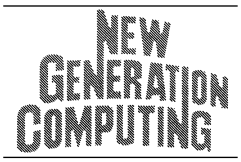
New Generation Computing, 34(2016)341-366
Ohmsha, Ltd. and Springer



©Ohmsha, Ltd. and
Springer Japan 2016

Constructing a Site for Publishing Open Data of the Ministry of Economy, Trade, and Industry — A Practice for 5-Star Open Data —

Yu ASANO,[1] Seiji KOIDE,[2] Makoto IWAYAMA,[1]
Fumihiro KATO,[2] Iwao KOBAYASHI,[3] Tadashi MIMA,[4]
Ikki OHMUKAI,[5] Hideaki TAKEDA[5]
[1]*Research & Development Group, Hitachi, Ltd.*
1-280 Higashi-Koigakubo, Kokubunji-shi, Tokyo 185-8601 JAPAN
[2]*Research Organization of Information and Systems, National Institute of Informatics*
2-1-2 Hitotubashi, Chiyoda-ku, Tokyo 101-8430 JAPAN
[3]*Scholex, Inc.*
2877-15 Nase-cho, Totsuka-ku, Yokohama-shi, Kanagawa 245-0051 JAPAN
[4]*Hitachi Consulting Co., Ltd.*
2-4-1 Koujimachi, Chiyoda-ku, Tokyo 102-0083 JAPAN
[5]*National Institute of Informatics*
2-1-2 Hitotubashi, Chiyoda-ku, Tokyo 101-8430 JAPAN
yu.asano.ko@hitachi.com, koide@nii.ac.jp,
makoto.iwayama.nw@hitachi.com, fumi@nii.ac.jp,
iwao@linkedopendata.jp, tmima@hitachiconsulting.co.jp,
i2k@nii.ac.jp, takeda@nii.ac.jp

Received 7 September 2015
Revised manuscript received 16 March 2016

Abstract We describe a procedure for constructing a website for publishing open data by focusing on the case of Open DATA METI, a website of the Ministry of Economy, Trade, and Industry of Japan. We developed two sites for publishing open data: a data catalog site and one for searching linked open data (LOD). The former allows users to find relevant data they want to use, and the latter allows them to utilize the found data by connecting them. To implement the data catalog site, we constructed a site tailored to the needs of the organization. Then we extracted a large amount of metadata from the individual open data and put it on the site. These activities would

In association with The Japanese Society for Artificial Intelligence

have taken a lot of time if we had used the existing methods, so we devised our own solutions for them. To implement the LOD searching site, we converted the data into LOD in the Resource Description Framework (RDF). We focused on converting statistical data into tables, which are widely used. Regarding the conversion, there were several kinds of missing information that we needed to associate with the data in the tables. We created a template for incorporating the necessary information for LOD in the original table. The conversion into LOD was automatically done using the template.

Keywords: Linked Open Data, Catalog Site, SPARQL Endpoint, Open DATA METI, Data Cube Vocabulary

§1 Introduction

Since the publication of the 5-star deployment scheme for open data[3] (See Table 1), the governments of many nations have been publishing their data as open data. In fiscal year (FY) 2014, the Cabinet Secretariat of Japan started full-scale operation of "DATA.GO.JP"[*1] which is an open data catalog site of Japanese government. More than 150 official organizations including prefectural and municipal governments have started to make their data open,[*2] and a road map[8] was published for promoting open data throughout Japan.

Table 1 5-Star Deployment Scheme for Open Data

★	Available on the web (whatever format) but with open license, to be Open Data
★★	Available as machine-readable structure data (e.g. Microsoft® Excel®)
★★★	The above plus, non-proprietary format (e.g. CSV)
★★★★	All the above plus, use open standards from W3C to identify things, so that people can point at your stuff (e.g. RDF)
★★★★★	All the above plus, link your data to other people's data to provide context (LOD)

There are two kinds of sites for publishing open data. One is a catalog site[*3,*4,*5,*6] which is a portal site having a catalog system for searching open data. The other is a site[*7,*8,*9] to publish open data in machine-readable format for ease of utilizing the retrieved open data. While the former sort of site handles any kind of format (1-5 star levels), the latter handles machine readable formats (3-5 star levels), such as CSV (Comma-Separated Values) or RDF (Resource

[*1] http://www.data.go.jp/
[*2] http://fukuno.jig.jp/2014/opendatajpstat
[*3] http://catalog.data.gov/dataset
[*4] https://data.gov.uk/data/search
[*5] http://www.data.go.jp/data/
[*6] http://datameti.go.jp/data/
[*7] http://eurostat.linked-statistics.org/sparql
[*8] http://datiopen.istat.it/sparql
[*9] http://datameti.go.jp/sparql

ngc34403 : 2016/9/26(10:13)

Description Framework). The latter site facilitates data retrieval and linking of multiple data.

The catalog site publishes information about open data together in one place. The information on each data consists of the title, category, data format, license, and URL (Uniform Resource Locator) of the data. This sort of information is called metadata. We propose six steps for constructing and maintaining catalog sites.

(1) Installing the catalog site
(2) Collecting the cataloging targets
(3) Gathering metadata
(4) Registering and updating metadata
(5) Understanding the users' experiences and needs
(6) Analyzing the data quality and improving it

Although some of the software for constructing catalog sites has recently become available, we would need to maintain the system, add new functions, and improve the registering and updating methods. Furthermore, the cost of gathering metadata would be high and should be reduced.

The latter sort of site, which publishes open data in machine-readable format, is constructed by converting open data into CSV or RDF. A linked open data (LOD) search site is an example. The site is constructed by converting open data into LOD and preparing an API to access or combine them. The conversion method depends on the type of data. For example, W3C (World Wide Web Consortium) proposed a standard method for statistical data. Furthermore, although conversion tools have also been developed,[5,13] their conversions are computationally costly; there is a need for more efficient conversion.

We were contracted by the Ministry of Economy, Trade and Industry (METI) to perform investigative research on open data. The contract ran from FY 2012 to FY 2013, and in 2012, we constructed a website to publish the data of the METI and the relevant ministries as open data. The site is called "Open DATA METI,"[*10] and it is the Japanese government's first effort at publishing open data. The Open DATA METI site actually consists of two sites: a catalog site and a LOD search site. In FY 2013, we expanded the data and functions, demonstrated how the LOD can be used, and created a number of use-cases of LOD.

In this paper, we describe our construction of this site for publishing open data of METI. In the next section, we describe the construction and maintenance of the catalog site. We also describe our method of efficiently gathering metadata from vast amounts of METI data and registering and updating this data. In Section 3, we describe how we constructed the LOD search site and how the LOD is used. We also explain our approach to convert statistical data into LOD and editing LOD by using an actual statistical table. Furthermore, we describe a demonstration of utilizing LOD. We finish by stating our conclusions.

[*10] http://datameti.go.jp/

§2 Open Data Catalog Site

At the beginning of this section, we describe the construction and maintenance of the catalog site. Subsequently, we introduce our method of efficiently gathering metadata from the existing data and registering and updating the metadata.

2.1 Constructing and Managing the Catalog Site

While there are preceding examples of catalog sites such as DATA.GOV (U.S. government site) and DATA.GOV.UK (UK government site), the procedure of their construction and the lessons learned have not been published. This paper describes the procedure for constructing and maintaining a catalog site and reports lessons learned in applying it to the case of Open DATA METI.

