



<http://training.theodi.org/InPractice>

---

David Tarrant · @davetaz

# Session 2

## Time planning in open data projects



# Outcomes

Understand how the 80/20 rule applies to open data projects

List the stages in an project that uses open data

Apply a number of tools quickly analyse data



# DATABLOG

Facts are sacred

<http://www.theguardian.com/news/datablog/2012/may/24/data-journalism-punk>



[Previous](#)

[Blog home](#)

[Next](#)

## Anyone can do it. Data journalism is the new punk

Can anyone be a data journalist? **Simon Rogers** on what we can learn from a 1977 diagram

• Another view: [What data can and cannot do](#) by Jonathan Gray

[Share](#) 24

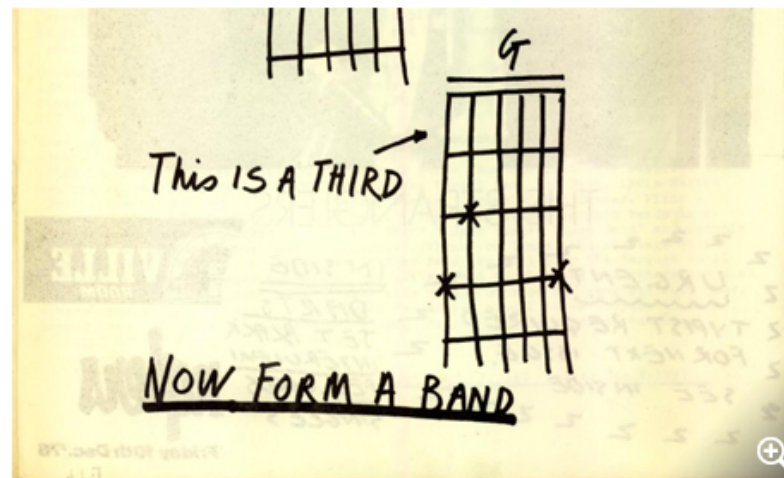
[Tweet](#) 18

[+1](#) 60

[Pin it](#) 1

[in Share](#) 5

[Email](#)



Page two of [Sideburns](#), January 1977



This is a chord... this is another... this is a third. NOW FORM A BAND



Posted by  
**Simon Rogers**  
Thursday 24 May 2012  
13.00 BST  
[theguardian.com](#)  
[Jump to comments \(8\)](#)

[Article history](#)

**Media**

[Data journalism](#) · [Open journalism](#)

**Technology**

# The 80/20 of data

In **any data project** you will spend 80% of your time preparing, analysing, cleaning, enriching etc... and 20% of the time actually producing the output.



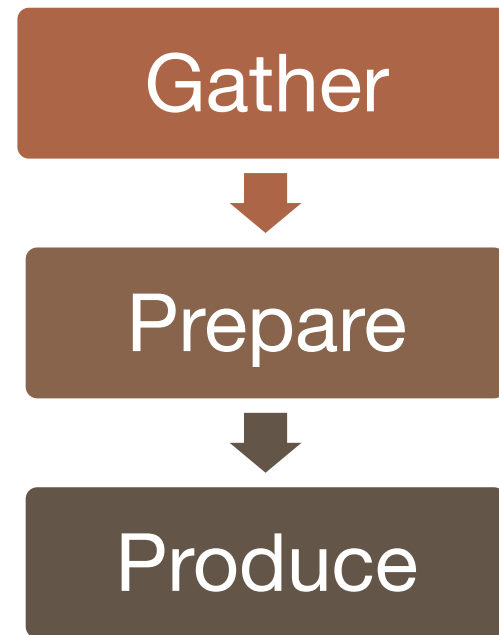
Guardian data blog takes a 70/30 view, but the point remains:

<http://www.theguardian.com/news/datablog/2011/apr/07/data-journalism-workflow>



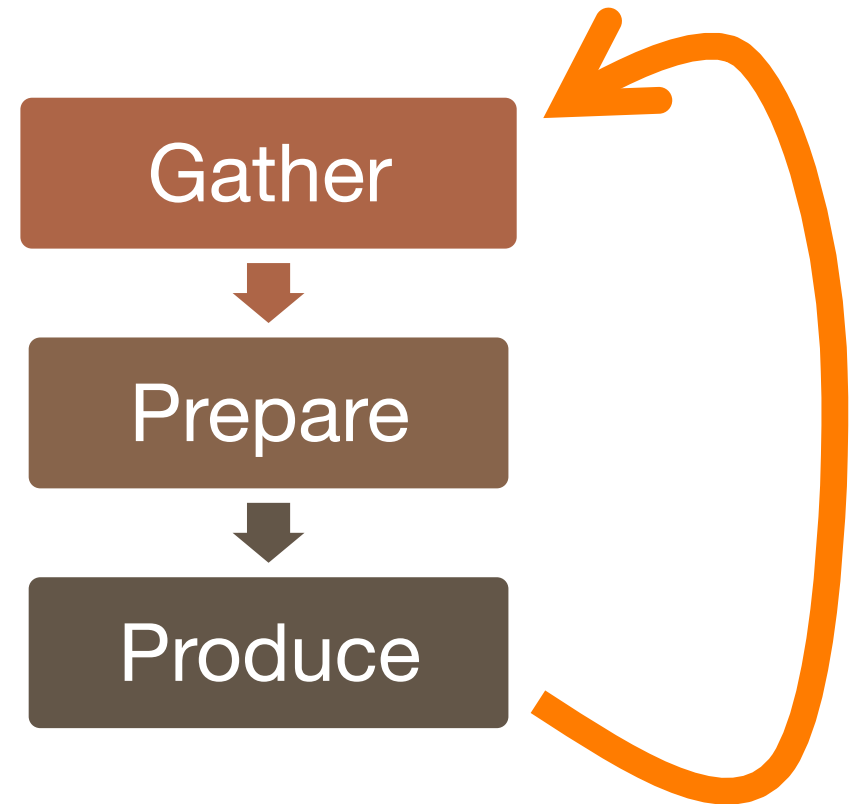
# The data percolator

# Data