

---

# Considerations of using "open" web data

Steven Maude  
The Sensible Code Company

# Truly open data

*“freely used, shared and built-on by anyone, anywhere, for any purpose”*

<https://blog.okfn.org/2013/10/03/defining-open-data/>

In practice, what we often have to work with is “data on the web”: **available, but not necessarily open.**



---

# Barriers

availability < *shared*  
licensing < *freely used, shared*  
discoverability < *shared*  
collection < *built-on*  
consistency < *built-on*  
longevity < *built-on*

---

# Availability





Doncaster Council

@MyDoncaster · 17 Nov 2017

4.7K

11K



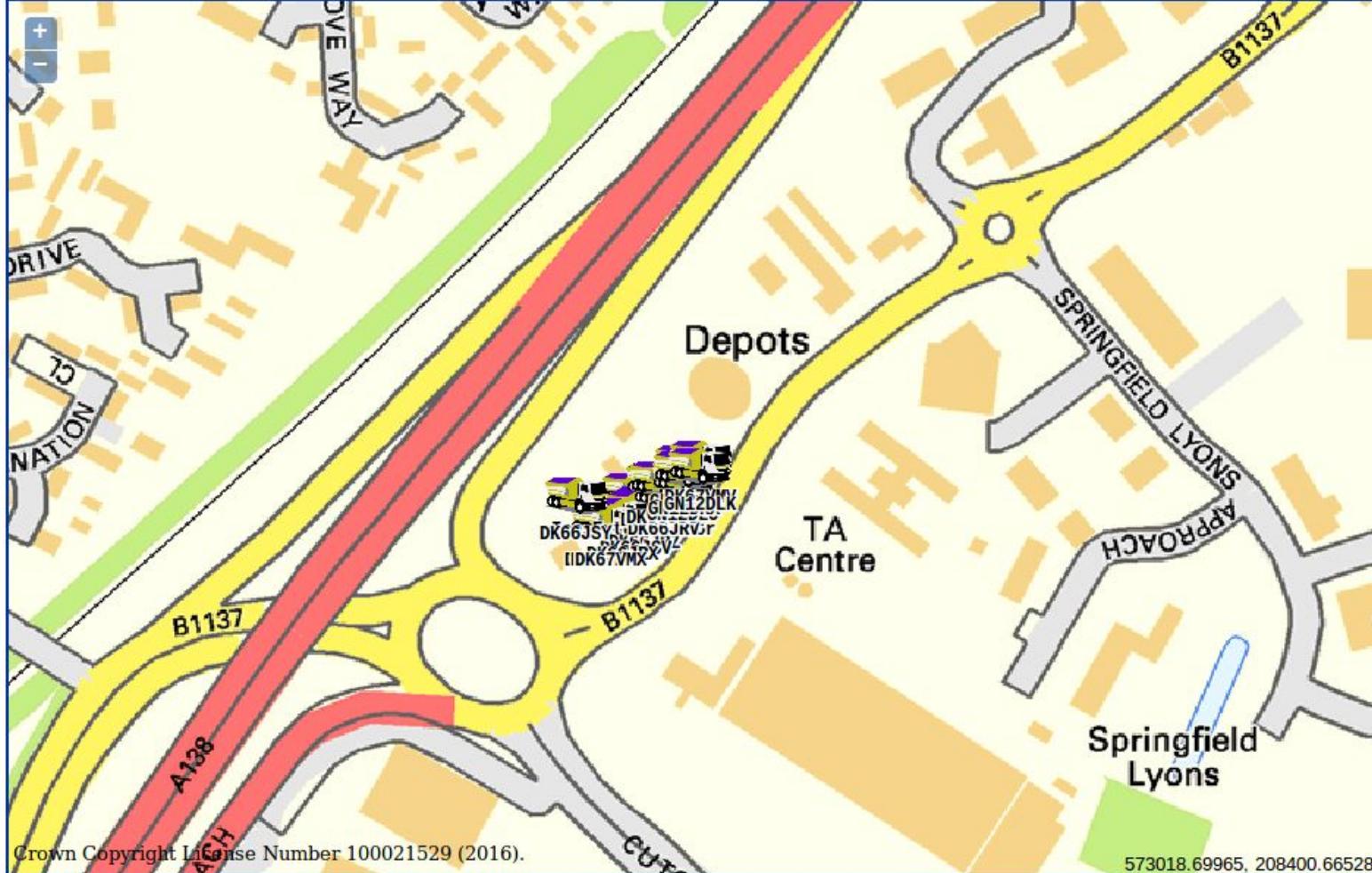
We are DELIGHTED to be able to confirm what our new gritters will be called, following the  
#DoncasterGrittingWorldCup final!

They will be called:

Gritsy Bitsy Teeny Weeny Yellow Anti-Slip Machiney

And

David Plowie







[Make a request](#)[Browse requests](#)[View authorities](#)[Read blog](#)[Help](#)[Sign in or sign up](#)

**Get answers from the government and public sector**

Make a request for information to a UK public authority: by law, they have to respond

Browse [450,558 requests](#) to [20,345 authorities](#)

**Want to know something?**

Start your own request

[Make a request »](#)

---

# Licensing

# Ninth Circuit Doubles Down: Violating a Website's Terms of Service Is Not a Crime

BY JAMIE WILLIAMS | JANUARY 10, 2018

Good news out of the Ninth Circuit: the federal court of appeals heeded EFF's [advice](#) and [rejected](#) an attempt by Oracle to hold a company criminally liable for accessing Oracle's website in a manner it didn't like. The court [ruled](#) back in 2012 that merely violating a website's terms of use is not a crime under the federal computer crime statute, the [Computer Fraud and Abuse Act](#). But some companies, like Oracle, turned to state computer crime statutes—in this case, California and Nevada—to enforce their computer use preferences.

