Statistics Japan

National Statistics Center
NSTAC

IAOS 2014 Conference – Meeting the Demands of a Changing World
Da Nang, Vietnam, 8-10 October 2014

Projects of Open Data for Official Statistics

Toshihiko AKATANI
Deputy Director,
Planning and Management Office,
National Statistics Center (NSTAC)

Table of Contents

I. Introduction

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

III. Project 2: Improvement of statistics GIS

IV. Project 3: Study of on-demand tabulation functions

V. Future work and some implications

VI. Conclusion

I. Introduction

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

III. Project 2: Improvement of statistics GIS

IV. Project 3: Study of on-demand tabulation functions

V. Future work and some implications

VI. Conclusion

G8 Open Data Charter (excerpt)

8. We therefore agree to follow a set of **principles** that will be the foundation for access to, and the release and re-use of, data made available by G8 governments. They are:

- Open Data by Default
- Quality and Quantity
- Useable by All
- Releasing Data for Improved Governance
- Releasing Data for Innovation

Picture: http://lod-cloud.net/

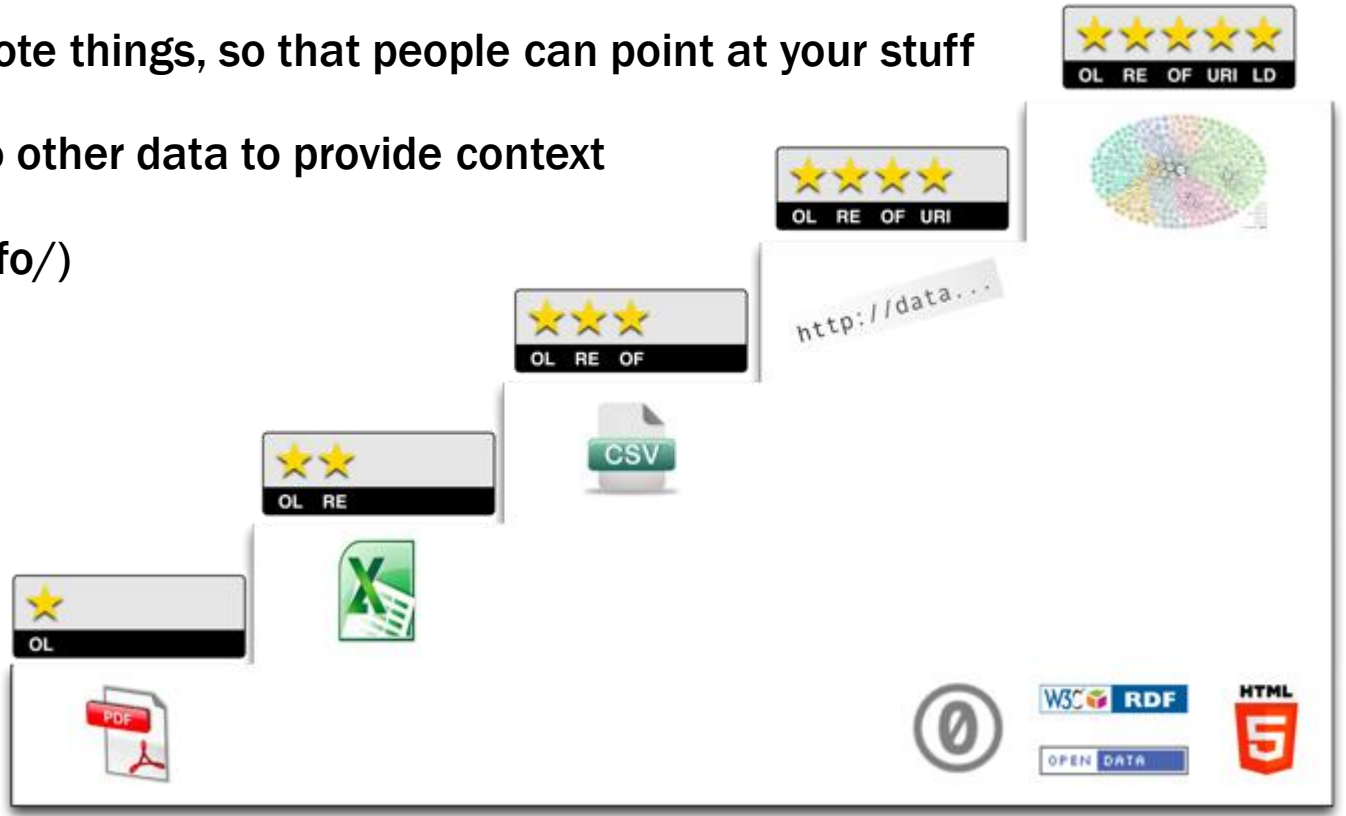
("G8 Open Data Charter", Published 18 June 2013)

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

- ★ 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

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



Japanese Government's Efforts on Open Data

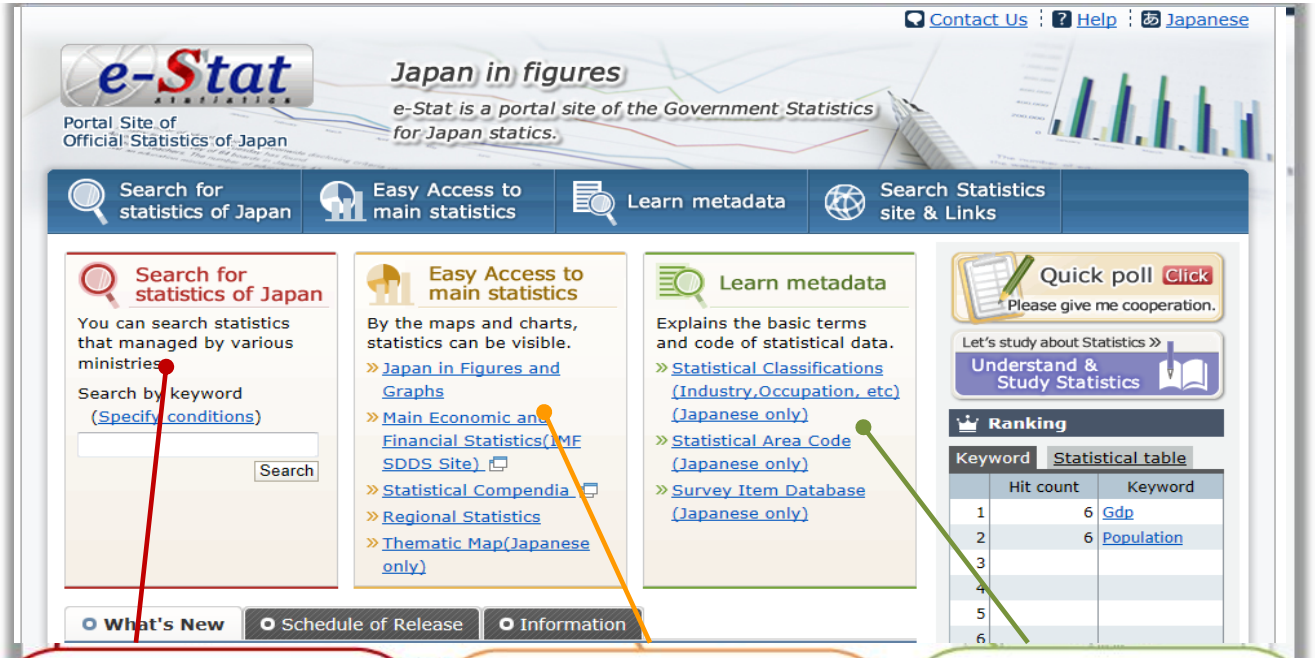
June 2013 Japan Revitalisation Strategy – JAPAN is BACK – (adopted by the Cabinet)

