World Grid Square Statistics and their Application to Data Analytics

Aki-Hiro Sato*, Shoki Nishimura**, Naoki Makita***, Tsuyoshi Namiki****, Hiroe Tsubaki*****

* Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University & Japan Science and Technology Agency, Yoshida Honmachi, Sakyo-ku, Kyoto 606-8501 JAPAN, sato.akihiro.5m@kyoto-u.ac.jp

** National Statistics Center, 19-1 Wakamatsu-cho, Shinjuku-Ku, Tokyo 162-8668 JAPAN, snishimura@nstac.go.jp

*** Office of Director-General for Policy Planning on Statistical Standards, Ministry of Internal Affairs and Communications, 19-1 Wakamatsu-cho, Shinjuku-Ku, Tokyo 162-8668 JAPAN, n.makita@soumu.go.jp

**** Statistics Bureau of Japan, Ministry of Internal Affairs and Communications, 19-1 Wakamatsu-cho, Shinjuku-Ku, Tokyo 162-8668 JAPAN, t.namiki@soumu.go.jp

***** National Statistics Center, 19-1 Wakamatsu-cho, Shinjuku-Ku, Tokyo 162-8668 JAPAN, htsubaki@nstac.go.jp

Abstract. Grid square statistics — a kind of spatial statistics — is useful as a tool for analyzing, comparing and synthesizing statistics regarding anonymization. We recently proposed a compatible extension of the Japanese national standard for grid square code (JIS X0410) for worldwide usage. The grid square approach can produce statistics of various types associated with geographic location anywhere in the world, identify a type of grid square statistics, and analyze relationships among various types of grid square statistics. This paper proposes how a data cube (linked open data) of World Grid Square Statistics can be constructed based on data collected from a range of sources, such as government statistics, satellite imagery, and point data from Internet providers. We present a definition of our World Grid Square coding system and a method of generating grid square statistics. We further propose a system for providing grid square statistics in a number of specific areas regarding their context and demonstrate their application to a data visualization and analytics platform.

1 Introduction

In 1960, Statistics Bureau of Japan began to study the definition of grid squares for computing spatial statistics from a horizontal perspective across other Japanese government ministries and departments. In the 1970s, they proposed grid square codes computed on the basis of geographical position (latitude and longitude). In 1973, "Standard Grid Square and Grid Square Code Used for the Statistics" was made as the Announcement No. 143 of the Administrative Management Agency (AMA) that hierarchically defines grid squares covering the entire land of Japan, and this definition was adopted by the Japanese Industrial Standards Committee in 1974. Adopting this definition, Japanese government ministries and departments and Japanese firms in industrial sectors have created grid square statistics for the purposes of planning, assessment, and evaluation in policy-making.

During more than 40 years of use in Japan, grid square codes and various types of grid square statistics have been provided by the government and by industrial organizations [1]. As one example, the Statistics Bureau of Japan, Ministry of Internal Affairs and Communications provides a GIS service supplying grid square statistics for population and economic census data [2]. Similarly, the National Land Information Division of the National Spatial Planning and Regional Policy Bureau (Ministry of Land, Infrastructure, Transport, and Tourism) provides public access to grid square statistics for numerical land information, transportation, and tourism [3].

Grid square statistics enable two different grid square statistics to be merged and integrated, and they can be retotaled following a given area. Because grid square statistics are created by geographical information (latitude and longitude), they have the advantage of not being influenced by temporal changes of administrative area.

In fact, Japanese grid squares defined in JIS X0410 are not based on any equal-area projection but on a conformal projection. Conformality of the Japanese grid squares is quite useful when we use the Japanese grid squares in an actual environment. The shape of the grid squares is tractable in a field by confirming the direction and position.

Since Japan adopted the World Geodetic System (ISO 6709) in 2002, all official geodetic measurements must be scaled in this way, with a 10-year moratorium under JIS X0410:2002. As a result, Japan's grid square coding system was also revised to conform with the World Geodetic System, and since 2012, all grid square statistics and data have been produced on that basis. In this sense, the Japanese grid square coding system is theoretically applicable to countries other than Japan.

According to the Global Statistical Geospatial Framework (GSGF) [4], there are five principles:

- **Principle 1** Accessible and Usable
- **Principle 2** Interoperable data and metadata standards
- **Principle 3** Common geographies for dissemination of statistics
- **Principle 4** Geocoded unit record data in a data management environment

- **Principle 5** Use of fundamental geospatial infrastructure and geocoding

Japan has three types of coding systems (JIS X0401, JIS X0402, and JIS X0410) in terms of common geographic areas for dissemination of statistics. JIS X0410 is one of them and provides a coding system to define grid squares based on geographical positions (latitude and longitude), identify a grid square by using numeric digits, and generate grid square statistics for dissemination of statistics. JIS X0401 is the To-Do-Fu-Ken (Prefecture) Identification Code to express 47 prefectures of Japan by using two numeric digits. This has a direct relationship with the Country subdivision code defined in ISO3166-2. JIS X0402 provides the identification code for cities, towns, and villages included in Japan by using five numeric digits. JIS X0401 and JIS X0402 form a hierarchical system similar to the NUTS classification (Nomenclature of territorial units for statistics) [5] for the economic territory of the European countries.

In order to guarantee interoperability among different organizations in public and private sectors, we have three Japanese industrial standards to identify geographic areas by using some numeric digits. This Japanese activity corresponds to Principle 3 and 4. In order to increase accessibility and usability, Japan has the portal site for official statistics named e-Stat [6] and the web-based GIS application platform named j STAT MAP [7]. This activity is associated with Principle 5.

In this paper, we explain how the Japanese grid square coding system can be extended to the World Grid Square coding system, and we introduce a multi-language web application to visualize and analyze statistics based on World Grid Square statistics. We show several examples of data analysis using these statistics and extract use cases of grid square statistics.

The rest of the paper is organized as follows. Section 2 provides a brief explanation of the World Grid Square coding system. Section 3 describes a visualization and analytics platform for World Grid Square statistics. Section 4 offers some examples of data sources and World Grid Square statistics. Section 5 explains how to use Linked Open Data (LOD) architecture to manage various types of grid square statistics. Section 6 addresses use cases of grid square statistics and some applications to data analytics. Finally, section 7 sets out conclusions and addresses future work.

2 World Grid Square Codes

In this section, we explain how to construct a World Grid Square coding system by adding upper two numeric digits to Japanese grid square codes (JIS X0410:2002). As shown in Figure 1, the World Grid Square system has a higher compatibility to the Japanese grid square coding system (JIS X0410). The World Grid Square coding system includes six levels of spatial resolution and enables identification of each grid square by its unique numeric digits (world grid square code) and JIS X0410. Table

1 shows the structure of World Grid Square codes. The first level is coded as six numeric digits, the second as eight numeric digits, the third as ten numeric digits, the fourth as eleven numeric digits, the fifth as twelve numeric digits, and the sixth level as thirteen numeric digits.