percolation: The stages of data analysis



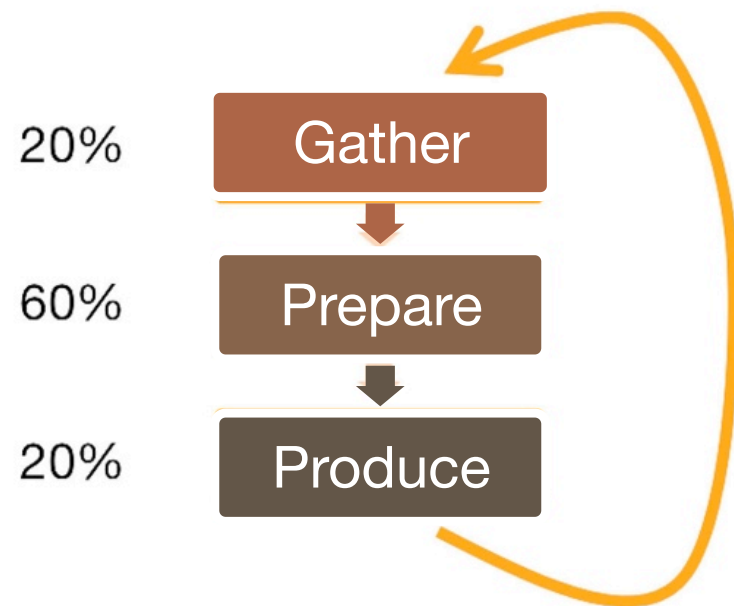
See also: the Data Journalism Handbook

# Data percolation: A model