

Frictionless Data

National platforms for SDG reporting

Jo Barratt

jo.barratt@okfn.org, @jobarratt



OPEN KNOWLEDGE
INTERNATIONAL



OPEN KNOWLEDGE

Vision

A world where **open knowledge** is ubiquitous, enabling **citizens** and **organizations** to create **insights** that drive **change** on global and local challenges, combat **injustice** and **inequality** and hold governments and corporations to account.



School of Data



OPEN KNOWLEDGE

	Election areas	Public Facilities	Annual Budget	Transport Timetables	Business Listings	Real-Time Transit	Procurement contracts	Air Quality	Agendas and decisions	Expenditure (detailed)	Crime statistics	Food safety inspections	Service Requests	Traffic accidents	Building permits	Business Permits	Total Score	
1	Helsinki					2016	2016		2016	2016	2016	2016	2016	2016	2016	2016	1315	
2	Vantaa	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	1110	
3	Espoo	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	1075	
4	Tampere	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	1030	
5	Turku	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	1020	
6	Kauniainen	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	1000	
7	Oulu																995	
8	Hämeenlinna	2014	2014	2014	2014	2014	2014	2014	2014	2014	+ Add	2014	2014	+ Add	2014	+ Add	750	
9	Jyväskylä	2014	2014	2014	2014	+ Add	2014	2015	+ Add	2014	2014	+ Add	+ Add	+ Add	2014	+ Add	+ Add	425
10	Mikkeli	+ Add	+ Add	2014	+ Add	+ Add	2014	+ Add	+ Add	2014	+ Add	+ Add	+ Add	+ Add	+ Add	+ Add	130	
11	Kuopio	+ Add	+ Add	+ Add	+ Add	+ Add	+ Add	+ Add	+ Add	+ Add	+ Add	+ Add	+ Add	+ Add	+ Add	+ Add	0	
11	Rovaniemi	(1)	(1)	(1)	+ Add	(1)	+ Add	+ Add	(1)	(1)	+ Add	(1)	(1)	(1)	+ Add	(1)	+ Add	0

index.ok.org



OpenTrials

All the data, on every trial in medicine

OpenTrials is a linked database for all the available information, on every trial ever conducted, with rankings for transparency and more. It is built and updated with your help.



For Patients

Find ongoing trials. Read trial results.
Contribute information about trials.



For Researchers

Find trial information.
Improve information about trials.



For Transparency

See which companies and drugs
are the most and least transparent.



A Global Network



ckan.org/instances

Widely adopted - over 20 national governments and 60+ local governments & cities

The screenshot shows the Africa Open Data website. At the top, it says "We are in BETA. Find a problem? Let us know." and "Log in Register". The main heading is "Africa Open Data" with a search bar. Below the heading, there's a featured article titled "Curating and opening up data all over Africa" with a sub-heading "This Platform will be the largest central repository for Government, Civil Society, Corporate and Donor Agency Data. This is an Ushahidi Initiative." There are also sections for "d bootcamp" and "Ghana Databotcamp". A search bar labeled "Search Your Data" is present with the example "eg. Gold Prices". At the bottom, there are sections for "Kenya Datasets from Kenya" and "Ghana Data Sets For the Republic of Ghana".

