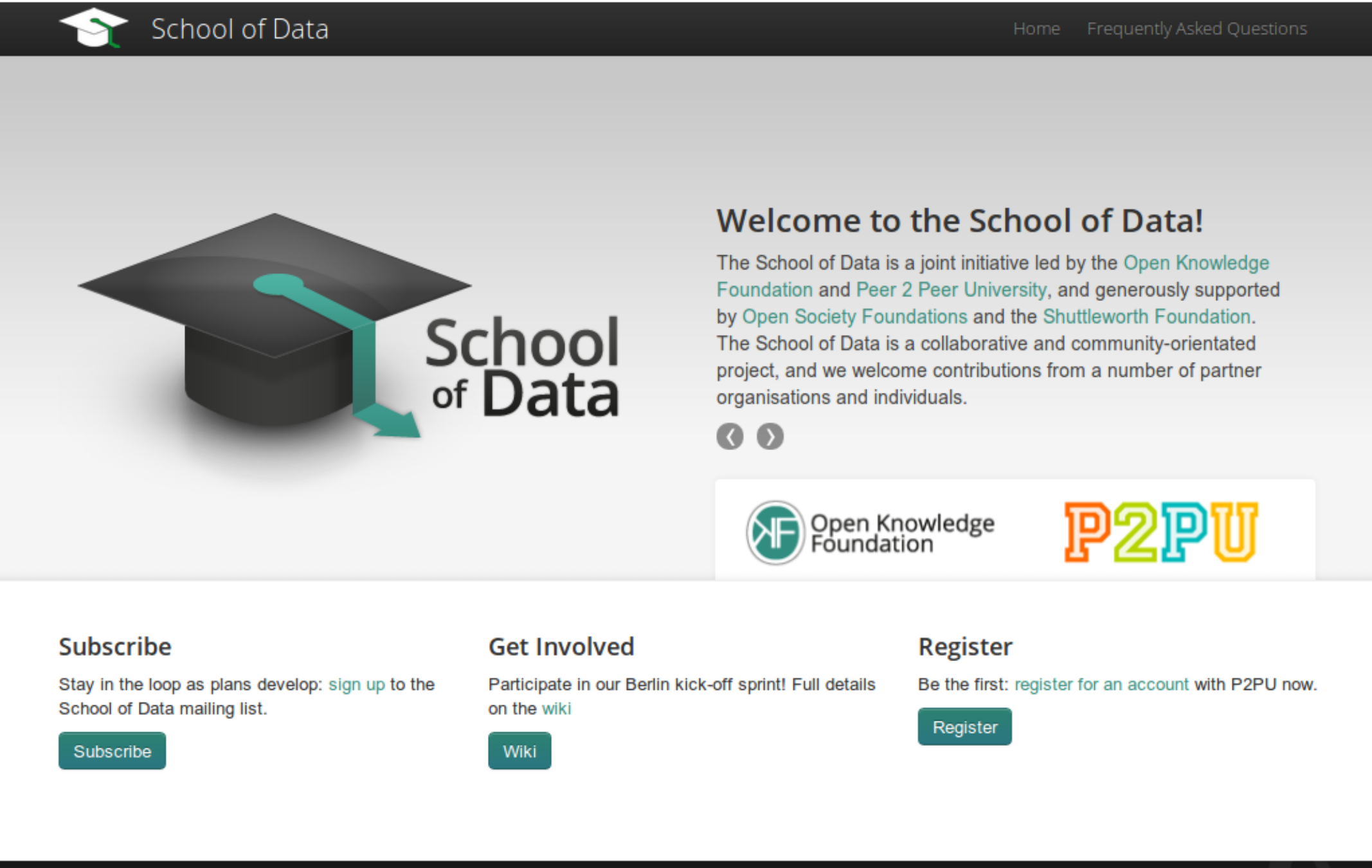
Total 133 | Map | Table | Source Data | Census Home | Code for this App

