Comparing the Distribution of Open Geospatial Information between the Cities of Japan and Other Countries

Toshikazu Seto and Yoshihide Sekimoto

Abstract

In recent years, the concept of "open data" and "open government" has attracted attention from government agencies and international organizations, although greater amounts and types of open data are necessary for the realization of open governance. Because open data vary in format, an analysis that indicates their usability must be conducted; thus, this study examined the present situation of open geospatial data by comparing the number of datasets and response formats released by national and local governments in Japan with the United States (US) and European Union (EU). It was found that open data provided in Japan range from only a few to dozens of datasets; however, format types are limited. In contrast, local governments in the US and EU are more open to the provision of data in various formats. Open data are moving toward incorporating real-time or flowing data, which are essential for the development and study of urban infrastructure.

T. Seto (Corresponding author)
Center for Spatial Information Science, the University of Tokyo,
Meguro-ku Komaba 4-6-1 Ce509, Tokyo Japan
Email: tosseto@csis.u-tokyo.ac.jp

Y. Sekimoto
Email: sekimoto@iis.u-tokyo.ac.jp

1. Introduction

In recent years, the concept of "open data" has attracted attention from government agencies in the United States, Europe, and Japan, as well as from international organizations such as the United Nations, Organization for Economic Co-operation and Development (OECD), and the World Bank. In June 2013, the Open Data Charter was promulgated at the G8 Lough Erne Summit (Gurin, 2014; Sui, 2014).

Open geospatial data have come to be regarded globally as crucial, as their purpose is not only to promote data distribution and public data, which is a primary objective for improving transparency and open governments (Goldstein and Dyson, 2013), but also to facilitate the utilization of open data in the English-speaking world. Moreover, application development has been promoted through "Ideathon" and "Hackathon" events; such activities provide a new method of attracting attention in order to encourage and facilitate citizen participation (McArthur et al., 2012; Kitchin, 2014).

Various policies and government strategies have been carried out since 2012 in Japan (Shoji, 2014), including the introduction of various guidelines and government standard terms and conditions in the past year in order to enhance guidance. On the other hand, since the format of open data currently varies, a cross-evaluation is necessary f to determine the usability of the data, especially in the case of geographical information comprised of latitudes and longitudes, as well as more readable mechanical data. Specifically, the format in which governments use or distribute data is important — in addition to desktop GIS and web GIS — as more general-purpose tools are also an important consideration for the possibility of using open data (Johnson and Sieber, 2012).

Against the backdrop of recent international trends in open data, this study examined the present situation and quantity of open geospatial data. In particular, we investigated open data trends with respect to the number of datasets and response formats based on a comparison of open data released by local governments in Japan with those released by local governments in the United States and European Union.

2. The distribution of open geographic information

With respect to Space Data Infrastructure (SDI), discussions of such platforms and concepts include case studies involving the FGDC (Federal Geographic Data Committee) and INSPIRE (by the EU Commission), which in many cases have been preoccupied with organizing public methods and licenses (Masser, 2011; Harvey, 2011). The Committee of Experts on Global Geospatial Information Management from 1968, 1974, 1980, and 1987 and the United Nations Secretariat have investigated open data in the past using a method of statistical evaluation, although the central issue was the organization of cadastral survey data and land use management of countries.

On the other hand, the distribution trend of open data has been investigated in recent years, including the Open Data Census 2013 (Global Open Data Index) of Open Knowledge and World Wide Web Foundation's Open Data Barometer of 2015. The former is mainly organized by country (recently by city), and has made efforts to score the situation of open data based on 10 open data fields, while the latter is scored by country using an index of 13 factors based on readability.

As described above, there have been a number of initiatives that apply open data indicators to countries, although a global survey of local governments, which are estimated to utilize a lot of geographic information on a detailed scale, has shown little progress. Therefore, it is possible to identify the open data situation of every Japanese local government in this study; we then compare this with overseas cities that have made advanced efforts towards open data.

In the case of Japan, most platforms were collected manually from web pages that published information about cities and were subjected to classification. For overseas cities, the focus was on continuously measuring the flow and automated collection; as a result, it was found that CKAN is the international standard platform. Socrata, on a developed trial basis, uses data acquisition programs to list collections about cities that have adopted Azavea (geospatial analysis system).

