Data Equity for Main Street

Handout 6: Data Quality (Optional)

When you decide to use open data, you will go to the site that has the data you need, and hopefully open the data set, and easily start using the data.

But, how do you know whether it's good quality data? How do you know whether your city is using best practices to present the data? What happens if you have problems using the data? What if your city's data aren't even available?

The questions below, from the Open Knowledge Foundation, will help you figure out the quality and completeness of the data you are looking at--if you see problems, you have something to back you up when you ask the city to fix them! And, if you don't see your city's data, we have a link where you can request them to make them open and available!

Open Data Scoring Table

Question	Details	Weighting
Openly licensed?	The licence must comply with the Open Definition which allows data to be freely used, reused and redistributed. The Open Definition provides a list of conformant licences. If the data uses one of these licences, it is openly licensed. Licences are commonly found in: ● the web page footer ● a link to Terms & Conditions ● the About section Some licences may allow re-use and redistribution but have not been assessed as conformant with the Open Definition. In this case, seek feedback on the Open Data Index discussion forum	30

Is the data machine readable?	All files are digital, but not all can be processed or parsed easily by a computer. In order to answer this question, you would need to look at the file type of the dataset. As a rule of thumb the following file types are machine readable: ● XLS ● CSV ● JSON ● XML The following formats are NOT machine readable: ● HTML ● PDF ● DOC ● JIF ● JPEG ● PPT If you have a different file type and you don't know if it's machine readable or not, ask in the Open Data Census forum	15
Is the data available for free?	The data is free if you don't have to pay for it.	15
Available in bulk?	Data is available in bulk if the whole dataset can be downloaded easily. It is considered non-bulk if the citizens are limited to getting parts of the dataset through an online interface. For example, if restricted to querying a web form and retrieving a few results at a time from a very large database.	10

Is the data provided on a timely and up to date basis?	Is the data current for the census year? You can determine or estimate when the data was last updated and its update frequency by reviewing: ● the metadata displayed for the data in an open data portal or web page ● the dataset title or filename e.g. Budget 2013-14 or Election_4July2015.csv ● metadata tags embedded in the web page that contains the data ● date values within the data to find the most recent date value ● the timestamp on the data file (although this may not be accurate) ● Some data is not updated on a regular basis. e.g. Pollutant emissions may be updated daily - while postal codes may not change for many years. You may need to use your judgement to determine if the data is timely and up to date. Document your rationale in the comments section. If you cannot determine a date, answer, "NO" i.e. the data is not timely or up-to-date.	10
Is the data available online?	Data is online if it can be accessed via the Internet (e.g. a website or open data portal). If the data has been emailed to you but is not accessible via the Internet, it is not considered to be available online.	5
Is data in digital form?	Data can be in a digital format, but not accessible online. For example: A country budget can be stored on a spreadsheet or otherwise on a private government network, but not on the Internet. This means that the data is digital, but not publicly available. If you know that the data is digital somewhere inside the government (e.g. a government official tells you so), then you should answer “YES” to this question and note in the comment section how you discovered the data is in digital form.	5

Publicly available?	Can the data be accessed by the public without restrictions? Data is considered publicly available when: ● It can be accessed online without the need for a password or permissions. ● If the data is in paper form, can be accessed by the public, and there is no restrictions on the number of photocopies that can be made. Data is **NOT** publicly available when: ● It is only made available after making a request. ● It was availiable because of FOIA. ● It can only be accessed by government officials.	5
Does the data exist?	Data must come from an official resource either issued directly by the government or by a third party officially representing the government. Data offered by companies, citizen initiatives or any non-governmental organisation do not count for the Index. If the government has given the right to publish the data to third parties, a submission with a link a to third party site is allowed. The third-party site must explicitly state that the data has been commissioned by the government. Check if the organization has an agreement with the government to be the official source and make a note in the comment section.	5

Reproduced from public domain data published by OFKN see *http://us-city.census.okfn.org/faq/*

Messy or Meaningless Data

Sometimes a data steward or publisher may have the best intentions, try to provide accurate, timely and open data, but they may mess up the dataset. Common examples of messy data include datasets where:

- Date fields are inconsistent (think “March 25th, 2015” vs. “3/25/15” vs. “25/03/2015”)
- After all the individual rows are listed, there’s an extra row included with the “Totals” entered as data rather than calculated. This has the effect of doubling all the totals when calculated.
- Shifted columns - halfway through a dataset the data that should be in one column gets bumped to the next column, This often happens when the data is uploaded from a word processor document table, or copied from a plain web page.

