The Outline of the Inter-Ministry Information System for Official Statistics and Advanced Use of Open Data

Kenichiro Kakihara

Statistics Bureau
Ministry of Internal Affairs and Communications

The outline of the Inter-Ministry Information System for Official Statistics (IMISOS)

Because the structure of statistics institutions in Japan has been decentralized, each ministry and agency used to operate similar systems relating to statistics.

- Therefore, the running cost of statistic information systems of each agency were duplicated.
- It was inconvenient that statistical tables were provided by each ministry and agency and that statistical users needed to access each system in the competent authorities.

Centralized the systems which were built by each individual ministry

Prepared the portal site which collects and integrates statistical information of all ministries

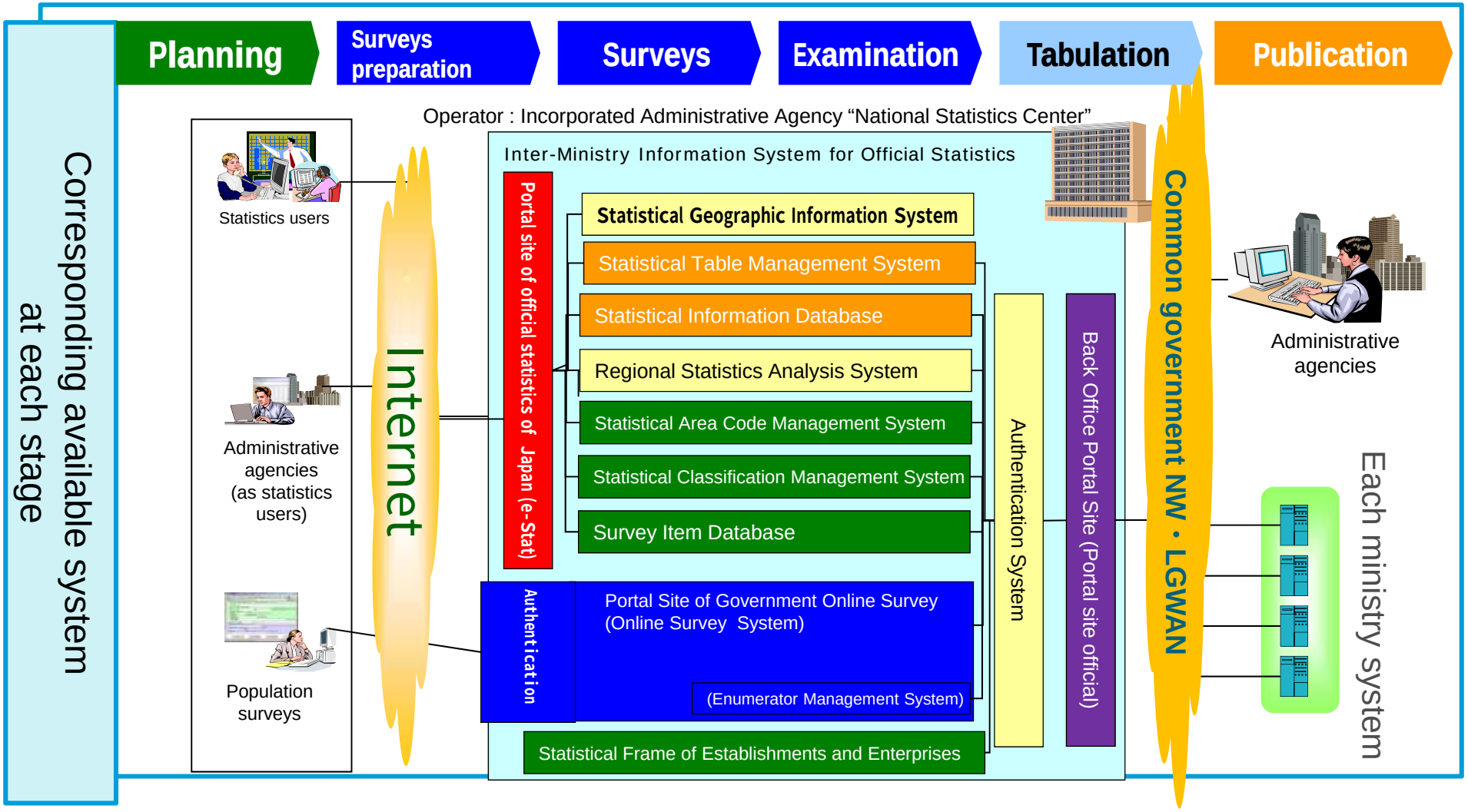
Building of the Inter-Ministry Information System for Official Statistics （since FY 2008）

i) To reduce running cost and building cost of statistical information systems

ii) Improvement of convenience for statistical users

The structure of the Inter-Ministry for Official Statistics System (IMISOS)

Each sub-system in the IMISOS is used at each stage; “Planning”, “Surveys preparation”, “Surveys”, “Examination”, “Tabulation” and “Publication” in public statistical work.