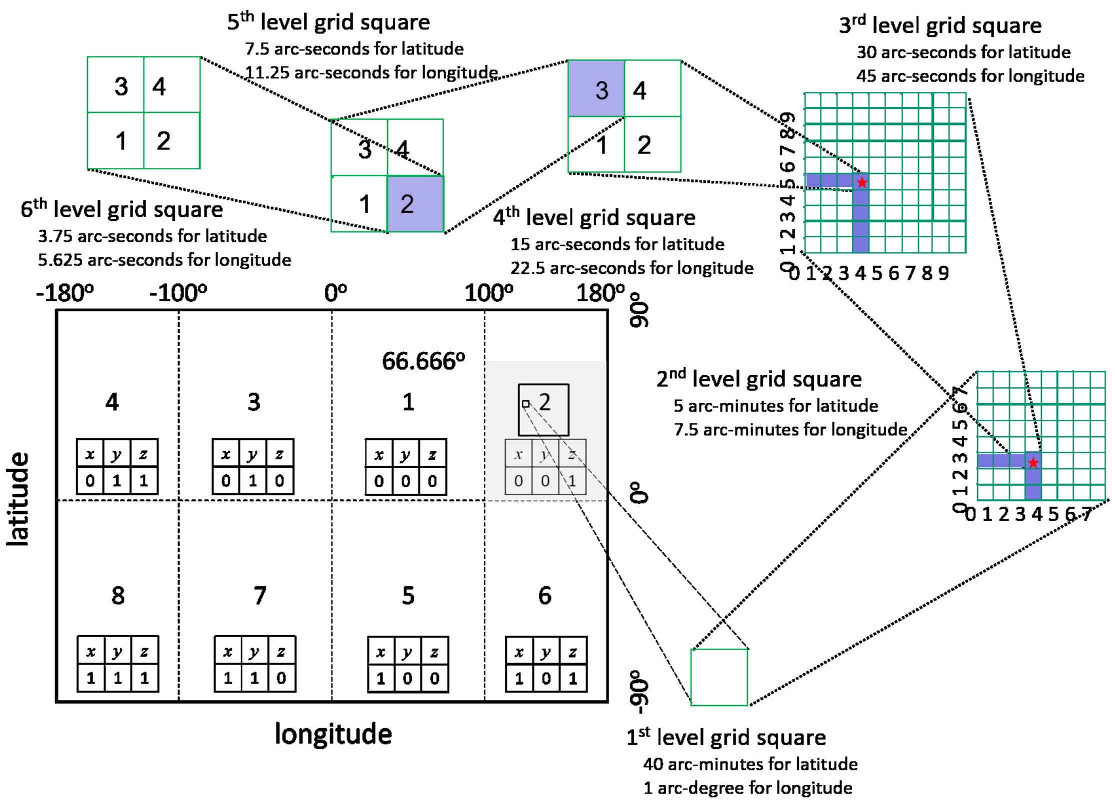

Figure 1: Conceptual illustration of World Grid Square codes. These codes are expressed by numeric digits in which length corresponds to spatial resolution. The first level of codes can be expressed as 6 numeric digits with 40 arc-minutes for latitude and 1 arc-degree for longitude. Second level codes can be expressed as 8 numeric digits with 5 arc-minutes for latitude and 7.5 arc-minutes for longitude. Third level codes can be expressed as 10 numeric digits with 30 arc-seconds for latitude and 45 arc-seconds for longitude. Fourth level codes can be expressed as 11 numeric digits with 15 arc-seconds for latitude and 22.5 arc-seconds for longitude. Fifth level codes can be expressed as 12 numeric digits with 7.5 arc-seconds for latitude and 11.25 arc-seconds for longitude. Sixth level codes can be expressed as 13 numeric digits with 3.75 arc-seconds for latitude and 5.625 arc-seconds for longitude.

No grid square overlaps with any other grid square, and each square's code can be calculated from geographic information (latitude and longitude). Consider three

Table 1: Description of the World Grid Square coding system, comprising 6 levels of spatial resolution.

Level	span for latitude	span for longitude	length of codes
1st level	40 arc-minutes	1 arc-degree	6 numeric digits
2nd level	5 arc-minutes	7.5 arc-minutes	8 numeric digits
3rd level	30 arc-seconds	45 arc-seconds	10 numeric digits
4th level	15 arc-seconds	22.5 arc-seconds	11 numeric digits
5th level	7.5 arc-seconds	11.25 arc-seconds	12 numeric digits
6th level	3.75 arc-seconds	5.625 arc-seconds	13 numeric digits

binary variables x, y, and z separating the earth into eight areas based on latitude and longitude, which construct the 0th-level grid square code. The three binary variables x, y, and z are given as follows:

1. Variable $x = 0$ if latitude is positive, otherwise $x = 1$
2. Variable $y = 0$ if longitude is positive, otherwise $y = 1$
3. Variable $z = 0$ for $|\text{longitude}| < 100°$, otherwise $z = 1$.

Finally, we define the 0th-level grid square code as

$$o = 2^2 x + 2y + z + 1. \tag{1}$$

From 0th-level grid square code o, we obtain x, y, and z, as

$$\begin{aligned} z &= (o - 1) \bmod 2, & (2)\\ y &= ((o - z - 1) \div 2) \bmod 2, & (3)\\ x &= (o - 2 \times y - z - 1) \div 4. & (4) \end{aligned}$$

Therefore, from a point data described as latitude and longitude, we can calculate grid square codes, including the position for six levels. Assume that p, q, r, u, v, w, s_2, s_4, and s_8 are integers. The 1st-level grid square code can be computed as

$$\text{1st-level grid square code} = \begin{cases} o00p0u & (p < 10, u < 10) \\ o0p0u & (10 \leq p < 100, u < 10) \\ op0u & (p \geq 100, u < 10) \\ o00pu & (p < 10, u \geq 10) \\ o0pu & (10 \leq p < 100, u \geq 10) \\ opu & (p \geq 100, u \geq 10) \end{cases}. \tag{5}$$

The 2nd-level grid square code can be described as

$$
\text{2nd-level grid square code} = \begin{cases} o00p0uqv & (p < 10, u < 10) \\ o0p0uqv & (10 \leq p < 100, u < 10) \\ op0uqv & (p \geq 100, u < 10) \\ o00puqv & (p < 10, u \geq 10) \\ o0puqv & (10 \leq p < 100, u \geq 10) \\ opuqv & (p \geq 100, u \geq 10) \end{cases}. \tag{6}
$$

The 3rd-level grid square code can be described as

$$
\text{3rd-level grid square code} = \begin{cases} o00p0uqvrw & (p < 10, u < 10) \\ o0p0uqvrw & (10 \leq p < 100, u < 10) \\ op0uqvrw & (p \geq 100, u < 10) \\ o00puqvrw & (p < 10, u \geq 10) \\ o0puqvrw & (10 \leq p < 100, u \geq 10) \\ opuqvrw & (p \geq 100, u \geq 10) \end{cases}. \tag{7}
$$