◎ Make the 2-year period between FY2014 and FY2015 the intensive period for taking measures
◎ Launch the data catalog website (trial version)
◎ Offer the world's top-level data sets to disclose (more than 10,000) by the end of FY2015

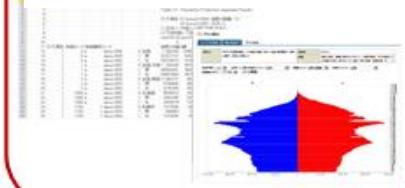
Data Catalog Site "data.go.jp" beta has launched (Dec 20, 2013)

Past Open Data Initiatives in Statistics

- The "Portal Site of Official Statistics of Japan (e-Stat)" established in FY2008 provides statistical tables of government agencies in a unified and integrated manner.
- Establish the database of "Fundamental Statistics" and other statistics.



Allows you to download statistics charts or prepare various graphs including population pyramid.



The use of regional statistics (statistics GIS) allows you to clearly understand what the region looks like.

Allows you to examine in detail questionnaires or their items for statistical surveys.

- No. of accesses:
  Approx. 18 million (FY2013)
  *Accesses by crawler not included
- Analysis of trends of users
  Personal users: Approx. 50%
  Private companies: Approx. 22%
  Gov. offices: Approx. 15%
  Unis and educational institutions: Approx. 10%
  Academic research institutes: Approx. 3%

- Statistics tables
  Approx. 500,000 tables for approx. 460 government statistics
- Statistical information database
  Approx. 60,000 tables for 40 Fundamental Statistics and 9 other statistics

Government's Open Data Initiatives in Statistics

(1) Data shall be provided in formats appropriate for machine reading. -> Provide data in formats (e.g. XML) that allow you to automatically reuse (e.g. process, edit) the data

- e-Stat provides the statistics tables for almost all countries (in Excel, CSV or other formats) and statistics data (in the XML format) for the major statistics. However, it requires users to manually operate the computer and download data to acquire the data or tables.
- It is necessary to provide the function that allows you to automatically and mechanically acquire statistics data and facilitate data processing or editing.

- Promote "advanced open data in statistics" including the establishment of API that allows you to automatically and mechanically acquire statistics data

(2) Data shall be released under the rules that enable secondary use. -> Reexamine the terms of use of websites, etc.

- It is necessary to reexamine the terms of use to widely allow for the secondary use of data.

- The government has adopted the template for the terms of use of websites covering open data, which is scheduled to be implemented in FY2014. (Set out the principles of clearly stating the source and allowing for secondary use including the use for commercial purposes. Any exceptions resulting from legal or other restrictions shall be clearly stated.)

Projects of Open Data for Official Statistics

◆ As the central statistical organization, Statistics Bureau of Japan (SBJ) and National Statistics Center (NSTAC) are promoting the following three themes which will upgrade the methods for disseminating voluminous and diversified statistical data to the next generation level, and enable their advanced use.

1. Development of an Environment for Advanced Use of Statistics by API

2. Improvement of Statistics GIS

3. Study of on-demand tabulation functions

◆ This will promote advanced use of statistics by public and private sectors; support for creation of services which generate new value added, and of innovative businesses; and so on.

I. Introduction

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

III. Project 2: Improvement of statistics GIS

IV. Project 3: Study of on-demand tabulation functions

V. Future work and some implications

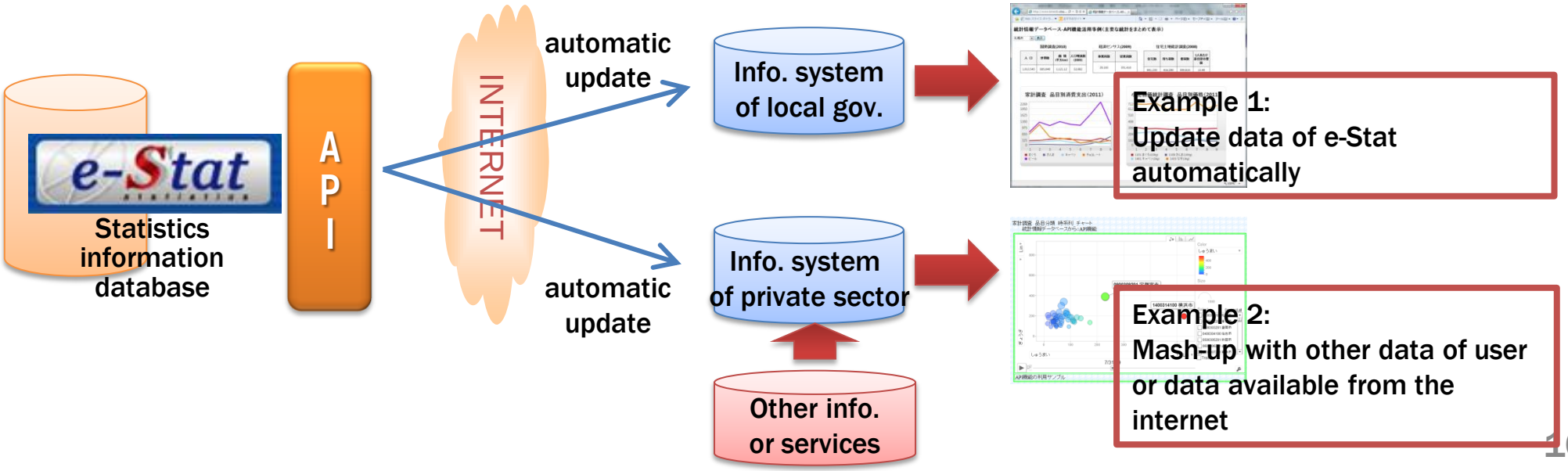
VI. Conclusion

Development of an Environment for Advanced Use of Statistics by API

- SBJ and NSTAC started a trial run of API functions for official statistics on 10 June 2013 in order to develop an environment for advanced use of statistical data.
- As of 31 July 2014, 2,131 users registered themselves, and the number is increasing steadily.
- To shift to high gear in 2014, functions and stress tests are being verified, and comments from users are being collected.