Since users search data in a catalog by using metadata, we have to consider the construction process from the standpoint of managing metadata. The followings are the six steps we propose.

(1) Installing catalog site: Construct the initial catalog site by using catalog software.
(2) Collecting cataloging targets: Clarify the standard of the target data, and collect data based on the standard.
(3) Collecting metadata: Normalize the cataloging targets by collecting metadata on them.
(4) Registering and updating metadata: Register the metadata in the catalog site. When data are updated, go back to (3) and collect the metadata of the updated data and register them in the catalog site. This step also includes adding metadata about new data and deleting metadata on unnecessary data.
(5) Understanding the users' experiences and needs: Monitor how the data is being used and get requests of data from users.
(6) Analyzing the data quality and improving it: Check that the registered data conform to the 5-star deployment scheme for open data and improve the star level as much as possible.

In the following sections, we explain these steps. Regarding (5), we constantly checked the statistics on page views, visiting users and data downloads. We also prepared bulletin boards and a data request form. As for (6), all the data in Open DATA METI satisfy the license requirements of open data. That is, almost all data have CC0 or CC-BY license, which allows derivative works and commercial use. The remaining data have CC-BY-ND license, which inhibits derivative works. The detail of each license is provided by Creative Commons site.[*11]

2.2 Installing the Catalog Site

For the catalog software of Open DATA METI, we used open-source soft-

[*11] http://creativecommons.org/licenses/

ware called CKAN.[*12] A major reason for selecting CKAN was that it has been used in many catalog sites including DATA.GOV.UK. Another reason was that a community led by the Linked Open Data Initiative (LODI) has already localized CKAN in Japanese. We also used WordPress[*13] for creating and maintaining the overall catalog site. The details of these activities are reported in a paper.[11)]

In FY 2012, we extended the original CKAN to support the CC-BY-ND license, inhibit dataset management by general users, and show counters of page views, visiting users, and data downloads.

In FY 2013, we added several new facilities to the catalog site, including a bulletin board, data preview, a "what's new" page, an English site, and keyword suggestions when searching.

2.3 Collecting Cataloging Targets

The major difficulty in collecting the cataloging target is how to judge whether each data satisfies the open data standard or not. In particular, it would take much time and effort to ask each data holder. In the case of Open DATA METI, METI assumed that the data published on its web site would automatically satisfy the open data standard because permission for it to be openly published had already been granted. Following this assumption, we first collected cataloging target candidates by crawling METI's web site and manually selected the appropriate ones.

Data are maintained in three layers in CKAN; groups, datasets, and resources. A group is a set of datasets, and a dataset is a set of resources. In the case of Open DATA METI, groups are genres such as whitepapers, statistics, and reports. Datasets are logical units of data in a group, such as "whitepaper on trade (FY 2014)." Resources are physical files such as HTML files and PDF files for the dataset.

At the end of FY 2012, Open DATA METI had 196 datasets focusing on whitepapers and statistics. In FY 2013, we added the datasets of five priority fields (whitepaper, disaster, geospatial, social mobility, and finance and contracts) defined in the "Open Government Data Strategy."[7)] We also added the following datasets; statistics, company information, datasets in National Institutes under METI, and other unreleased data in METI. In February 2014, the number of datasets reached 2,242 (about forty thousand resources). Most of the added datasets were selected from "high value" dataset, which is shown in the "G8 Open Data Charter."[9)]

2.4 Collecting Metadata

Users search data in a catalog by using metadata. To make this search more accurate and thorough, every data should have necessary and sufficient metadata. In the case of Open DATA METI, each data has 10-20 kinds of metadata, and there are more than 40,000 data (datasets and resources). It would have taken too much time to manually collect the metadata, so we developed a

[*12] http://ckan.org/

[*13] https://ja.wordpress.org/

tool that can automatically extract them. In this section, we introduce the set of metadata used in Open DATA METI and explain the tool we developed.

METI and our group defined 46 kinds of metadata (see Appendix A). Here, together, we referred to the "Data Catalog Vocabulary"[*14] from W3C and the metadata used in DATA.GOV and DATA.GOV.UK. We also aligned our metadata and the metadata defined in DATA.GO.JP, which is the Japanese government's catalog site opened in 2013. The metadata that only Open DATA METI has are about the inquiry address (division, e-mail address, and phone number).

Almost all metadata appear on the web pages related to the target data. However, it is difficult to automatically extract such metadata by using, for example, web scraping techniques,[4] because the variation of the web pages is very large.

We developed a web scraping tool to extract metadata from a specific style of web page. Concretely speaking, this tool extracts metadata about figures and tables from the web pages of a whitepaper. The extracted metadata are "title," "URL," and others. Figure 1 shows how these metadata appear in a web page.

Fig. 1 Metadata Extraction from a Web Page[*15]

Since there might be errors in automatically extracted metadata, we have to check and correct these before publication. To reduce the cost of doing so, we tuned our tool so that the precision of the extraction was 100% in our evaluated cases. This means that we did not have to check the metadata that our tool ex-

[*14] http://www.w3.org/TR/vocab-dcat/

[*15] http://www.meti.go.jp/report/tsuhaku2012/2012honbun/index.html

tracted. We only had to check the metadata which the tool failed to extract. For precision, we manually wrote extraction rules for regular expressions (examples are shown in Fig. 2). In addition, we filtered out noise by using error patterns obtained by error analysis. The followings are the error patterns we used.

A) The extracted "title" is a null string.
B) The same "title"s are extracted in sequence.
C) There are HTML tags in the extracted "title."
D) "format" cannot be extracted.
E) "mimetipe" cannot be extracted.
F) "resource_type" cannot be extracted.

To deal with the different formats of the various whitepapers, we localized the format-dependent steps in the tool into only those which use the extraction rule described above. Figure 2 shows an example of rules for extracting metadata (title and URL) from HTML of the figure in Fig. 1. The tag information (i.e., "<tag attr=val>") is used to identify the position of metadata, and the regular expression is used to extract/check the metadata. To process a new format of whitepaper, all we have to do is to prepare a new extraction rule for that format. We evaluated our tool by measuring the time needed to extract metadata on figures and tables with and without it. The result was that our tool could reduce the time to 1/40th of that without using it. Our tool could handle 4,509 (about 12%) among the 37,251 resources in Open DATA METI.

HTML

```
<div class="img_name">第1-1-1-1 図　世界実質GDP 成長率の推移</div>
<div><img src="../image/i01010101.gif" alt="第1-1-1-1 図 世界実質GDP 成長率の推移"></div>
<p class="excel"><span>Excel形式のファイルは<a href="../excel/i01010101.xls">こちら</a></span></p>
```

Rules

```
{
    "caption_tag"   : "div",
    "caption_attr"  : "class",
    "caption_val"   : "img_name",
    "caption"       : ".*?[図表]+[¥¥s  ]+.*",
    "caption_num"   : ".*?[図表]+[¥¥s  ]+",
    "url_tag"       : "p",
    "url_attr"      : "class",
    "url_val"       : "excel",
    "url"           : ".*¥¥.(xls|xlsx|ppt|pptx)"
}
```

Fig. 2 HTML of Title and URL, and Rules for Extracting Them

2.5 Registering and Updating Metadata

The collected metadata are registered in catalog software, in our case, CKAN. In CKAN, there are two methods for registering metadata: an interactive one through web pages and a batch method through APIs. The interactive method has an advantage of being able to handle metadata intuitively, but

registering thousands of metadata interactively is unrealistic.