The 4th-level grid square code is computed from

$$
\text{4th-level grid square code} = \begin{cases} o00p0uqvrws_2 & (p < 10, u < 10) \\ o0p0uqvrws_2 & (10 \leq p < 100, u < 10) \\ op0uqvrws_2 & (p \geq 100, u < 10) \\ o00puqvrws_2 & (p < 10, u \geq 10) \\ o0puqvrws_2 & (10 \leq p < 100, u \geq 10) \\ opuqvrws_2 & (p \geq 100, u \geq 10) \end{cases}. \tag{8}
$$

The 5th-level grid square code is computed from

$$
\text{5th-level grid square code} = \begin{cases} o00p0uqvrws_2s_4 & (p < 10, u < 10) \\ o0p0uqvrws_2s_4 & (10 \leq p < 100, u < 10) \\ op0uqvrws_2s_4 & (p \geq 100, u < 10) \\ o00puqvrws_2s_4 & (p < 10, u \geq 10) \\ o0puqvrws_2s_4 & (10 \leq p < 100, u \geq 10) \\ opuqvrws_2s_4 & (p \geq 100, u \geq 10) \end{cases}, \tag{9}
$$

and the 6th-level grid square code is computed from

$$
\text{6th-level grid square code} =
$$

$$\begin{cases} o00p0uqvrws_2s_4s_8 & (p < 10, u < 10) \\ o0p0uqvrws_2s_4s_8 & (10 \leq p < 100, u < 10) \\ op0uqvrws_2s_4s_8 & (p \geq 100, u < 10) \\ o00puqvrws_2s_4s_8 & (p < 10, u \geq 10) \\ o0puqvrws_2s_4s_8 & (10 \leq p < 100, u \geq 10) \\ opuqvrws_2s_4s_8 & (p \geq 100, u \geq 10) \end{cases} . \tag{10}$$

In the above equations, all integers p (p is two or three digits), q (q is one digit), r (r is one digit), u (u is one or two digits), v (v is one digit), w (w is one digit), s_2 (s_2 is one digit), s_4 (s_4 is one digit), and s_8 (s_8 is one digit) were calculated from latitude and longitude as follows:

$$\begin{cases} p := \lfloor (1-2x)\text{latitude} \times 60 \div 40 \rfloor, \\ a := \{(1-2x)\text{latitude} \times 60 \div 40 - p\} \times 40, \\ q := \lfloor a \div 5 \rfloor, \\ b := (a \div 5 - q) \times 5, \\ r := \lfloor b \times 60 \div 30 \rfloor, \\ c := (b \times 60 \div 30 - r) \times 30, \\ s_{2u} := \lfloor c/15 \rfloor \quad (s_{2u} \text{ is one digit}), \\ d := (s_{2u}/15 - s_{2u}) \times 15, \\ s_{4u} := \lfloor d/7.5 \rfloor \quad (s_{4u} \text{ is one digit}), \\ e := (d/7.5 - s_{4u}) \times 7.5, \\ s_{8u} := \lfloor e/3.75 \rfloor \quad (s_{8u} \text{ is one digit}), \\ u := \lfloor (1-2y)\text{longitude} - 100z \rfloor, \\ f := (1-2y)\text{longitude} - 100z - u, \\ v := \lfloor f \times 60 \div 7.5 \rfloor, \\ g := (f \times 60 \div 7.5 - v) \times 7.5, \\ w := \lfloor g \times 60 \div 45 \rfloor, \\ h := (g \times 60 \div 45 - w) \times 45, \\ s_{2l} := \lfloor h/22.5 \rfloor \quad (s_{2l} \text{ is one digit}), \\ i := (s_{2l}/22.5 - s_{2l}) \times 22.5, \\ s_{4l} := \lfloor i/11.25 \rfloor \quad (s_{4l} \text{ is one digit}), \\ j := (s_{4l}/11.25 - s_{4l}) \times 11.25, \\ s_{8l} := \lfloor j/5.625 \rfloor \quad (s_{8l} \text{ is one digit}), \\ s_2 := s_{2u} \times 2 + s_{2l} + 1, \\ s_4 := s_{4u} \times 2 + s_{4l} + 1, \\ s_8 := s_{8u} \times 2 + s_{8l} + 1 \end{cases} . \tag{11}$$

However, we need to determine a position of a grid square from its grid square code. Such a transformation can be described as follows:

If we have the 1st-level grid square code opu (o(1 digit), p(3 digits), and u(2 digits)), then latitude and longitude at its northwestern corner can be computed as

$$\text{latitude} = (1-2x)\{(p - x + 1) \times 40 \div 60\}, \tag{12}$$
$$\text{longitude} = (1-2y)(100 \times z + u + y). \tag{13}$$

If we have the 2nd-level grid square code $opuqv$ (o(1 digit), p(3 digits), u(2 digits), q(1 digit), and v(1 digit)), then latitude and longitude at its northwestern corner can be computed as

$$\begin{aligned}
\text{latitude} &= (1-2x)\{p \times 40 \div 60 \\
&+ (q - x + 1) \times 5 \div 60\}, \qquad (14)\\
\text{longitude} &= (1-2y)(100 \times z + u \qquad (15)\\
&+ (v + y) \times 7.5 \div 60).
\end{aligned}$$

When the 3rd-level grid square code is given as $opuqvrw$ (o(1 digit), p(3 digits), u(2 digits), q(1 digit), v(1 digit), r(1 digit), and w(1 digit)), latitude and longitude at its northwestern corner can be computed as

$$\begin{aligned}
\text{latitude} &= (1-2x)(p \times 40 \div 60 + q \times 5 \div 60 \\
&+ (r - x + 1) \times 30 \div 3600), \qquad (16)\\
\text{longitude} &= (1-2y)(100 \times z + u + v \times 7.5 \div 60 \\
&+ (w + y) \times 45 \div 3600). \qquad (17)
\end{aligned}$$

When the 4th-level grid square code is given as $opuqvrws_2$ (o(1 digit), p(3 digits), u(2 digits), q(1 digit), v(1 digit), r(1 digit), w(1 digit), and s_2(1digit)), latitude and longitude at its northwestern corner can be computed as

$$\begin{aligned}
\text{latitude} &= (1-2x)(p \times 40 \div 60 + q \times 5 \div 60 \\
&+ r \times 30 \div 3600 \\
&+ ((s_2 - x) \bmod 2) \times 15 \div 3600), \qquad (18)\\
\text{longitude} &= (1-2y)(100 \times z + u + v \times 7.5 \div 60 \\
&+ w \times 45 \div 3600 \\
&+ \lfloor (s_2 + y - 1)/2 \rfloor \times 22.5 \div 3600). \qquad (19)
\end{aligned}$$