Section 3 discusses open data initiatives in Japan (e.g., urban population and number of datasets) weighted by category. In Section 4, we picked chose advanced open data in 10 cities throughout the world to quantitatively compare the number of data sets and applied formats on the basis of mechanical extraction from an open data catalog. In this section, we also used publicly available city data from the OECD and World Bank to compare features of the city, such as population and some indicator, in an attempt to better understand shared features of cities that are aggressively adopting open data. In Section 5, we describe challenges and future

prospects of acquiring the amount of open data distributed, and summarize the results of the analysis.

3. The development of open geographic information within local governments in Japan

At present, 98 local governments in Japan have published open data, including Sabae City of Fukui Prefecture, which was the first to publish data in January 2012 and helped to form the government policy that was adopted in July 2012. The open data roadmap was formulated in May 2013, marking the beginning of advanced efforts toward open data by local governments.

In August 2013, data was published at municipal levels, and open data published by new cities throughout 2014 increased almost every month (Fig. 1, 2). Since October 2014, the number of cities turning toward open data has increased rapidly, including a small municipality in Fukui Prefecture that was considered in the wake of Sabae, which publishes open data jointly. The number of cities using open data has continued increasing constantly to date. In addition to government policies, the Code for Japan is also becoming proactive, and more opportunities for Ideathon(s) and Hackathon(s) are emerging.

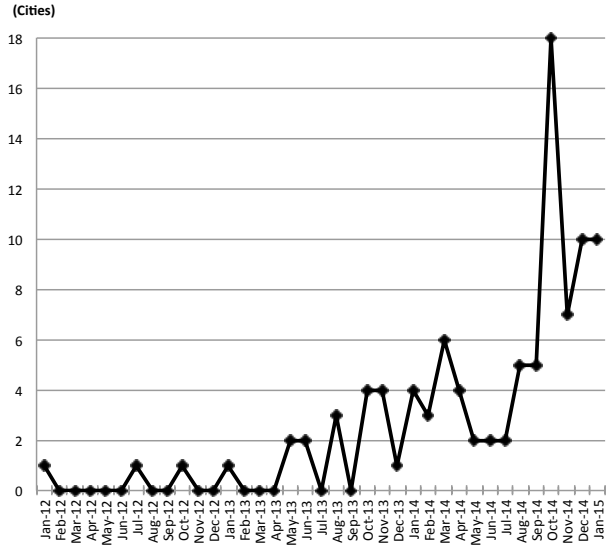

Fig. 1. The transition of published open data in Japanese local governments

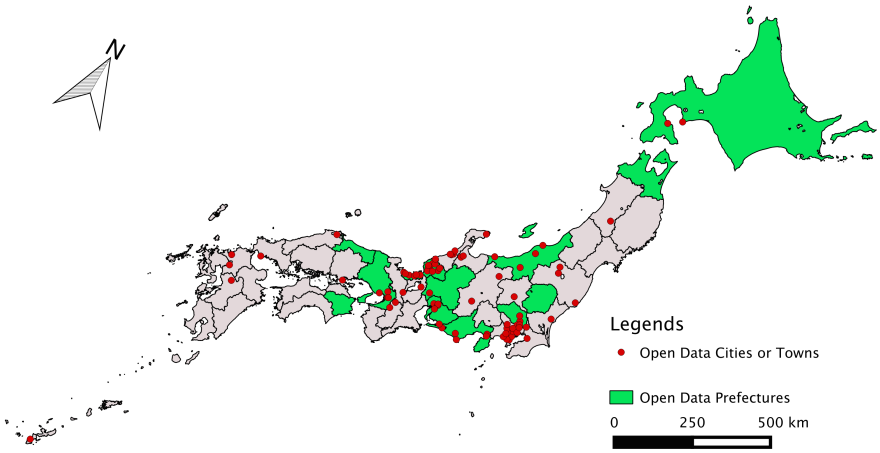

Fig. 2. The location of open data cities among Japanese local governments

Although used in many local governments within Japan, open data portals such as CKAN and Socrata are not nearly as common around the rest of the world. Moreover, the introduction of the CKAN platform has shown little progress, as it is limited to Sabae city of Fukui prefecture, Fukuoka City, and Kitamoto city of Saitama prefecture.