Africa Open Data

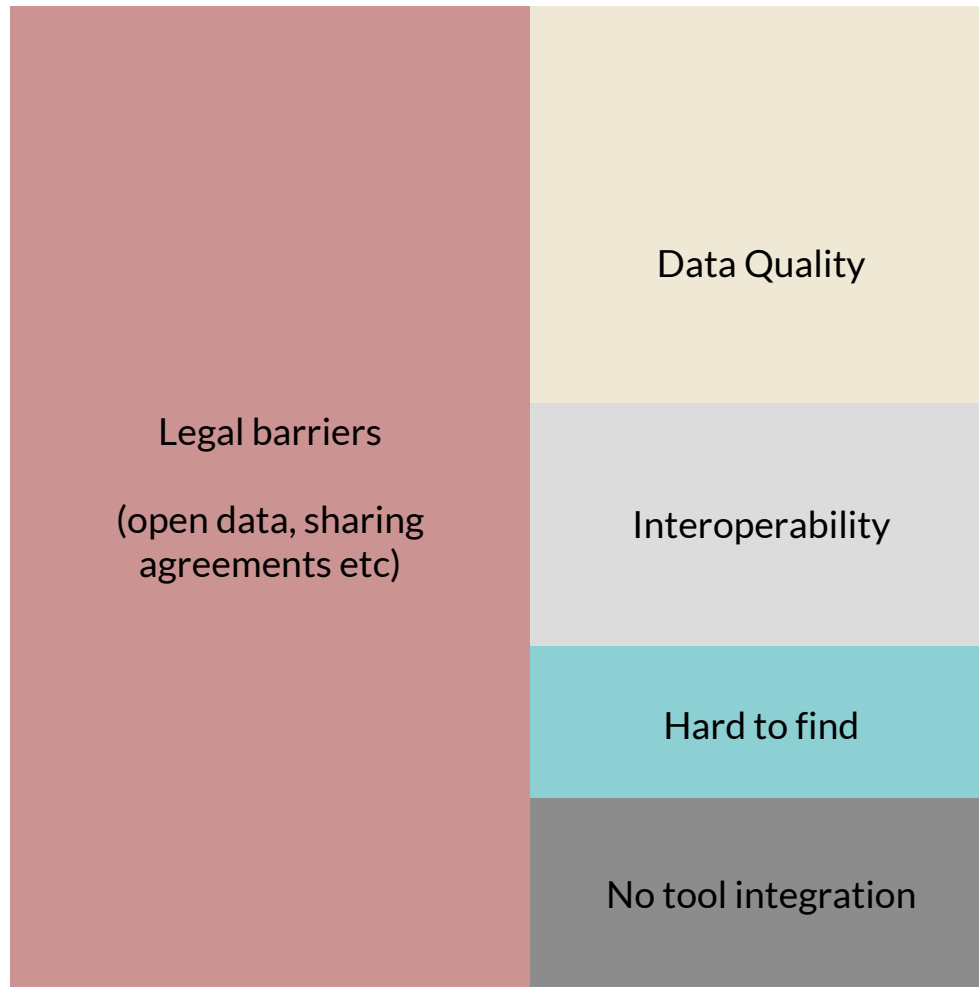
The screenshot shows the Buenos Aires Data website. The header includes "Buenos Aires Ciudad" and "En todo está vos". The main heading is "Buenos Aires Data" with a sub-heading "Plataforma de datos públicos y transparencia de la Ciudad de Buenos Aires". Below the heading, there's a search bar and a section titled "Un nueva versión de Buenos Aires Data." with bullet points: "Acceso programático vía RESTful API a los conjuntos de datos.", "Visualización de los conjuntos de datos en formato de lista, gráficos y mapas interactivos.", "Agregar nuevos conjuntos de datos nuevos.", and "Nuevo modo de acceder y usar los datasets." At the bottom, there are three icons: "Datos" (Data), "Aplicaciones" (Applications), and "Laboratorio" (Laboratory).

Buenos Aires Data

The screenshot shows the City of Ottawa Open Data Catalog website. The header includes "OPEN DATA" and "City of Ottawa". The main heading is "City of Ottawa Open Data Catalog". Below the heading, there's a search bar and a section titled "Browse Datasets" with a sub-heading "Browse by Department Help". The main content area shows a list of datasets with titles like "311 Service Request Volumes", "Address Points (main and subordinates)", "Ball Diamonds", "Basket Ball Courts", "Beaches", "Bicycle Parking - King & Posts", and "Bicycle trio counters (automated)". On the right side, there's a "Tags" section with a list of tags and their counts: "recreation (12)", "budget (3)", "transport (2)", "library (2)", "finance (2)", "cycling (2)", "bus (2)", "ward (1)", "vendors (1)", and "tree (1)".

City of Ottawa Open







Greater London Authority Spending

0	Report of						
1	Reporting	1					
2	Start	1st April, 2013					
3	End Date:	27th April, 2013					
4	Financial	2012 / 13					
5							
6	Vendor ID	Vendor Name	Cost	Expenditure	Document	Amount	Clearing
7	15500130	LONDON	544075	Grants to External	5107789349	12,700,000.00	02 Apr
8	15500097	LONDON	544075	Grants to External	5107841602	12,500,000.00	22 Apr
9	10022138	HARINGEY	544075	Grants to External	5107841553	9,787,182.00	12 Apr
10	10006080	LONDON	544075	Grants to External	5107844319	6,895,149.00	10 Apr
11	18000042	FAMILY MOSAIC	544076	HSG Grants to	1900017728	4,825,000.00	05 Apr
12	15500051	LONDON	544075	Grants to External	5107849492	4,536,577.00	18 Apr
13	18000048	HYDE HOUSING	544076	HSG Grants to	1900017672	4,523,837.00	05 Apr
14	18000066	EAST THAMES	544076	HSG Grants to	1900017812	4,053,595.00	05 Apr
15	18000049	THE GUINNESS	544076	HSG Grants to	1900017739	3,997,854.00	05 Apr
16	18000009	PEABODY TRUST	544076	HSG Grants to	1900017759	3,880,412.00	05 Apr
17	18000045	LONDON AND	544076	HSG Grants to	1900017881	3,640,000.00	08 Apr
18	18000066	EAST THAMES	544076	HSG Grants to	1900017808	3,572,500.00	05 Apr
19	18000045	LONDON AND	544076	HSG Grants to	1900017871	3,260,000.00	08 Apr
20	18000078	BERKELEY	544078	HSG Grants to Non-	1900017683	3,174,765.00	05 Apr