Our approach was to use spreadsheets for adding/updating/deleting metadata, and use the APIs for automatically reflecting the metadata on the spreadsheets to CKAN. Handling metadata on spreadsheets is effective because many users are familiar with spreadsheet software and it is easy for them to maintain the data by using the various functions provided by the software. A problem here is that there are two versions of metadata, one is on the spreadsheets and the other is on CKAN, and users have to carefully synchronize them. In this section, we explain the tool we developed for easily synchronizing the metadata between spreadsheets and CKAN.

There are two issues related to the synchronization.

Issue A: There is no function for validating whether each metadata on a spreadsheet is well-formed or not. This may cause a conflict because ill-formed metadata on the spreadsheet cannot be registered with CKAN.

Issue B: In registering metadata with CKAN by using the APIs, we have to select the appropriate API depending on whether the metadata is new or already registered with CKAN. This means that we have to record the history of metadata registrations and modifications on a spreadsheet; i.e., we cannot overwrite the metadata on the spreadsheet. A simple solution is to forcibly override the metadata and register them as new ones. However, this would also change the URIs (Uniform Resource Identifier) of the related resources. This is not recommended from the standpoint of the URI permanence of the same resource.

As for issue A, we implemented the metadata checking function in spreadsheet software. Users can modify ill-formed metadata when they enter them on the spreadsheet. All the metadata on the spreadsheets are thus guaranteed to be registered with CKAN, and there is no risk of the metadata conflicting. At the same time, we implemented functions to help when entering metadata. One is for alerting users of the absence of the metadata and the notational variants of the metadata. Another is for selecting metadata from those already entered.

As for issue B, we created a new API layer, called the utility layer (see Fig. 3), which wraps CKAN's original APIs and provides more user friendly

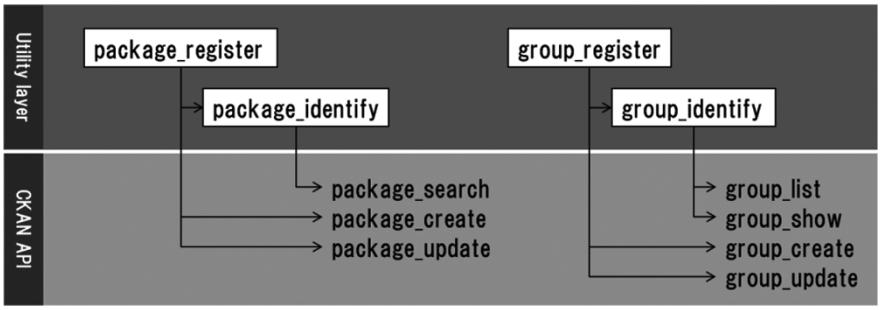


Fig. 3 Utility Layer that Wraps CKAN API

functions. For example, we made a new API called "package_register," which checks whether the specified set of metadata is already registered with CKAN and automatically selects one of the two original APIs depending on the check result.

The following steps explain how "package_register" works.

Step 1: The API checks whether a specified dataset is already registered with CKAN. In the case of Open DATA METI, there is no unique ID (corresponding to the URI) for each dataset because of an operational reason. Without the ID, the API has to search the dataset in CKAN by using the title string of the dataset.[*16] This is what the sub-API "package_identify" does.

Step 1-1: If the specified dataset has not been registered with CKAN, the API just calls "package_create," which is CKAN's API for registering a new dataset. The API ends here.

Step 1-2: If the specified dataset has already been registered with CKAN, the API compares the metadata of the specified dataset and the metadata of the existing dataset in CKAN.

Step 1-2-1: If there is a difference, the API calls "package_update," which is CKAN's API for updating an existing dataset. The API ends here.

Step 1-2-2: If there is no difference, the API ends here.

Note that "package_register" automatically selects the appropriate API of CKAN, and if it is not necessary, it does not call any API. By using "package_register," users can handle metadata on spreadsheets without caring about the synchronization issue.

2.6 Future Issues

Collecting metadata is the most costly part of constructing and maintaining a catalog site. For Open DATA METI, we reduced this cost by automatically extracting metadata about figures and tables from whitepapers. However, the effect of this measure was limited, and the scope of automatic metadata extraction will have to be enlarged. We discuss this issue below.

Unlike formatted metadata such as dates and phone numbers, free formatted metadata such as titles are difficult to extract. Generally speaking, we could assume that the anchor text to a resource is a candidate title of the resource. Unfortunately, this assumption leads to there being many exceptions. Figure 4 shows some examples of the exceptions. In the figure, all the anchor texts, such as "(Part 1) (PDF: 2,112KB)," "(Part 2) (PDF: 1,764KB)" and so on, do not represent the content of the corresponding resources. In these examples, we have to find additional strings that appropriately represent the content of the resources. In many cases, such strings are located around the anchor text and a merging of the found string and the anchor text becomes a good title string. In Fig. 4, by combining "Section 2 Towards enhancing the competitiveness of

[*16] For this search, every title string in a catalog should be unique. This is a restriction of Open DATA METI.

ngc34403 : 2016/9/26(10:13)

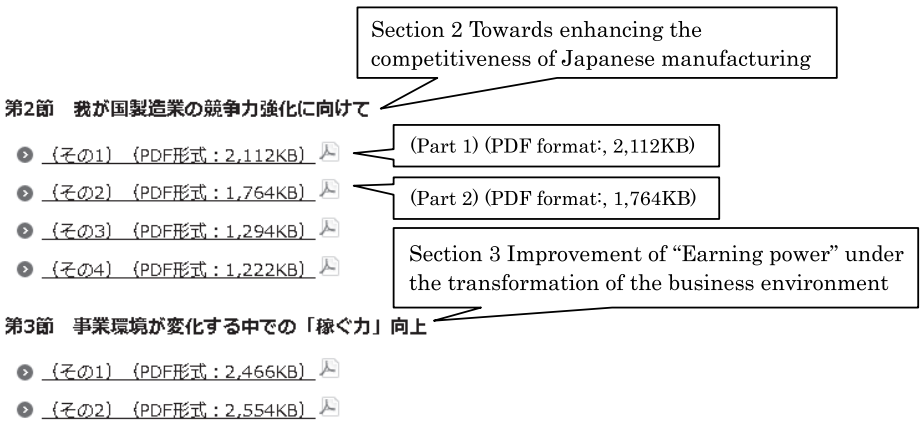


Fig. 4 Examples of Inappropriate Anchor Texts as Titles

Japanese manufacturing" and "(Part 1)," we arrive at a good title, i.e., "Section 2 Towards enhancing the competitiveness of Japanese manufacturing (Part 1)." In the case of Open DATA METI, we manually searched these strings and created the title strings. Automation of this task is required.

Another difficulty in extracting metadata from web pages is that the metadata does not always appear on the same page where the target resource is described. For example, phone numbers for the public to make inquiries are often written on the page for an organization, while the data that the organization publishes are located under that page. In this case, we have to find the organization's page from the data page.

Lastly, some metadata do not appear on web pages at all. In the case of Open DATA METI, not many of the publication dates, frequencies of updates, and licenses are specified on the web pages. These metadata should thus be attached manually. To optimize this task, the open data publishing task should be embedded into the existing job workflow. For example, the publishing date metadata can easily be collected in the web publishing workflow. Workflow optimization from the standpoint of handling metadata is a crucial issue for sustainable publishing of open data.