On the other hand, Shizuoka Prefecture utilizes the open source platform CMS by NetCommons (Fig. 3) to publish all open data of small municipalities in the prefecture. Another example is Tokamachi-town of Niigata Prefecture, which takes advantage of the private portal site LinkData.org (Fig. 4), an open data platform developed by Tetsuro Toyoda's RIKEN laboratory. LinkData.org contains a dataset of registration and related apps that can be archived, and links the various efforts surrounding open data, such as ideas. By performing uploads in text or tabular, it is possible to automatically convert to multiple formats, including RDF (Turtle). However, Tokyo Metropolitan Government has never provided an open data and the any platform 2014.

Thus, although open data published at local levels of government in Japan is actively progressing, with the exception of municipalities that have published an application programming interface (API) in CKAN, it has become mechanically clear that it is difficult to perform a quantitative evaluation; therefore, we created a detailed, analog open data list on a trial basis, which was maintained until October 2014. These data were published in the CC-BY through a web page. Our target of 52 cities was reached, constituting 2,040 datasets.

Fig. 3. The open data portal site of Shizuoka using NetCommons

Fig. 4. LinkData.org used by Tokamachi-town of Niigata Prefecture

The classified data categories of Japanese open data are shown in Table 1. According to this table, population and statistical data account for more than 20% of total open data, while the next most common category, public relations, accounts for about 16% of government information. Many of these documents are in formats such as PDFs, but some information, such as the positions of administrative facilities, has been published in a

geographic data format.

The geographic data format is primarily distributed for disaster prevention, education, and tourism sectors, as much of the original data and urban planning diagrams contain positional information regarding facilities. While hazard maps have also been actively published in recent years, they have often remained as PDFs or in image formats. Much of the geographic data (often those that are output from web GIS) that are routinely managed by local governments are open data published on center web pages.

On the other hand, open data published in cities and beyond has been increasing since October 2014, owing to the fact that the number of datasets has almost doubled, and an analysis of the public trend toward open data was carried out in January 2015. There were 4,261 datasets among the 98 local governments, and we carried out our analysis by classifying the local municipalities by population size (Table 2).

Table 1. Category of open data types used by Japanese local governments

	Usual Formats	Geodata Formats	Total	%
Population, Statistics	455	6	461	22.60
Public Relations	253	81	334	16.37
Disaster Prevention	45	188	233	11.42
Education	115	113	228	11.18
Tourism	18	135	153	7.50
Infrastructure	70	75	145	7.11
Agriculture	101	14	115	5.64
Welfare	64	46	110	5.39
Traffic	59	32	91	4.46
Industry	63	21	84	4.12
Environment	52	15	67	3.28
Town Planning	9	5	14	0.69
Crime prevention	5		5	0.25
Total	1,309	731	2,040	100.00

Table 2. Quantity of open data from Japanese local governments

Cities /Formats	Population	ODCities /All Cities	Datasets	Avg.	Location Information	%
Prefecture	>50k	8/47	961	120.1	320	33.3
Government Ordinance City	>50k	10/20	981	98.1	115	11.7
Core City	>30k	9/43	562	62.4	128	22.8
City and Town	<30k	71/1,742	1,757	24.7	483	27.5
Total	–	98	4,261	–	1,046	24.5

At the prefecture level, the ratio of location information is greater than 30%, and in many cases it is progressing in the background of the introduction of web GIS. Half of ordinance-designated cities are utilizing open data; however, the proportion of position information in their datasets is small. Although there are a large number of open data cities, only about 4% of them use position information.

Cities and towns have a high percentage of RDFs because municipal open data is often provided utilizing LinkData.org, as shown in Fig. 4. As more populous cities have one to two stars of open data, cities and towns have an interest in the introduction of such through smartphone applications and web services; therefore, they are more active in open data distribution.

In addition, Figure 5 expresses the relationship between population size and the amount of datasets published by local governments (excluding prefectures) (Fig. 5). A number of open data are provided in a small city outside of Fukuoka. Such a tendency, as seen in the example of Sabae City, can be seen as attempts to revitalize local areas through data-driven efforts based on open data.

