

Workshop on the Communication of Statistics

20 April 2015

27–29 April 2015, Washington D.C., United States of America

Projects of Open Data for Official Statistics of Japan

Toshihiko Akatani (National Statistics Center, Japan)

Abstract:

As central National Statistical Offices (NSOs) in Japan, the Statistics Bureau of Japan (SBJ) and the National Statistics Center (NSTAC) are making progress in the following projects, for such purposes as to promote advanced use of statistics by the public and private sectors, and to support both the creation of novel data services that generate new added value and the creation of innovative business.

1. Development of an environment for advanced use of statistics by API

The API functions, introduced in “e-Stat” (the portal site of official statistics of Japan; developed by SBJ and operated by NSTAC) in October 2014, enable acquisition of the 57 types of governmental statistics stored in the e-Stat database in machine-readable formats.

The functions enable automatic retrieval of the statistical data in e-Stat, as well as more advanced data analytics coordinating other data held by users.

2. Improvement of statistics GIS

In January 2015, a new statistics GIS, “jSTAT MAP –Small area analytics on maps–” was released on e-Stat. It has new features of importing and analyzing any data owned by users, compiling statistics in any selected area, and creating statistical reports in any selected area.

With respect to the API functions in particular, SBJ and NSTAC have been developing an environment for the use of API for all governmental statistics, by effectively exploiting the e-Stat database (launched in 2008).

For the future, the further compilation, in a database, of statistical data not produced by the central NSO, the development of metadata, multilingualization, and conversion to Linked Open Data (LOD) are issues that need to be tackled.

SBJ and NSTAC will continue to lead the government’s initiatives as top runners in the field of Open Data.

Keywords: open data, application programming interface, geographic information system

Projects of Open Data for Official Statistics of Japan

Toshihiko AKATANI

Deputy Director,

Corporate Development Office, National Statistics Center

19-1 Wakamatsu-cho, Shinjuku-ku, Tokyo, 162-8668, Japan

Tel: +81-3-5273-1206

e-mail: takatani@nstac.go.jp

1. Introduction

With respect to the provision of official statistics, utilization of the Internet is already becoming popular in each National Statistics Office (NSO). However, against the background of rapid progress of information and communication technologies (ICT), the potential new ways to utilize statistics data are expanding. Therefore, addressing this issue is one of the top priorities for NSOs.

In particular, promoting the utilization of official data—i.e. facilitation of “Open Data”—is recently expected as an important tool for the creation of new businesses and services, realization of public service through cooperation between the public and private sectors, and assurance of transparency and accountability of government. The statistics sector is needed to play a forward-looking role and lead as a top runner of Open Data in the government, not only within each country but also internationally.

Considering the above situation, the Statistics Bureau of Japan (SBJ) and National Statistics Center (NSTAC) launched a project named “Gateway to Advanced and User-friendly Statistics Services” (GAUSS) in 2011, to promote advanced use of statistics by the public and private sectors and support for creation of services that generate new added value and innovative businesses. This project intends to upgrade the methods for disseminating voluminous and diverse statistical data to the next generation level, and enable their advanced use.

The project includes following themes:

Project 1: Development of an environment for advanced use of statistics by API

Project 2: Improvement of statistics GIS

This paper overviews each project and provides a glimpse of future work and implications for NSOs.

2. Project 1: Development of an environment for advanced use of statistics by API

The first project is “Development of an environment for advanced use of statistics by API.” API stands for “Application Programming Interface,” and it enables software components to interact with each other.

The API functions produced by SBJ and NSTAC offer automatic retrieval of the statistical data stored in the database of “e-Stat” (the portal site of official statistics of Japan; developed by SBJ and operated by NSTAC) via programs. In the past, e-Stat users retrieved statistics data by manually searching and downloading tables.