Full screen version - CSV to HTML powered by DataPipes

https://www.london.gov.uk/sites/default/files/gla_migrate_files_destination/Mayor%20202013-14%20-%20P1%20-%20Combined.csv

Greater London Authority

This isn't about a single data file but an entire dataset. The dataset is the UK's Greater London Authority spend data:

<http://www.london.gov.uk/mayor-assembly/gla/spending-money-wisely/budget-expenditure-charges/expenditure-over-250>

The good news is that, as per UK government policy, the GLA provides regular data on spending over £250, and, furthermore, the data is in simple [CSV files](#).

That's where the good news ends. Here are primary issues:

- Data is in **many separate files** (65 as per summer 2013) with no consistent naming scheme so downloading them requires [scraping](#) (or tedious click and save)
- Amongst the 65+ CSV files one find approximately **25+ different structures** (structure means things like: number of blank or non-data lines at top of files, blank columns, varying headers). This therefore requires [elaborate coding](#) to consolidate the data into a single form. You can see a full summary of all the different file structures in this [data wrangle](#)

Additional specific errors:

- Almost every file does not start with actual data headers but with random metadata (see file to left as an example)

<http://okfnlabs.org/bad-data/ex/gla-spending/>



OPEN KNOWLEDGE



FRictionLESS DATA

STANDARDS AND TOOLING

- Introduce a significant, measurable improvement in how data is shared, consumed, and analyzed.
- Make it easier to maintain and improve data quality.



UNStats / SDG

Watch 5 Star 0 Fork 1

- Code
- Issues 0
- Pull requests 0
- Projects 0
- Wiki
- Insights

Branch: master SDG / DataPackages / Goal /

Create new file Upload files Find file History

soho501 Update datapackage.json		Latest commit 56eb87a on 21 Dec 2017
..		
1	Update datapackage.json	a month ago
10	Adding Goal Data Packages	a month ago
11	Adding Goal Data Packages	a month ago
12	Adding Goal Data Packages	a month ago
13	Adding Goal Data Packages	a month ago
14	Adding Goal Data Packages	a month ago
15	Adding Goal Data Packages	a month ago
16	Adding Goal Data Packages	a month ago
17	Adding Goal Data Packages	a month ago
2	Adding Goal Data Packages	a month ago
3	Adding Goal Data Packages	a month ago

<https://github.com/UNStats/SDG/tree/master/DataPackages/Goal>

goodtables.io

See how your data is doing

Sign in with Github

The screenshot shows the goodtables.io interface for a file named 'some-file.csv'. The interface includes a sidebar with navigation icons, a main content area with two tables, and a right-hand notification panel. The first table has a 'Missing Header' error highlighted in red. The second table has a 'Duplicate Header' error highlighted in red. The notification panel shows three 'Pushed on master' events with status indicators (red 'x' for error, green checkmark for success).

github/okfn/test-data/master/**some-file.csv** ⁴ Table 1 of 3

1 x **Missing Header**

1	id	name		name
>				

1 x **Duplicate Header**

1	id	name	name
>			

Pushed on master
7 minutes ago 2928dd

Pushed on master
2 days ago 6298fe

Pushed on master
3 days ago 76564d

Pushed on master
5 days ago 2768de

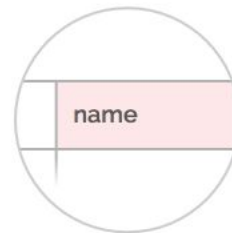
Continuous data validation for spreadsheets



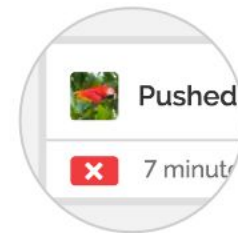
Monitor files in multiple formats, including CSV, Excel, LibreOffice, and more.



Checks the data structure and contents of your tables for potential issues.



See errors in familiar context, so they can be easily isolated and resolved.



