Frictionless Data

National platforms for SDG reporting

Jo Barratt

jo.barratt@okfn.org, @jobarratt

OPEN KNOWLEDGE INTERNATIONAL

OPEN KNOWLEDGE

Vision

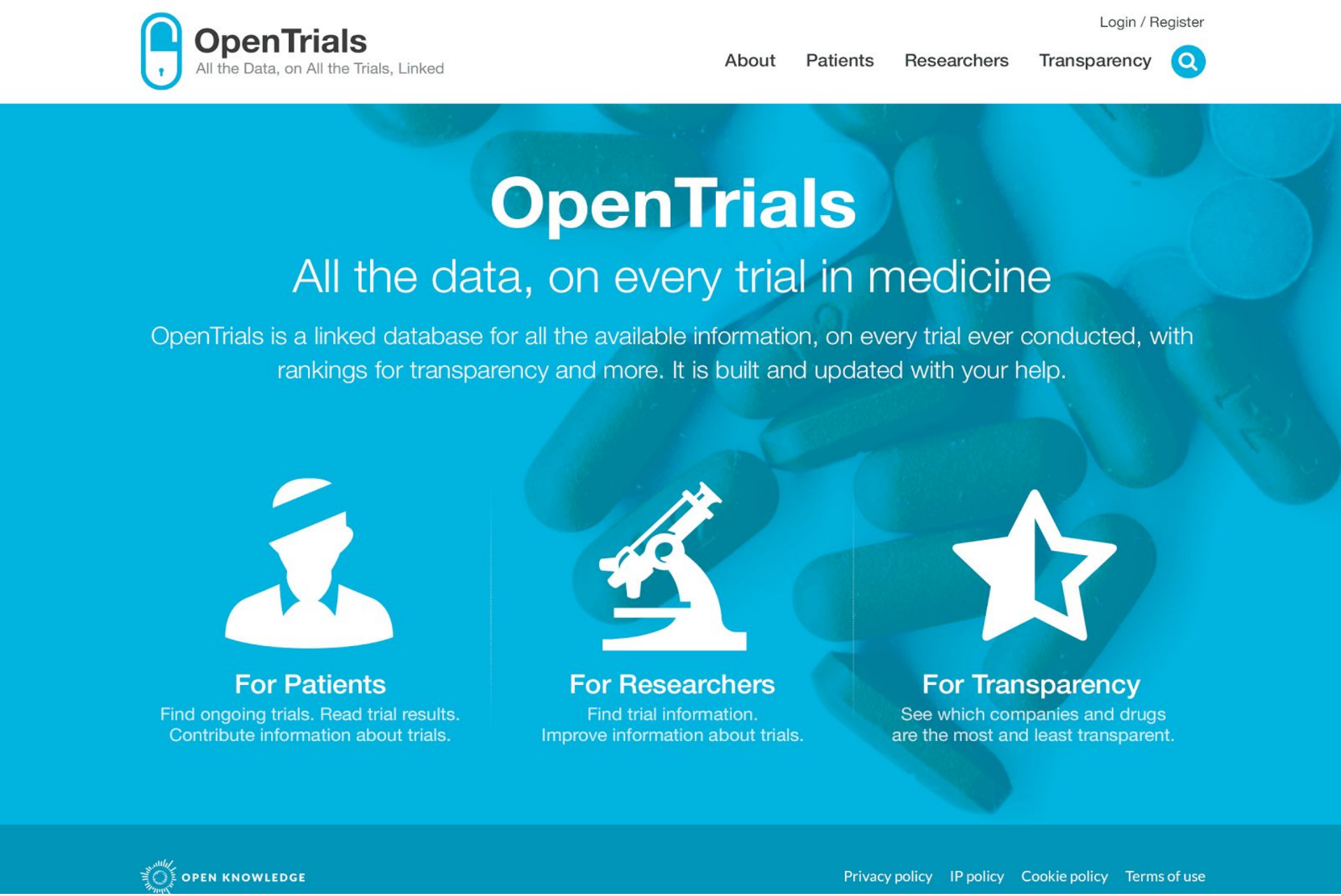
A world where **open knowledge** is ubiquitous, enabling **citizens** and **organizations** to create **insights** that drive **change** on global and local challenges, combat **injustice** and **inequality** and hold governments and corporations to account.

