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

European Data Forum, Copenhagen

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

Licensed under cc-by v3.0 all jurisdictions
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
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 and to ensure the data is being used appropriately.

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.
The **Open Definition** sets out principles to define ‘openness’ in relation to content and data and can be summed up in the statement that:

“A piece of content or data is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike.”

In addition this site hosts the **Open Software Service Definition (OSSD)** which defines ‘openness’ in relation to online (software) services. It can be summed up in the statement that:

“A service is open if its source code is Free/Open Source Software and non-personal data is open as in the Open Definition.”

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)

<table>
<thead>
<tr>
<th>Year</th>
<th>Civilian noninstitutional population</th>
<th>Total</th>
<th>Percent of population</th>
<th>Employed</th>
<th>Percent of population</th>
<th>Agriculture</th>
<th>Nonagricultural industries</th>
<th>Number</th>
<th>Percent of labor force</th>
</tr>
</thead>
<tbody>
<tr>
<td>1940</td>
<td>99,980</td>
<td>55,640</td>
<td>55.7</td>
<td>47,520</td>
<td>47.6</td>
<td>9,540</td>
<td>37,980</td>
<td>8,120</td>
<td>14.6</td>
</tr>
<tr>
<td>1941</td>
<td>99,900</td>
<td>55,910</td>
<td>56.0</td>
<td>50,250</td>
<td>50.4</td>
<td>9,100</td>
<td>41,250</td>
<td>5,560</td>
<td>9.9</td>
</tr>
<tr>
<td>1942</td>
<td>98,640</td>
<td>56,410</td>
<td>57.2</td>
<td>53,750</td>
<td>54.5</td>
<td>9,250</td>
<td>44,500</td>
<td>2,660</td>
<td>4.7</td>
</tr>
<tr>
<td>1943</td>
<td>94,640</td>
<td>55,540</td>
<td>58.7</td>
<td>54,470</td>
<td>57.6</td>
<td>9,080</td>
<td>45,390</td>
<td>1,070</td>
<td>1.9</td>
</tr>
<tr>
<td>1944</td>
<td>93,220</td>
<td>54,630</td>
<td>58.6</td>
<td>53,660</td>
<td>57.9</td>
<td>8,950</td>
<td>45,010</td>
<td>570</td>
<td>1.2</td>
</tr>
<tr>
<td>1945</td>
<td>94,090</td>
<td>53,860</td>
<td>57.2</td>
<td>53,820</td>
<td>56.1</td>
<td>8,590</td>
<td>44,240</td>
<td>1,040</td>
<td>1.9</td>
</tr>
<tr>
<td>1946</td>
<td>103,470</td>
<td>51,520</td>
<td>50.8</td>
<td>55,250</td>
<td>53.6</td>
<td>8,820</td>
<td>46,430</td>
<td>2,270</td>
<td>3.9</td>
</tr>
<tr>
<td>1947</td>
<td>106,018</td>
<td>60,168</td>
<td>56.8</td>
<td>57,812</td>
<td>54.5</td>
<td>8,256</td>
<td>49,557</td>
<td>2,356</td>
<td>3.9</td>
</tr>
</tbody>
</table>

Persons 14 years of age and over

<table>
<thead>
<tr>
<th>Year</th>
<th>Total</th>
<th>Percent of population</th>
<th>Employed</th>
<th>Percent of population</th>
<th>Agriculture</th>
<th>Nonagricultural industries</th>
<th>Number</th>
<th>Percent of labor force</th>
</tr>
</thead>
<tbody>
<tr>
<td>1947</td>
<td>101,827</td>
<td>59,350</td>
<td>58.3</td>
<td>57,038</td>
<td>56.0</td>
<td>7,890</td>
<td>49,148</td>
<td>2,311</td>
</tr>
<tr>
<td>1948</td>
<td>103,068</td>
<td>60,621</td>
<td>58.8</td>
<td>58,343</td>
<td>56.6</td>
<td>7,629</td>
<td>50,714</td>
<td>2,276</td>
</tr>
<tr>
<td>1949</td>
<td>103,994</td>
<td>61,286</td>
<td>59.9</td>
<td>57,651</td>
<td>55.4</td>
<td>7,650</td>
<td>49,993</td>
<td>3,637</td>
</tr>
<tr>
<td>1950</td>
<td>104,995</td>
<td>62,208</td>
<td>59.2</td>
<td>58,918</td>
<td>56.1</td>
<td>7,160</td>
<td>51,758</td>
<td>3,288</td>
</tr>
<tr>
<td>1951</td>
<td>104,621</td>
<td>62,017</td>
<td>59.2</td>
<td>59,961</td>
<td>57.3</td>
<td>6,726</td>
<td>53,235</td>
<td>2,055</td>
</tr>
<tr>
<td>1952</td>
<td>105,231</td>
<td>62,138</td>
<td>59.0</td>
<td>60,250</td>
<td>57.3</td>
<td>6,500</td>
<td>53,749</td>
<td>1,883</td>
</tr>
<tr>
<td>1953 (1)</td>
<td>107,056</td>
<td>63,015</td>
<td>58.9</td>
<td>61,179</td>
<td>57.1</td>
<td>6,260</td>
<td>54,919</td>
<td>1,634</td>
</tr>
<tr>
<td>1954</td>
<td>106,321</td>
<td>63,643</td>
<td>58.8</td>
<td>63,109</td>
<td>55.5</td>
<td>6,205</td>
<td>53,904</td>
<td>3,532</td>
</tr>
<tr>
<td>1955</td>
<td>109,681</td>
<td>65,021</td>
<td>59.3</td>
<td>62,170</td>
<td>56.7</td>
<td>6,450</td>
<td>55,722</td>
<td>2,852</td>
</tr>
<tr>
<td>1956</td>
<td>110,954</td>
<td>66,552</td>
<td>60.0</td>
<td>63,799</td>
<td>57.5</td>
<td>6,283</td>
<td>57,514</td>
<td>2,750</td>
</tr>
<tr>
<td>1957</td>
<td>112,265</td>
<td>66,929</td>
<td>59.6</td>
<td>64,071</td>
<td>57.1</td>
<td>5,947</td>
<td>58,123</td>
<td>2,859</td>
</tr>
<tr>
<td>1958</td>
<td>113,727</td>
<td>67,639</td>
<td>59.5</td>
<td>61,036</td>
<td>55.4</td>
<td>5,586</td>
<td>57,450</td>
<td>4,602</td>
</tr>
<tr>
<td>1959</td>
<td>115,329</td>
<td>68,369</td>
<td>59.3</td>
<td>61,630</td>
<td>56.0</td>
<td>5,565</td>
<td>59,065</td>
<td>3,740</td>
</tr>
</tbody>
</table>
```python
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:-1]
        # 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
```
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