	Election Results (national)	Company Register	National Map (Low resolution: 1:250,000 or better)	Government Budget (National, high level, not detailed)	Government Spending (National, transactional level data)	Legislation (laws and statutes) - National	National Statistical Data (economic and demographic information)	National Postcode/ZIP database	Public Transport Timetables	Environmental Data on major sources of pollutants (e.g. location, emissions)
United Kingdom	YYYYY?	YYYNNN	YYYYY?	YYYYYY OPEN DATA	YYYYYY OPEN DATA	YYYNYY	YYYYYY OPEN DATA	YYYYYY OPEN DATA	No info	YYYY??
Brazil	YYNNYN	No info	No info	YYYY?Y	No info	No info	No info	No info	No info	No info
Australia	YYYYYY OPEN DATA	YYNNNN	YYYYYY OPEN DATA	YYNNYY	YYYNYN	YYNNY?	YYYYYY OPEN DATA	YYYYYY OPEN DATA	YYYYYY OPEN DATA	YYYYYN
Netherlands	YYYYYY OPEN DATA	YYNNNN	No info	No info	No info	No info	No info	YYYYYN	No info	No info
Iceland	YYNNYN	YYYNYN	YYYNNN	YYYNYN	YYYNNN	YYYYYN	YYY?Y?	YYYYYN	No info	No info
Denmark	YYYYY?	No info	No info	YYYYY?	No info	YYYYY?	YYYYY?	YYYYY?	No info	No info
Czech Republic	YYYNYN	YYYNYN	YYYYNN	NNNNNN	YYYNYN	YYYN?N	YYYNYN	YYYNYN	YYYNNN	No info
Norway	No info	No info	YYYYYN	No info	No info	No info	YYYYY?	No info	No info	YYYYYY OPEN DATA
Croatia	YYNNYN	YYNNYN	No info	YYYYYN	No info	No info	YYNNYN	No info	No info	No info
Greece	YYNNYN	No info	No info	No info	No info	No info	YYYYYY OPEN DATA	No info	No info	No info