§3 LOD Search Site

In this section, we describe how we constructed the LOD search site and how the LOD is used. We also explain our approach to convert statistical data into LOD by using an actual statistical table. Furthermore, we demonstrate applications of utilizing LOD.

In order to utilize open data, we have to convert open data into LOD and provide an interface that allows users to search them. The LOD search site can be constructed as following steps.

(1) Installing LOD search site: Construct an initial LOD search site by using software that allows users to register and search LOD.

(2) Selecting LOD targets: Decide the LOD targets based on a standard, which is, for example, the popularity.
(3) Converting data into LOD: Convert data into LOD.
(4) Registering and updating LOD: Register and update the LOD on the LOD search site.

We used Virtuoso[*17] to build the LOD search site of the Open DATA METI project. We chose it because we had experience in using it to construct a LOD system for biological information.[15] Regarding steps (1) and (4), we used the default functions of Virtuoso. Regarding step (2), we selected statistical tables, which can be combined, as LOD targets. In section 3.1, we describe step (3) especially in relation to statistical data. In section 3.2, we explain our approach to promoting utilization of the published LOD. We describe future tasks in section 3.3.

3.1 Conversion of Statistical Tables into LOD

3.1.1 Conversion of Statistical Tables into LOD and Its Problems

We selected commonly used statistical tables as the LOD targets from various kinds of open data in the data catalog.

The conversion of statistical tables into LOD entails converting statistical data in a table format such as CSV into RDF for connecting various statistical data. RDF represents data by using three components (triples), which consist of subject, predicate, and object. For example, the data "The prefectural capital of Tokyo is Shinjuku." is expressed by using a triple, which consists of "Tokyo" as a subject, "prefectural capital" as a predicate, and "Shinjuku" as an object. In RDF, a URI is associated with each of these components, i.e., "Tokyo," "prefectural capital," and "Shinjuku." The same URI is used for all the words that represent the same concept even if the words appear in different tables. This conversion allows users to handle statistical data without having worry about differences between words referring to the same thing.

Each statistical data may include hundreds of thousands of cells. Therefore, manual conversion into LOD would be a hard work. Although we could use tools such as OpenRefine[*18] for converting tables into RDF, we would first need to transform the original table into a template table structure. The transformation method depends on the table structures. Users have to prepare each transformation method for each table structure. To overcome this problem, we propose a method to embed the necessary information for LOD in the original statistical tables.[1,2] For example, URIs of dimensions and measures are the necessary information. More detail of the necessary information is described in section 3.1.4. The statistical table including the necessary information can be automatically converted into LOD. Therefore, our approach allows users to edit data in the original table format and convert the edited statistical data into

[*17] http://virtuoso.openlinksw.com/
[*18] http://openrefine.org/

LOD.

3.1.2 RDF Data Cube Vocabulary

In this section, we briefly explain the RDF Data Cube Vocabulary,[19] which is used for expressing statistical data in LOD. The vocabulary is mean to publish multi-dimensional data in such a way that the data are linked to related data. It is a W3C recommendation standard, and it makes processing data, such as filtering, aggregation, and integration of data, easy.

By using RDF Data Cube Vocabulary, each set of multi-dimensional data is expressed based on the following cube model.

(1) The dimension and its value (to identify the observation)
(2) The measure (the phenomenon being observed)
(3) The attribute (the unit of measure)

The model is used to express actual measured values, which are called observed values.

Multi-dimensional data, especially statistical data, is often expressed in the form of a 2D table. The components of the cube model mostly appear in the row or column heading portions of the table. However, this is not always the case. Some components may not appear in the table. We illustrate the corresponding row or column heading portion of each component by using example tables, Table 2 and Table 3. Table 2 shows populations in two dimensions, i.e., prefecture and year. Table 3 shows populations and areas in these two dimensions. In these tables, the observed values would be the values in cells other than the row heading portion (e.g. row 1 in Table 2) or the column heading portions (e.g. column 1 in Table 2). The populations in Table 2 and Table 3 were extracted from an investigation conducted by the Japanese Ministry of International Affairs and Communications and the areas are from an investigation conducted by the Ministry of Land, Infrastructure, Transport and Tourism.

In Table 2 and Table 3, "2010" and "2005" are values of the dimension "year." Only the values of this dimension, not the dimension itself appear in these tables. Human beings can make guesses about the missing information

Table 2 Populations by Prefecture and Year

	2010	2005
Saitama	7,194,556	7,054,382
Chiba	6,216,289	6,056,462
Tokyo	13,159,388	12,576,611
Kanagawa	9,048,331	8,791,587

Table 3 Populations and Areas by Prefecture and Year

	Population (persons)		Area (km^2)
	2010	2005	2010
Saitama	7,194,556	7,054,382	3,767.92
Chiba	6,216,289	6,056,462	5,081.91
Tokyo	13,159,388	12,576,611	2,102.95
Kanagawa	9,048,331	8,791,587	2,415.86

on the basis of the content or title of the tables. For example, we can guess that "year" is a dimension of Table 2 and Table 3 based on the values "2010" and "2005." However, explicit descriptions of this information are necessary for automatic processing of multi-dimensional data. Generally speaking, dimensions are often missing in tables.

The measures of Table 3 are "population" and "area," and these appear in the row heading portion of the table. On the other hand, the measure "population" does not appear in the Table 2. Human beings are able to guess the measure "population" based on the title of the table. A table with multiple measures usually contains the names of measures to distinguish observed values of one measure from different observed values of another measure. A table with a single measure, such as Table 2, often omits the measure name. In Table 3, "persons" is an attribute of the measure "population." "km^2" is an attribute of the measure "area." Attributes may not appear in tables.

3.1.3 Statistical Data in RDF

In this section, we explain the conversion using the RDF Data Cube Vocabulary. To publish statistical data based on the RDF Data Cube Vocabulary, we have to define a data structure (schema) and express observations (instances) based on the structure. In this paper, we focus on the expressions of observations.

There are two approaches to handle multiple measures. The first is to handle a single observed value as one observation. Each observation is expressed by related information on each observed value, which appears in each cell of the table. In the expression, the measure dimension, which is the extra dimension for measures, is used for identifying the measure of the observed value. The related information consists of dimensions, measures, observations, and attributes. This approach is able to support tables with complex structures because information is expressed in each cell of the table. The second approach is called multi-measure observation. This approach handles multiple observed values with the same dimension as an individual observation. It allows multiple observed values to be attached to one observation.

We chose the former approach in consideration of tables with complex structures. Each attribute is defined with the corresponding measure. As such, we do not need to define an attribute for each observation and we can reduce the quantity of expressions. Figure 5 is an example of expressing the cell of column 2, row 5 in Table 3, which refers to Tokyo's population in 2010. Each triple consists of a target cell as a subject, each arrow as a predicate, and each linked value as an object. For example, a triple including arrow 4 in Fig. 5 expresses that the value of the dimension "year" (the predicate whose URI is "eg:refYear") of the target cell (the subject) is "2010" (the object). The prefix "qb" identifies a namespace "http://purl.org/linked-data/cube#" and the prefix "eg" is an arbitrary namespace.