School of Data

OPEN KNOWLEDGE

index.ok.org

OPEN KNOWLEDGE



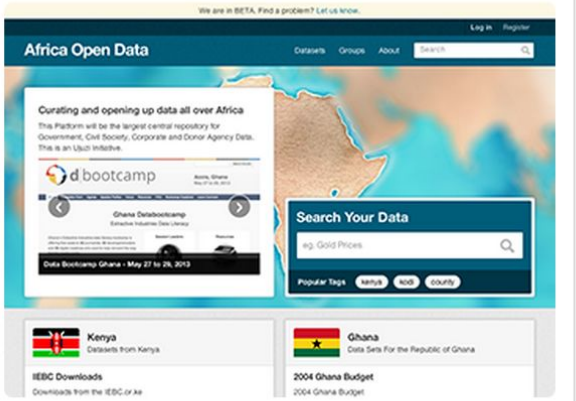
OPEN KNOWLEDGE

A Global Network

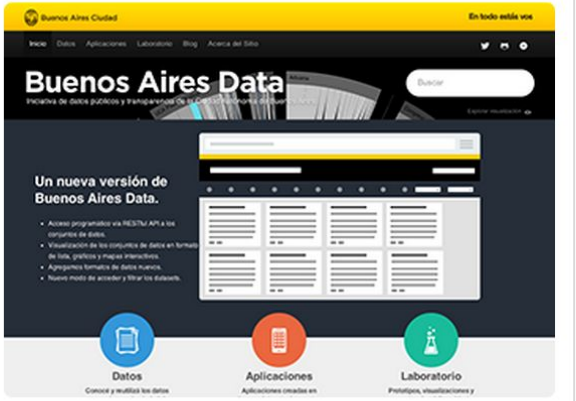
ckan

ckan.org/instances

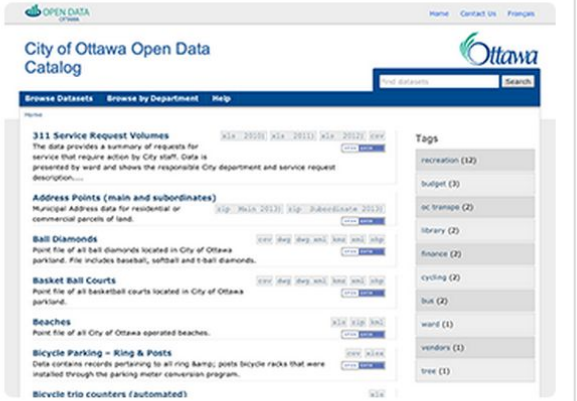
Widely adopted - over 20 national governments and 60+ local governments & cities