of data preparation and analysis



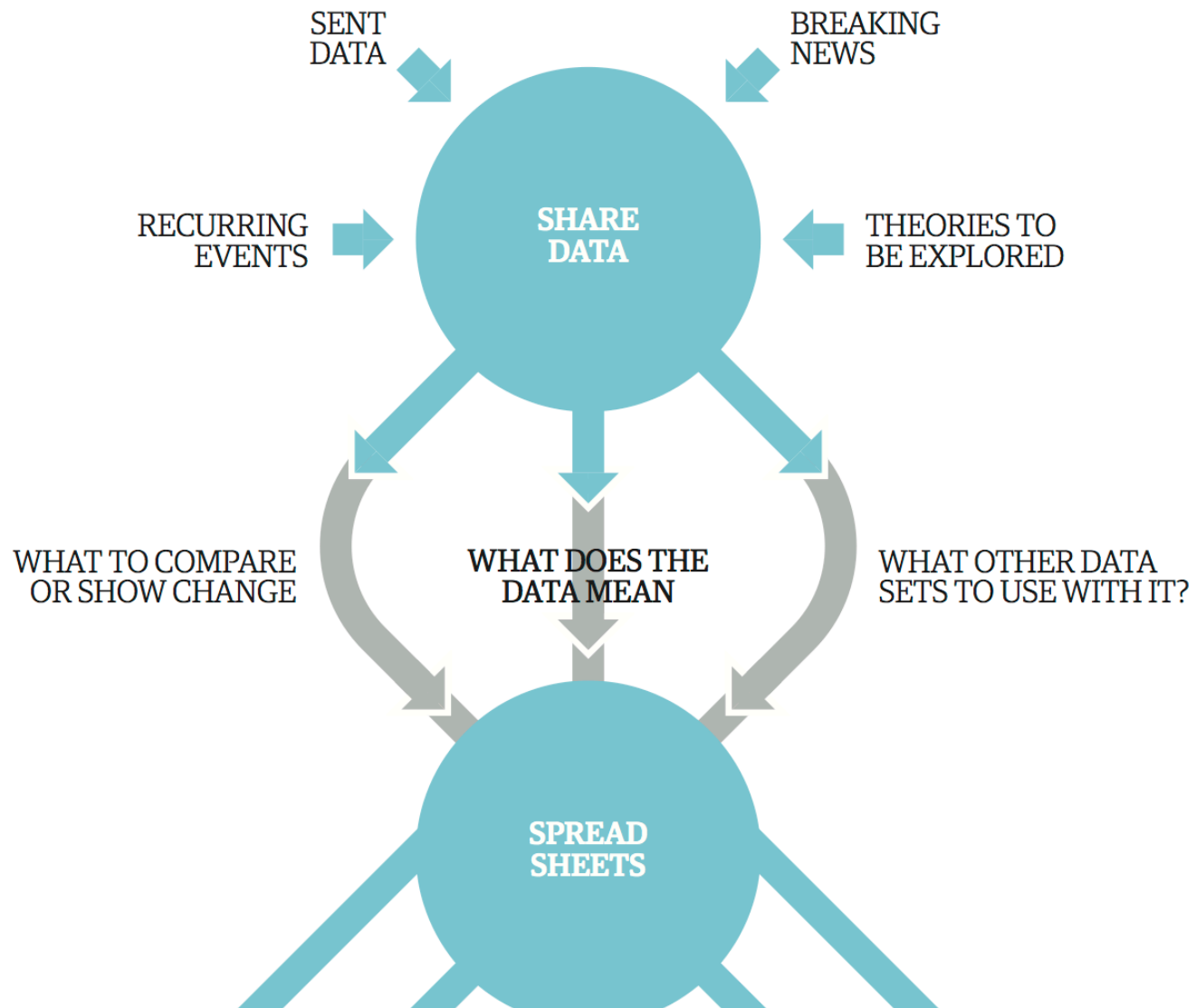


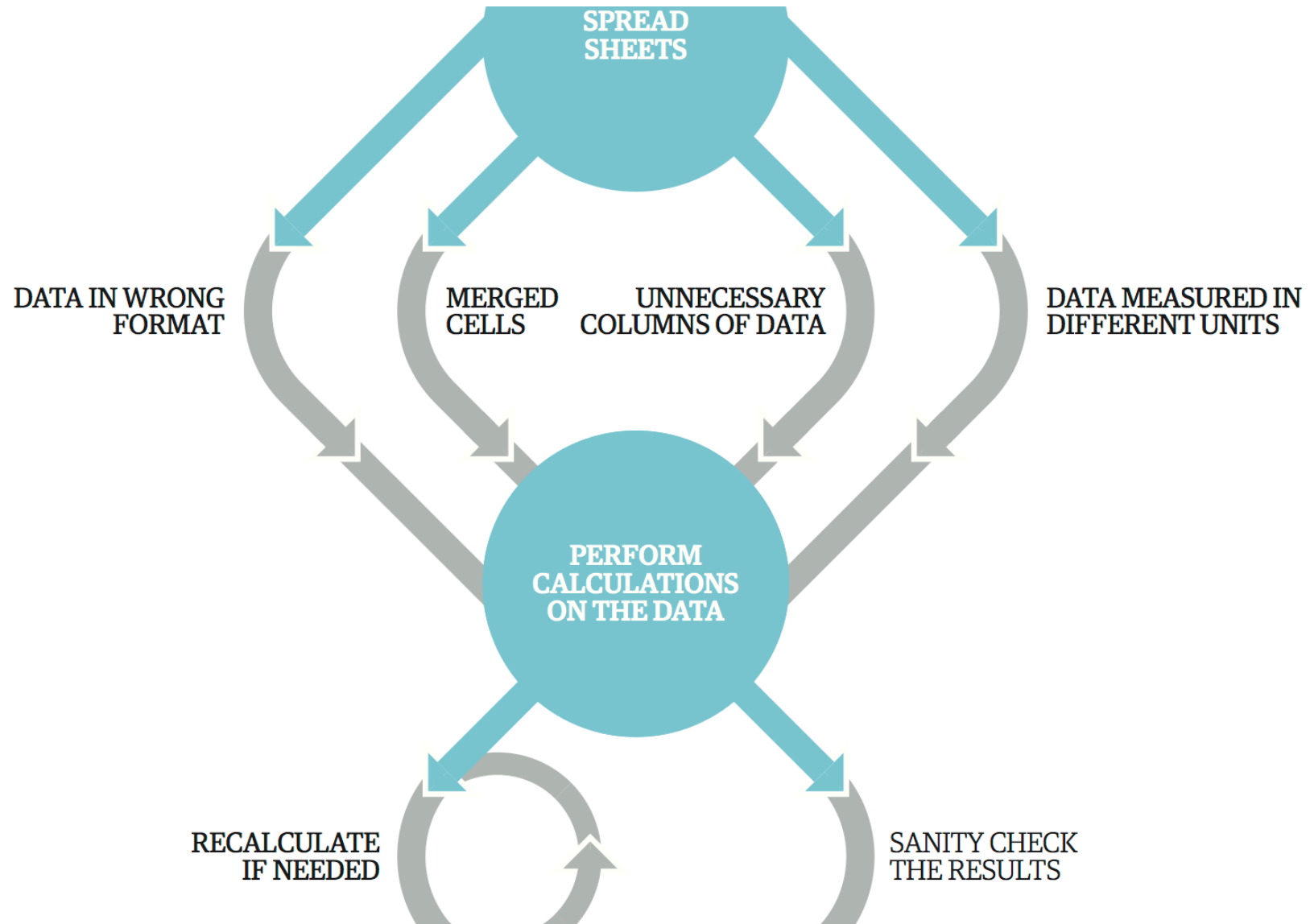
# How should I budget my time?