The Portal Site of Official Statistics of Japan

Portal site of statistical tables of government agencies

- The “Portal Site of Official Statistics of Japan (e-Stat)” which began in FY2008 provides statistical tables of government agencies in a unified and integrated manner.
- Establishes a database of major statistics including fundamental statistics.

●Statistics tables
Approx. 720,000 tables
（as of Sep30,2017）

●No. of accesses:
Approx. 30 million (FY2016)
*Accesses by crawlers not included

The specific efforts of the government for Open Data

"The undeveloped resource：Data" makes "a new value"

Declaration to be the World's Most Advanced IT Nation(Cabinet decision)

■ **"Open Data"; Private sector can access data in the public sector**

Data is released to the private sector on the Internet in machine-readable formats under the rule that enables secondary use.
Through sharing and utilization of information with various people, the new value will be created.

■ **"Aims of Governmental Open Data Strategy"**

- To ensure the administrative transparency
- To realize the administrative service which can fulfill various needs
- To vitalize the Japanese economy and to improve the industrial competitiveness

⇒ Official statistics is the front-runner in the field of open data.

Efforts for Open Data in Statistics

- Open Data have been implemented traditionally in the field of official statistics.
- Not simple Open Data but also efforts for advancement and sophistication are required there.
- Specifically, as the central statistical organization, the Statistics Bureau of Japan(SBJ) and National Statistics Center (NSTAC) are promoting the following three themes.

1 **Development of API function**

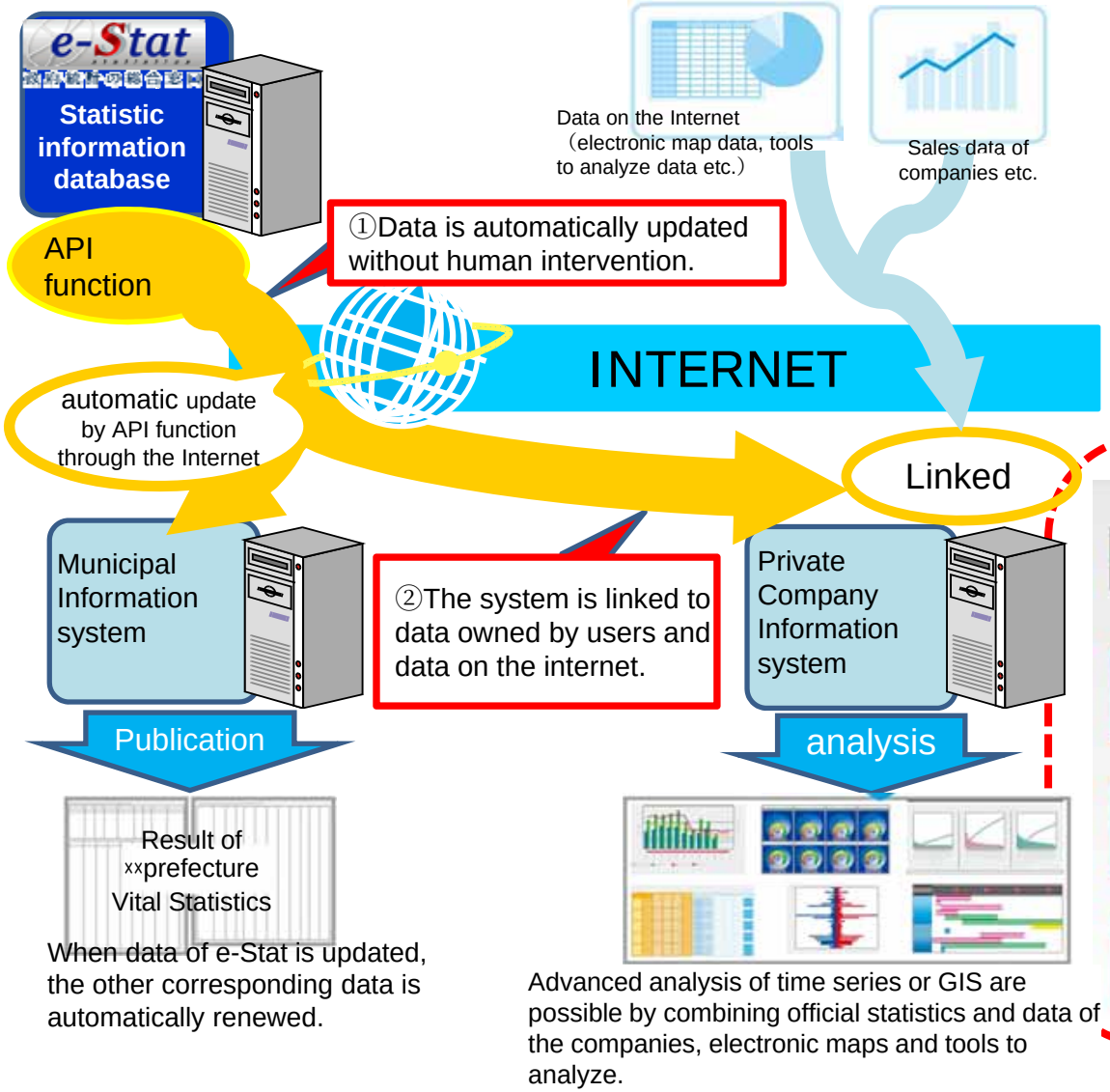
2 **Enhancement of Statistics GIS Function （jSTAT MAP）**