When the 5th-level grid square code is given as $opuqvrws_2s_4$ (o(1 digit), p(3 digits), u(2 digits), q(1 digit), v(1 digit), r(1 digit), w(1 digit), s_2(1 digit), and s_4(1 digit)), latitude and longitude at its northwestern corner can be computed as

$$\begin{aligned}
\text{latitude} &= (1-2x)(p \times 40 \div 60 + q \times 5 \div 60 \\
&+ r \times 30 \div 3600 \\
&+ ((s_2 - 1) \bmod 2) \times 15 \div 3600 \\
&+ ((s_4 - x) \bmod 2) \times 7.5 \div 3600), \qquad (20)\\
\text{longitude} &= (1-2y)(100 \times z + u + v \times 7.5 \div 60 \\
&+ w \times 45 \div 3600 \\
&+ \lfloor (s_2 - 1)/2 \rfloor \times 22.5 \div 3600 \\
&+ \lfloor (s_4 + y - 1)/2 \rfloor \times 11.25 \div 3600). \qquad (21)
\end{aligned}$$

When the 6th-level grid square code is given as $opuqvrws_2s_4s_8$ (o(1 digit), p(3 digits), u(2 digits), q(1 digit), v(1 digit), r(1 digit), w(1 digit), s_2(1 digit), s_4(1 digit), and s_8(1 digit)), latitude and longitude at its northwestern corner can be computed as

$$\begin{aligned}
\text{latitude} &= (1-2x)(p \times 40 \ \div 60 + q \times 5 \div 60 \\
&+ r \times 30 \div 3600 \\
&+ ((s_2 - 1) \bmod 2) \times 15 \div 3600 \\
&+ ((s_4 - 1) \bmod 2) \times 7.5 \div 3600 \\
&+ ((s_8 - x) \bmod 2) \times 3.75), \qquad (22) \\
\text{longitude} &= (1-2y)(100 \times z + u + v \times 7.5 \div 60 \\
&+ w \times 45 \div 3600 \\
&+ \lfloor (s_2 - 1)/2 \rfloor \times 22.5 \div 3600 \\
&+ \lfloor (s_4 - 1)/2 \rfloor \times 11.25 \div 3600 \\
&+ \lfloor (s_8 + y - 1)/2 \rfloor \times 5.625). \qquad (23)
\end{aligned}$$

Because the world grid square code is highly compatible to Japanese grid square code, we can create a sequence of numeric digits to express a World Grid Square by adding two numeric digits to the upper side of a Japanese grid square code. We provide open libraries to compute grid square codes with four computer languages (R, JavaScript, PHP, and python) from [8]. In our library, we define the following functions:

- calculate northwestern geographic position of the grid (latitude, longitude) from meshcode: `meshcode_to_latlong_grid(meshcode)`
- calculate northwestern geographic position of the grid (latitude, longitude) from meshcode: `meshcode_to_latlong_NW(meshcode)`
- calculate southwestern geographic position of the grid (latitude, longitude) from meshcode: `meshcode_to_latlong_SW(meshcode)`
- calculate northeastern geographic position of the grid (latitude, longitude) from meshcode: `meshcode_to_latlong_NE(meshcode)`
- calculate southeastern geographic position of the grid (latitude, longitude) from meshcode: `meshcode_to_latlong_SE(meshcode)`
- calculate a basic (3rd level) grid square code (10 digits) from a geographic position (latitude, longitude): `cal_meshcode(latitude,longitude)`
- calculate a 1st-level grid square code (6 digits) from a geographic position (latitude, longitude): `cal_meshcode1(latitude,longitude)`

- calculate a 2nd-level grid square code (8 digits) from a geographic position (latitude, longitude): `cal_meshcode2(latitude,longitude)`
- calculate a 3rd-level grid square code (10 digits) from a geographic position (latitude, longitude): `cal_meshcode3(latitude,longitude)`
- calculate a 4th-level grid square code (11 digits) from a geographic position (latitude, longitude): `cal_meshcode4(latitude,longitude)`
- calculate a 5th-level grid square code (12 digits) from a geographic position (latitude, longitude): `cal_meshcode5(latitude,longitude)`
- calculate a 6th-level grid square code (13 digits) from a geographic position (latitude, longitude): `cal_meshcode6(latitude,longitude)`.

The total number of 1st-level grid squares over the world is $360 \times 180 \times 3/2 = 97,200$. The total number of 2nd-level grid squares is 64 times larger than the total number of 1st-level grid squares. The total number of 3rd-level grid squares is 100 times larger than the total number of 2nd-level grid squares. The total number of 4th-level grid squares is 4 times larger than the total number of 3rd-level grid squares. The total number of 5th-level grid squares is 4 times larger than the total number of 4th-level grid squares. The total number of 6th-level grid squares is 4 times larger than 5th-level grid squares. We can also approximate the total number of grid squares in land since Earth's total land mass is 29.1998% of its total surface. Thus, the total number of each level grid squares over the world can be estimated in Table 2.

Table 2: The total number of each level grid squares over the world.

Layer type	# of grid squares	# of grid squares in land (approximation)
1st level	97,200	28,383
2nd level	622,080	2,838,220
3rd level	622,080,000	181,646,111
4th level	2,488,320,000	726,584,445
5th level	9,953,280,000	2,906,337,778
6th level	39,813,120,000	11,625,351,113

The shape of the grid square is not a complete square but a trapezoid with some slight curvature. The northern west-to-east span is normally different from the southern west-to-east span. Denoting the northern west-to-east span as W_1, the southern west-to-east span as W_2, and the north-to-south span as H, we can approximate the area of a given grid square as $A = (W_1 + W_2)H/2$. The area estimation of each level of grid squares is sensitive to geodetic datum such as OSGB36, GRS80, WGS 84, and EGM2008. We set the geodetic datum as WGS 84 throughout our investigation.

Since the area of a grid square is independent of longitude but dependent on latitude, we have an interest in the dependence of the area on the latitude. For example, the area of the 3rd-level grid square depends on its latitude. The area is less than 1.28 km^2 around the north and south poles and is equal to about 1.28 km^2 at the equator. The maximum area of grid squares at each level is given on the equator line (latitude = 0). The maximum area at the 1st-level grid square is 8191.83 km^2, at the 2nd-level is 128.00 km^2, at the 3rd-level is 1.28 km^2, at the 4th-level is 0.32 km^2, at the 5th-level is 0.08 km^2, and at the 6th-level is 0.02 km^2. These area estimations can cover shortcomings of the non-equal area property of the World Grid Squares.

Adopting the method provided by the Statistics Bureau, Ministry of Internal Affairs and Communications for creating grid square data for administrative areas in Japan [9], we produced World Grid Square data for administrative areas in more than 252 countries, using a project web page from the Research Institute for World Grid Squares [10]. Figure 2 shows examples of World Grid Square data for administrative areas for three countries. Grid squares included completely or partially in the same administrative area are indicated by the same color.