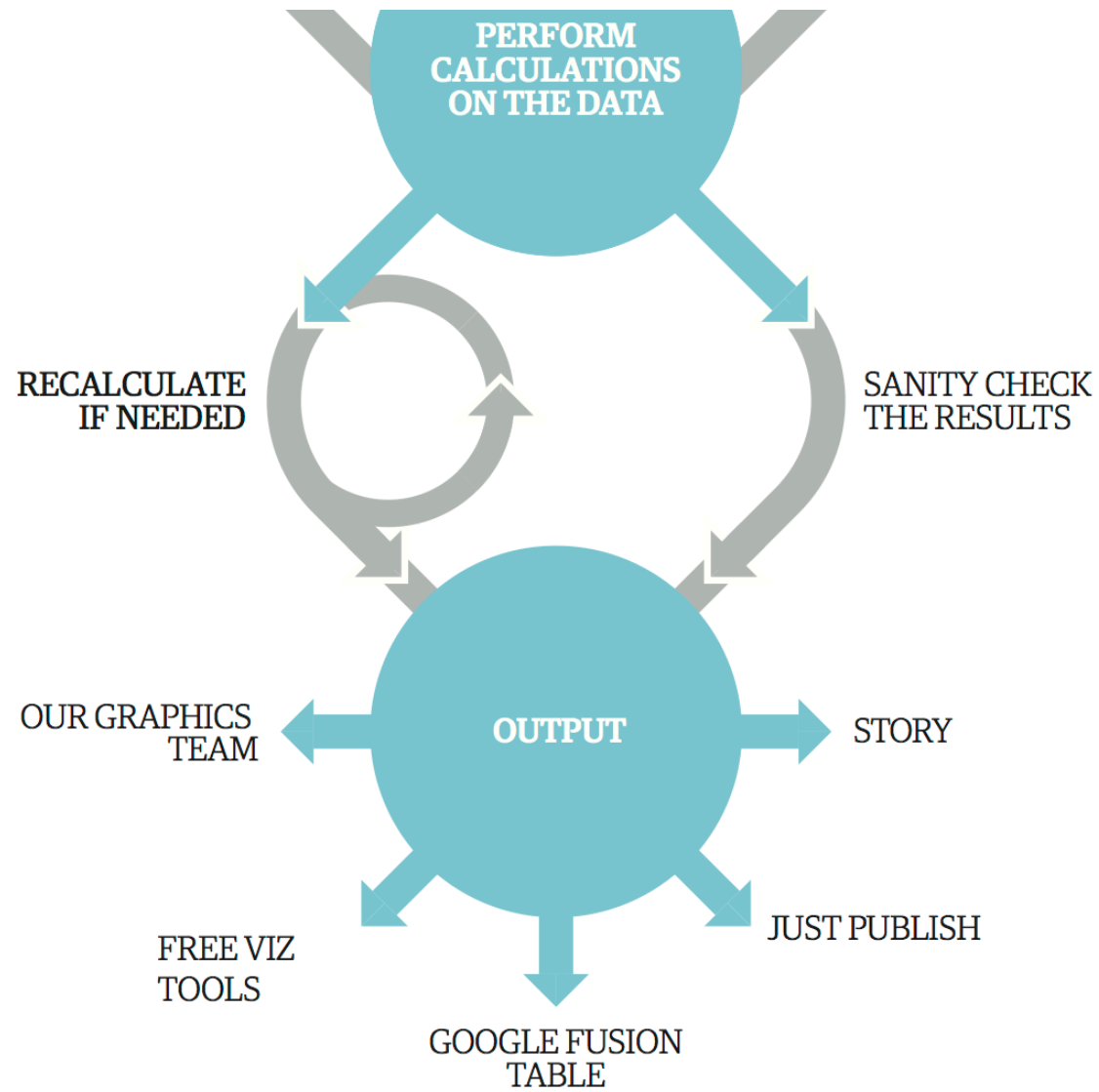


Finding a story is a creative process.

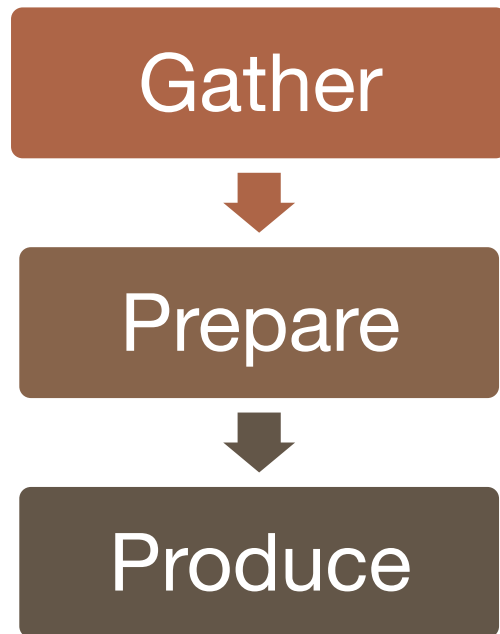
Let it percolate!