Outline of API function

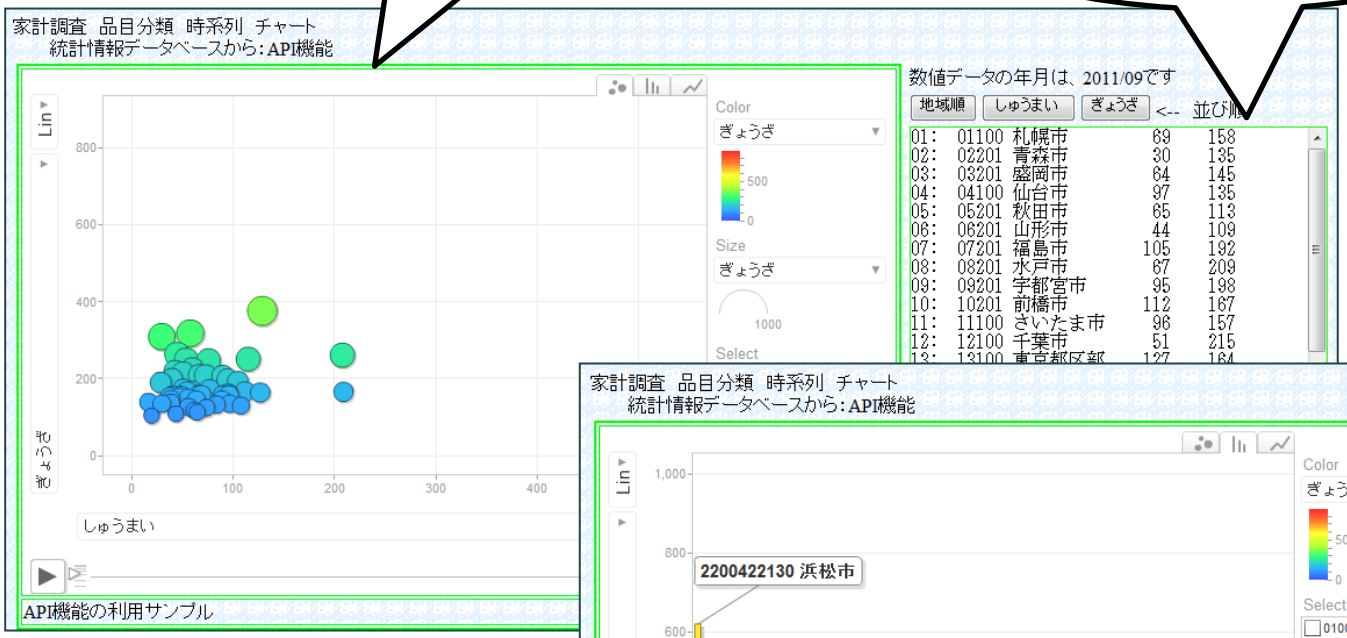
API (Application Programming Interface) function has been newly added to e-Stat enabling the conversion of statistics data to machine readable data.