Figure 6 shows the corresponding RDF triples of Fig. 5. Here, the URI of Table 3 is "eg:dataset-02" and the URI of the target cell is "eg:dataset-02-

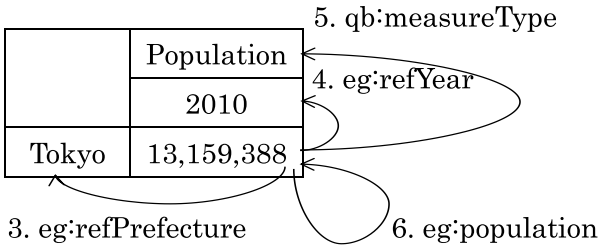


Fig. 5 RDF Graph of Tokyo's Population in 2010

```
1: eg:dataset-02-000003 a qb:Observation ;
2:   qb:dataset eg:dataset-02 ;
3:   eg:refPrefecture eg:prefecture-13 ;
4:   eg:refYear 2010 ;
5:   qb:measureType eg:population ;
6:   eg:population 13159388 .
```

Fig. 6 RDF Triples of Tokyo's Population in 2010

000003." Lines 1-2 describe that the cell is an observed value and is included in Table 3. Lines 3-6 describe triples with subjects that are the cell, predicates that are the URIs indicated by the arrows in Fig. 5, and objects that are the values given to the points indicated by the arrows.

3.1.4 Proposed Template for Statistical Data in RDF

In this section, we show how to embed necessary information for LOD in the original statistical table. The conversion of a portion of a statistical table (Fig. 5) into RDF triples (Fig. 6) needs additional information that does not appear in the table as follows.

A) URI of dimension: As mentioned before, dimensions often do not appear in statistical tables.
B) URI of value of dimension: Give a corresponding URI to each value of a dimension in a statistical table.
C) URI of measure: Give a corresponding URI to each measure in a statistical table. In tables with a single measure, the measures may not appear in the tables.
D) Distinction between the values of dimensions and measures: These are usually undistinguished in tables. For example, Table 3 does not have information about the first row being for measures (population, area) and the second row being values of dimensions (2010, 2005).

To convert statistical data in a table into RDF triples, we create a new template from the original table for inputting the necessary information. The template in Fig. 7 is made from Table 3. All the necessary information can be input in the template. In the template, the gray portions are added to the original table. We show how to make a template and input the above information

ngc34403 : 2016/9/26(10:13)

			Population (persons)		Area (km^2)
			eg:population		eg:area
			2010	2005	2010
		eg:refYear	2010	2005	2010
	eg:refPrefecture	#			
Saitama	eg:prefecture-11		7, 194, 556	7, 054, 243	3, 767. 92
Chiba	eg:prefecture-12		6, 216, 289	6, 056, 462	5, 081. 91
Tokyo	eg:prefecture-13		13, 159, 388	12, 576, 601	2, 102. 95
Kanagawa	eg:prefecture-14		9, 048, 331	8, 791, 597	2, 415. 86

[Step 1] [Step 2]

[Step 1] [Step 2]

Fig. 7 Creating a New Template from the Original Table 3 Content

from A) to D) below.

A new template is created from the original table in two steps.

Step 1: Insert one row below each row of the column heading portion and insert one column on the right side of each column of the row heading portion. Each cell in the rows/columns inserted takes a URI or original string of the upper/left cell that refers to a value of the dimension or a measure. For example, in the first row in Fig. 7, two measures, "Population" and "Area," appear. The URIs of these measures should be input to the corresponding cells in the second row. In the first column, prefectures such as "Saitama" and "Chiba" are values of a dimension. The URIs of these values should be input to the corresponding cells in the second column. This information corresponds to the above item B). From the perspective of LOD, URIs should be used for the values of the dimensions. If the original strings are used as the values of the dimension, each string is input to the added column/row.

Step 2: Insert one row between the column heading portion and the data portion, and insert one column between the row heading portion and the data portion. Each cell in a row/column inserted takes a URI of the dimension if the column/row including the cell is for values of the dimension. For example, the values of the dimension "year" are input in the fourth row in Fig. 7. Therefore, the URI of the dimension "year" is input to the cell of column 3 row 4 in Fig. 7. This information corresponds to the above item A). When the original table only has a single measure, the URI of this measure is given to the intersection cell, which is marked by "#," because most tables with single measures omit the measure. In the case of Fig. 7, the table has more than two measures, and consequently, the cell should be blank.

Our template has cells for all the necessary information to represent the data in RDF. By giving this template and the URI of the original table, the corresponding RDF triples such as Fig. 6 can be generated automatically. In this paper, we do not show the steps for generating the RDF triples.

Our template has the following characteristics.

- The data structure of the original table is maintained. Users can edit

statistical data as they do on the original table, and the editing results are immediately reflected in the RDF.
- URIs of the same group are listed in a row/column and their labels are also listed in the adjacent row/column. These characteristics for overview and correspondence enable users to input the necessary URIs easily and consistently.
- Necessary data for RDF are embedded in the table. Users and programs can easily manage data without referring to additional data.
- Users can use any spreadsheet software for handling our template because it is a simple table format.

3.1.5 Application of Template to an Actual Statistical Table

In this section, we illustrate an application example using an actual statistical table. As an experiment, we converted six statistical tables published by METI into RDF triples by using our template. We successfully converted them into about 3 million triples. Figure 8 is an example of applying our template to an actual statistical table, which has complex column/row headings. The dimensions and measures of this table are defined as follows.[16)]

Statistical table 1. 平成 22 年工業統計 市区町村編[*19]
(Industrial statistics of municipality 2010)

Dimensions: 市区町村 (Municipality), 産業中分類 (Industrial middle classification), 調査年 (Survey year)

Measures: 事業所数 (Number of establishments), 従業者数 (Number of employees), 現金給与総額 (Value of total cash wages and salaries), 原材料使用額等 (Value of raw materials), 製造品出荷額等 (Value of manufactured goods shipments), 粗付加価値額 (Grass value added), 有形固定資産年末現在高 (Value of tangible fixed assets at year end)

Each of the above dimensions and measures is represented by using a URI. For example, the dimension "municipality" is represented as "ktsh:refMinicipality," and the measure "number of establishments" is represented as "ktsh:numberOfEstablishments." In Fig. 8, columns/rows whose color is gray are added to the original table in order to input URIs as the necessary information for RDF. Figure 8 looks like the original table by making the added portions invisible. The prefixes such as "ktsh,"[*20] "area-code,"[*21] and "sangyo-code"[*22] in the names of the dimensions and measures are existing namespaces.

We should use standard URIs prepared by an authoritative organization for the municipality and industrial classification. However, such standards were

[*19] http://www.meti.go.jp/statistics/tyo/kougyo/result-2/h22/kakuho/sichoson/xls/h22-k6-data-j.xls (sheet 2000)

[*20] http://datameti.go.jp/lod/kougyou-toukei/

[*21] http://datameti.go.jp/scheme/standard-area-code/

[*22] http://datameti.go.jp/scheme/jsic/2007/

ngc34403 : 2016/9/26(10:13)