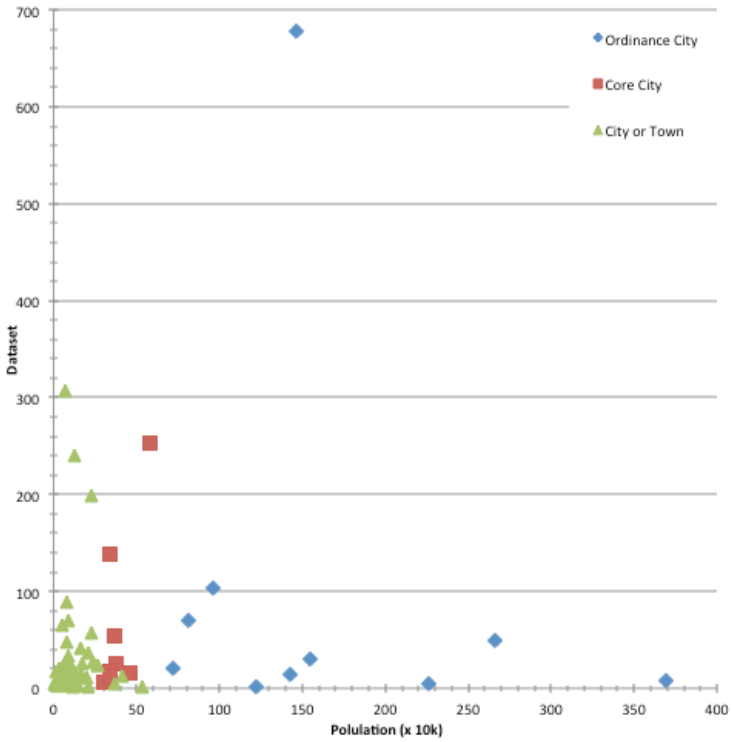

Fig. 5. The population size and datasets of open data cities

4. Comparison with open data in international cities

With respect to international open data, organizations such as Open Knowledge, the World Bank, and OECD (among others) provide summarized guidance and research reports on open data by country. A quantitative analysis at an urban level has not been seen so far for open data platforms that are globally distributed such as CKAN and Socrata, although attempts have been made to analyze Github.

Most open data portals around the world use the CKAN, and more than 120 open data cities have been introduced; 40 of which are city level (Fig. 6). Therefore, 40 cities were targeted, which represent the distribution of cities from United Nations population data and open dataset numbers.

Many cities with a high number of datasets for the population ratio are European cities, especially Helsinki, Finland, although Amsterdam, Netherlands, and other cities such as Glasgow, Scotland, also opened large amounts of data. In addition, the United States widely publishes open data, and with the exception of some cities such as Houston, Texas, and Denver, Colorado, is the majority of these data are published using the Socrata platform. In Philadelphia, Pennsylvania, most open data are distributed via Github (launched by a web GIS platform called Azavea). As a part of this study, 11 typical cities that use the open data platform CKAN (including Fukuoka City of Japan) were subjected to an inter-city comparison (Fig. 7).