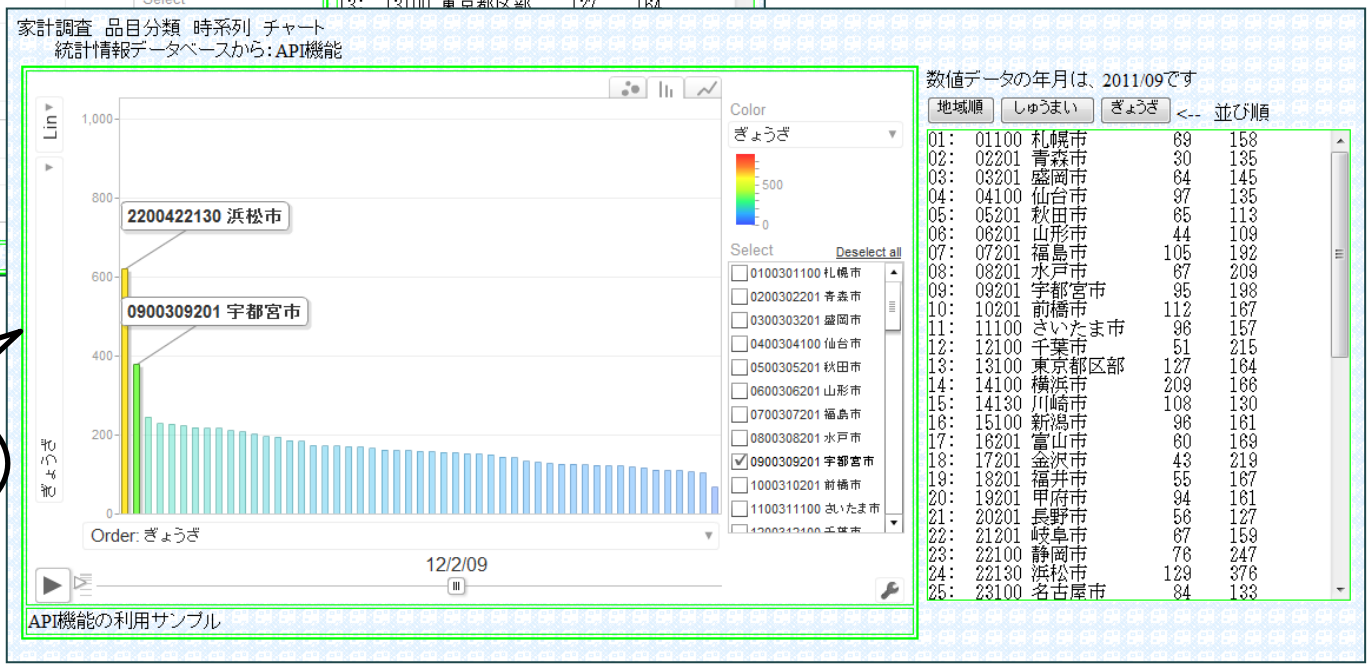


Analytical example using API functions

Presenting data in a bubble chart

Presenting the latest monthly data every time the data base is updated





数値データの年月は、2011/09です

地域順 しゅうまい ぎょうざ <-- 並び順

01:	01100	札幌市	69	158
02:	02201	青森市	30	135
03:	03201	盛岡市	64	145
04:	04100	仙台市	97	135
05:	05201	秋田市	65	113
06:	06201	山形市	44	109
07:	07201	福島市	105	192
08:	08201	水戸市	67	209
09:	09201	宇都宮市	95	198
10:	10201	前橋市	112	167
11:	11100	さいたま市	96	157
12:	12100	千葉市	51	215
13:	13100	東京都区部	127	164
14:	14100	横浜市	209	166
15:	14130	川崎市	108	130
16:	15100	新潟市	96	161
17:	16201	富山市	60	169
18:	17201	金沢市	43	219
19:	18201	福井市	55	167
20:	19201	甲府市	94	161
21:	20201	長野市	56	127
22:	21201	岐阜市	67	159
23:	22100	静岡市	76	247
24:	22130	浜松市	129	376
25:	23100	名古屋市	84	133

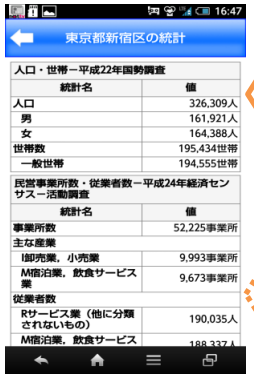
Presenting data in a bar chart

The statistics data of consumer spending on Chinese dumplings (results of Family Income and Expenditure Survey) by prefectural capital or ordinance-designated city is acquired by API and superimposed with the program available on the Internet (mash-up).

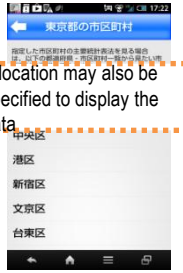
Official App by SBJ on Smartphone (App on Statistics)

- A trial version of “App on Statistics” (for Android) was released on 15 April 2014
- The app interlocks the statistical API functions with the smartphone GIS to display the statistics data of your current location, and provides other functions that allow you to feel familiar with statistics data.

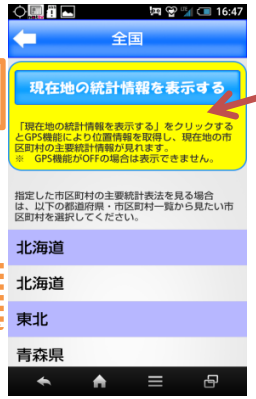
City Stat



Displays statistics data of your location using GPS



A location may also be specified to display the data



Pocket Statistics

Displays a list of statistics data on each item

Link to SBJ top page

Statistics Clock

Displays the statistical information and quiz of the day

Major statistics data available so far (SBJ data):

- Population/households (Population Census)
- Number of private establishments and employees (Economic Census for Business Activity)
- Major prices (Retail Price Survey)
- Monthly income and expenditure (Family Income and Expenditure Survey), etc.

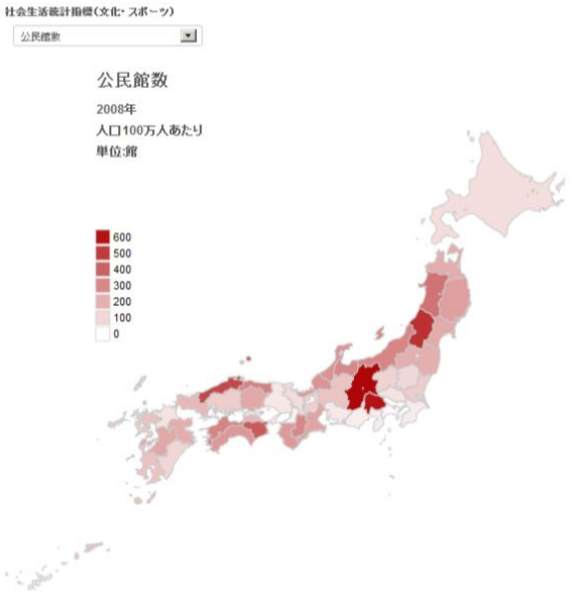
*Use of the API functions allows you to always acquire the latest statistics data.

Applications of API Functions

Mapping prefectures with respect to indicators related to sports and culture based on STATISTICAL OBSERVATIONS OF PREFECTURES.

（NIKKEI (Japan Economic Newspaper) Open Data Information Portal Site）

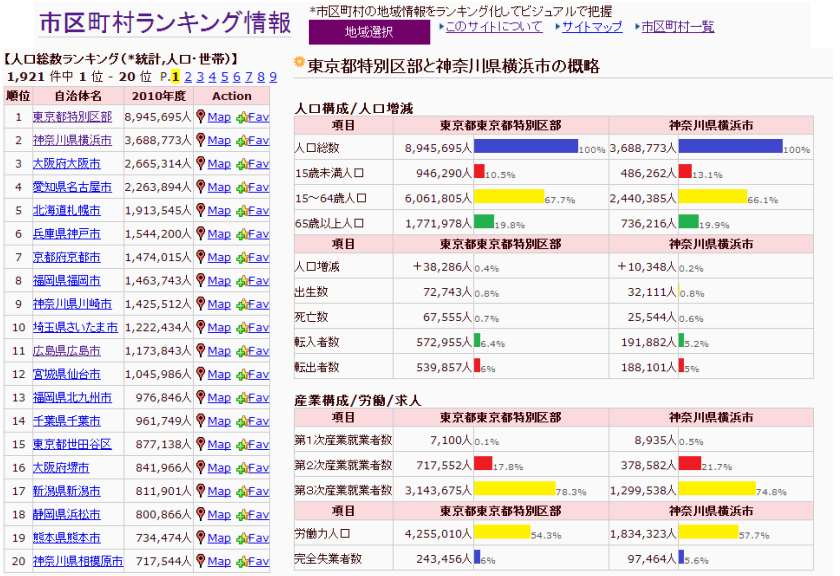
http://opendata.nikkei.co.jp/article/201306statdb/



Ranking all the municipalities of the whole country with respect to area information