3 **Statistics LOD**

- These will promote advanced use of statistics by public and private sectors. And we support creation of service which generate new added value, innovative businesses and so on.

Development of an Environment for Advanced Use of Statistics by API

API function has been provided since October 2014



When data of e-Stat is updated, the other corresponding data is automatically renewed.

Advanced analysis of time series or GIS are possible by combining official statistics and data of the companies, electronic maps and tools to analyze.

By using the API function

(Application
Programming
Interface)

Statistical data on e-Stat is available like database of your own company.

For smart phones and tablets, We developed the application which includes API function.
Now it's available.

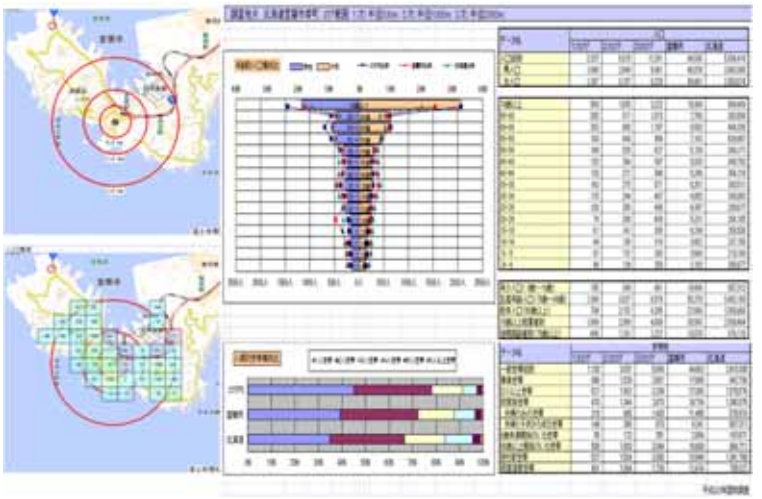
Enhancement of Statistics GIS Function （jSTAT MAP）

“jSTAT MAP(map sub-regional analysis function)” has been provided since January 2015.

Statistical GIS(Geographic Information System) which can visually display geographic information has already been provided on e-Stat, but this is further enhanced one.