Expenditure on Grand Total (2010)

- 17.7b - 23.5b
- 23.5b - 27.5b
- 27.5b - 31.8b
- 31.8b - 43.3b
- 43.3b - 47b
- 47b - 105.3b (£)

The Daily Bread  Country & Regional Analysis  Departmental Spending  About
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.

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

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.
## Dados em destaque

### Informações Legislativas da Câmara dos Deputados
O Projeto Dados Abertos é uma evolução do serviço SIT-Câmara – Serviço de Integração Tecnológica da Câmara dos Deputados – que permite a integração, ...

### Microdados do Exame Nacional do Ensino Médio - Enem
O Ministério da Educação apresentou uma proposta de reformulação do Exame Nacional do Ensino Médio (Enem) e sua utilização como forma de seleção ...

### Malha geométrica dos municípios brasileiros
A Malha Municipal Digital do Brasil é um produto cartográfico do IBGE, elaborado pela Coordenação de Estruturas Territoriais, a partir do Arquivo ...

<table>
<thead>
<tr>
<th>Conjunto de dados</th>
<th>Mês</th>
<th>Totais</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malha geométrica dos municípios brasileiros</td>
<td>206</td>
<td>207</td>
</tr>
<tr>
<td>SICONV - Convênios e Contratos de Repasse da ...</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>Sistema de Informações Organizacionais do Governo ...</td>
<td>95</td>
<td>96</td>
</tr>
<tr>
<td>Microdados do Sistema Nacional de Avaliação da ...</td>
<td>74</td>
<td>74</td>
</tr>
<tr>
<td>Anuário Estatístico de Acidentes de Trabalho - AEAT</td>
<td>70</td>
<td>80</td>
</tr>
</tbody>
</table>
Ministerial gifts, hospitality, travel and meetings with external organisations in Cabinet Office / May-July 2010 - Rt Hon David Cameron MP meeting in CSV Format

<table>
<thead>
<tr>
<th>Minister</th>
<th>Date</th>
<th>Name of External Organisation</th>
<th>Purpose of meeting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prime Minister, The Rt Hon David Cameron MP</td>
<td>May, 2010</td>
<td>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</td>
<td>To discuss Big Society</td>
</tr>
<tr>
<td></td>
<td>May, 2010</td>
<td>Rupert Murdoch</td>
<td>General meeting</td>
</tr>
<tr>
<td></td>
<td>June, 2010</td>
<td>World Bank</td>
<td>To discuss business issues</td>
</tr>
<tr>
<td></td>
<td>June, 2010</td>
<td>Bob Geldof</td>
<td>To discuss development issues</td>
</tr>
<tr>
<td></td>
<td>June, 2010</td>
<td>Ratan Tata</td>
<td>To discuss business issues</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>To discuss new</td>
</tr>
</tbody>
</table>
Welcome to the Data Hub!

Find data

Find datasets

the Data Hub contains 2277 datasets that you can browse, learn about and download.

Share data

Add your own datasets to share them with others and to find other people interested in your data.

Create a dataset »

Collaborate

Find out more about working with open data by exploring these resources:
- GetTheData.org
- DataPatterns.org
- Open Data Manual

Who else is here?

LOD Cloud

This group catalogs data sets that are available on the Web as Linked Data and contain data links pointing at other Linked Data sets. The descriptions of the data sets in this group are...