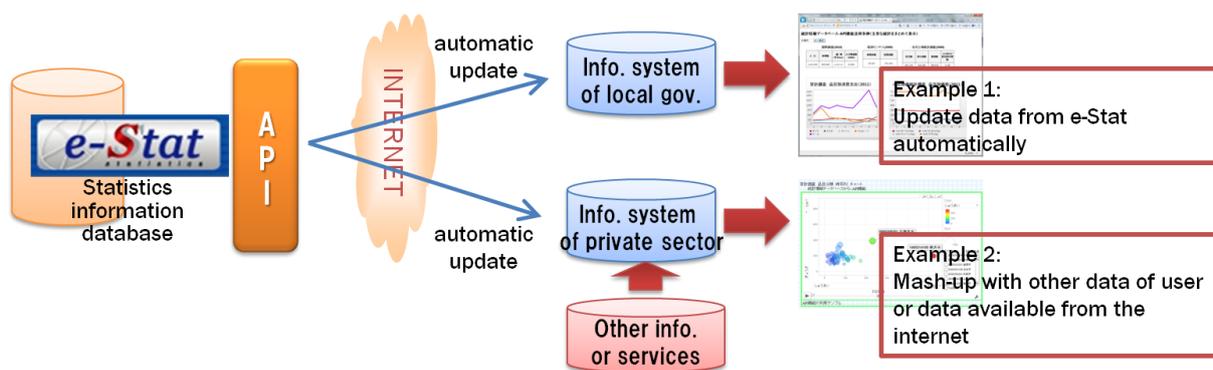


Figure 1: Overview of the API functions

With the API functions, for example, local governments and companies could streamline a cumbersome routine job of updating statistics data in their information system from e-Stat.

Moreover, the API functions would enable more advanced data analytics, coordinating other programs available from the Internet.

Here is one of the examples powered by the API functions:

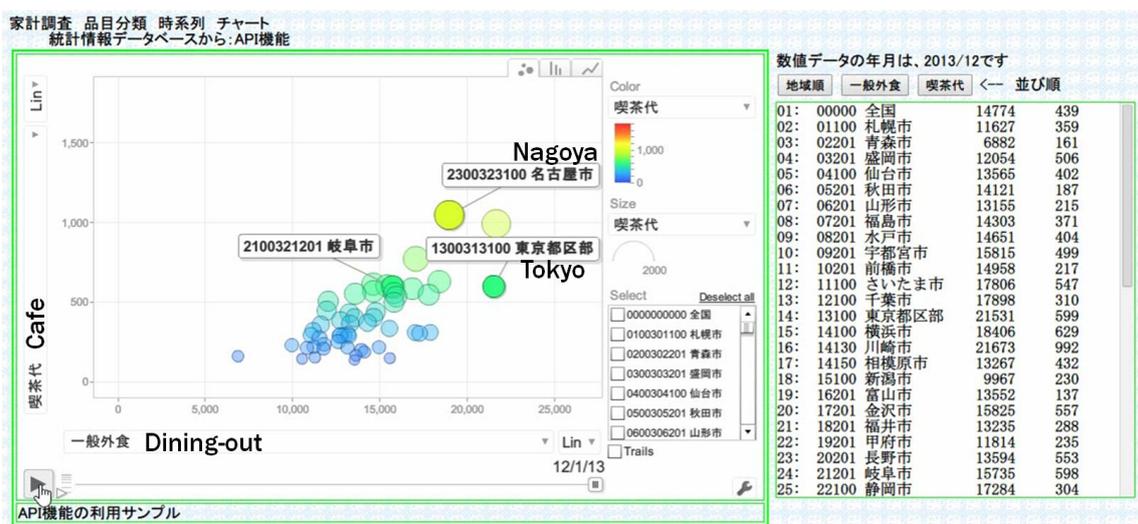


Figure2: Example of analysis powered by the API functions

The bubbles in the scattered chart represent statistics data of 52 major cities in Japan derived from the Family Income and Expenditure Survey from January 2007 to December 2013. The horizontal axis shows average expenditure for dining-out of the households in the respective cities and the vertical axis shows that for cafes (a part of dining-out).

The dynamics of monthly expenditure are displayed with the motion of each bubble by adopting an Open API (Google motion chart). Among big cities such as Tokyo that have increasing demand in the restaurant industry, Nagoya has a unique preference for cafes, which is known as a custom of fancy “breakfast specials.”

The motion-bubble scatter chart demonstrates not only new ways for expression of statistics but also indicates

how easy it is to realize such new expressions by “mash-up” (collaboration with statistics APIs and other programs). This is one of the most attractive points of the API functions.

A trial run of the API functions has started on 10 June 2013, which provides access to official statistics held by SBJ (Population Census, Family Income and Expenditure Survey and so on; 22 statistics, 32,000 tables).

Following the trial run, based on analysis of use, understanding needs, and validating the system, the API functions have been formally incorporated in e-Stat since October 31, 2014, to make all ministries’ statistical data (School Basic Survey, Patient Survey and so on; 57 statistics, 70,000 tables) available in machine-readable formats.