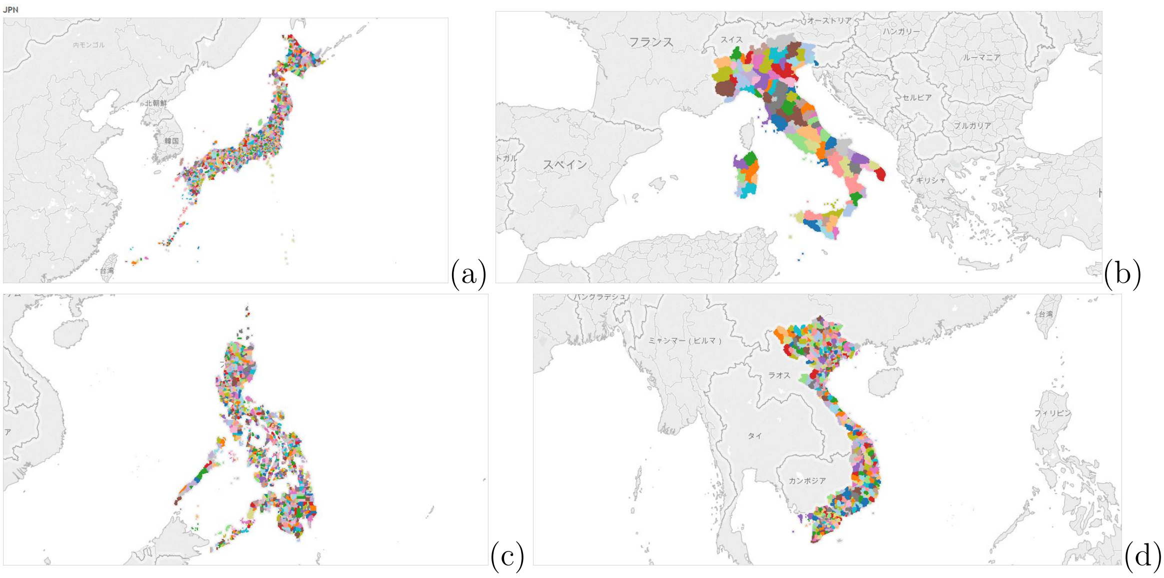

Figure 2: Examples of World Grid Square data for administrative areas in (a) Japan, (b) Italy, (c) the Philippines, and (d) Vietnam. Grid squares included in the same administrative area are shown in the same color.

Additionally, World Grid Square statistics for elevations (based on satellite data) provided by the Japan Aerospace Exploration Agency (JAXA) and World Grid

Square statistics for night-time light intensity provided by NASA have been released through our project page [10]. The location, extent and brightness of night-time lights can be used as primary data for measuring human artificial activities [11, 12, 13]. Figure 3 shows examples of World Grid Square statistics for elevation in Japan. One record includes minimum, median, mean, and maximum elevations in a grid square. Furthermore, Global Soil Information can be used for creating grid square statistics about selected soil properties [14].

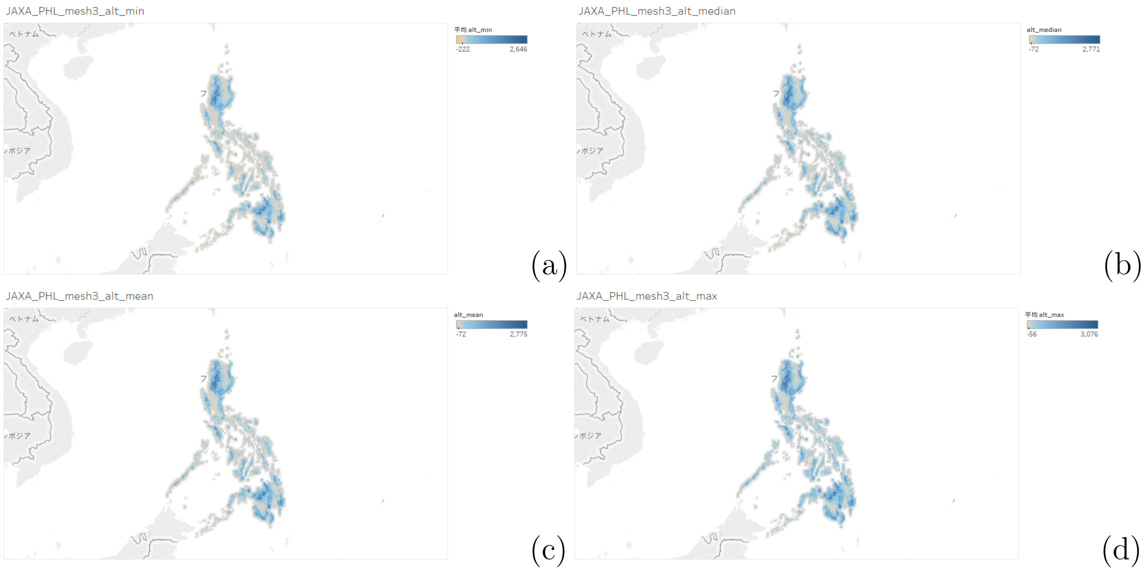

Figure 3: Examples of World Grid Square statistics for elevations in the Philippines: (a) minimum elevations, (b) median elevations, (c) mean elevations, and (d) maximum elevations. Grid square statistics for elevations were generated from Daichi (ALOS) 30-m elevation DEM data produced by JAXA.

3 Visualization and analytics platform for World Grid square statistics

This section describes the visualization and analytics platform that we are developing. This platform is implemented on a cloud service using three types of virtual servers (database server, Web server, and data collection server). The platform provides data and analytics functions implemented in Javascript, and users can access the service from their own Web browser. This enables use of any operating system, which is a reasonable assumption for a multi-operational platform. As the platform can also provide data through its WebAPI, users can access our service from their own computer software. Automatic data extraction and computation can be realized using the WebAPI.

Figure 4 shows the proposed visualization and analytics platform for World Grid Square statistics, which are stored separately in a number of databases and segmented into different tables by country. A management unit selects adequate tables of World Grid Squares and extracts the statistics and data as grid squares indicated by latitude, longitude, and a user-assigned size. In the case of HTML output, the grid square statistics extracted by the management unit for World Grid Square statistics are converted and displayed as a table. In the case of API output, the grid square statistics extracted by the management unit are displayed in a computer-readable format.

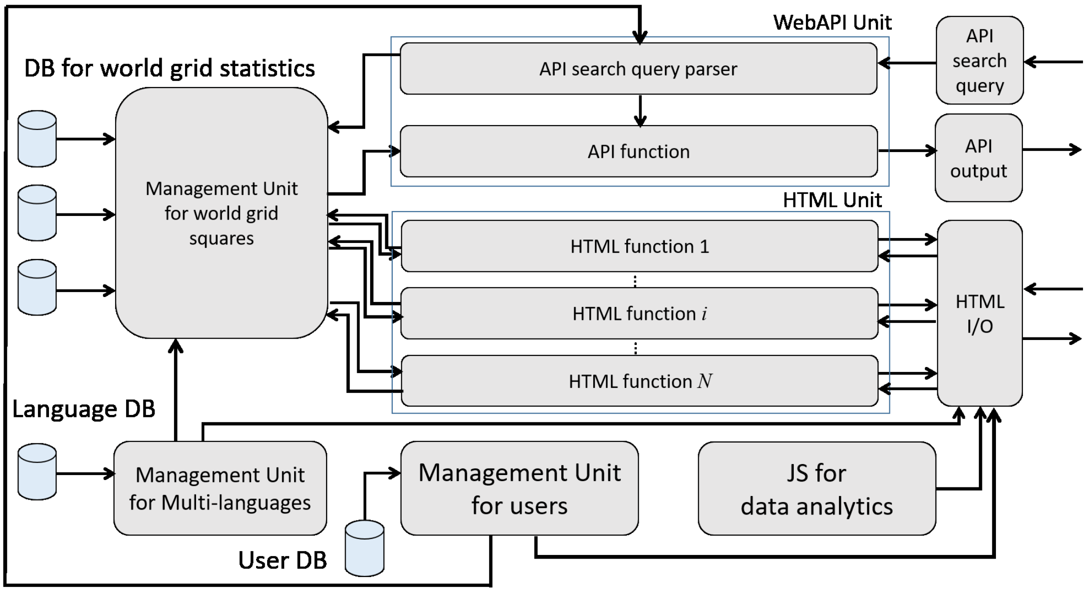

Figure 4: A conceptual illustration of the visualization and analytics platform for World Grid Square statistics.