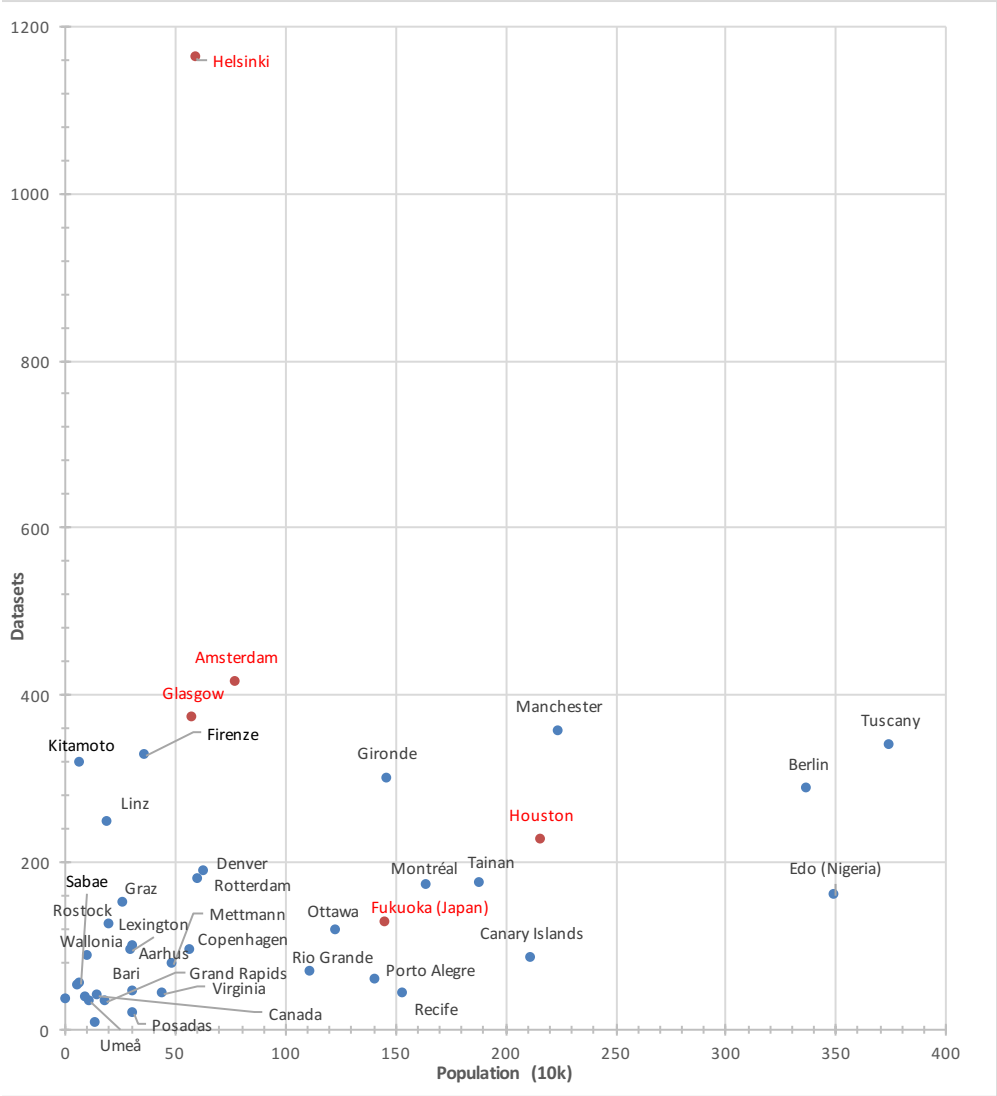

Fig. 6. International comparison of open data cities using CKAN

Table. 3. Basic characteristics of international open data cities

City name	Platform	Total population of metro area (10k)	Urbanized area (km2)	Population density (people per km2)	Employment (10k person)
Amsterdam	CKAN	236.1	593	837.3	119.5
Glasgow	CKAN	94.8	262	1,107.5	44.5
Helsinki	CKAN	145.6	790	229.2	72.8
Houston	CKAN	562.9	4,827	264.8	249.6
Fukuoka	CKAN	252.4	380	2,187.3	118.3
Boston	Socrata	363.9	2,129	753.5	177.5
San Francisco	Socrata	684.8	3,649	400.7	293.7
Chicago	Socrata	931.5	6,303	533.4	431.4
Dallas	Socrata	414.5	3,033	343.5	183.8
New York	Socrata	1,653.9	5,191	1,673.7	758.2
Philadelphia	Github	402.5	2,104	930.2	183.6

International open data cities provide data in multiple formats (Table 4, 5); therefore, a divergence in the number of data lists (datasets) and the actual data (data resources) is observed. Open data cities using CKAN have between 1.1 and 5.4 different formats on average per dataset; in particular, Houston has real data constituting 42 policy categories provided in 19 formats, and provides the largest number of SHP files.

Although cities that use Socrata are often located in the United States, data items are recorded in 1.5 to 2.4 different formats on average worldwide. This is due to the technical specifications of Socrata, which generally uses four formats: CSV, JSON, XML, and RDF. Otherwise, the Google Maps API is available for the visual interface, including maps and graphs.

The amount of geographic information as a percentage of open data is different for each city, with an especially high proportion of more than 80% in Glasgow. However, New York, New York; San Francisco, California; and Chicago, Illinois use varying formats; in particular, New York and San Francisco have developed formats such as Geo JSON and KML because these formats are suitable for web applications.

On the other hand, Amsterdam and Helsinki used CKAN for their portal sites, and geographic information on them is substantial because it is linked with external applications such as web GIS. Similarly, Philadelphia also added to the actual data provided in Github, as it supports the provision of data in a variety of APIs as a catalog.