Since then, the API functions have attracted a lot of data users, from mass media to individual people, who have developed hundreds of application programs and released them on the internet, such as “Comparison of various data in each district on a map,” “Web site about the ranking of prefectures,” “Website which uses graphs to visualize trends of population and businesses based on time series data,” and “Mobile application about official statistics.” Every case contains developer’s ingenuity to utilize statistics data from users’ various viewpoints.

In this way, providing API functions enables big portions of statistical data to be machine-readable, and may contribute to outstanding effects on social and economic activities.

3. Project 2: Improvement of Statistics GIS

Statistics GIS (Geographic Information System) is a website that provides an integrated demonstration of statistics information and geographic information to comprehend statistical data on a map. Now, e-Stat has been providing Statistics GIS as a “Thematic Map.” The GAUSS project has developed a new function to enhance the Statistics GIS, by adding a new feature to import any data owned by users in an arbitrary area over the website. On January 20, 2015, a new statistics GIS “jSTAT MAP –Small area analytics on maps–” started offering the above functions.

Below is an example of analysis using jSTAT MAP.

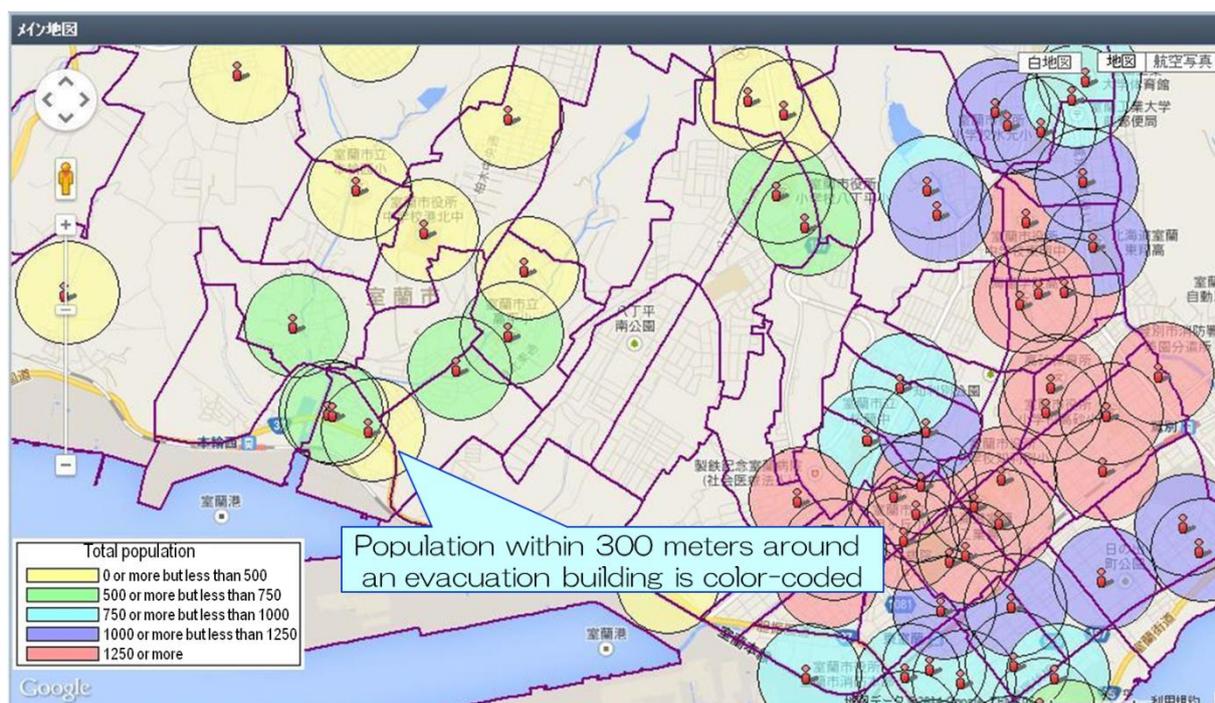


Figure 3: Example of analysis using jSTAT MAP

In the analysis, municipal information on evacuation buildings or shelters within a city in the case of a disaster, which is published by the local government as a part of the Open Data policy, is incorporated into the Statistics GIS (plots in human form represent evacuation buildings). If you have address information for each building, you can import the data to the map by geo-coding, which converts the address into the coordinates of longitude and latitude.

Each color-coded circle shows the population within 300 meters around an evacuation building. The detailed population of each area is calculated by proportional estimation using small area results of Population Census and is displayed on each circle.