Africa Open Data

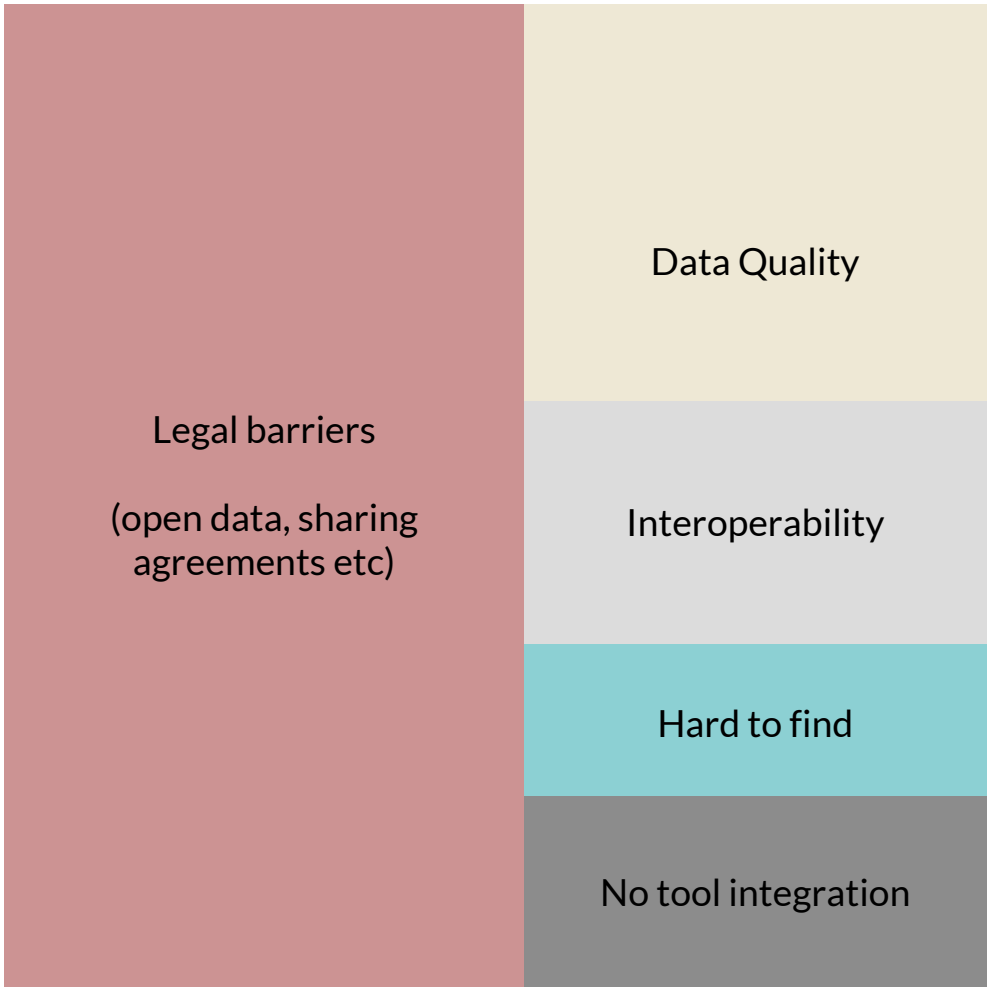


Buenos Aires Data



City of Ottawa Open

OPEN KNOWLEDGE



Treemap of issues …

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%27%202013-14%20-%20P1%20%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

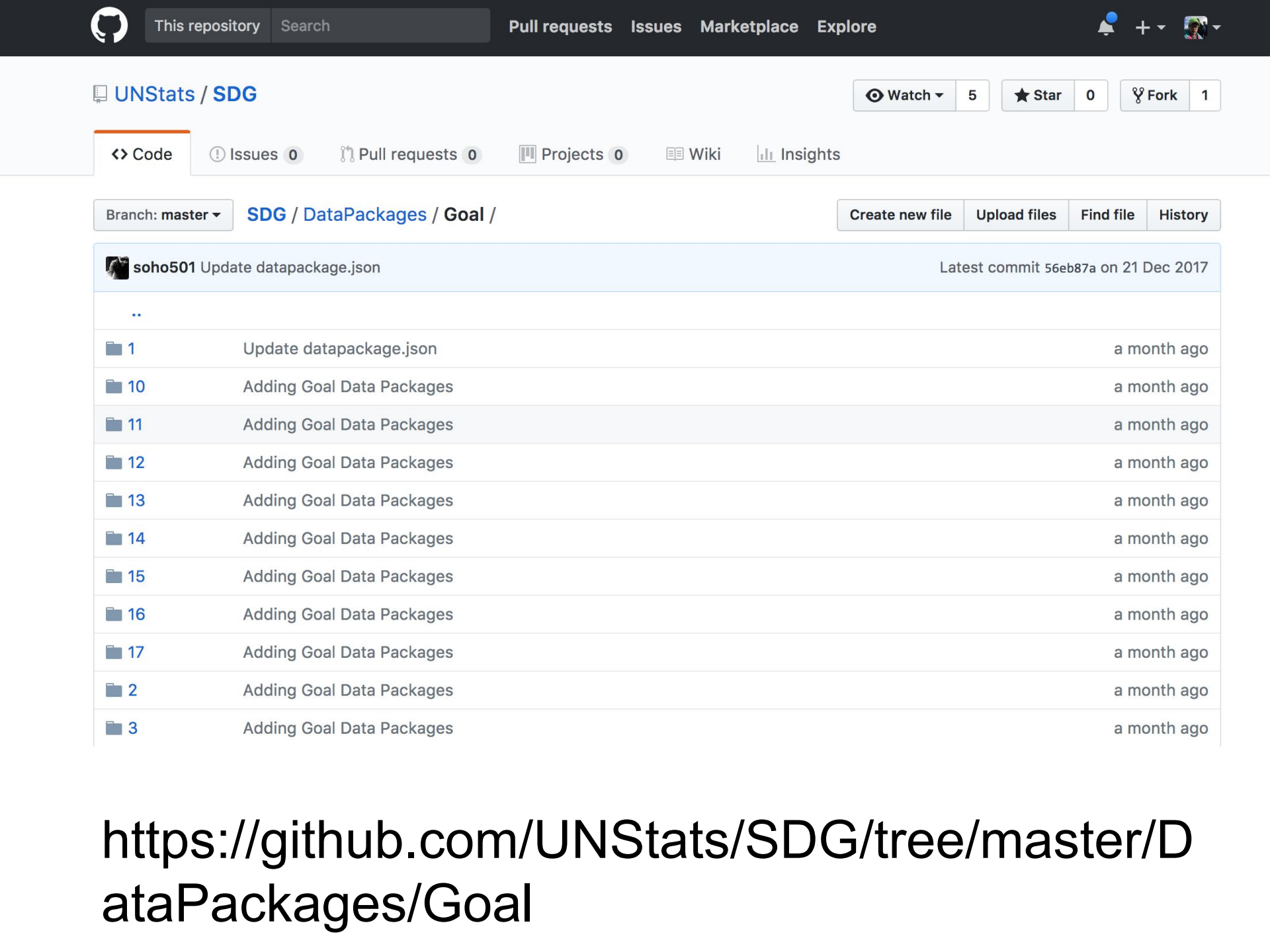
Additional specific errors:

- 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

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

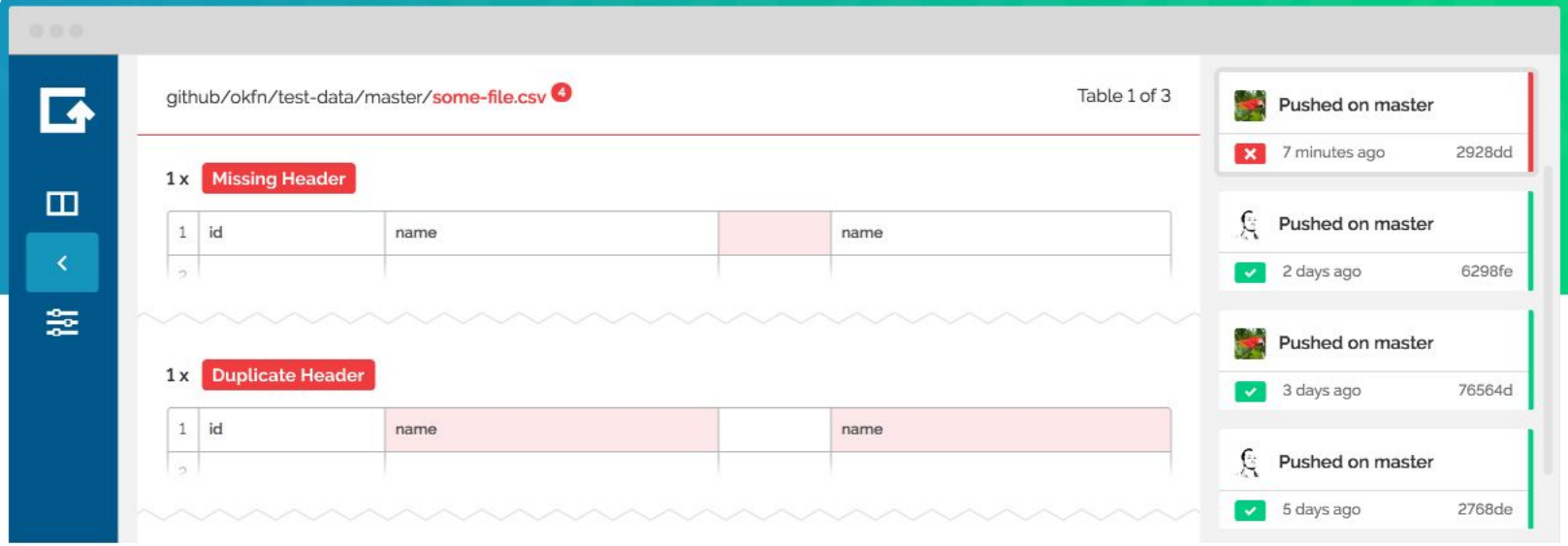


https://github.com/UNStats/SDG/tree/master/DataPackages/Goal

See how your data is doing

Sign in with Github

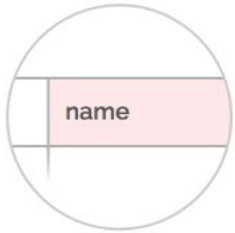
goodtables.io



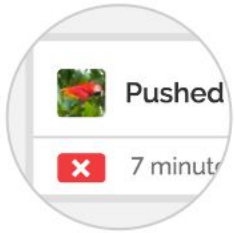
Continuous data validation for spreadsheets

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.