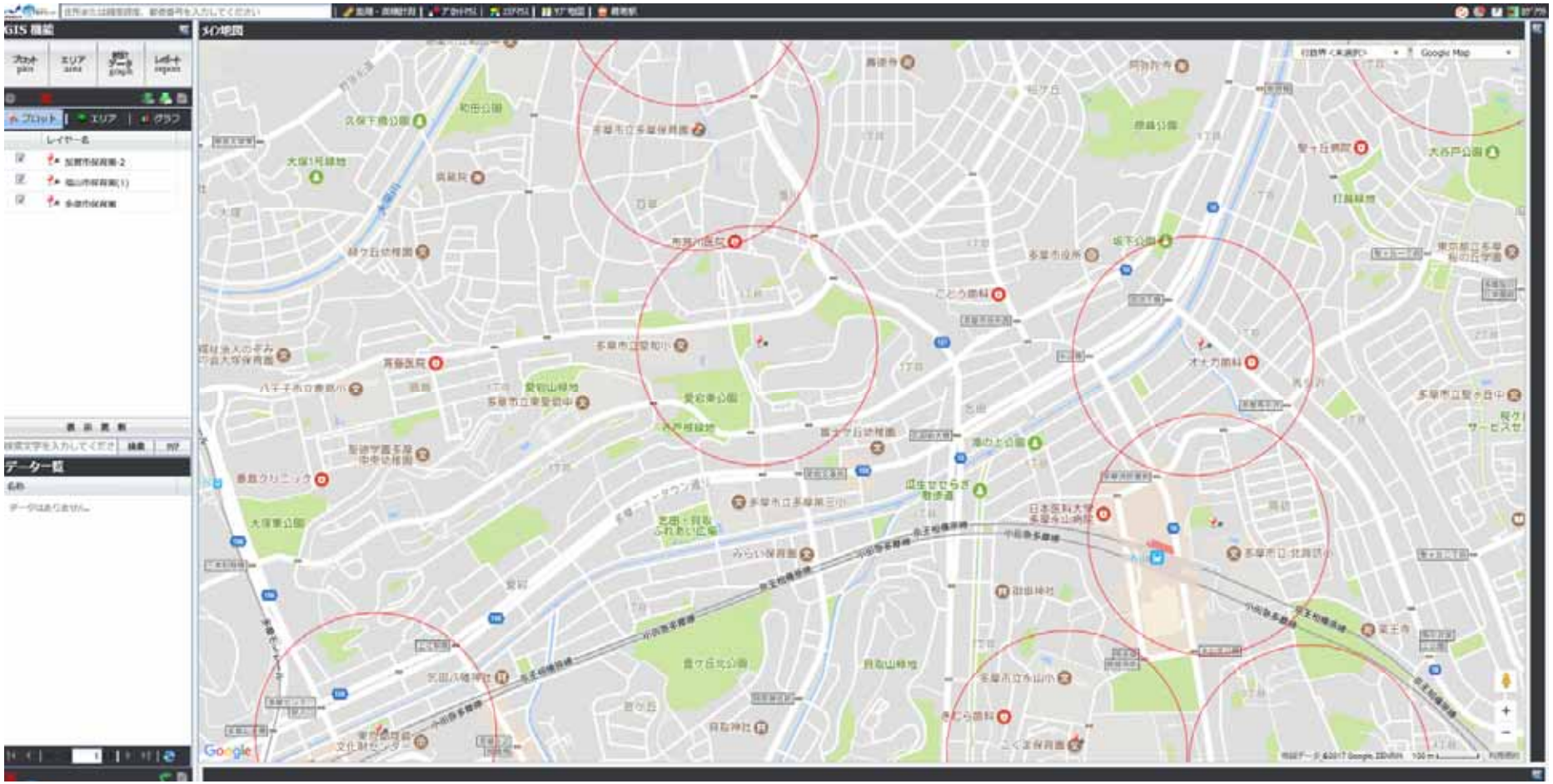
“Functions”
- **i**) It can import user’s data and analyze it.
- **ii**) It can display statistics data on an arbitrarily designated area.
- **iii**) Analytical results on the selected zone are output as a report.

The function to allow the acquisition and analysis of data possessed by users, and the utilization of data in any specified area visually,