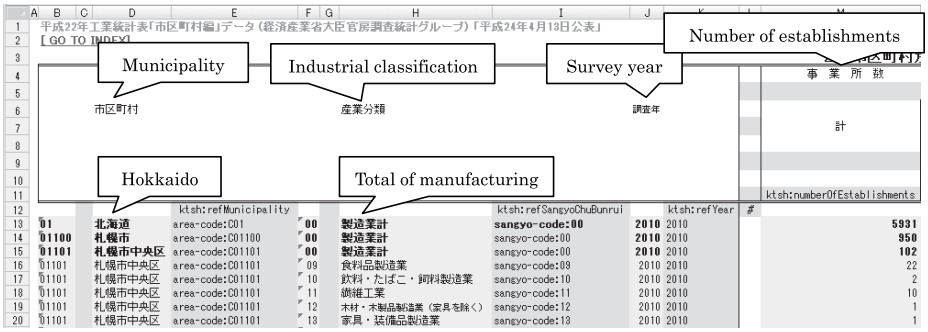


Fig. 8 Example Application of Proposed Template (Position of Industrial Statistics of Municipality)

not available at the time, so we prepared tentative URIs for the municipality and industrial classification based on the standard area code and the Japanese industrial code, which is published by the Ministry of Internal Affairs and Communications. When a hierarchical structure exists between codes, we related them. For example, an industrial classification consists of four hierarchical structures; a large classification, a middle classification, a small classification, and a fine classification. The hierarchical relations are expressed using the relation "skos:broader/narrower" of the Simple Knowledge Organization System (SKOS). Making hierarchical relations between concepts allows users to get a value of a broader concept by aggregating the values of its narrower concepts. Additionally, our data are linked to external data by adding more than four thousand links in total. For example, the URI for Hokkaido is linked to URIs for Hokkaido in DBpedia Japanese[*23] and the URI for the area codes provided in Gateway to Advanced and User-friendly Statistics Service[*24] by using "owl:sameAs," which is a property of Web Ontology Language (OWL)[18] and it indicates that two objects are the same. In addition, we prepared company codes and XBRL data of major companies because introduction of enterprise identification numbers in Japan began in 2015.

3.1.6 Comparison with Existing Work

In this section, we compare our approach with the existing ones. Several tools have been developed to clean up data in tables or to convert them into RDF. OpenRefine,[*18] RDF Refine,[*25] Han's tool,[5] and Salas's tool[13] are such tools. OpenRefine is used to clean up table data before converting it into RDF. For example, it can be used for unifying forms of the date and fluctuations in descriptions. Three other tools are provided as a plug-in in OpenRefine. These tools can automatically convert statistical data with the necessary data into RDF. However, although RDF Refine and Han's tool are powerful, they

[*23] http://ja.dbpedia.org/

[*24] http://statdb.nstac.go.jp/system-info/api/api-spec/

[*25] http://refine.deri.ie/

ngc34403 : 2016/9/26(10:13)

are designed for only simple tables that have a single row heading. To use these tools on complex tables, users have to convert complex tables into simple ones before converting them into RDF. Therefore, these tools cannot convert statistical tables with complex structures into RDF without first converting their original table structures.

Salas's tool is a dedicated tool to convert statistical tables into RDF. With this tool, users can import an original statistical table and input the necessary data for converting it into RDF. Our approach, instead, inputs the necessary data for RDF in the original table. Our approach can be applied to any spreadsheet software because the only necessary operation is adding rows and columns. Our approach allows users to edit data and convert data into RDF on their favorite spreadsheet software.

3.2 Utilization of Statistical LOD

Statistical LOD is LOD composed of statistical data. In this section, we describe an example of utilization of statistical LOD and our activities for promoting utilization of statistical LOD.

3.2.1 Example of Utilization of Statistical LOD

Once a number of statistical tables have been converted into LOD, users can perform crossover searches by selecting the dimensions or measures. The statistical LOD can be combined with other LOD published by other organizations. So far, six statistical tables of METI have been converted into LOD. Portions of two of these tables are shown below.

Statistical table 2. 平成 22 年工業統計 細分類別統計表[*26]
(Industrial statistics of prefecture 2010)
Dimensions: 都道府県 (Prefecture), 産業細分類 (Industrial fine classification), 調査年 (Survey year)
Measures: The same measure as Statistical table 1.

Statistical table 3. 平成 22 年工業統計 産業編[*27]
(Industrial statistics of industry sector 2010)
Dimensions: 都道府県 (Prefecture), 産業中分類 (Industrial middle classification), 調査年 (Survey year)
Measures: 在庫額 (Value of stocks), 有形固定資産額 (Value of tangible fixed assets), リース契約による契約額及び支払額 (Value Of lease payments)

For example, let us suppose a user wants to investigate industrial trends in Tokyo and searches the values of manufactured goods shipments and the values of stocks for each industrial middle classification. The former values appear

[*26] http://www.meti.go.jp/statistics/tyo/kougyo/result-2/h22/kakuho/saibunrui/xls/h22-k8-data-j.xls (Sheet:1003)

[*27] http://www.meti.go.jp/statistics/tyo/kougyo/result-2/h22/kakuho/sangyo/xls/h22-k3-data-j.xls (Sheet:3220)

in Statistical table 2. Statistical table 2 includes the values of manufactured goods shipments by prefecture and an industrial fine classification. Statistical table 3 includes the values of stocks by prefecture and an industrial middle classification. The hierarchies of the industrial classifications of these values are different. For them to be combined, the values need to be unified so as to have the same dimensions, which are prefecture and the industrial middle classification. Therefore, first, the values of the manufactured goods shipments by prefecture and the industrial middle classification are calculated based on the search results for the values of manufactured goods shipments by prefecture and the industrial fine classification from Statistical table 2 by using the hierarchical relation among industrial codes. The values of stocks can be found as a result of searching Statistical table 3. The final result (the values of manufactured goods shipments and values of stocks) is gotten by combining these results. It can be given by one SPARQL query (see Appendix B).

3.2.2 Development of Web API

We provided a SPARQL endpoint so that the LOD can be searched. However, SPARQL is not an easy-to-use query language, and only expert users can get the required data by using SPARQL. To solve this problem, we developed Web APIs for typical search patterns so that many users can get data easily. The Web APIs are follows.

A) Dataset list (http://datameti.go.jp/datasets)
To get a list of datasets in LOD.

B) Label search (http://datameti.go.jp/search/*label*)
To get a list of resources, whose object "rdfs:label" includes the input keyword "label" appearing at the end of URI.

C) Codes/Classifications (http://datameti.go.jp/scheme/*path*)
To get codes/classifications such as the standard area code and industrial classification. Their URIs are designed by using "http://data.go.jp/scheme/...". For example, the URI for Sapporo city is "http://datameti.go.jp/scheme/standard-area-code/C01100".

D) Dataset (http://datameti.go.jp/lod/*path*)
To get each LOD dataset, the URIs of the datasets are designed by using "http://datameti.go.jp/lod/...". For example, the URI of Statistical table 3 is "http://datameti.go.jp/lod/kougyou-toukei/h22-k3-data-j-3220/h24-2-2-1".

There are two ways to get data with a specified format such as JSON, XML, CSV, TSV, and HTML with the above APIs. The first way is to specify a required MIME type in the Accept header of the HTTP request. The other way is to add a required file extension at the end of the URI. When users specify both of them, the latter way takes priority. In addition, URIs starting from "http://datameti.go.jp/{scheme,lod}/ " fall within "303 URIs forwarding to Different Documents" of Cool URIs.[14] Each URI is redirected to "http:datameti.go.jp/{scheme,lod}/page/path" by specifying a MIME type in

the Accept header of the HTTP request.

3.2.3 Organization of Idea-thon[*28]

Showing people in public and private sectors the value of open data will be important for promoting its utilization. For that purpose, we organized an idea-thon. Fifty-three participants were divided into eight groups and discussed ideas. The participants included five people from METI and four people from the Small and Medium Enterprise Agency. They generated the following ideas, e.g., match a social problem with companies who have solutions for the problem, analyze businesses by visualizing their company data, encourage enterprises to publish open data, and streamline existing works by utilizing standard codes such as the Japanese article number code, which refers to products in Japan.