goodTables.io

Manage Sources | Logout

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
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
No jobs yet

test-gtio-3
No jobs yet

OPEN KNOWLEDGE
UK SPEND PUBLISHING DASHBOARD

About Pricing FAQ

OPEN KNOWLEDGE

UK Spend Publishing Dashboard

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

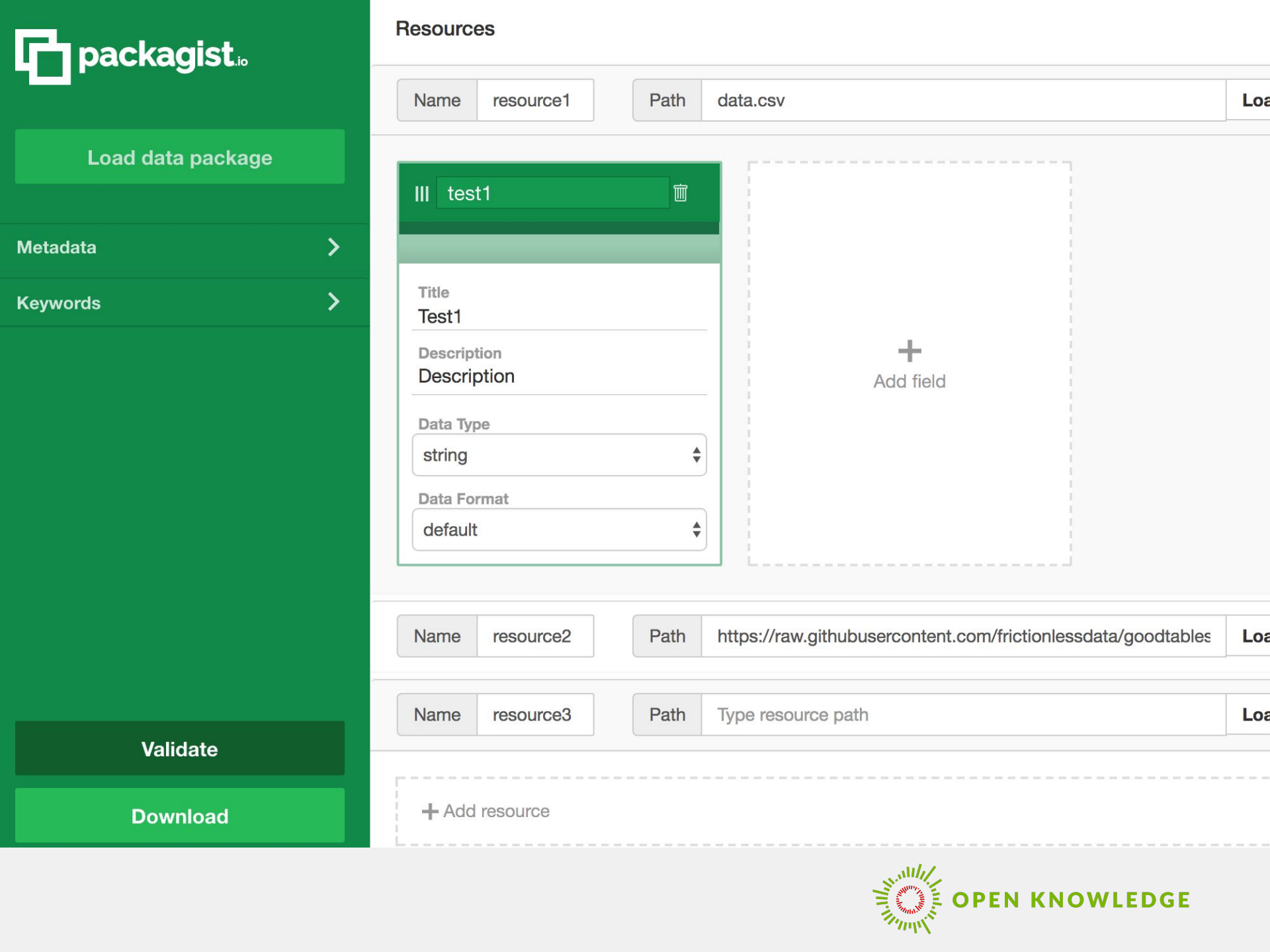
0	0	33	2394
CORRECT (%)	SCORE (%)	PUBLISHERS	DATA FILES