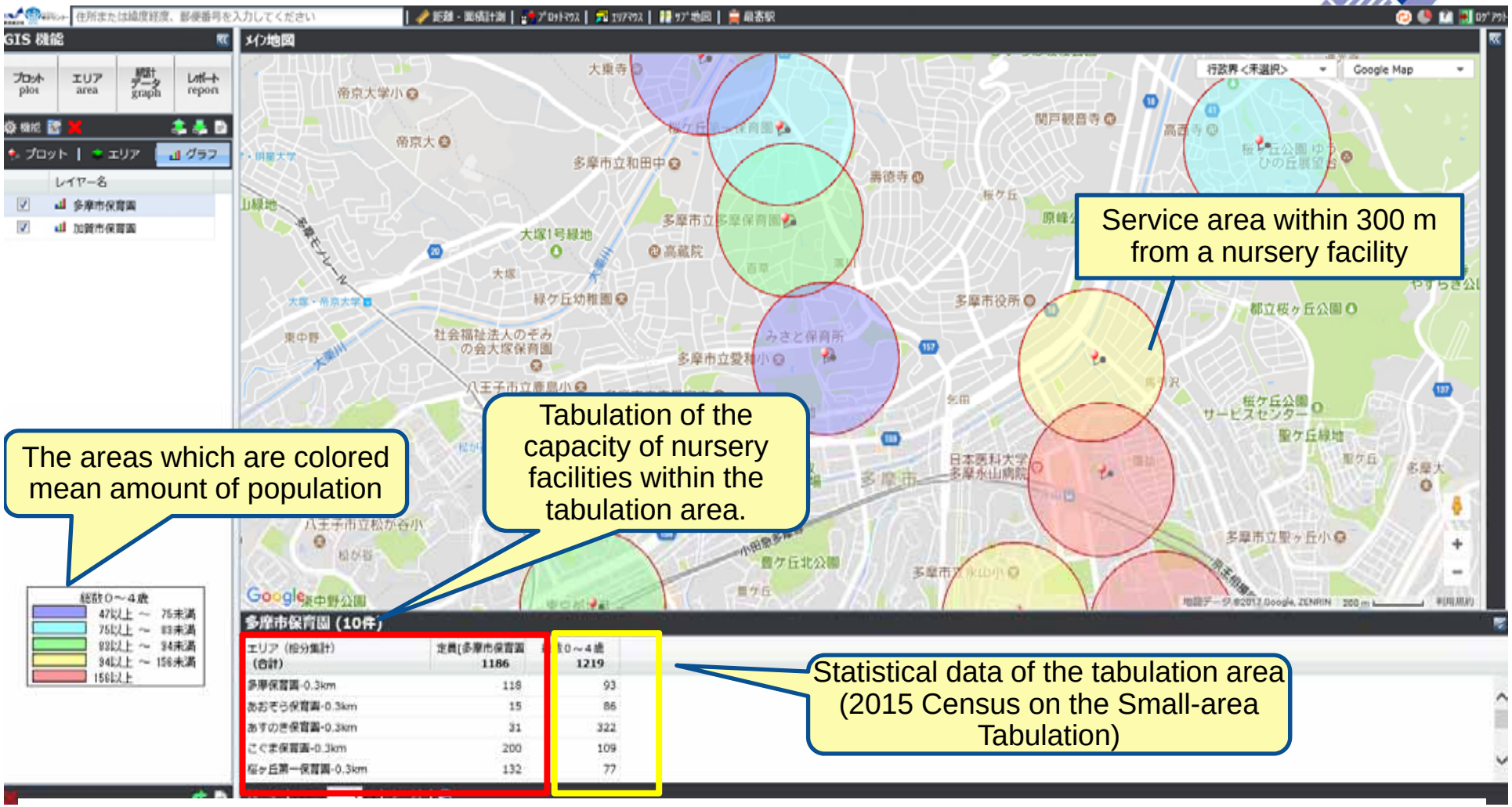
Just by specifying a central point, analytical results such as age structure in the concentric zone is compiled as a report in EXCEL format.

Paradigm of jSTAT MAP (1)



- This map is based on “Google Map”.
- You can point facilities on this map at once by importing a CSV-format list which contains its name, address and so on.
- The red pins shown on the screen indicate nursery schools.
- The circles around nursery schools indicate concentric circles with a radius of 500 m.

Paradigm of jSTAT MAP (2)



The population of 0 – 4-year-old children within a radius of 300 m from each nursery facility is indicated by importing the information of nursery schools, etc. to statistics GIS. This enables statistical understanding of the state of excess or deficiency of nursery facilities.