（Corporation M&A bank）

http://city.ma-bank.net/



I. Introduction

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

III. Project 2: Improvement of statistics GIS

IV. Project 3: Study of on-demand tabulation functions

V. Future work and some implications

VI. Conclusion

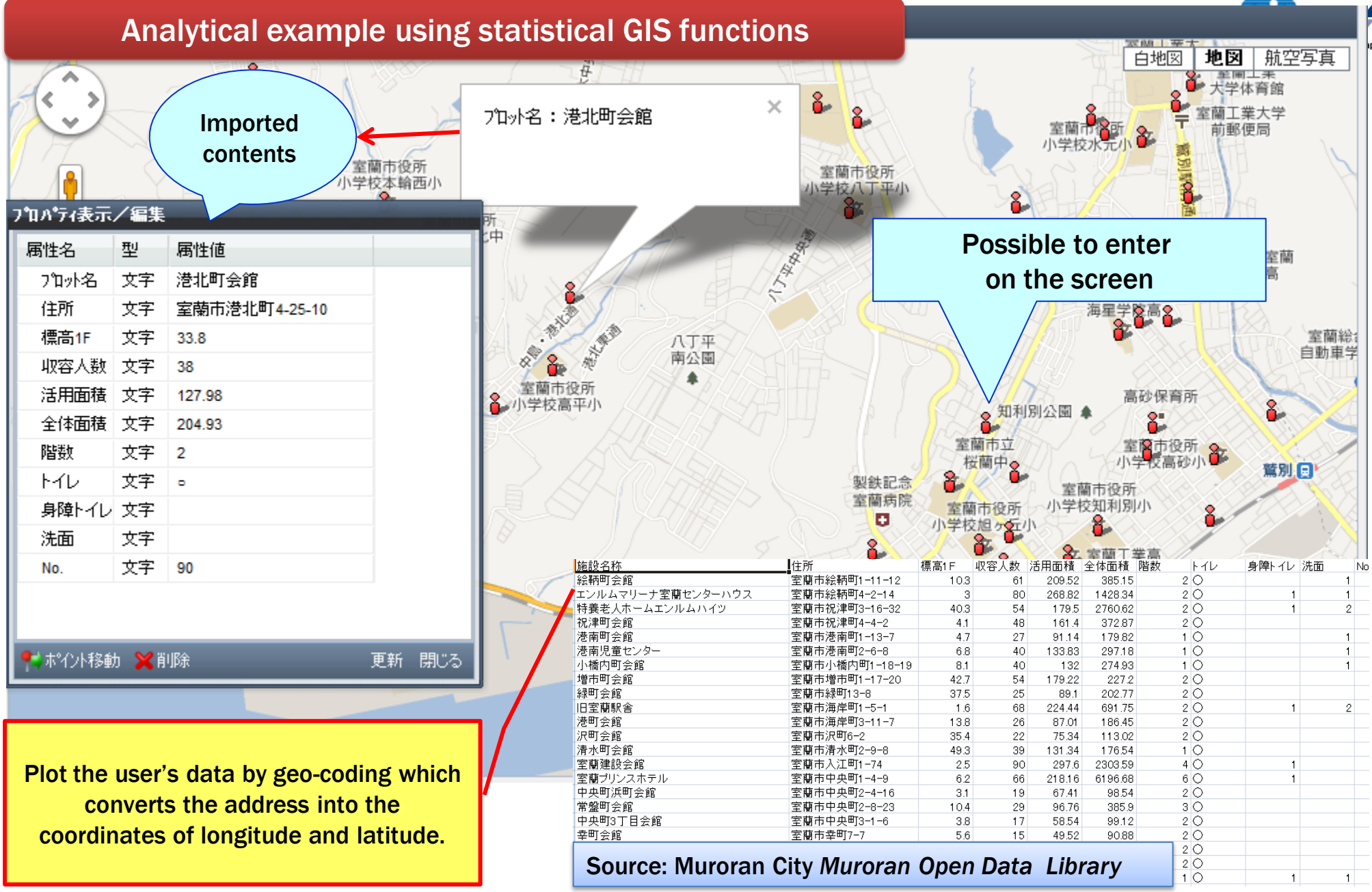
Improvement of Statistics GIS

- To improve Statistics GIS on e-Stat, there is a new system under development that enables retrieving data held by users and analysis of statistics data in an arbitrarily designated area.

The addition of functions enabling
①retrieving data held by user
②compiling statistics data in an arbitrarily designated area

- On 18 October 2013, a trial run started under a user registration system.
（Functions and stress tests are being verified through provisional dissemination of statistics provided by SBJ. 1,077 users are registered as of 31 July 2014）