http://census.opengovernmentdata.org/

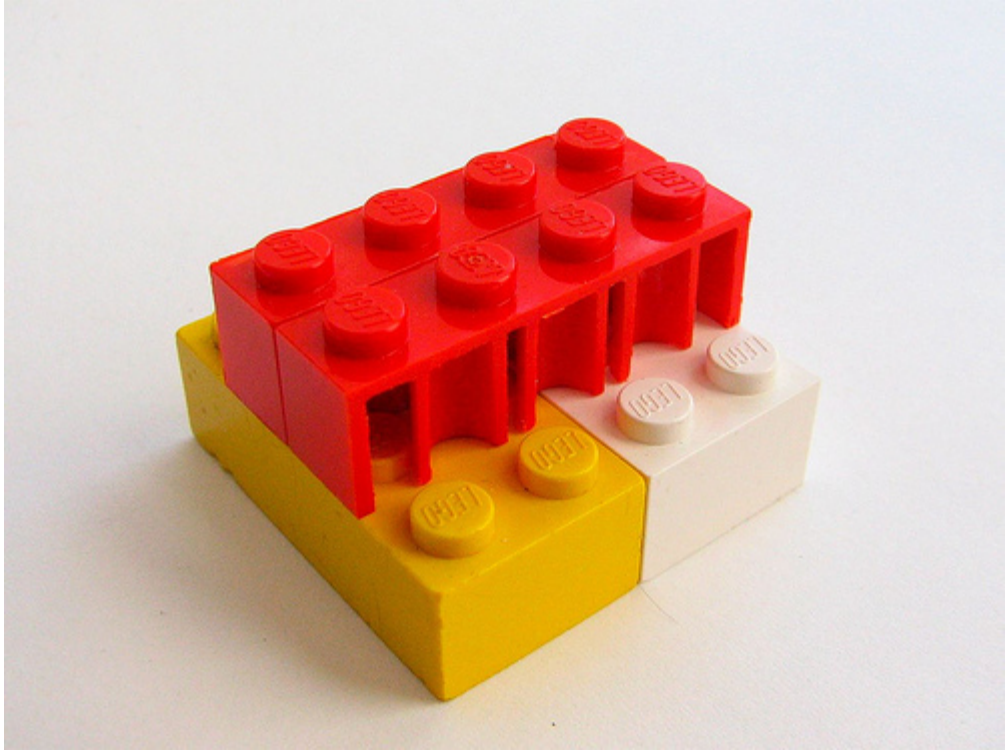
Machine Readable

Education and Skills



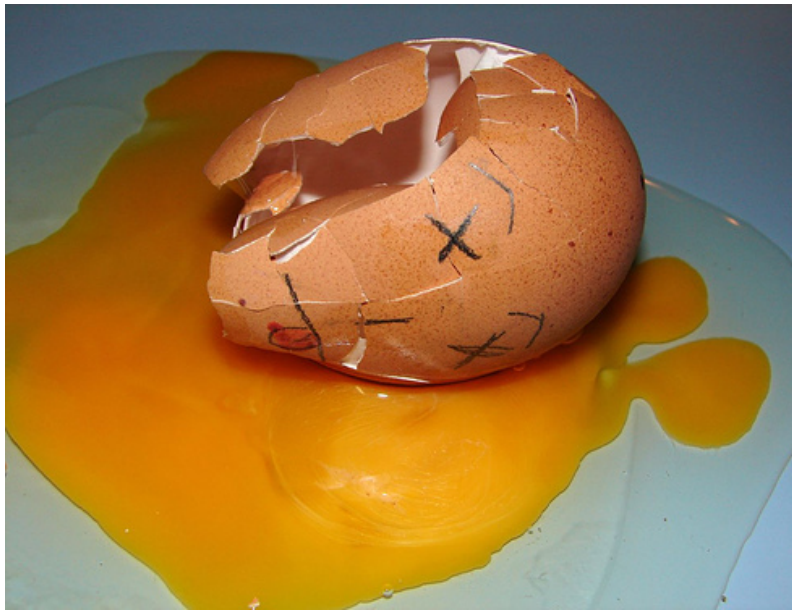
Scale
(how do we?)

We Compentize to Scale



We Want and Need to Integrate

Without Open Data this will Fail!



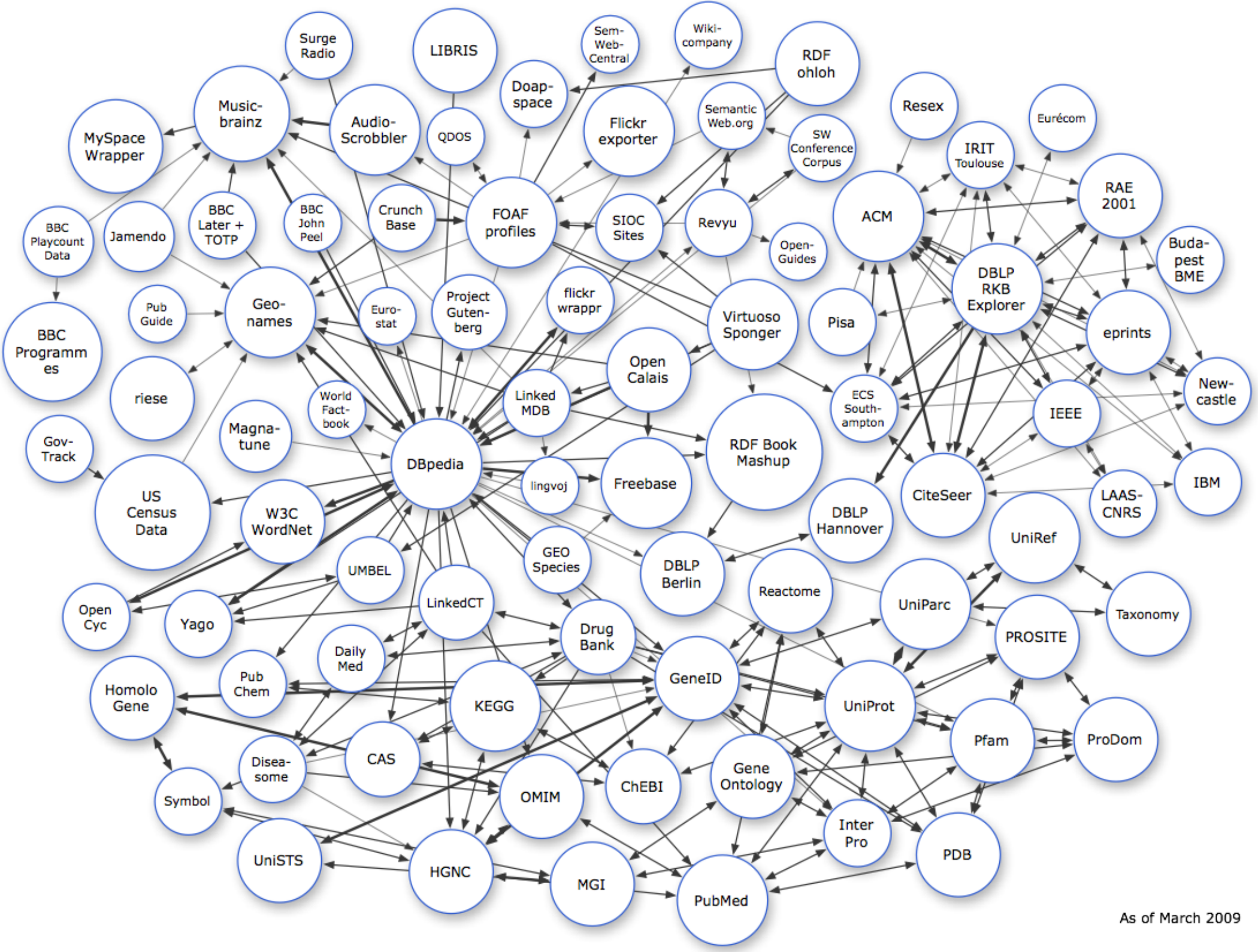
Small Data

vs

Big Data

It's about small pieces loosely joined not one ring to rule them all!