A management unit includes a database for users and selects a language based on user settings. A multi-language management unit selects terms according to the user's language setting; field names and expressions on the platform are translated by a dictionary into the user-selected language. The language DB (dictionary) is constructed by an API of automatic language translation functions provided by Microsoft Azure. We create words manually in English and Japanese and translate them automatically into other languages using the translation API. A human translator who knows English or Japanese but whose mother tongue is other than English or Japanese translates words into their own language, validating and translating them manually.

At present, our platform supports nine languages (English, Japanese, German, Italian, Spanish, Korean, Vietnamese, Chinese, and Polish). The visualization and analytics platform is available from [15]. Figure 5 shows screen-shots of the data visualization and analytics platform. The grid square data on population census (left) and forests (right) in Kyoto, Japan can be seen in the bottom screen-shot. We can select various types of grid square statistics on the data visualization and analytics platform.

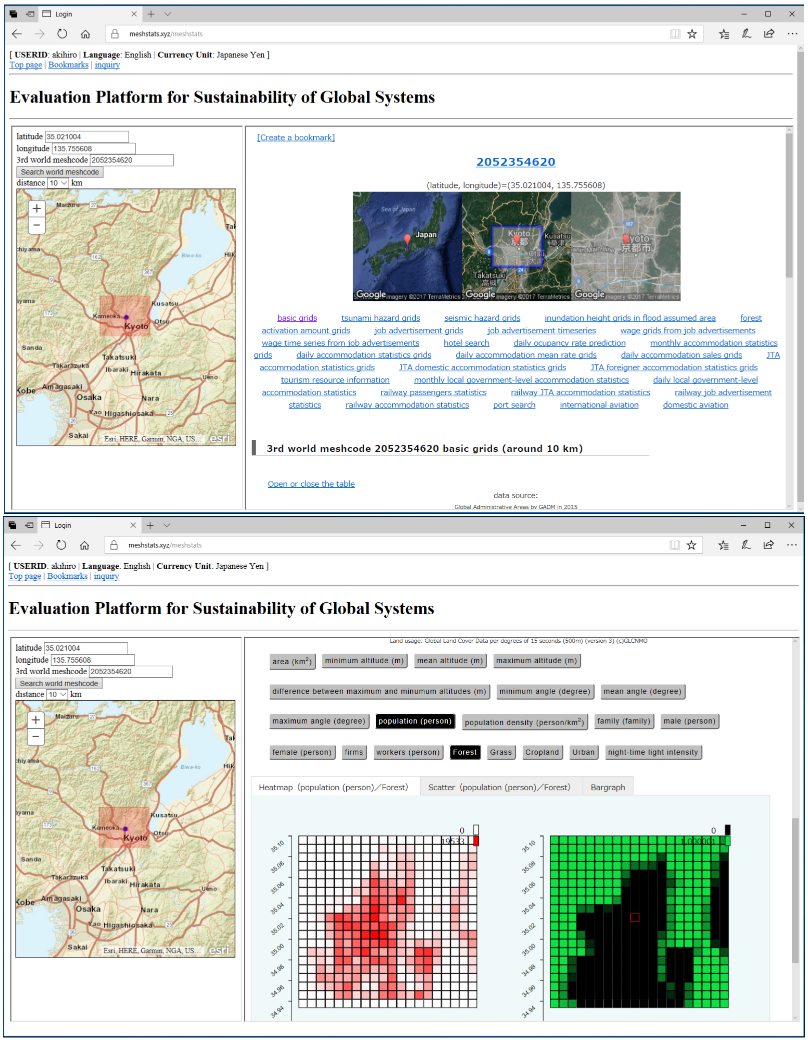

Figure 5: Screen-shots of the data visualization and analytics platform.

4 Probable data sources

For present purposes, we identified three types of data source: government statistics, satellite imagery, and point data collected from the Internet. For example, Global Soil Information Facilities provide girded data originated from satellite imagery [16], and OpenStreetMap provides point data about objects [17]. These data can be used as primary data to produce World Grid Square statistics.

We further found four types of procedure for creating World Grid Square statistics.

1. Convert grid square statistics provided as part of government statistics into World Grid Square statistics.

2. Aggregate girded data and compute World Grid Square statistics.

3. Compute grid square statistics from data that include geographical information (latitude and longitude).

4. Generate World Grid Square statistics from polygon data by checking their inclusiveness.

Some World Grid Square statistics have been already provided from the visualization and analytics platform: Population census by Statistics Bureau of Japan, Ministry of Internal Affairs and Communications; Economic census by Statistics Bureau of Japan, Ministry of Internal Affairs and Communications; Accommodation Survey by Japan Tourism Agency, Ministry of Land, Infrastructure, Transport, and Tourism; tsunami hazard by NOAA tsunami run-up catalogue data and digital elevation model by JAXA; seismic hazard provided by Japan Seismic Hazard Information Station (J-SHIS); Anticipated Flood depth class by Ministry of Land, Infrastructure, Transport and Tourism; hotel opportunities and job advertisement by Recruit Web Service. We currently collect World Grid Square statistics mainly in Japan; statistics and data for other countries were produced from GADM and satellite data sources (digital elevation model, land cover classification (20 classes) by Geospatial Information Authority of Japan (GLCNMO), and night-time light intensity). We can increase the number of types of World Grid Square statistics in countries other than Japan by adding further grid square statistics and data to the platform.

5 How to produce world grid square statistics based on LOD

Since 2016, National Statistics Center of Japan has provided government statistics of Japan based on Linked Open Data (LOD) at an official statistics portal called e-Stat [18].

LOD are presented using standard technologies based on World Wide Web Consortium (W3C) recommendations and are created using the Resource Description Framework (RDF). Normally, LOD are searched using SPARQL Language (SPARQL). The e-Stat portal has released various government statistics expressed as multi-dimensional data, using RDF formats employing the RDF Data Cube Vocabulary recommended by W3C.

The RDF Data Cube Vocabulary recommends that statistics should comprise four elements: dimension, measure, observation, and attribute. *Dimension* expresses the categories in which statistics are created, such as region, time, gender, and age. *Measure* shows the quantities of statistics created, such as persons, price, and indices. *Observation* is numerical, expressing the statistics themselves. *Attribute* expresses additional information such as unit and conditions of observation. Figure 6 shows instances of the four elements (dimension, measure, observation, and attribute).