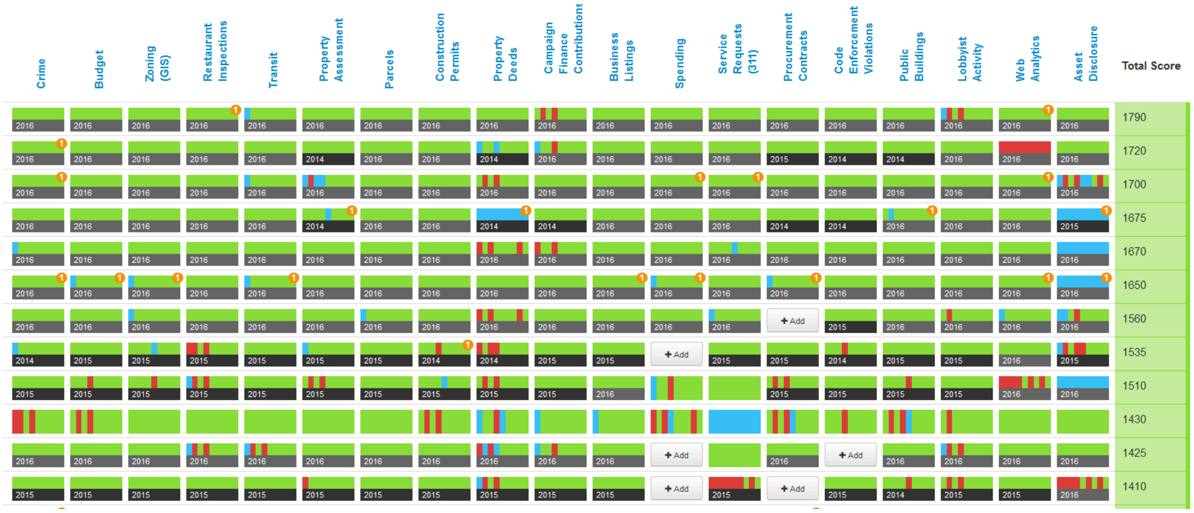
For more (and more horrible) examples of messy data and meaningless data, check out the wall of shame in this project: http://okfnlabs.org/bad-data/ … or this wonky but amusing video from Socrata Connect 2015 https://youtu.be/7F6z_jt6iXs

Report cards on well-known open data portals

Most of the well-known open data portals in the world have already been found and scored in each major subject matter category using the rubric above by advocates in the technology and data science world. The following provide links and snapshots; for the current grade of your favorite city or state, take a look at the live sites. If your city or state is not on the list yet, you can add them by sending in an update.

US City census (by Open Knowledge Foundation)

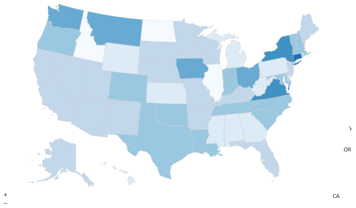
http://us-city.census.okfn.org/



Crime	Budget	Zoning (GIS)	Restaurant Inspections	Transit	Property Assessment	Parcels	Construction Permits	Property Deeds	Campaign Finance Contributions	Business Listings	Spending	Service Requests (311)	Procurement Contracts	Code Enforcement Violations	Public Buildings	Lobbyist Activity	Web Analytics	Asset Disclosure	Total Score
2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	1790
2016	2016	2016	2016	2016	2014	2016	2016	2014	2016	2016	2016	2016	2015	2014	2014	2016	2016	2016	1720
2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	1700
2016	2016	2016	2016	2016	2014	2016	2016	2014	2014	2016	2016	2016	2014	2014	2016	2016	2016	2015	1675
2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	1670
2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	1650
2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	+ Add	2015	2016	2016	2016	2016	1560
2014	2015	2015	2015	2015	2015	2015	2014	2014	2015	2015	+ Add	2015	2015	2014	2015	2015	2016	2015	1535
2015	2015	2015	2015	2015	2015	2015	2015	2015	2015	2016			2015	2015	2015	2015	2016	2016	1510
																			1430
2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	+ Add		2016	+ Add	2016	2016	2016	2016	1425
2015	2015	2015	2015	2015	2015	2015	2015	2015	2015	2015	+ Add	2015	+ Add	2015	2014	2015	2015	2016	1410

