Featured Articles

# Latest Trends and Future Outlook for Open Data: Making Public-sector Data Available to the Private Sector

Daisuke Hasegawa
Yu Asano, Ph.D.

*OVERVIEW: There has been considerable activity in recent years in both Japan and elsewhere in open data initiatives, whereby public-sector data held by national, local, and other government agencies is made available for secondary use. In its strategy entitled, "Declaration to be the World's Most Advanced IT Nation," the Japanese government is also pursuing open data, setting a target of making available a level of data that is equal to other developed nations by the end of FY2015. Along with supplying an Open Data Solution that covers all of the processes needed to implement open data at public institutions and supports all activities from the planning associated with the introduction of open data to the preparation, publishing, and use of the data itself, Hitachi is also working on the research and development of the next generation of technologies for improving the usefulness of data.*

## INTRODUCTION

"OPEN data" means making data with a high level of public interest that is held by national, local, and other government agencies available on the Internet so that it can be accessed by the general public, and the activities associated with doing this. Unlike past information release practices, open data initiatives involve making data available in data formats and with licenses to use that facilitate its secondary use. Government initiatives in this area got underway with Japan's 2012 Open Government Data Strategy. The Japanese government is actively pursuing open data initiatives, including having identified such endeavors in the 2013 "Declaration to be the World's Most Advanced IT Nation" as being among activities that will help establish the type of society it aspires to realize, and a further expansion in activities is evident, facilitated by public interest.

This article describes trends in open data, in the public and private sectors, what Hitachi is doing to support progress on open data, advanced technologies for making data more useful, and the outlook for open data in the future.

## TRENDS IN OPEN DATA AND INITIATIVES

### Trends in Making Open Data Available

In the early 2000s, open data initiatives aimed at improving government transparency spread around the world, starting in the UK, with government data catalogs already available in 45 different countries[1]. With reference to the G8 Open Data Charter adopted in June 2013, Japan launched its DATA.GO.JP government data catalog in October 2014. Public-sector data that is in high demand such as statistical or geospatial information held by central government ministries and agencies was released as open data, with cross-agency searching and collation available through DATA.GO.JP.

In local governments of Japan, early-adopter local government agencies have preempted central government initiatives by providing open data, particularly data of relevance to dealing with local issues such as welfare, tourism, and political expenditures. In Fukui Prefecture, examples can be seen of work that aims to facilitate the use of data in ways that extend across all the municipalities in the prefecture, such as having municipalities release data in a common format. Already more than 100 local government entities are working on open data (as of the end of June 2015), and these are increasing at an accelerating rate, with the expectation of further expansion in the future prompted by government measures for encouraging wider adoption (see Fig. 1).

The USA and the UK, meanwhile, which are recognized as leaders in open data, have released 141,218 and 25,664 datasets respectively through their government data catalogs (as of the end of June

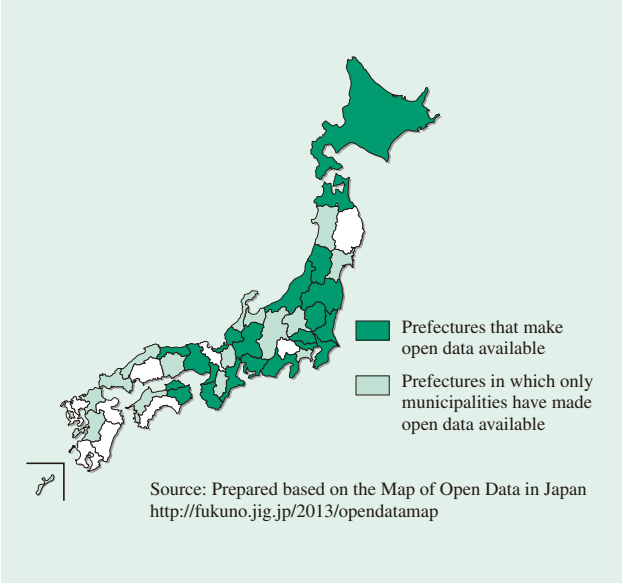


*Fig. 1—Spread of Open Data Initiatives by Local Government. The map shows local governments that have already made open data available.*

2015)[2), (3)]. Similarly, the Digital Accountability and Transparency Act of 2014 (Data Act) passed by the USA in May 2014 obliges the publication of financial and expenditure data for government agencies using specified formats and field names.