It is difficult to assess how our work was really made use of at this time. However, we accomplished our purpose of public and private sectors sharing an image of the future. We also identified two important things for promoting open data. Firstly, public sectors should have a policy of using open data for promoting civic participation. Secondly, a venue should be provided for dialogue among public and private sectors. Through the idea-thon, we identified a variety of possibilities for open data that were not expected by the data providers.

3.2.4 Organization of Visualize-a-thon[*29]

We organized a visualize-a-thon, which is an event where participants visualize and analyze data, in this case, the present conditions of the Japanese economy by creating visualizations that used statistical LOD of Open DATA METI. Although the published statistical LOD consisted of only six datasets,

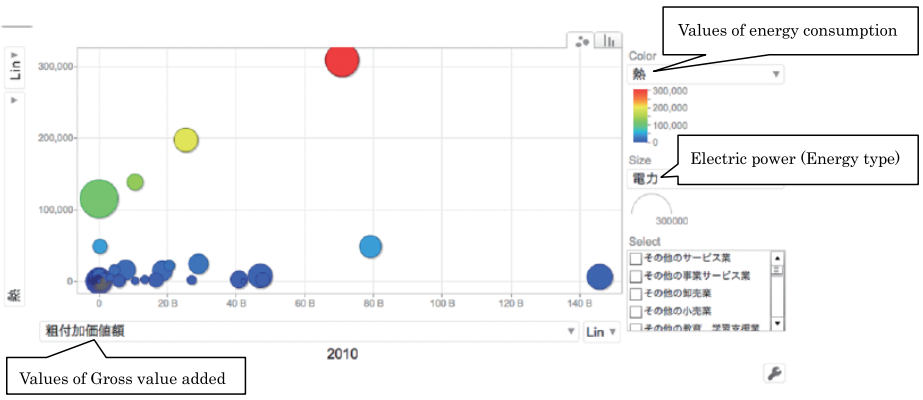


Fig. 9 Correlation between "Energy consumption" and "Gross value added" by industrial classification

[*28] "Idea-thon" is a coined word consisting of "idea" and "marathon." This is an event in which various participants discuss ideas on certain topics. The goal is for them to share their ideas.

[*29] "Visualize-a-thon" is a coined word consisting of "visualize" and "marathon." This is an event to visualize data in ways that make it easy to analyze and understand them.

the fifty-one participants managed to create many SPARQL queries and visualizations. The details of this event are described in a report.[6] For example, Fig. 9 shows the correlation between energy consumption and gross value added by each industrial category.

3.3 Future Issues

It is difficult to give appropriate URIs for measures and dimensions. From the viewpoint of LOD, it is desirable to use common URIs for the same measures and dimensions. However, the URIs of Open DATA METI are not standard ones for industrial classifications and municipalities. In addition, as yet, there are no systems or services which can easily search them. Establishing a common basis for maintaining standard URIs is thus a challenge that should be dealt with in the future.

Moreover, there are many cases in which machine reading of statistical tables is difficult because tables are meant to be read by persons, not machines. For example, multiple tables may be included in a file, or the values of different dimensions may appear in the same row. In the former case, one could divide up a file into smaller ones that each includes only one table before being converted into LOD. In the latter case, rows/columns could be added for different dimensions, but grouping the values of the same dimensions would be a complicated task. A possible solution would be publishing statistics in line with the guideline[12] for creating data in table format.

§4 Conclusion

We constructed the catalog site and LOD search site of Open DATA METI according to the 5 star deployment scheme for open data. We described the construction and maintenance of the catalog site. We also showed utilities for collecting, registering, and updating the metadata. After that, we described the conversion of statistical tables into LOD for the LOD search site. We proposed a new approach to input the necessary information for LOD in the original table and automatically convert the table into LOD.

There are three main problems to maintain the spread of open data. The first problem is that it costs a lot to construct the site. It is necessary to publish a certain amount of data from the beginning to show the effect of publishing open data. The high initial cost for the site construction is a problem. Some local governments do not use existing software for maintaining their catalog sites. We hope that these local governments will shift to existing software having low operational costs.

The second problem is that catalog software has not yet matured. We needed to develop some functions because CKAN (version 1.8) was still in development at the time. For example, we limited the data registration functionality available to general users, because there was no classification of the users' role. We also developed a function for registering metadata of multiple datasets and resources in one lump. In addition, although CKAN has a group/dataset/resource hierarchy for structuring data, a deeper hierarchy may

ngc34403 : 2016/9/26(10:13)

be necessary. Furthermore, we need more engineers who know how to use open-source software for constructing catalog sites in order to let the whole country share the benefits of open data at a low cost.

Third, existing web sites and sites for publishing open data (catalog site and LOD search site) often handle the same or closely related data. However, the workflows for these different sites are not connected with each other; there is a need for a coherent seamless workflow. To solve this problem, a workflow could be made in such a way that the cost of publishing open data is minimized. In fact, we have analyzed and optimized such a workflow.[6] Here as well, there is a pioneering study[17] on economically publishing open data with an integrative system.

Our experience of developing Open DATA METI served as a reference model for DATA.GO.JP by the Cabinet Secretariat and for catalog sites of local governments. DATA.GO.JP switched over from a trial version to a full-scale edition in 2014, and the system for promoting open data of the government is being put into place. Sharing experiences, lessons, and results provided through these kinds of activities will be important in the future.

Acknowledgements

This study was part of METI's open data research project, running from 2012 to 2013. We would like to express our gratitude to all the people engaged in the survey, research, and development. Particularly, we would like to thank the members of METI and Hitachi Systems, Ltd.

References

1) Asano, Y., et al, "Template for Converting Statistical Data to RDF," *in Proc. of the 12th Forum on Information Technology (FIT2013), 12, 2*, pp. 361–362, 2013.
2) Asano, Y., et al, "A Template for Handling Statistical Data in RDF," *in Proc. of Second International Workshop on Semantic Statistics (SemStats2014)*, 2014.
3) Berners-Lee, T., "Linked Data," http://www.w3.org/DesignIssues/LinkedData.html, 2006.
4) Chang, C. H., et al., "A Survey of Web Information Extraction Systems," *IEEE Transactions, 18, 10*, pp. 1411–1428, IEEE, 2006.
5) Han, L., et al., "RDF123: From Spreadsheets to RDF," *Lecture Notes in Computer Science, 5318*, pp. 451–466, 2008.
6) Hitachi Consulting Co., Ltd., "Report of Research about Open Data Promotion," Ministry of Economy, Trade, and Industry, http://www.meti.go.jp/meti_lib/report/2014fy/E004103.pdf, 2014.
7) IT Strategic Headquarters, "Open Government Data Strategy," http://japan.kantei.go.jp/policy/it/20120704/text.pdf, 2012.
8) IT Strategic Headquarters, "Roadmap for Electronic Government Open Data Promotion," http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou3.pdf, 2013.