State Open Data Census (by US Open Data)

https://census.usopendata.org/#

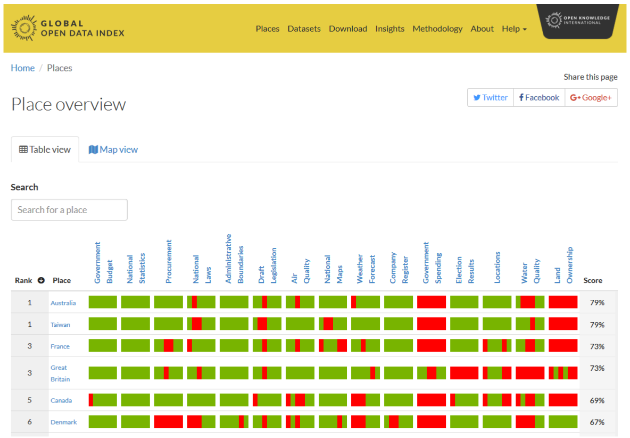


Address Points	Checkbook	Companies	Incarceration	Legislation	Population Projections	Real Estate
F	F	F	F	F	B−	F
F	B	A−	F	F	B−	F
F	B	C	F	F	B	F
A−	B−	D	F	F	B−	F
F	F	F	F	B	B	F
A	A−	A+	F	F	A+	F
A	A	F	A+	A	A+	A+

Link to this handout: https://goo.gl/bssYpz

Global Open Data Index (by Open Knowledge Foundation)

https://index.okfn.org/place/



Place overview

Rank	Place	Government Budget	National Statistics	Procurement	National Laws	Administrative Boundaries	Draft Legislation	Air Quality	National Maps	Weather Forecast	Company Register	Government Spending	Election Results	Locations	Water Quality	Land Ownership	Score
1	Australia	🟩	🟩	🟩	🟩🟥🟩	🟩	🟩🟥🟩	🟩🟥🟩	🟩	🟥🟩	🟩	🟥	🟩	🟩	🟩🟥🟩	🟥	79%
1	Taiwan	🟩	🟩	🟩	🟩🟥🟩	🟩	🟩🟥🟩	🟩	🟩🟥🟩	🟩	🟩	🟥	🟩	🟩	🟩🟥🟩	🟥	79%
3	France	🟩	🟩	🟩🟥🟩	🟥🟩	🟩	🟩🟥🟩	🟩🟥🟩	🟥🟩🟥	🟩🟥🟩	🟩	🟥	🟩	🟥🟩🟥🟩	🟩🟥🟩	🟥	73%
3	Great Britain	🟩	🟩	🟩🟥🟩	🟩🟥🟩	🟩	🟩🟥🟩	🟩	🟩	🟩🟥🟩	🟩	🟩🟥🟩	🟥	🟥🟩🟥	🟥	🟥🟩🟥🟩🟥	73%
5	Canada	🟥🟩	🟩	🟩	🟩	🟩	🟥🟩	🟥🟩🟥🟩	🟩	🟥🟩	🟩	🟥	🟥🟩	🟥🟩🟥	🟥🟩🟥🟩	🟥	69%
6	Denmark	🟩	🟩	🟥	🟥🟩	🟩🟥🟩	🟩🟥🟩	🟥🟩🟥🟩	🟩🟥🟩	🟥🟩	🟩🟥🟩	🟥	🟩	🟩	🟥🟩	🟩	67%

How to get your jurisdiction on the list:

All the major community scorecards have posted links that let you "Add a New Location" or "Submit an Update": for convenience, here they are as of the publication date of this curriculum:

Suggest a new city for the City Census: http://us-city.census.okfn.org/faq#missing-place
Request your own hosted city open data census scorecard
http://census.okfn.org/en/latest/#get-your-local-open-data-census-here

Suggest an edit to the States open data census:
https://github.com/opendata/Open-Data-Census/issues/new

Use the table above and the local knowledge of your library staff, patrons and state experts to inform your submission. When you do that submission, it's a good idea to let the administrator of the portal know you're praising (or critiquing) their work. See the suggested text of a letter in Class 4D - Making the Data Better

Link to this handout: https://goo.gl/bssYpz