## Benefits of Making Open Data Available

As noted above, considerable progress is being made on open data initiatives. The significance of work by national, local, and other government agencies on open data include the following benefits.

(1) Greater transparency in government

The benefits of national and local government agencies releasing budget and other data associated with administrative activities include fulfilling their obligation of accountability to the tax-paying public, raising public confidence in their administration, and raising public interest in participating in government.

(2) More efficient government services

National and local government agencies release observational data and other data relating to public activity in forms that make it easy to manipulate. Private sector service providers undertake certain governmental tasks associated with providing information, publicizing policies, and so on by using this data to develop and supply applications and other services that take account of things like public needs and usability.

(3) Economic revitalization

National and local government agencies release data that private sector service providers and other agencies would find difficult to obtain on their own, such as statistical and geographical data. This promotes economic activity through the use of the data by private sector service providers for things like developing new businesses, more efficient marketing, and enhancing productivity.

Furthermore, work on making open data available is not restricted to government agencies alone. In 2014, for example, Tokyo Metro Co., Ltd. made realtime information on railway services available as open data, and used it to stage an "Open Data Utilization Contest." By making its information publically available as open data, numerous developers participated in the contest, leading to the creation of 281 new applications[4].

## Examples of Businesses that Use Open Data

Table 1 lists some of the numerous examples of new businesses that use open data in those nations that are leaders in the field. Uses include assessing real estate values and the trustworthiness of companies.

TABLE 1. Example Uses of Open Data
*The table lists some open data practices adopted outside Japan.*

| Category | Summary of service or application | How open data is used | How it is funded |
|---|---|---|---|
| Real estate | Service providing real estate sales data to real estate intermediaries | Collects information relevant to real estate prices, such as local environment, employment, education, and weather | • Service fee |
| Company information | Database of approximately 85 million companies from 103 countries (registered address, financial information, etc.) | Collects company registration data in a database and also allows crawling of company websites | • Selling data<br>• Service fee (API usage fee) |
| Finance | Mobile app for informing credit card holders of suspicious transactions | Derives algorithms that detect suspicious transactions from historical data on fraud and complaints | • Service fee |
| Transportation | Mobile app for identifying the quickest and easiest (least congested, etc.) route to a destination | Looks up details of transportation operator services when searching for routes, and app users submit congestion and other data | • Advertising |

API: application programming interface

## HITACHI'S OPEN DATA SOLUTION

### Steps Leading up to Release of Open Data

Initiatives by public agencies such as national and local governments for making open data available require activities such as identifying the data they hold and establishing the infrastructure for releasing the data. The steps that need to be worked through to achieve this are described in detail in the Open Data Guide v1.0 published by the General Incorporated Association "Vitalizing Local Economy Organization by Open Data & Big Data" (see Fig. 2) and the "Introduction to Open Data: A Beginners' Guide for Local Government Agencies" published in February 2015 by the Cabinet Secretariat.

The Open Data Guide v1.0 specifies that the situational review step should include a data stocktake to determine things like the types and quantities of data held by the organization and to identify which departments manage it. Similarly, the preparation step involves establishing the data catalog and other infrastructure for making data available, and the publication and maintenance step involves ensuring the ongoing availability of the data.

However, having government agencies establish the services described above will likely increase the workloads associated with managing and publishing the large quantities of data held by each agency. To reduce this burden, they will need assistance with establishing the infrastructure for making open data available. Hitachi launched its Open Data Solution in July 2014 based on its past business experience and research and development work on open data. The following section describes this solution.

### Features of Open Data Solution

The Open Data Solution supplied by Hitachi covers all of the processes needed to implement open data at public institutions and supports all activities from planning to the preparation, publishing, and use of open data. The Open Data Solution provides eight services (see Fig. 2). The following describes the main service options.

(1) Current situation review and analysis service

When getting started with open data, the first issue is to choose which data to make available. To make this decision, it is first necessary to undertake a quantitative assessment of things like the types and quantities of data held by the organization, its formats, and whether or not it can be released. On this basis, data that is in high demand from users (public and companies) needs to be made available in formats and via methods that are easy to use.

Along with conducting a stocktake of the data held by the organization to clarify these information assets, the current situation review and analysis service combines this stocktake with the results of a needs analysis to prioritize which data to release and formulate plans for how to do so. It also identifies the organizational and technological issues associated with doing this and investigates how to overcome them to formulate an action plan.