LOD Cloud has 311 datasets.

Library Linked Data

Group for Library Linked Data Policy described at: http://esw.w3.org/TaskForces/CommunityProjects/LinkedLCloud (partial) available at...

Bibliographic Data

bibliographic metadata from libraries and related institutions.

Linking Open Data

A group for Linking Open Data datasets. The initial import of data for this group was done in October 2009 from the list of RDF datasets dumps provided by the W3C Linking Open Data...

Linking Open Data has 83 datasets.

Climate Data

Weather, temperature, carbon, water, soil and all other kinds of climate related open data. Including (but not limited to) data about: the atmosphere the weather...

Climate Data has 59 datasets.

OpenSpending

Datasets to be imported to the OpenSpending.org site. Packages listed here will automatically be available for selection in the OpenSpending web importer.
Recline Data Explorer and Data Library

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.

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

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.
<table>
<thead>
<tr>
<th>code</th>
<th>name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZW</td>
<td>Zimbabwe</td>
</tr>
<tr>
<td>ZM</td>
<td>Zambia</td>
</tr>
<tr>
<td>YE</td>
<td>Yemen</td>
</tr>
<tr>
<td>EH</td>
<td>Western Sahara</td>
</tr>
<tr>
<td>WF</td>
<td>Wallis And Futuna</td>
</tr>
<tr>
<td>VI</td>
<td>Virgin Islands, U.S.</td>
</tr>
<tr>
<td>VG</td>
<td>Virgin Islands, British</td>
</tr>
<tr>
<td>VN</td>
<td>Viet Nam</td>
</tr>
<tr>
<td>VE</td>
<td>Venezuela, Bolivarian Republic Of</td>
</tr>
<tr>
<td>see</td>
<td>Vatican City State</td>
</tr>
</tbody>
</table>
Where We Are
Huge Growth in Last Few Years

Especially for Government Data
Data Catalogs Around the World as of May 2012

http://datacatalogs.org/
http://datahub.io/dataset/datacatalogs-org
Open Government Partnership

Parceria para Governo Aberto

Conferência Anual
Annual Meeting
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, we also often think about how we can make life easier for other startups. Today, we’re doing both — a little announcement, and hopefully some help for other startups that are thinking about the same things.

So, the announcement:
Starting today, we’re embracing the OpenStreetMap movement, so all the tiles you see when you go to foursquare.com will look a tiny bit different (we think the 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
<table>
<thead>
<tr>
<th></th>
<th>Election Results (national)</th>
<th>Company Register</th>
<th>National Map (Low resolution: 1:250,000 or better)</th>
<th>Government Budget (National, high level, not detailed)</th>
<th>Government Spending (National, transactional level data)</th>
<th>Legislation (laws and statutes) - National</th>
<th>National Statistical Data (economic and demographic Information)</th>
<th>National Postcode/ZIP database</th>
<th>Public Transport Timetables</th>
<th>Environmental Data on major sources of pollutants (e.g. location, emissions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>United Kingdom</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>No info</td>
<td>YYYY</td>
</tr>
<tr>
<td>Brazil</td>
<td>YYYY</td>
<td>YYYY</td>
<td>No info</td>
<td>YYYY</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
</tr>
<tr>
<td>Australia</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
</tr>
<tr>
<td>Netherlands</td>
<td>YYYY</td>
<td>YYYY</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
<td>YYYY</td>
<td>YYYY</td>
<td>No info</td>
<td>No info</td>
</tr>
<tr>
<td>Iceland</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>No info</td>
<td>No info</td>
</tr>
<tr>
<td>Denmark</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>No info</td>
<td>No info</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
<td>No info</td>
<td>No info</td>
</tr>
<tr>
<td>Norway</td>
<td>No info</td>
<td>No info</td>
<td>YYYY</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
<td>YYYY</td>
<td>YYYY</td>
<td>YYYY</td>
</tr>
<tr>
<td>Croatia</td>
<td>YYYY</td>
<td>YYYY</td>
<td>No info</td>
<td>YYYY</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
</tr>
<tr>
<td>Greece</td>
<td>YYYY</td>
<td>YYYY</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
<td>No info</td>
</tr>
</tbody>
</table>

http://census.opengovernmentdata.org/
Machine Readable
Welcome to the School of Data!

The School of Data is a joint initiative led by the Open Knowledge Foundation and Peer 2 Peer University, and generously supported by Open Society Foundations and the Shuttleworth Foundation. The School of Data is a collaborative and community-orientated project, and we welcome contributions from a number of partner organisations and individuals.

Subscribe
Stay in the loop as plans develop: sign up to the School of Data mailing list.

Get Involved
Participate in our Berlin kick-off sprint! Full details on the wiki

Register
Be the first: register for an account with P2PU now.
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
(?)
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 webservice for data transformation

Contents

- Data Protocols Manifesto
- Changes and Syncing
  - SLEEP
  - CouchDB
  - MVCC and WAL for
  - General Overview
- Data Query Protocol
  - Introduction
  - Proposal
  - Existing Work
- Refining Protocol
  - Refine API
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)
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]