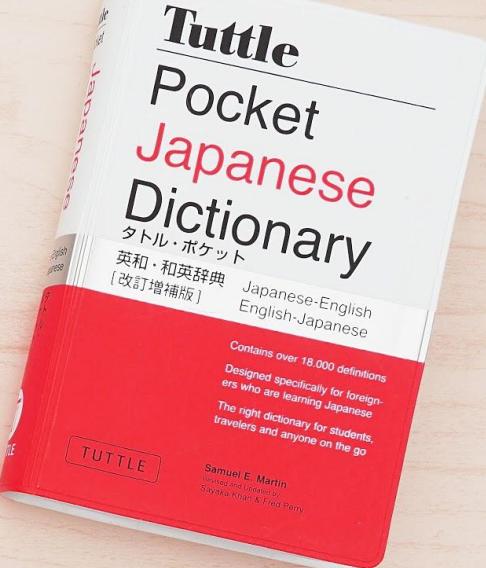
This decision shores up the good precedent from 2012 and makes clear—if it wasn't clear already—that violating a corporate computer use policy is not a crime.

*Oracle v. Rimini* involves Oracle's terms of use prohibition on the use of automated methods to download support materials from the company's website. Rimini, which provides Oracle

---

# Discoverability





---

# Collection and formats

Headers	Cookies	Params	Response	Timings	Stack Trace
<b>Request URL:</b> http://www.essexhighways.org/Rings/Default.aspx/GetTrackingData					
<b>Request method:</b> GET					

Headers	Cookies	Params	Response	Timings	Stack Trace
▼ Filter properties					
▼ JSON					
d: [{"assetRegistration": "SV57EDL", "assetName": "SV57EDL", "assetId": "560884", "assetGroupName": "De-Fitted", "inputs": [], "privacy": false, "assetType": "GRITTER"}, {"assetRegistration": "YJ62KCE", "assetName": "YJ62KCE", "assetId": "559782", "assetGroupName": "De-Fitted", "latitude": 54.131046, "longitude": -1.506838, "inputs": [], "privacy": false, "assetType": "GRITTER"}, {"assetRegistration": "GN12DGZ", "assetName": "GN12DGZ", "assetId": "545339", "assetGroupName": "Springfield Depot", "date": "2015-01-20T14:18:25", "status": "behaviourExcessi...", "assetType": "HGV"}, {"assetRegistration": "DK60UPE", "assetName": "DK60UPE", "assetId": "543613", "driverName": "Tommy Smith", "driverGroupName": "Picketts Lock Fleet", "assetGroupName": "Picketts Lock Fleet", "date": "2018-01-20T10:52:01", "status": "driving", "speed": 14.0, "heading": 50.0, "latitude": 51.598694, "longitude": -0.091143, "road": "Turnant Road", "roadNumber": "", "city": "N17, London", "postCode": "N17 6", "country": "UK", "formattedAddress": "Turnant Rd, N17, London, N17 6, UK", "inputs": [], "privacy": false, "assetType": "TIPPER"}]					



## OUR PRODUCTS



Used by thousands of people every day to instantly convert their PDFs into Excel files.

PDFTables uses artificial intelligence techniques and highly scalable infrastructure to deliver high-speed conversions.

[Learn more](#)



### QuickCode

Our original, pioneering product — an online Python and R data analysis environment.

QuickCode is perfect for numerically-literate economists, statisticians and data managers who are new to coding.

[Learn more](#)



### DataBaker

Built for the [ONS](#) to wrangle a mountain of spreadsheets into clean, data tables.

DataBaker helps users code recipes that turn tricky Excel spreadsheets into normalised, time series data.

[Learn more](#)



### TableBuilder

Makes safe-to-publish tables of information from large confidential data in a snap.

TableBuilder's speed, based on new hardware options, allows interactions with survey data as if it were open data.

[Learn more](#)



Making public transit data universally accessible.

Public transit agencies are experts at running large, complex transit networks, but they might have less experience building apps. We believe that common formats for exchanging public transit information is the solution, so third-party developers can build the innovative transit apps. These formats allow public transit agencies to publish their transit data and developers to write applications that consume that data in an interoperable way.

The following documentation provides specifications for exchanging both static and realtime public transit information, including information about stops, routes, schedule, service alerts, and other details.

## General Transit Feed Specification (GTFS)

The General Transit Feed Specification (GTFS) can be used to share *static* public transit data.

[READ ABOUT GTFS](#)

## GTFS Realtime

The GTFS-realtime Specification is an extension to GTFS that can be used to share *realtime* public transit data.

[READ ABOUT GTFS-REALTIME](#)



---

# Consistency





---

# Longevity





## Blog

# The story of getting Twitter data and its “missing middle”



by [Francis Irving](#) on August 21, 2014 under Products • 8 Comments

We've [tried hard](#), but sadly we are not able to bring back our Twitter data tools.

Simply put, this is because Twitter have no route to market to sell low volume data for spreadsheet-style individual use.

The screenshot shows the Twitter search interface. At the top, there are two search bars: one with a magnifying glass icon labeled "Search for Tweets" and another with a Twitter bird icon labeled "Search for Tweets". Below these is a main search bar with the placeholder "Search for Tweets". Underneath the search bar is a button labeled "Search". A tooltip or callout box is overlaid on the bottom right of the search area, containing text about using Twitter's advanced search operators. The text in the tooltip reads: "You can use Twitter's advanced search operators in your query. e.g. data OR journalism, -estratascarf, to:bengeldartre".

# Case by case; no general solution

But worthwhile to try: richness in combining data sources to build or discover new things.



# Thanks :)

Non-credited images either my own browser screenshots or from unsplash.com.