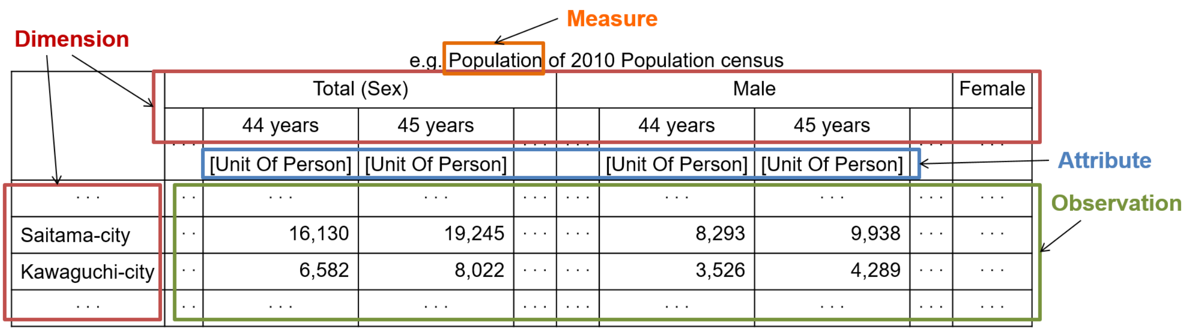

Figure 6: Definition of terms in LOD for statistics.

The observation in each cell can be expressed by its own URI, and attributes such as units and conditions are linked to the vocabulary databases. The National Statistics Center of Japan has released part of the government's statistics (specifically, population census) in RDF format through e-Stat.

Grid square statistics have not yet been released as LOD. However, the same procedure as for existing trials of government statistics released in RDF formats is applicable in releasing government statistics in grid square form. The National Statistics Center plans to release grid square statistics that include the population census as LOD. As a grid square code is equivalent to a regional unit, the grid square code is identified as an entity belonging to dimension. We can describe a World Grid Square code as a unique URI, and the World Grid Square code can be linked to basic (geographic) information. Figure 7 shows a schematic of a relation diagram expressing World Grid Square statistics as LOD. Data definition can be standardized by providing RDF formatted LOD as URI. Moreover, as a World Grid Square code can be expressed as a unique URI, World Grid Square statistics can readily be linked to data other than government statistics. This enables World Grid Square codes to be used as a primary key to extract data from several databases.

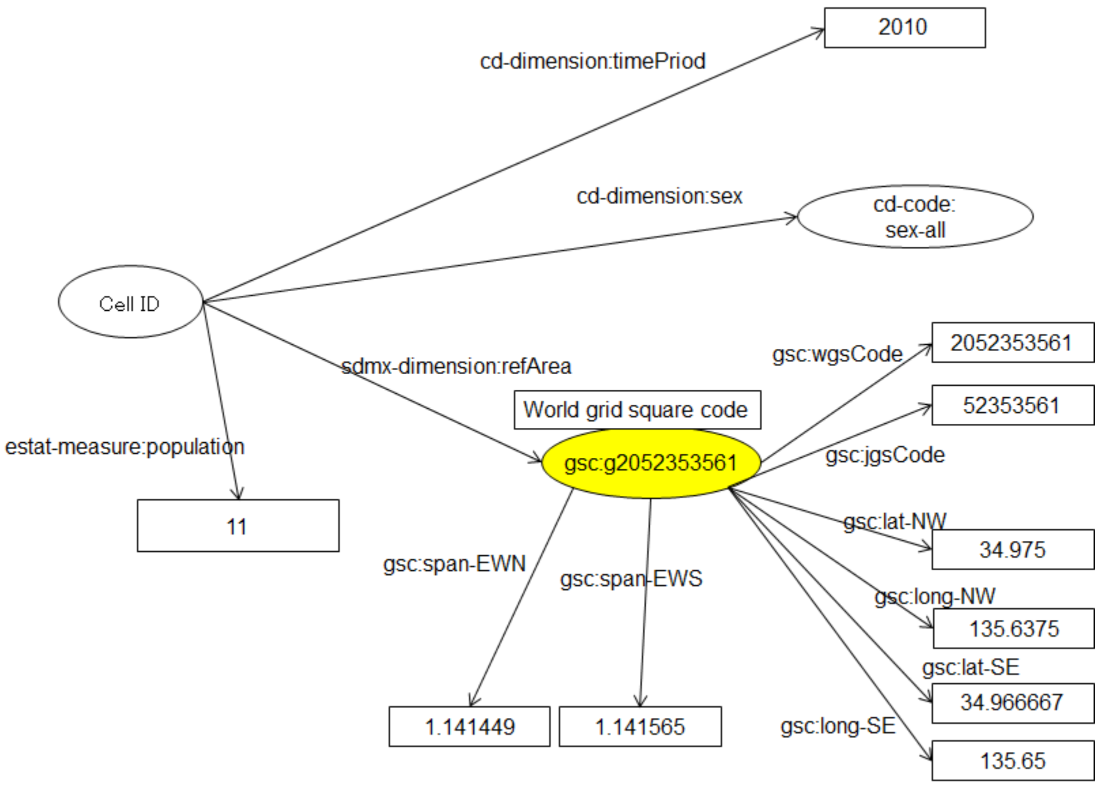

Figure 7: RDF description of the world gird square.

6 Case studies

It is assumed that World Grid Square statistics can be used in the following six cases.

Data linkage and data processing: We can link different grid square statistics (linkage) and use them in operations among different grid square statistics, synthesizing new grid square statistics from several original data types.

Mapping: We can visualize grid square statistics on a map for use in analyzing our focal area.

Data creation on given areas: We can generate statistics on a given area by recalculating grid square statistics for that area.

Identifying effective areas: As grid squares make it easy to measure distances among grid squares, World Grid Square statistics can be used to calculate demand within a given distance.

Defining observation areas: Grid square statistics can be used to define an area for collection of data or samples.

Unit for numerical simulation: Grid square statistics can be used to conduct numerical simulations for a unit such as diffusion processes, percolation models, and migration processes.

Using Japanese grid square statistics, successive studies have measured economic activities [19, 20, 21] and assessed risks of natural disasters [22, 23, 24]. Grid square data on seismic hazards are provided by the Japan Seismic Hazard Information Station (J-SHIS) [23]. We include several examples of data analysis using Japanese grid square statistics. For instance, Figure 8 shows grid square statistics for the number of job advertisements in each third level grid square. By performing a cross-sectional analysis of the number of workers and job advertisements in a given area, we identified a power-law relationship between the two. Figure 9 shows grid square data for anticipated water heights of inundation zone areas, based on polygon data provided by the Ministry of Land, Infrastructure, Transport, and Tourism [24].