What's LOD

Five Levels of Open Data

Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5 star deployment scheme for Open Data

(Source: http://5stardata.info/)

★	make your stuff available on the Web (whatever format) under an open license
★★	make it available as structured data (e.g., Excel instead of image scan of a table)
★★★	use non-proprietary formats (e.g., CSV instead of Excel)
★★★★	use URIs to denote things, so that people can point at your stuff
★★★★★	link your data to other data to provide context **(LOD)**

Standardized technology
"World Wide Web Consortium(W3C)"

Definition of data

Resource Description Framework(RDF)
W3C Recommendation
1.0(22 February 1999) / 1.1(25 February 2014)

Search of data

SPARQL Query Language(SPARQL)
W3C Recommendation
1.0(15 January 2008) / 1.1(21 March 2013)

Definition of multi-dimensional data

RDF Data Cube Vocabulary
W3C Recommendation
(16 January 2014)

Convert to RDF

RDB to RDF Mapping Language (R2RML)
W3C Recommendation
(27 September 2012)

What's LOD

Change from "link to file" to "link to data"

Link to File

Create URI to page
(http://www.e-stat.go.jp/pages.html)

Create URI to file
(http://www.e-stat.go.jp/xls0001.xls)

Link to Data

Sex / Age / Standard area code	Total (Sex)			Male		
	...	44 years [Person]	45 years [Person]	...	44 years [Person]	45 years [Person]
...	...	...	...	...	...	...
Saitama-city	...	18,130	19,245	...	8,293	8,936
Kawaguchi-city	...	6,592	6,022	...	3,626	4,299
...	...	...	...	...	...	...

Create URI to data
(http://data.e-stat.go.jp/lod/.../C11201)

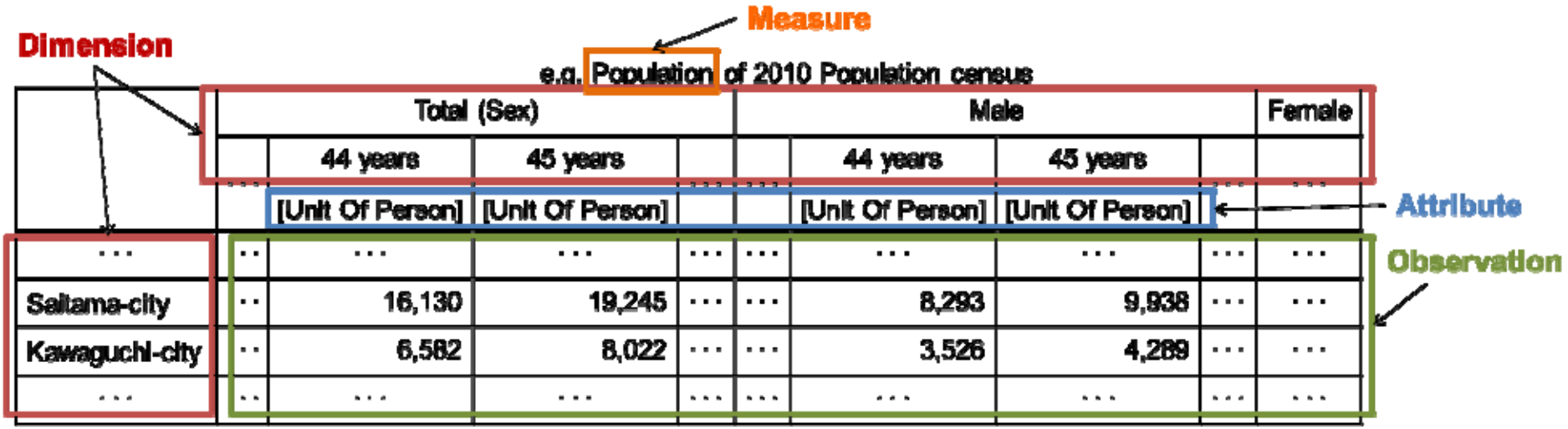
Create URI to data
(http://data.e-stat.go.jp/lod/.../obs00001)

Statistical LOD of e-Stat(e-Stat LOD)

The e-Stat LOD by using RDF Data Cube Vocabulary

The RDF Data Cube Vocabulary is compatible with the cube model that underlies SDMX(Statistical Data and Metadata eXchange).