If the local government publishes the capacity of each building, it will be possible to compare the capacities and the populations. Furthermore, confirming the capacity of a neighborhood park may enable the identification of the most appropriate evacuation route in the area. Also, an estimation of population is available not only in circled areas but also in arbitrarily designated areas.

Visualization like this will help to make various considerations easy, including which building vulnerable people, such as elderly people, babies and toddlers, should evacuate to and which building should accept a request for support from the local government. This will enable governments to develop a more detailed policy regarding disaster reduction and sightseeing than they have now, exploiting jSTAT MAP.

Aside from making it easier to utilize statistics, the enhanced Statistics GIS will provide a major advantage in that Open Data published by other ministries in central governments or local governments and statistical data can be systematically linked. Statistics GIS is expected to have the role of stimulating and activating Open Data initiatives in other administrative sectors.

4. Characteristics of Open Data policy in an official statistics sector of Japan

As mentioned in the Introduction, the statistics sector needs to play a forward-looking role and lead as a top runner of Open Data in the government, not only within each country but also internationally. What this means specifically is thought to be the following, based on Japan’s experience.

First, there is the fact that a one-stop service of official statistics was laid out many years before data catalogue sites were developed in several countries recently. In Japan, a data catalogue site “data.go.jp” has been operated as full-scale service from October 1, 2014. However, “e-Stat,” the portal site of official statistics in Japan, was launched on April 1, 2008. It could be argued that the official statistics sector was completely playing the role of Open Data even before the term “Open Data” was coined.

In addition, regarding the API function, SBJ and NSTAC have accomplished the development of an environment for use of API focused on all governmental statistics, effectively utilizing the database of e-Stat. In e-Stat, governmental statistics are stored not only in Excel (xls) and CSV files but also in database formats. The API function introduced this time has the biggest advantage in that we can access necessary information via the existing database of e-Stat and retrieve data in certain formats, such as XML and JSON. This avoids the necessity to develop a database separately for installing the API function and enables the API function to cover statistics produced by ministries other than NSOs.

Meanwhile, most administrative data in sectors other than official statistics stored in data.go.jp are in PDF format, which is not machine-readable. Compared with this situation, it is recognized that the situation of data development in the official statistics sector is in one of the most advanced.

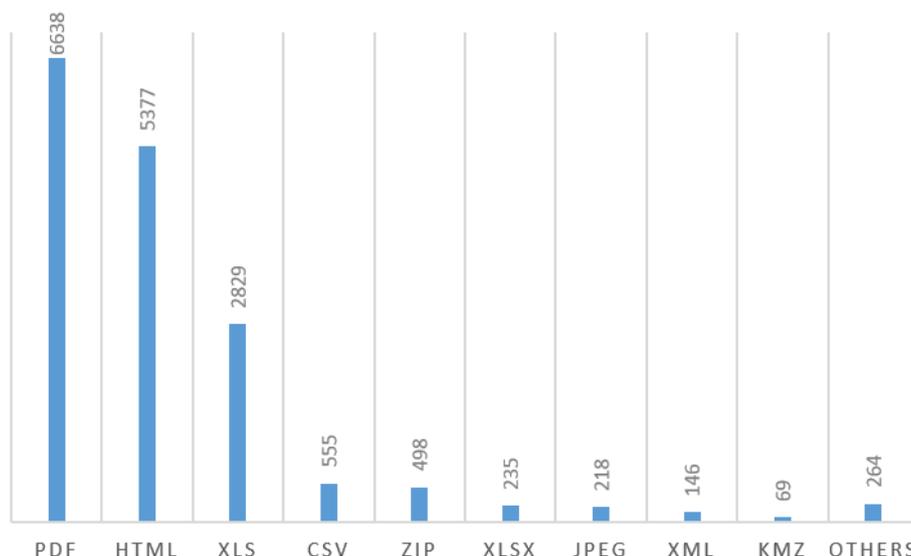


Figure 4: Datasets available in data.go.jp (as of 13 March 2015)

Factors which led to this situation may include the fact that the official statistical system of Japan is highly decentralized. Instead of permitting each ministry to conduct surveys and produce the statistics it needs, the enhancement of the control tower function as a policy coordinator and the development of one-stop service have been promoted for a long time. This could have contributed to the current Open Data policy of Japan.

Based on this present status, reference to a link to e-Stat is regarded as registration of official statistics data itself in data.go.jp.

5. Ideathon, Hackathon

In Japan, a contest of ideas and programs exploiting Open Data, called “LOD Challenge,” has been held under the academic-industry partnership from 2011. SBJ and NSTAC have participated in this contest as “Data Provision Partners” since 2013 and raise outstanding examples of use of statistics data from the public.