Analytical example using statistical GIS functions



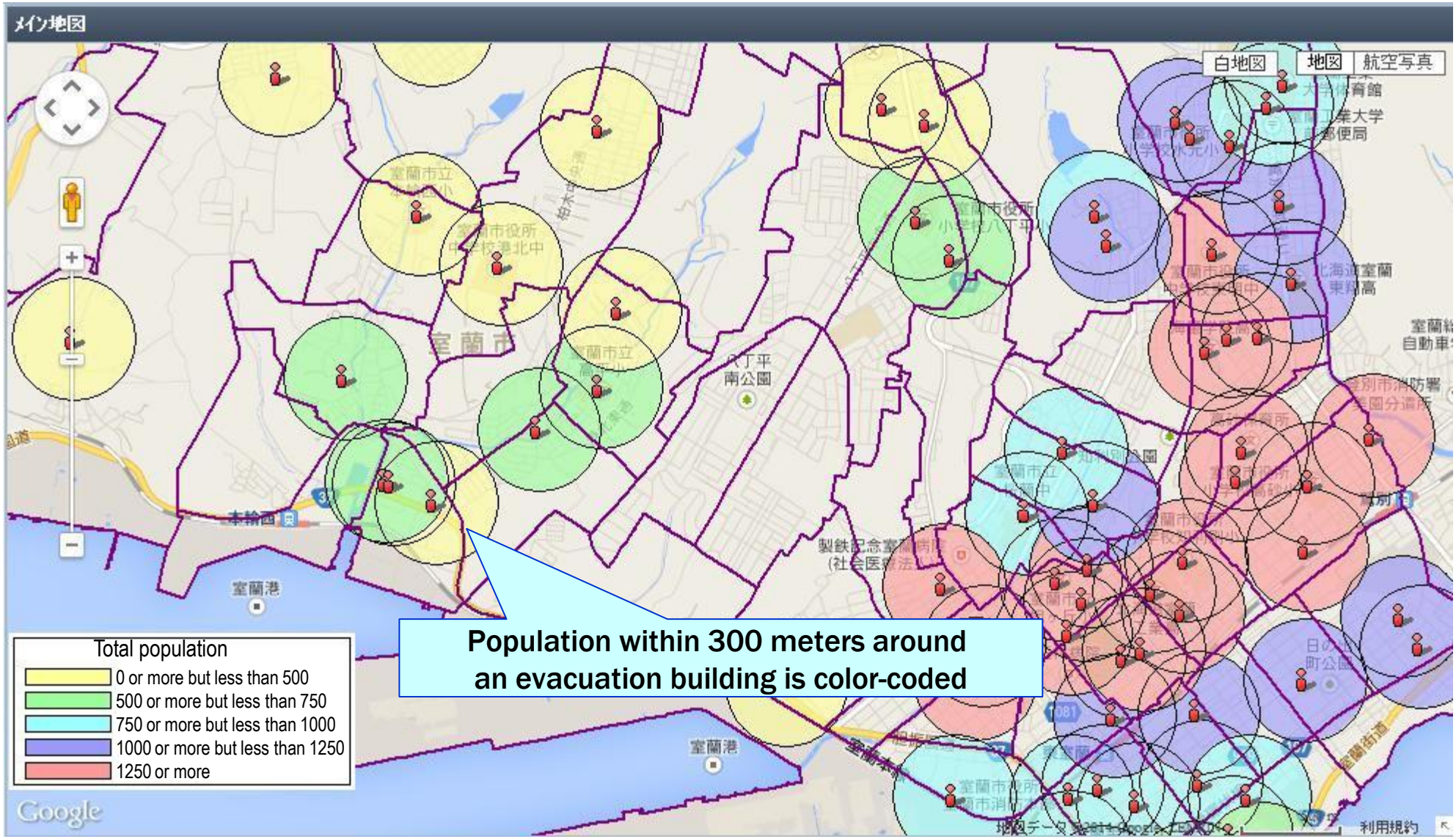
施設名称	住所	標高1F	収容人数	活用面積	全体面積	階数	トイレ	身障トイレ	洗面	No
絵鞆町会館	室蘭市絵鞆町1-11-12	10.3	61	209.52	385.15	2	○		1	
エンルムマリーナ室蘭センターハウス	室蘭市絵鞆町4-2-14	3	80	268.82	1428.34	2	○	1	1	
特養老人ホームエンルムハイツ	室蘭市祝津町3-16-32	40.3	54	179.5	2760.62	2	○	1	2	
祝津町会館	室蘭市祝津町4-4-2	4.1	48	161.4	372.87	2	○			
港南町会館	室蘭市港南町1-13-7	4.7	27	91.14	179.82	1	○		1	
港南児童センター	室蘭市港南町2-6-8	6.8	40	133.83	297.18	1	○		1	
小橋内町会館	室蘭市小橋内町1-18-19	8.1	40	132	274.93	1	○		1	
増市町会館	室蘭市増市町1-17-20	42.7	54	179.22	227.2	2	○			
緑町会館	室蘭市緑町13-8	37.5	25	89.1	202.77	2	○			
旧室蘭駅舎	室蘭市海岸町1-5-1	1.6	68	224.44	691.75	2	○	1	2	
港町会館	室蘭市海岸町3-11-7	13.8	26	87.01	186.45	2	○			
沢町会館	室蘭市沢町6-2	35.4	22	75.34	113.02	2	○			
清水町会館	室蘭市清水町2-9-8	49.3	39	131.34	176.54	1	○			
室蘭建設会館	室蘭市入江町1-74	2.5	90	297.6	2303.59	4	○	1		
室蘭プリンスホテル	室蘭市中央町1-4-9	6.2	66	218.16	6196.68	6	○	1		
中央町浜町会館	室蘭市中央町2-4-16	3.1	19	67.41	98.54	2	○			
常盤町会館	室蘭市中央町2-8-23	10.4	29	96.76	385.9	3	○			
中央町3丁目会館	室蘭市中央町3-1-6	3.8	17	58.54	99.12	2	○			
幸町会館	室蘭市幸町7-7	5.6	15	49.52	90.88	2	○			
						2	○			
						2	○			
						1	○	1	1	

Plot the user's data by geo-coding which converts the address into the coordinates of longitude and latitude.

Source: Muroran City *Muroran Open Data Library*

In the analysis, municipal information on evacuation buildings or shelters within a city in case of a disaster is incorporated into the statistical GIS functions to display an estimate of the population within the area of an evacuation site.

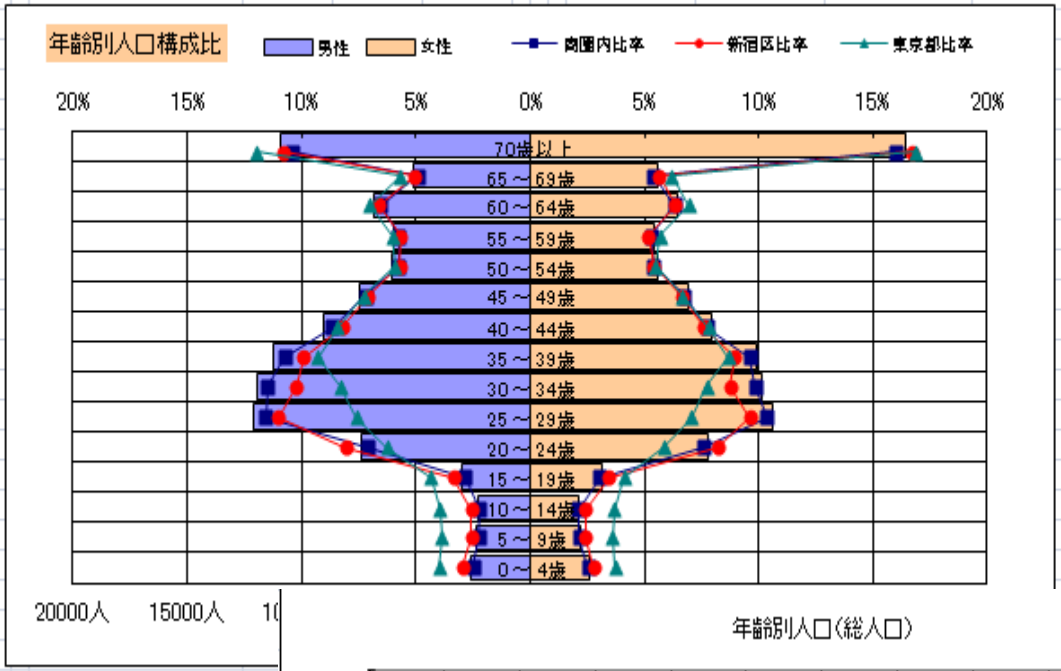
Analytical example using statistical GIS functions



In the analysis, municipal information on evacuation buildings or shelters within a city in case of a disaster is incorporated into the statistical GIS functions to display an estimate of the population within the area of an evacuation site.