Each observation is described by using dimensions, measure, and attribute



Search for e-Stat LOD

To search for e-Stat LOD, use SPARQL.
With SPARQL, you can cross-search multiple dataset.

```
SELECT  ?area ?numberOfHouseholds ?numberOfFacilities
WHERE { ?o1 pc-measure-2010:numberOfHouseholds ?numberOfHouseholds;
          pc-dimension-2010:typeOfHouseholdByPresenceOfAgedHouseholdMembers
             pc-code-2010:typeOfHouseholdByPresenceOfAgedHouseholdMembers-total;
          cd-dimension:standardAreaCode ?area_code.
        ?o2 ssds-measure-2016:J230121 ?numberOfFacilities;
          cd-dimension:standardAreaCode ?area_code.
        ?area_code sacs:administrativeClass sacs:Prefecture ;
        rdfs:label ?area.
        FILTER ( lang(?area)= "en" )
} ORDER BY ?area_code
```

Multiple dataset can be cross-searched

Population census

Population- prefectures

	Total	Male	Female
...	...	...	...
Tokyo-to	158,956	132,471	...
Kanagawa-ken	69,445	55,673	...
...	...	...	...

dataset

Economic census for business frame

numberOfEstablishments- prefectures

	Total	A	A01
...	...	...	...
Tokyo-to	158,956	132,471	...
Kanagawa-ken	69,445	55,673	...
...	...	...	...

dataset

System of social and demographic statistics

numberOfBirth- prefectures

	Birth	...	...
...	...	...	...
Tokyo-to	158,956	...	...
Kanagawa-ken	69,445	...	...
...	...	...	...

dataset

Statistics Dashboard

Provide major statistics created by governmental ministries and agencies, as well as local governments, on a dashboard* for easy visual understanding and simple data utilization by users (released in May 12 2017)

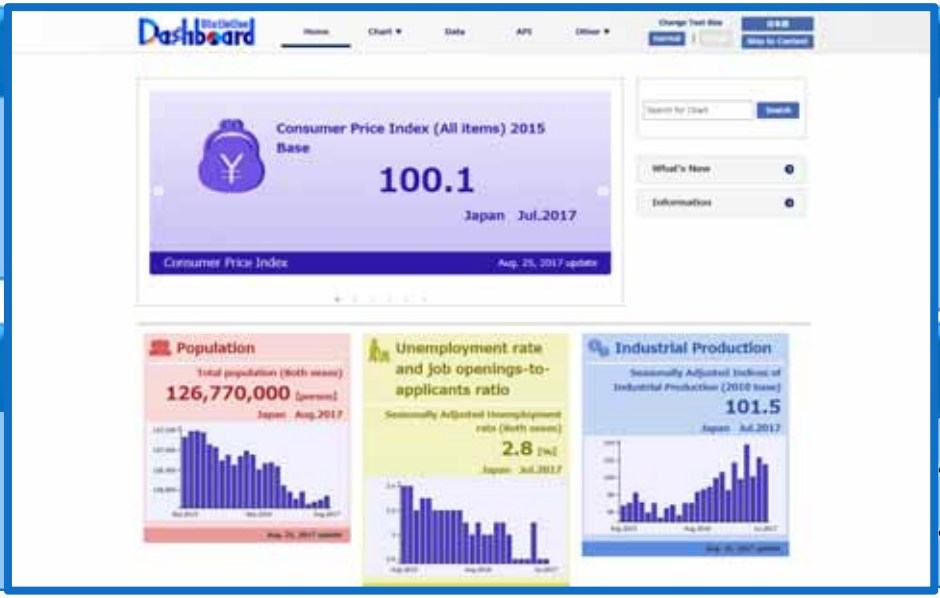
* Dashboard: Tool that summarizes various statistical data and displays a set of graphs and charts based on the processed data

Visualizing public statistics

- Provides major statistics in a visually comprehensible form
- Even usable by non-specialist users without specialized knowledge

Operability

- Provides data in graphs that are easy to manipulate
- Anyone can use and analyze data with ease



Accumulating a variety of data

- Accumulates large amounts of data
- Visualizes a wide range of data about both the current state and temporal changes in each prefecture as well as national indices

Accommodating user needs

- Priority display of data with higher priority by area
- Easier access to key indicators

Secondary use of Dashboard database data

- Public statistics are accessible without going through the Statistics Bureau website
- Diversified means of accessing statistics

<Use of statistical data by local governments>
Easier analysis of current state and temporal changes; Allows for comparative analysis with national and other region's statistics

<Use of statistical data by private organizations>
Using open public data from regional and business perspectives to create new ideas

<Use of statistical data in educational settings>
More practical lessons can be given, including exercises and introductions of analytical examples using public data

STAT DASH Grand Prix 2016

We carried out "STAT DASH Grand Prix 2016" with National Statistics Center to collect ideas which can utilize statistics data and API more efficiently.