Also, SBJ and NSTAC held Ideathon and Hackathon internally in 2013. Excellent works in the contest were adopted for a smartphone application, called “App On Statistics,” by SBJ and this application has been provided to the public since April 2014.



Figure 5: Interface of App On Statistics

6. Future work

As described above, various initiatives have been already done for Open Data of official statistics of Japan. Nonetheless, the following issues are recognized as future important work to upgrade the initiatives to a higher level.

6.1 Compilation database of official statistics produced by non NSOs

Since all governmental statistics stored in a database format in e-Stat are within the scope of API functions, the next task is further compilation in databases.

Although there are 57 types of statistics in a database format, more than 500 types of governmental statistics are covered by e-Stat. Therefore, only about 10% of statistics in e-Stat are compiled in a database. However, not all ministries which are not NSOs have skills and know-how of compilation for existing statistics.

In the future, SBJ and NSTAC will support such ministries to expand the scope of statistics available in API

functions.

6.2 Facilitation of metadata

When making statistics data machine-readable, also making metadata machine-readable will contribute to an appropriate understanding of statistics.

However, metadata of official statistics in Japan is not always standardized. Furthermore, to make metadata really available, it is necessary to clarify the name of metadata as well as a definition for it. With respect to this issue, each country has to work in a coordinated manner, in collaboration with international initiatives such as SDMX (Statistical Data and Metadata eXchange).

Also note that NSTAC currently conducts a trial run of provision of statistical data as SDMX.

6.3 Multilingualization

The viewpoint of international cooperation is essential for Open Data policy. International organizations such as OECD and IMF are promoting data provision as SDMX and so on, mentioned in 6.2. With regard to upgrading Open Data, it is necessary to take note to provide statistical data in an internationally-comparable form.

In Japan, many statistics data and metadata are provided only in Japanese and are not adopted for international data retrieval. Hence, Japan's NSOs have a big challenge of multilingualization, in particular translation into English.

6.4 Coping with LOD

As Tim Berners-Lee (known as the inventor of the World Wide Web) advocates, there are the following five levels of Open Data.

API functions now available are at level three. In the future, transformation to the highest level will be facilitated to enable interaction with various data on the web.

NSTAC conducted a trial run of provision of municipality code data as LOD (Linked Open Data) from December 6, 2013. It is planned to start consideration regarding how to provide statistical data itself as LOD.

Stars	Description	Examples
★	Available online, openly licensed, in any electronic format	Statistics in non-machine-readable tables (GIF or JPEG), unstructured HTML, or embedded in PDF reports. Data is publicly available, but difficult to search, and must be re-entered by hand.
★★	Available online, openly licensed, in common electronic formats.	Data files in proprietary formats such as Excel, SPSS, SAS or STATA, which require special software and training.

★★★	Available online, openly licensed, in non-proprietary electronic formats.	Data files in open formats such as CSV, JSON, XML, DDI, SDMX, or structured ASCII
★★★★	All of the above, plus use of unique URIs (unique Internet identifiers) to identify and define data.	Data files in Linked Data formats such as RDF, allowing data to be interlinked with other data files easily.
★★★★★	All of the above, and links to other data to provide context	Dataset pages provide machine-readable metadata, and links to standard definitions and related information

Source: Open Data Challenges and Opportunities for National Statistical Offices (January 18, 2014, Distributed at UNSC Side Event: Open Data for NSOs)

Figure 6: Five Star Open Data

Keeping in mind that official statistics are national and social infrastructure, NSOs are required to realize more convenient and advanced use of statistics, as well as to contribute to social and economic development, addressing Open Data policies.

SBJ and NSTAC will continue to lead the government's initiatives as top runners in the field of Open Data.

References:

- [1] e-Stat (Portal Site of Official Statistics Of Japan),
<http://www.e-stat.go.jp/>
- [2] Projects of Open Data for Official Statistics, Statistics Bureau of Japan, 2013,
<http://www.stat.go.jp/english/info/news/pdf/1957z1.pdf>
- [3] DATA.GO.JP
<http://www.data.go.jp>
- [4] The World Bank, Open Data Challenges and Opportunities for National Statistical Offices (Distributed at UNSC Side Event: Open Data for NSOs), 2014,
https://unstats.un.org/unsd/statcom/statcom_2014/UNSC-OpenDataAndNSOs.pdf