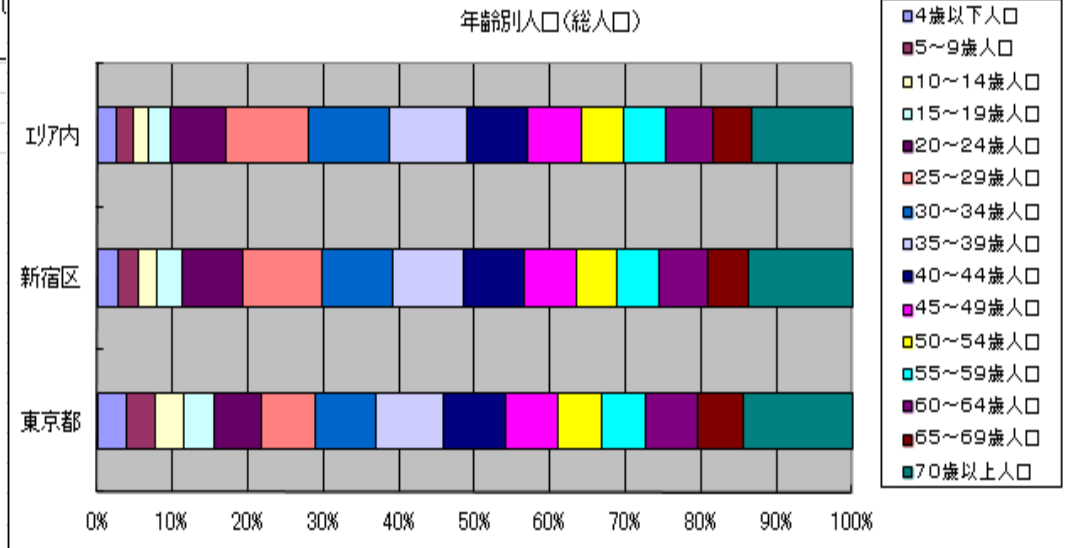
Examples of GIS Function （”Rich Report”）

調査地点　東京都新宿区西新宿2丁目　エリア範囲　1次：半径500m　2次：半径1000m　3次：半径2000m



データ名	人口				
	1次エリア	2次エリア	3次エリア	新宿区	東京都
人口総数	4,807	43,920	213,500	326,309	13,159,388
男人口					
女人口					
70歳以上					
65-69					
60-64					
55-59					
50-54					
45-49					
40-44	379	3,522	16,956	25,211	1,053,232
35-39	492	4,571	21,093	30,007	1,164,057
30-34	536	4,845	22,087	30,186	1,038,768
25-29	561	4,888	22,684	32,768	949,354
20-24	338	3,001	15,173	25,904	785,911
15-19	116	1,174	6,074	10,791	546,573
10-14	62	744	4,466	7,943	492,799
5- 9	70	789	4,523	7,952	484,303
0- 4	110	995	5,186	9,105	500,269

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



データ名		エリア内	
		値	割合(%)
総人口		213,500	
年齢別人口(総人口)	4歳以下人口	5,186	2.4
	5～9歳人口	4,523	2.1
	10～14歳人口	4,466	2.1
	15～19歳人口	6,074	2.8
	20～24歳人口	15,173	7.1
	25～29歳人口	22,684	10.6
	30～34歳人口	22,087	10.3
	35～39歳人口	21,093	9.9
	40～44歳人口	16,956	7.9
	45～49歳人口	14,428	6.8
	50～54歳人口	11,582	5.4
	55～59歳人口	11,450	5.4
	60～64歳人口	13,270	6.2
	65～69歳人口	10,693	5.0
	70歳以上人口	27,310	12.8

I. Introduction

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

III. Project 2: Improvement of statistics GIS

IV. Project 3: Study of on-demand tabulation functions

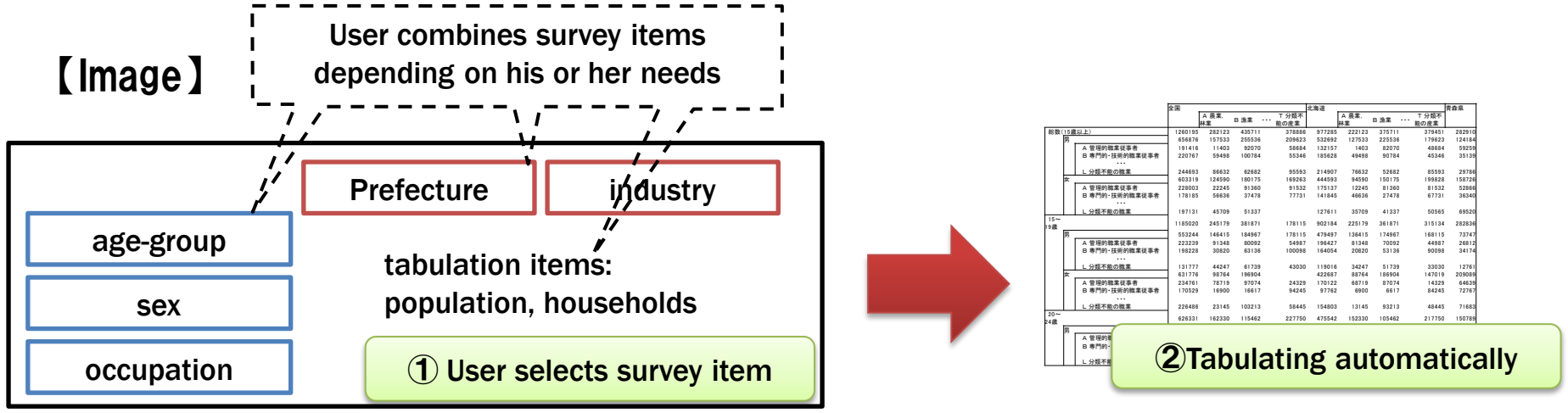
V. Future work and some implications

VI. Conclusion

Study of On-demand Tabulation Functions

- A new statistics delivery service which puts out statistics table automatically when a user selects survey item is currently under study. This service will possibly be used by the public as well as the academic sector.
- It is expected that this new function will answer various needs such as academic research, enabling arbitrary cross tabulation which is not included in existing tables.

*For practical use, there are some issues to be solved regarding operational and institutional aspects and confidentiality.