Table 4. Quantity of open data from international open data cities

City Name	Platform	Datasets	Resources	Category	Format	Avg. Datasets	GIS Datasets	GIS Datasets (%)	GIS Format Types
Amsterdam	CKAN	157	479	18	9	3.1	78	16.3	4
Glasgow	CKAN	372	512	13	13	1.4	429	83.8	12
Helsinki	CKAN	1,163	1,292	19	13	1.1	79	6.1	4
Houston	CKAN	225	357	42	19	1.6	199	55.7	6
Fukuoka	CKAN	126	684	7	4	5.4	11	1.6	1
Boston	Socrata	319	487	11	5	1.5	46	9.4	4
San Fransisco	Socrata	826	1,748	11	4	2.1	577	33.0	4
Chicago	Socrata	1,011	2,140	16	4	2.1	520	24.3	4
Dallas	Socrata	55	96	5	4	1.7	9	9.4	4
New York	Socrata	3,686	5,120	12	4	1.4	625	12.2	4
Philadelphia	Github	228	536	15	8	2.4	196	36.6	6
Total		8,168	13,451	–	–	–	2,769	–	–

Table 5. Format types among international open data cities

Cities /Formats	Usual Formats						Geodata Formats		Others Apps	Total
	Document ★	XLS ★★	CSV ★★★	XML ★★★	API ★★★★	RDF ★★★★★	SHP ★★	JSON・KML・GML ★★★★		
Amsterdam	52	17	31	0	12	0	30	11	326	479
Glasgow	15	56	337	6	0	0	40	54	4	512
Helsinki	86	922	27	17	0	0	63	29	148	1,292
Houston	50	91	10	3	6	0	185	10	2	357
Fukuoka	46	610	22	0	0	0	0	0	6	684
Boston	207	0	56	0	0	56	56	112	0	487
San Fransisco	0	0	437	437	0	437	0	437	0	1,748
Chicago	0	0	535	535	0	535	0	535	0	2,140
Dallas	0	0	24	24	0	24	0	24	0	96
New York	0	0	1,280	1,280	0	1,280	0	1,280	0	5,120
Philadelphia	183	3	13	9	110	0	118	5	95	536
Total	639	1,699	2,772	2,311	128	2,332	492	2,497	581	13,451

5. Conclusions

This study found that, in terms of quantity, open data provided in Japan range from only a few to dozens of datasets; however, format types are limited. In contrast, local governments in the US and EU are more open to the provision of data in various formats. Some local governments in the EU have adopted the open-source data portal platform CKAN, and provide open geospatial data in various formats. In the US, local governments use content management systems that allow data to be organized by format; it was found

that the proportion of geospatial data was relatively high. Local governments using Socrata offered a greater variety of open data pertaining to map reduction and visualization compared with those using CKAN.

To support the distribution of open data, further study is necessary in regard to data characteristics that were not addressed in this study (e.g., positional accuracy and update frequency); moreover, building applications are critical. In Japan and New York, open data released in recent years include real-time data on expansion and space utilization rather than static geospatial information, which are essential to dynamic simulations and the development and study of urban infrastructure. In general, there is a continuing need for further comparative studies on the utility of open geospatial data in decision-making.

Acknowledgements

This research was supported by JSPS KAKENHI Grant Numbers 25870907, 25244042, 15K03008.

References

Goldstein, B. and Dyson, L. eds. (2013). Beyond transparency: Open data and the future of civic innovation. San Francisco: Code for America Press.

Gurin, J. (2014). Open data now: The secret hot startups, smart investing, savvy marketing and fast innovation, New York: McGraw-Hill Education.

Harvey, F. (2011). Spatial data infrastructure for cadastres: Foundations and challenges, in Nyerges, T.L., Couclelis, H. and McMaster, R. (eds.), The SAGE handbook of GIS and society, London: SAGE.

Johnson, P.A and Sieber, R.E. (2011). Motivations driving government adoption of the Geoweb, GeoJournal, 77(5), 667–680.

Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences, London: SAGE.

Masser, I. (2011). Emerging frameworks in the information age: The spatial data infrastructure phenomenon, in Nyerges, T.L., Couclelis, H. and McMaster, R. (eds.), The SAGE handbook of GIS and society, London: SAGE.

McArthur, K., Lainchbury, H. and Horn, D (2012). Open data hackathon, how to guide, http://goo.gl/W2sr2

Shoji, M. (2013). Utilizing Open Data: 8. Preparation for the reuse of data in Japan, IPSJ Magazine, 54(12), pp.1244–1247.

Sui, D. (2014). Opportunities and impediments for open GIS, Transactions in GIS, 18(1), 1–24.