Automatically validate on every change, helping you to keep your data error free.

✖ frictionlessdata/example-goodtables.io 🔄
 2 days ago #2

✔ frictionlessdata/ADB-User-Study 🔄
 2 days ago #3

ⓘ amercader/co2-ppm 🔄
 10 days ago #

✔ amercader/newcastle-libraries-data 🔄
 10 days ago #42

✖ amercader/car-fuel-and-emissions 🔄
 a month ago #1

✖ test-gtio-1 a
 2 months ago #2

ⓘ amercader/badgelona 🔄
 No jobs yet

ⓘ amercader/ne-data-test 🔄
 No jobs yet

ⓘ amercader/solr-node-client 🔄
 No jobs yet

ⓘ frictionlessdata/datapackage-ckan-js 🔄
 No jobs yet

ⓘ test-gtio-2 a
 No jobs yet

ⓘ test-gtio-3 a
 No jobs yet



UK Spend Publishing Dashboard

How are government departments living up to their commitments to publish timely, high quality data on spending?

0

CORRECT (%) ⓘ

0

SCORE (%) ⓘ

33

PUBLISHERS

2394

DATA FILES

Publishers

Title	Type	Homepage	Email	Correct ⓘ	Score (all time)	Last file	Score ⓘ
Cabinet Office	Ministerial department	↗	✉	0	50 %	August 2014	90 %
Department for Business, Innovation and Skills	Ministerial department	↗	✉	0	41 %	February 2015	90 %
Department of Health	Ministerial department	↗	✉	0	56 %	July 2015	90 %
Foreign and Commonwealth Office	Ministerial department	↗	✉	0	36 %	April 2015	90 %
Ministry of Defence	Ministerial department	↗	✉	0	59 %	June 2015	90 %
Department for Communities and Local Government	Ministerial department	↗	✉	0	27 %	April 2012	70 %
Department for Environment, Food and Rural Affairs	Ministerial department	↗	✉	10	57 %	June 2015	60 %
Scotland Office	Ministerial department	↗	✉	5	57 %	October 2012	30 %
Department for Work and Pensions	Ministerial department	↗	✉	0	45 %	June 2015	20 %
Attorney General's Office	Ministerial department	↗	✉	0	0 %	March 2010	0 %

Resources

Load data package

Metadata >

Keywords >

Validate

Download

Name	resource1	Path	data.csv	Load
------	-----------	------	----------	------

test1

Title
Test1

Description
Description

Data Type
string

Data Format
default

+
Add field

Name	resource2	Path	https://raw.githubusercontent.com/frictionlessdata/goodtables	Load
------	-----------	------	---	------

Name	resource3	Path	Type resource path	Load
------	-----------	------	--------------------	------

+ Add resource

DATA
HUB

Data online made simple

The image shows a workflow for uploading data to DataHub. On the left, a terminal window displays the following commands and output:

```
bash
~/my-data $ ls
datapackage.json  data.csv
~/my-data $ data push
> Uploading [=====] 100% 8.5s
> Push complete (9.61MB) [8509ms]
> Ready! https://datahub.io/your-username/my-data
(copied to clipboard)
~/my-data $
```

On the right, a browser window shows the URL `https://datahub.io/you/your-data`. Below the URL are three horizontal lines representing a menu or list. At the bottom, there are two icons: a grid representing a table and a line graph representing a chart.





**Rufus Pollock explains
the problem
Frictionless Data
addresses through
simple specifications
and software.**



Processing, publication, access, insight. The future of data is frictionless.

- Project website: <http://frictionlessdata.io/>
- Specifications: <http://frictionlessdata.io/specs/>

Pilot



Data Management for TEDDINET

Frictionless Data specifications for data legacy

Pilot



eLife

Investigating the applicability of the goodtables library to facilitate data validation in the field of Life and Biomedical Sciences and make a case for the reusability of data shared with eLife as additional or supporting files.

Pilot



Western Pennsylvania Regional Data Center

Using ckanext-validation extension to highlight quality of datasets in CKAN-based open data portals.

Pilot



University of Cambridge - Retinal Mosaics

Investigating the applicability of the Data Package concept to facilitate data reuse in the field of Computational Neuroscience.

Pilot



UK Data Service

Using Frictionless Data software to assess and report on data quality and make a case for generating visualizations with ensuing data and metadata.

Pilot



Causa Natura - Pescando Datos

Using goodtables to improve quality of data to support fisher communities and advocacy groups.



OPEN KNOWLEDGE



OPEN KNOWLEDGE
INTERNATIONAL