Machine Integrable (?)


Surge Radio
LIBRIS
Sem-Web-Central
Wiki-company
RDF ohloh
Doap-space
Music-brainz
Audio-Scrobbler
QDOS
Flickr exporter
Semantic Web.org
SW Conference Corpus
Resex
Eurécom
IRIT Toulouse
MySpace Wrapper
BBC Playcount Data
Jamendo
BBC Later + TOTP
BBC John Peel
Crunch Base
FOAF profiles
SIOC Sites
Revyu
Open-Guides
ACM
RAE 2001
Buda-pest BME
DBLP RKB Explorer
Pisa
eprints
Pub Guide
Geo-names
Euro-stat
Project Guten-berg
flickr wrappr
Virtuoso Sponger
BBC Programmes
riese
World Fact-book
Linked MDB
Open Calais
ECS South-ampton
IEEE
New-castle
Gov-Track
Magna-tune
DBpedia
RDF Book Mashup
Freebase
lingvoj
CiteSeer
LAAS-CNRS
IBM
US Census Data
W3C WordNet
DBLP Hannover
UniRef
UMBEL
GEO Species
DBLP Berlin
Reactome
UniParc
Taxonomy
Open Cyc
Yago
LinkedCT
Drug Bank
PROSITE
Daily Med
Pub Chem
GeneID
Homolo Gene
KEGG
UniProt
Pfam
ProDom
Disea-some
CAS
Gene Ontology
Symbol
OMIM
ChEBI
Inter Pro
UniSTS
HGNC
MGI
PubMed
PDB
As of March 2009

Keep it Simple ...

Data Protocols

Search

Data Protocols

Simple data protocols (and pointers to existing ones) for doing collaborative, distributed development of data.

Things like:

- Revisioning of databases and datasets including diffing and merging
- Protocols for sharing and syncing data changes
- Web-oriented query protocols for data
- Data packaging, publication and installation
- Webhooks and webservices for data transformation

Table of Contents

Contribute

Contributions, comments and corrections are warmly welcome
They can be submitted via one o
following routes:

Contents

DERI Vocabularies

DERI Vocabularies is a URI space for RDF Schema vocabularies and OWL ontologies maintained at DERI, the Digital Enterprise Research Institute at NUI Galway, Ireland. The site is operated by DER
Linked Data Research Centre. *Read more...*

Schema.org CSV

Author: Michael Hausenblas

Mapping vocabulary for CSV files that use Schema.org terms in the header.

Namespace URI: http://purl.org/NET/schema-org-csv#
Terms: 3 Classes, 2 Properties

ADMS API Vocabulary

Author: Gofran Shukair

Namespace URI: http://vocab.deri.ie/admsapi#
Terms: 1 Class, 3 Properties

Simple Data Format (SDF)

This document defines a simple data publishing format (Simple Data Format) for publishing and sharing data.

Status: Draft

Contribute

Comments, suggestions and discussion welcome - see sidebar for various options on how to contribute including mailing list, twitter and issue tracker.

Key Design Features and Principles

The format's focus is on simplicity and web usage – that is, usage online with access and transmission *over HTTP*. In addition the format is focused on data that can be presented in a tabular structure and in making it easy to produce (and consume) this format from spreadsheets and relational databases.

The key features of this format are the following:

- CSV (comma separated variables) as the base data format
- JSON (with CSV alternative) as the base format for schema definition
- JSON (with CSV alternative) as the base format for metadata definition
- Usage of linked data / semantic web attributes for schema definition via the JSON-LD standard
- Support for normalization (i.e. splitting of data into multiple CSV file tables and definition of links between files)

Contribute

Contributions, comments and corrections are warmly welcome
They can be submitted via one of following routes:

1. A patch to the git repo (fork
   pull recommended) – best fo
   textual corrections and
2. The mailing list – best fo

MicroSchemas

A Set of **Simple** Cross-Country Data Formats

Transport (time-tables), Locations, Spending etc

Conclusion

Open Data is Platform not a Commodity

Let's Build on It, Not Sell It!

Increasing Amounts of Data

Be Problem and Application Driven

(Rather than Data and Technology Driven)

Thank-you!

Rufus Pollock
@rufuspollock[.org] - @okfn[.org]