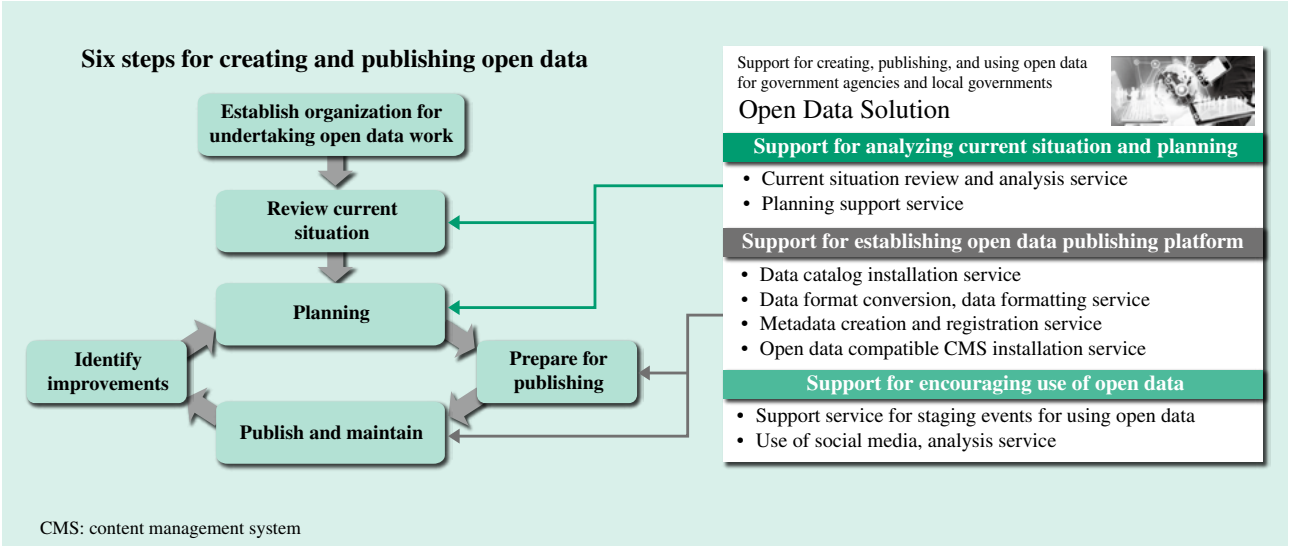


*Fig. 2—Steps Associated with Making Open Data Available, and Hitachi's Solution.*
*The steps for making open data available are listed on the left and the options available through Hitachi's Open Data Solution are shown on the right. The figure also shows the interrelationships.*

(2) Data catalog installation service

Standard practice when making open data available is to set up a data catalog to make it easy for users to find the information they require from among the large quantity of data offered. The data provider inserts metadata (bibliographic data) about the open data in the data catalog. Data users can access the data catalog to search the available data using key words or to check data summaries to determine whether the data is useful to them prior to downloading.

The data catalog installation service sets up and operates data catalogs using the CKAN open data software adopted by the US and UK government.

An example application at the Cabinet Secretariat is the implementation of the DATA.GO.JP data catalog of the Japanese government that commenced operation in October 2014. DATA.GO.JP currently makes available 13,038 data sets (as of the end of June 2015).

(3) Content management system (CMS) installation service for open data

Adding data to a data catalog requires that staff spend more time preparing and entering metadata. In considering staff workflow data, providers find it desirable that the entry of metadata into the data catalog be conducted in parallel with the adding of content to the website.

The CMS installation service for open data is able to update the data catalog at the same time as content is added to the website by using the automatic open data catalog integration function of the content management system that supports the maintenance of local government websites.

## ADVANCED TECHNOLOGIES FOR OPEN DATA

### Publication in Data Formats that Facilitate Secondary Uses

As in the examples described above, there is evidence of the growing use of open data in business. It is recognized that providing open data in formats that facilitate secondary uses is important for encouraging its use. To achieve this, steps are being taken in leading nations to provide linked open data (LOD), a format that facilitates machine processing (use by computer programs).

The properties that characterize LOD are use of the resource description framework (RDF), a data format that enables the meaning of data to be determined during machine processing, and the tagging of data with unique uniform resource identifiers (URIs) to define references to external data. The use of LOD facilitates things like searching across multiple datasets and integrated analysis of data regardless of whether it is sourced from inside or outside the organization.

### Technique for Converting Statistical Tables to LOD

Statistics are one form of data held by government that is already used in a variety of different ways. This section describes a Hitachi technique for efficiently converting statistical tables to LOD.