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 %

http://uk-25k.datadashboards.io/

OPEN KNOWLEDGE


packagist.io
Load data package
Metadata
Keywords
Validate
Download
Resources
Name
resource1
Path
data.csv
Loa
test1
Title
Test1
Description
Description
Data Type
string
Data Format
default
Add field
Name
resource2
Path
https://raw.githubusercontent.com/frictionlessdata/goodtables
Loa
Name
resource3
Path
Type resource path
Loa
Add resource
OPEN KNOWLEDGE

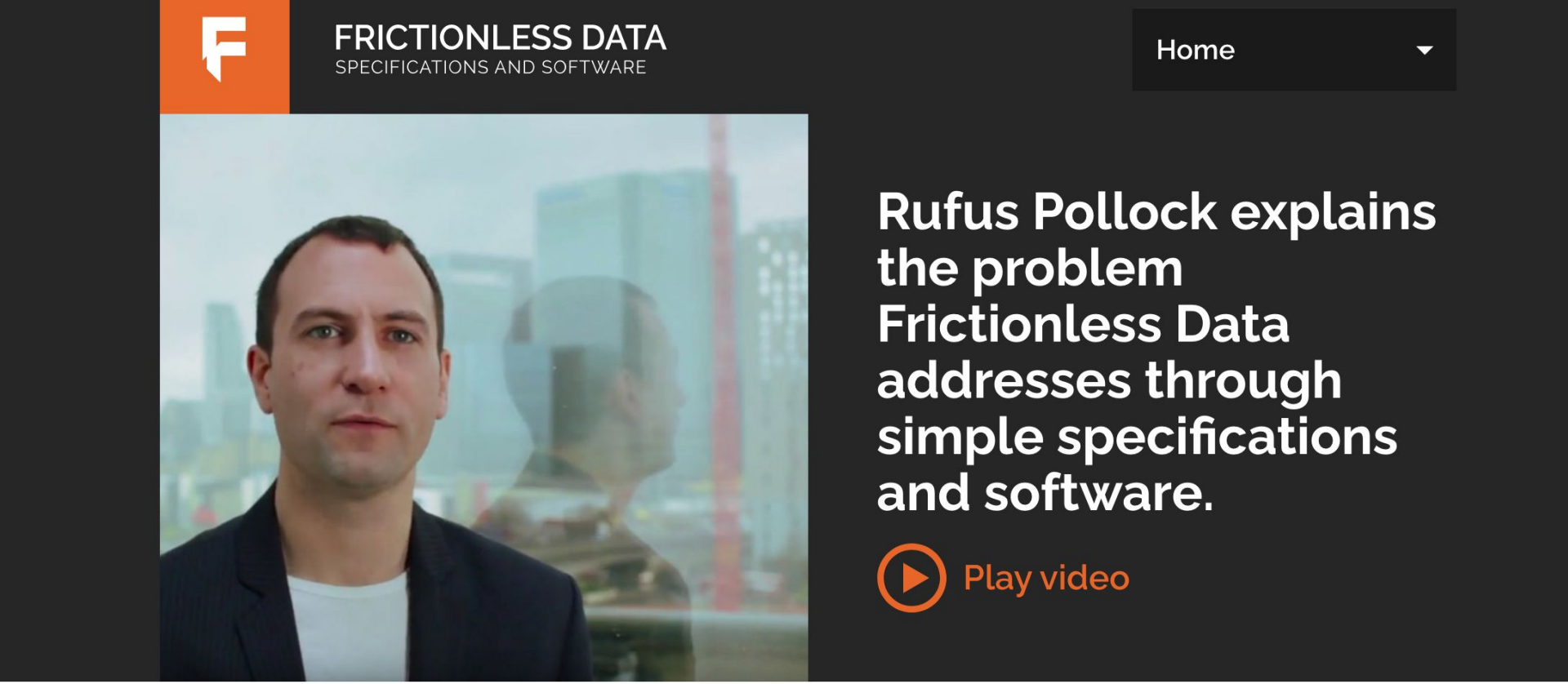
DATA HUB

DATASETS AWESOME REQUESTS BLOG DOCS PRICING PUBLISH CHAT

DATA HUB

Data online made simple

OPEN KNOWLEDGE



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

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

OPEN KNOWLEDGE

Pilot

Data Management for TEDDINET

Frictionless Data specifications for data legacy

Pilot

Western Pennsylvania Regional Data Center

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

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

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

University of Cambridge - Retinal Mosaics

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

Pilot

Causa Natura - Pescando Datos

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

OPEN KNOWLEDGE
INTERNATIONAL