Mechanism of On-demand Tabulation Functions

1. NSTAC develops Hyper data cube using microdata

Microdata

development

Hyper data cube

•Tabulation
•Disclosure Control

INTERNET

2. Users designate survey items

Users

3. On-demand tabulation functions apply tabulation and disclosure control from Hyper data cube and output secured statistical tables automatically

独立行政法人 統計センター NSTAC

Initiatives of On-demand Tabulation Function

Base data	Outline	Adoption example
Multidimensional tables or data cubes	Providing tabulation prepared in advance, tabulating in combinations by assumable needs	“StatLine,” Statistics Netherlands “Interactive Data Dissemination System,” Census and Statistics Department of Hong Kong
Hyper data cube (High level multidimensional table)	One table tabulated in all survey items and classifications	“Census CDATA Online,” Australian Bureau of Statistics
Microdata	Search and tabulation by microdata	“Advanced Query System,” US Census Bureau “Census Table Builder,” Australian Bureau of Statistics

I. Introduction

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

III. Project 2: Improvement of statistics GIS

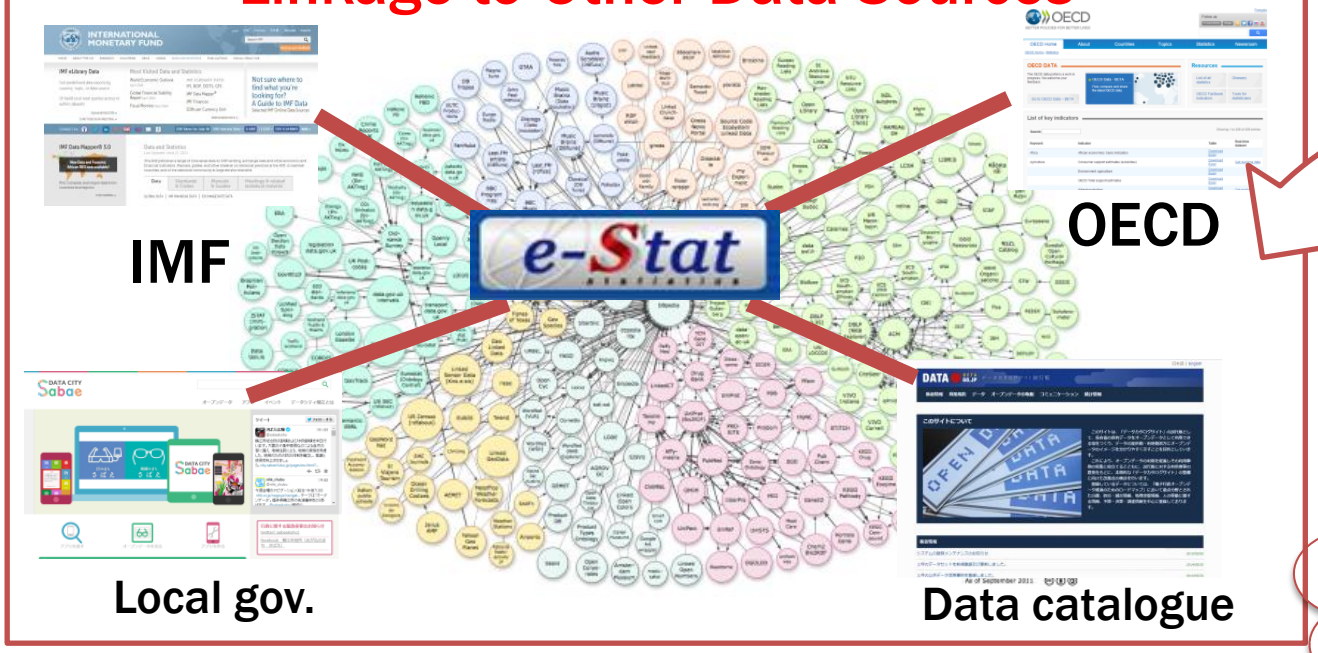
IV. Project 3: Study of on-demand tabulation functions

V. Future work and some implications

VI. Conclusion

Future Work and Some Implications

Linkage to Other Data Sources

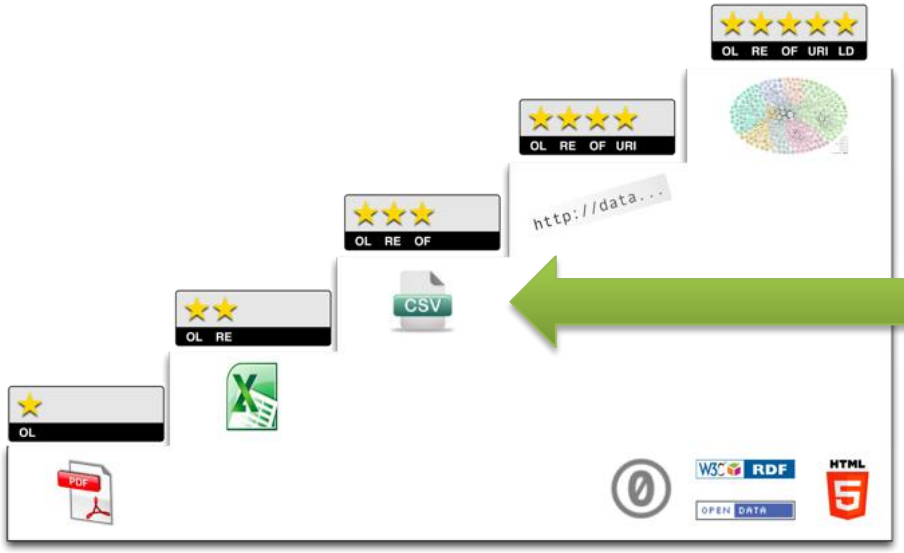


the Highest Level
(Five stars)

In the Future

API functions
(Three stars)

At Present



I. Introduction

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

III. Project 2: Improvement of statistics GIS

IV. Project 3: Study of on-demand tabulation functions

V. Future work and some implications

VI. Conclusion

Conclusion

- Statistics sector leads Open Data policy as a top runner
- Machine readable format such as API would enable more advanced data analytics within less burden of retrieving data
- Renovation of statistics GIS would stimulate and activate various Open Data initiatives
- On-demand tabulation function will fulfill various needs as a new form of secondary use of official statistics

Thank you for your attention!

Toward enhancement of open data for official statistics

Gateway to Advanced and User-friendly Statistics Services (GAUSS)

http://statdb.nstac.go.jp/

*You can access the GAUSS website from the SBJ or the NSTAC homepage. As of October 2014, the GAUSS website is in Japanese only.