# How should I budget my time?



1.1 **FIND** reliable data sources

1.2 Understand your **RIGHTS**

1.3 Visualise and **UNDERSTAND** your data

2.1 **CLEAN** your data

2.2 **TRANSFORM** it where useful

2.3 **COMBINE** it with other data sets

3.1 **REDUCE** and find the story

3.2 Think and understand the **CONTEXT**

3.3 Do your results pass a **SENSE-CHECK?**

# Time planning

Gather

Produce

Prepare

2.1 CLEAN

2.2 TRANSFORM

2.3 COMBINE

2.4 ENRICH

2.5 ANALYSE

# Tools

**2.1 CLEAN**

**2.2 TRANSFORM**

**2.3 COMBINE**

**2.4 ENRICH**

**2.5 ANALYSE**

- Open Refine
- Microsoft Excel
- Scripts / Code
- Publishing platforms
- Datawrapper.de

# Introducing Open Refine

Google Refine 2.0 - Introduction (1 of 3) (vide...

Google refine government IT contracts Permissions Mass edit 2350 cells in column Type of Contract Undo New / Open ... Export Help

Facet / Filter Undo / Redo 5200 rows Extensions: | Freebase

Refresh Reset All Remove All

Type of Contract change

591 choices Sort by: name count cluster

Agreement 30  
HITSS Task Order 32  
CPAF 39  
T&M w/ FFP: Time & Materials w/  
Firm Fixed Price mix 29  
Time and Material 37  
Firm Fixed 28  
Term 25  
Firm Fixed Price 34  
Labor Hours 21  
Seed 31  
FFP LDE: Firm Fixed Price Level

Show as: rows records Show 5 10 25 50 rows

	Contract ID	Contractor Name	Type of Contract	Date of Award	Start Date	End Date	Total value of Contract	Contract Awarded
1.	1939	ADAP SOFTWARE EXPRESS INC DELL MARVETRAD L.P.	Microsoft Enterprise Agreement	04/01/2009	04/01/2009	06/30/2011	1.952	yes
2.	1940	8842 SOFTWARE DISTRIBUTION INCORPORATED	Rarely Service Desk Maintenance	04/01/2009	04/01/2009	03/01/2010	0.891	yes
3.	1941	GOVCONNECTION INCORPORATED	Class Shielded	05/01/2009	05/01/2009	04/30/2011	0.307	yes
4.	1942	ITS CORPORATION	Time & Materials	12/01/2008	01/01/2009	12/31/2011	20	yes
5.	7430	ISNET INTERNATIONAL CORPORATION		05/04/2009	05/05/2009	07/03/2009	0.040167	yes
6.	1945			01/26/2009	01/26/2010	06/30/2010	0.738	yes
7.	1946	IT FEDERAL SALES LIMITED LIABILITY COMPANY		10/01/2009	10/01/2009	09/30/2010	0.349	yes
8.	1947		Firm Fixed Price	03/09/2009	10/01/2009	06/30/2010	0.884	yes
9.	1948		Firm Fixed Price	11/05/2009	11/05/2009	05/30/2010	0.362	yes
10.	1949	REDHAWK IT SOLUTIONS LLC	Firm Fixed Price	01/22/2009	01/01/2010	12/31/2010	0.913	yes

0:00 / 6:48 YouTube

<http://openrefine.org>






# Datawrapper.de

## Datawrapper

An open source tool helping anyone to create simple, correct and embeddable charts in minutes.

[Login / Sign Up](#) [Try it now!](#)



## Easy to use

This little tool reduces the time needed to create a correct chart and embed it into any website from hours to seconds. It makes charting easy, and helps you avoiding common pitfalls.

## Full control

Datawrapper is 100% free software. You can use our free hosted service, or [install it on your own server](#) for maximum control. Written in PHP, Datawrapper is easy to install, modify and extend.

## Customizable

We encourage you to [customize the charts layout](#), fonts and colors to make them integrate nicely in your website. Strengthen your own brand, not ours.



# Exercise

Using datawrapper.de. What insight can you find from the following dataset.

You might need to clean it.

<http://bit.ly/fcowine>



# Outcomes

Understand how the 80/20 rule applies to open data projects

List the stages in an project that uses open data

Apply a number of tools quickly analyse data



Thank-you