The conversion of statistical tables to LOD requires that data in table formats such as comma-separated values (CSV) or XLS be converted to RDF and tagged with URIs. In RDF, data is represented by three elements (subject, predicate, and object) collectively called a “triple.” Data expressing that the population of Japan is 127.3 million, for example, is represented by a triple in which Japan is the subject, population is the predicate, and 127.3 million is the object. Here, the terms “Japan” and “population” are represented by unique URIs. The same URI is used to indicate a particular concept regardless of the label used. Using the same URI in different LOD allows those LOD to be linked. This enables efficient crossover search and use of the data by users.

As statistical tables may contain hundreds of thousands of values or more, manually converting them to LOD would be very labor-intensive. Accordingly, there is potential for using a tool to support conversion to LOD. However, past tools have only been able to convert data stored in specific structures, and as statistical tables have many different forms, this results in extra work for data structure conversion. This created the challenge of how to convert statistics to LOD without imposing extra work.

In response, Hitachi developed an LOD conversion support tool that is made up of templates for the efficient conversion of statistical tables to LOD and a program for performing the conversion automatically[5), 6)]. The templates reduce the amount of work involved in converting data structures because they are suitable for use with statistical tables that have a variety of different structures. The tool also enables compatibility with external LODs to be achieved efficiently because it includes a function for recommending the corresponding external URI for each term. The technique was used on six statistical tables published by the government as open data, which demonstrated that it could be used to convert tables with millions of values to LOD, making the data easy to search and interlink.

## OUTLOOK FOR OPEN DATA IN THE FUTURE

As noted above, agencies such as national and local governments in Japan are working to establish an open data environment at a level equivalent to other developed nations. Nevertheless, progress to date has not been sufficient to achieve things like economic revitalization or more sophisticated public services through open data initiatives.

To create an environment in which greater use can be made of open data, Hitachi believes it is essential to make progress in a synergistic manner by working through a cycle of three steps: (1) providing data through public institutions, (2) providing services through private-sector service providers and others, and (3) providing feedback to public institutions from the public, companies, and others.

Based on the solution described above, Hitachi supports the provision and use of open data by public institutions seeking to engage in open data initiatives and by private-sector service providers who want to use open data, from identifying the challenges to formulating and implementing solutions. Hitachi is also developing the solution further and conducting research into technology with the aim of enhancing the efficiency of customers' existing operations and improving business services.

## CONCLUSIONS

This article has described trends and initiatives in open data, Hitachi's solution and technology for overcoming the challenges faced by data providers, and the outlook for open data in the future. Hitachi aims to implement related products and services through Social Innovation with national, local, and other government agencies involving open data and through collaborative creation with private-sector companies involving business applications.

### REFERENCES

(1) OPEN GOVERNMENT, DATA.GOV (Jun. 2015), http://www.data.gov/open-gov/
(2) Search for a Dataset, DATA.GOV (Jun. 2015), http://catalog.data.gov/dataset
(3) Datasets, DATA.GOV.UK (Jun. 2015), https://data.gov.uk/data/search
(4) Tokyo Metro Co., Ltd. News Release, "Open Data Initiatives at Tokyo Metro," (Mar. 2015), http://www.tokyometro.jp/news/2015/article_pdf/metroNews20150327_36-194.pdf in Japanese.
(5) Y. Asano et al., "Template for Converting Statistical Data to RDF," 12th Forum on Information Technology, 12(2), F-034 (Aug. 2013), http://ci.nii.ac.jp/vol_issue/nels/AA1242354X_ja.html in Japanese.
(6) Y. Asano et al., "A Template for Handling Statistical Data in RDF," Second International Workshop on Semantic Statistics (SemStats 2014) (Oct. 2014).

## ABOUT THE AUTHORS

Daisuke Hasegawa
*Public Strategy Planning Department, Innovative Business Operation, Government & Public Corporation Information Systems Division, Information & Telecommunication Systems Company, Hitachi, Ltd. He is currently engaged in business planning for information systems in the public sector.*

Yu Asano, Ph.D.
*5th Group, Intelligent Information Research Department, Center for Technology Innovation – Systems Engineering, Research & Development Group, Hitachi, Ltd. She is currently engaged in the research and development of basic technology for the implementation of open data. Dr. Asano is a member of The Japanese Society for Artificial Intelligence (JSAI) and the Information Processing Society of Japan (IPSJ).*