9) "G8 Open Data Charter," https://www.gov.uk/government/uploads/system/uploads/ attachment_data/file/207772/Open_Data_Charter.pdf, 2013.
10) Ministry of Economy, Trade, and Industry, "OpenDataMETI Application Profile V1.0," http://datameti.go.jp/wp-content/uploads/2014/01/OpenDATAMETI- ApplicationProfile_v1.0.pdf, 2014.
11) Koide, et al, "An LOD Practice - Lessons and Learned from Open Data METI," *in Proc. of 2013 Linked Data in Practice Workshop*, 2013.
12) "The Basic Idea about the Data Published by the Ministries for the Promotion of Secondary Use (Guideline) (Appendix)," http://www.kantei.go.jp/jp/singi/it2/densi/kettei/gl_betten.pdf, 2013.
13) Salas, P. R., et al., "Publishing Statistical Data on the Web," *in Proc. of IEEE Sixth International Conference on Semantic Computing*, pp. 285–292, IEEE Press, 2012.
14) Sauermann, L., et al., "Cool URIs for the Semantic Web," W3C Interest Group Note, http://www.w3.org/TR/cooluris/, 2008.
15) Takeda, H., et al., "Towards LOD of Species to Build the Biology Information Infrastructure," *in Proc. of the 26th Annual Conference of the Japanese Society for Artificial Intelligence*, pp. 1–4, 2012.
16) Takeda, H., et al., "Presentation of Statistical Data and their Relationships LOD," *in Proc. of the 27th Annual Conference of the Japanese Society for Artificial Intelligence*, 2013.
17) Taylor, S., et al., "Reasoning Driven Configuration of Linked Data Content Management Systems," *in Proc. of the 3rd Joint International Semantic Technology Conference (JIST2013), Lecture Notes in Computer Science, 8388*, pp. 429–444, 2014.
18) W3C, "OWL Web Ontology Language Reference," http://www.w3.org/TR/owl-ref/, 2004.
19) W3C, "The RDF Data Cube Vocabulary," http://www.w3.org/TR/vocab-data-cube/, 2014.

Appendix A: Metadata in Open DATA METI

Metadata for group, dataset, and resource in Open DATA METI is shown in the following table. The details are published in a document.[10)]

group	name, title, title_en, description, image_url, state, themes, dataset, publisher, homepage
dataset	name, title, publisher, contact_point, contact_email, contact_tel, creator, contributor, frequency, license_id, rights, notes, tag1, tag2, state, version, release_date, resources, url, spatial_geographical_coverage, temporal_coverage-from, temporal_coverage-to, supplementation
resource	name, url, description, format, size, release_date, last_modified, license, rights, resource_type, mimetype, mimetype_inner, language

Appendix B: Example of SPARQL query

An example query for searching the values of manufactured goods shipments and the values of stocks for each prefecture and industrial middle classi-

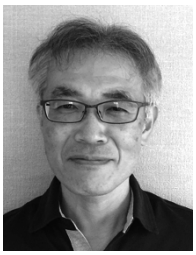
fication is shown below. Each of the variables "?pref," "?industry_mid," "?shipments" and "?stocks" corresponds to each of the prefectures, the industrial middle classifications, the values of manufactured goods shipments and the values of stocks. The query can be evaluated at the endpoint[*9] provided by METI.

```
PREFIX ktsh: <http://datameti.go.jp/scheme/kougyou-toukei-schema/>
select distinct ?pref ?industry_mid ?shipments ?stocks
where {
  {  select distinct ?pref_uri ?industry_mid_uri (SUM(?shipments_fine) AS ?shipments)
     where {
       ?obs1 ktsh:refPrefecture ?pref_uri.
       ?obs1 ktsh:valueOfManufacturedGoodsShipments_by10ThousandYen ?shipments_fine.
       FILTER(str(?shipments_fine) != "X")
       ?obs1 ktsh:refSangyoSaiBunrui ?industry_fine_uri.
       ?industry_small_uri skos:narrower ?industry_fine_uri.
       ?industry_mid_uri skos:narrower ?industry_small_uri.
     } GROUP BY ?industry_mid_uri ?pref_uri
  }
  ?obs2 ktsh:refSangyoChuBunrui ?industry_mid_uri.
  ?industry_mid_uri rdfs:label ?industry_mid.
  ?obs2 ktsh:refPrefecture ?pref_uri.
  ?pref_uri rdfs:label ?pref.
  ?obs2 ktsh:valueOfStocks_goodsInProgress_atYearEnd_byMillionYen ?stocks.
}
```

Yu Asano, Ph.D.: She is a researcher at Intelligent Information Research Department in Hitachi, Ltd. She received Ph.D. (2012) degrees from Department of Information Science and Technology, the University of Hokkaido. She is a member of Japanese Society for Artificial Intelligence and Information Processing Society of Japan. She is also an executive committee member of Linked Open Data Challenge Japan. Main research interests are the Semantic Web technology and natural language processing.

Seiji Koide, Ph.D.: He is a researcher at National Institute of Informatics (NII) Japan, and a board member of NPO Linked Open Data Initiative, a CEO of Ontolonomy, LLC. He received the B. Eng., M. Eng. from Nagoya University, Japan, and Ph.D. degrees from the Graduate University for Advanced Studies (Sokendai), in 1970, 1972 and 2011, respectively. He engaged in R&D of Rolling and Plasticity firstly and Artificial Intelligence secondly at IHI (Ishikawajima-Harima Heavy Industry, Co. LTD.). He is a permanent member of the Japan Society of Mech. Eng, a member of the Japan Society of Artificial Intelligence, Information Processing Society of Japan (IPSJ), Japan Society for Software Science and Technology. He is a member of AWG (Application Working Group) of the editorial committee in IPSJ.

ngc34403 : 2016/9/26(10:13)



Makoto Iwayama, Ph.D.: He is a senior researcher at Hitachi, Ltd. He received his Ph.D. (1992) from Tokyo Institute of Technology. His research interests are natural language processing and information retrieval. He is a member of ACM, Japanese Society for Artificial Intelligence (JSAI), Information Processing Society of Japan (IPSJ), and the Association for Natural Language Processing.

Fumihiro Kato: He is a researcher at Research Organization of Information and Systems (ROIS). He received his B.A. (2002) and M.M.G. (2004) from Keio University. His research interests include Web technologies like Linked data. He is a member of the Japanese Society for Artificial Intelligence.

Iwao Kobayashi: He is an Information Architect. His specialized fields are ICT and Community Development. CEO of Scholex since 2011. Co-founder of Laboratory urban DECODE since 2015. Vice Chief Director of NPO Linked Open Data Initiative since 2012.

Tadashi Mima: He is a director of Hitachi Consulting Co., Ltd. He had withdrawn from Doctorial program in Informatics with the completion of the program, The Graduate University for Advanced Studies. He is a Japanese committee member of ISO/IEC JTC 1/SC 27 WG5. He is a member of the Institute of Electronics, Information and communication Engineers (IEICE).

ngc34403 : 2016/9/26(10:13)

Ikki Ohmukai, Ph.D.: He received his Ph.D. degree in informatics from the Graduate University for Advanced Studies in 2005. He joined National Institute of Informatics in 2005 and has been an associate professor since 2009. His research interests are the semantic web and social media. He is a member of IPSJ and JSAI.

Hideaki Takeda: He is a professor at National Institute of Informatics (NII) Japan, and a professor at the Graduate University for Advanced Studies (Sokendai). He received the B. Eng., M. Eng. and Dr. Eng. degrees from the University of Tokyo, Japan, in 1986, 1988 and 1991, respectively. He worked at Norwegian Institute of Technology and Nara Institute of Technology prior to joining the current institution. He has been the Sumitomo endowed professor in the University of Tokyo between 2005 and 2010. His interest includes Semantic Web, Social Web, and Community-based systems.