The purposes are to improve service of “Portal Site of Official Statistics of Japan (e-Stat)” and to disseminate advanced use of official statistics.

There were two categories at this contest.

The amount of the application was 69 works.

- Development of administrative service : 37 works
- Enlightenment of utilizing statistic data : 32 works

There were 12 works which passed primary examination (each 6 works in each categories). After the presentation, The Minister of Internal Affairs and Communications selected The Minister of Internal Affairs and Communications Award.

The Winners of The Minister of Internal Affairs and Communications Award

Development of administrative service

Proposal of portal site which is named “e-Stat Junior” for primary and secondary students.

Kwansei Gakuin Senior High School Mathematical Sciences Club

The proposal is to provide portal site “e-Stat Junior” which can support statistic study for primary and secondary students.

“e-Stat Junior” provides statistic data which is based on “Course of Study”. Also it is made by not using professional terms and by adding kana (which enable children to read kanji easily) on kanji.

▲Screenshot of “e-Stat Junior”

Enlightenment of utilizing statistic data

Open source data visualization platform for open data - Development of E2D3 (Excel to D3.js) and Creation of education project of data literacy by E2D3

Development team of E2D3 ver.0.7

The proposal is to provide open-source software “Visualization platform made by everyone” which can show statistic data comprehensibly and visually by Excel.

“E2D3” will support improving data literacy (abilities to find necessary data, to coordinate and to comprehend) at various fields such as primary schools, universities or enterprises.

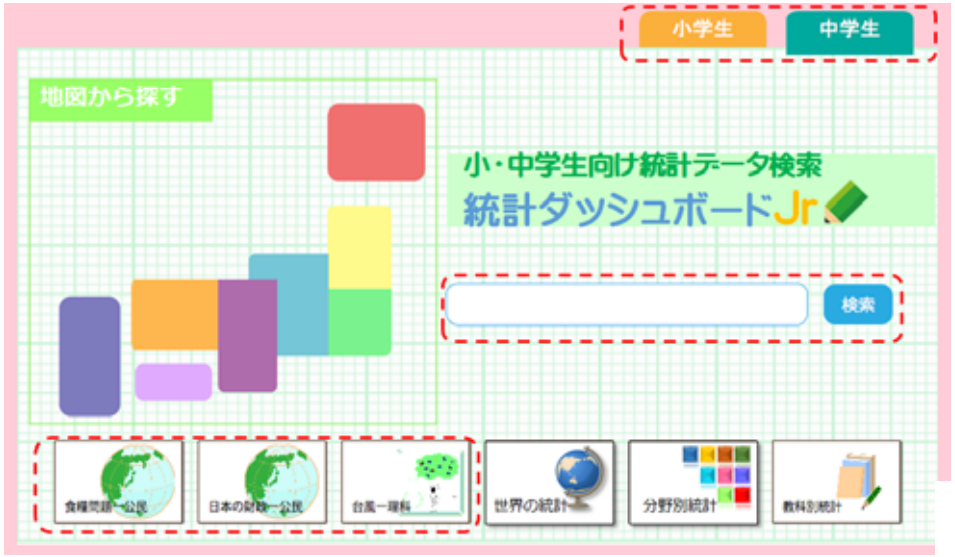
▲A graph made by E2D3

Development of the Statistics Dashboard for Elementary and Junior High School Students

In response to the content proposed by “e-Stat Junior,” which received the Minister of Internal Affairs and Communications Award, the statistics dashboard Jr. (tentative name) is in the process of development for the purpose of promoting the understanding and use of statistics by elementary and junior high school students (scheduled to be released in 2018).

- Indicating statistical data by region
- Searching statistical data with the keywords used in schoolbooks
- Indicating statistical data by school year and by subject, etc.

A statistics site easy to understand even for elementary and junior high school students, avoiding the use of technical terms.



International Development of ICT in the Official Statistics

Supporting the introduction according to the demand of each country with a focus on 5 subsystems of the “government statistics sharing system” owned by Japan.

Statistical table management system

e-Stat

Statistics geographical information system

Statistical information database

Online survey system

Thank you for your attention!

API functions
http://www.e-stat.go.jp/api/

jSTAT MAP –Small area analytics on maps–
https://jstatmap.e-stat.go.jp/

Statistics Dashboard
http://data.e-stat.go.jp/dashboard/?language=en