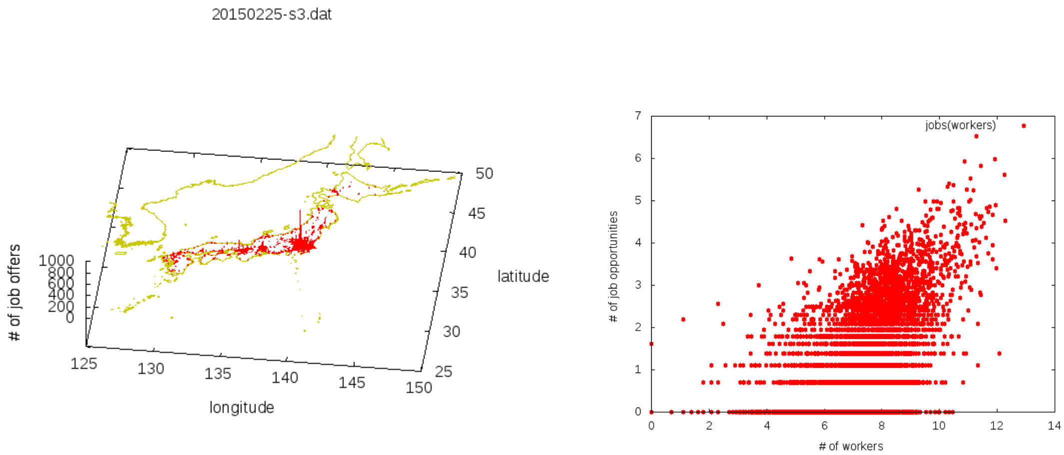

Figure 8: (a) The relationship between the number of job advertisement on each 3rd level grid square statistics on 25 Feb. 2015 in Japan. (b) The double logarithmic scatter plots between the number of workers and the number of job advertisements in Kyoto, Japan.

Figure 9 shows grid square data on anticipated water heights of inundation zone areas. This grid square data are produced by polygon data of 2010 anticipated inundation zone data provided by Ministry of Land, Infrastructure, Transport, and Tourism [24].

7 Conclusion

We have discussed how data analytics can employ worldwide spatial statistics based on our recently proposed World Grid Square coding system. We have briefly elabo-

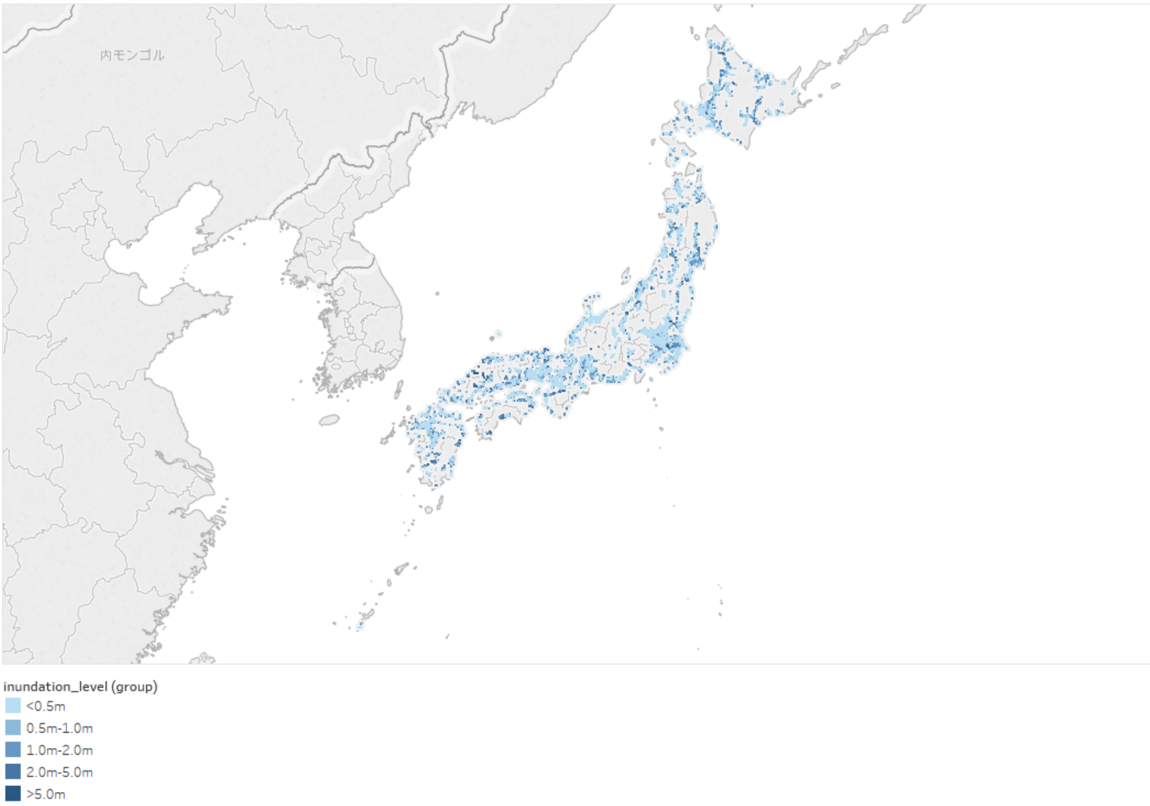

Figure 9: 3rd level grid square statistics for anticipated water heights of inundation zone areas in 2010.

rated a definition of World Grid Square statistics and some examples of their application to administrative areas and elevations. We also addressed use cases for data production based on World Grid Square statistics, referring to some case studies of data applications.

The paper introduced our data visualization and analytics platform. Named MESHSTATS, it provides multi-language visualization, data analytics and quantification functions for several domains. In developing functions for the proposed platform, we used a CAPD cycle comprising data acquisition, data collection, data analysis, interpretation, and deployment. We need to develop a function based on domain knowledge and use cases for applications, and to increase the number of data providers (data producers) and data consumers (data users).

In future work, we will focus on increasing the variety of data providers by organizing these. To this end, we need to develop software to compute grid square statistics from primary data and to share knowledge to produce World Grid Square statistics. We would emphasize the importance of organizing stakeholders who can provide primary data and/or produce World Grid Square statistics from their own data, create use cases for data applications of the world grid squares, implement functions on the platform, and evaluate work flows, including data analysis of World Grid Square statistics in actual environments.

Acknowledgement

This work is financially supported by JST PRESTO Grant Number JPMJPR1504, Japan. The grid square data and statistics were also computed by using computational resources provided by Institute for Mathematical Statistics based on HPCI project (hp160060).

References

[1] Grid Square Statistics, Statistics Bureau, Ministry of Internal Affairs and Communications, `http://www.stat.go.jp/english/data/mesh/`

[2] Geospatial statistics, Statistics Bureau, Ministry of Internal Affairs and Communications, `http://www.e-stat.go.jp/SG1/estat/toukeiChiri.do?method=init`

[3] Land Numerical Information download service, Ministry of Land, Infrastructure, Transport, and Tourism, `http://nlftp.mlit.go.jp/ksj/index.html`

[4] European Forum for Geography and Statistics [Online] Available: `http://www.efgs.info/information-base/production-model/global/`

[5] NUTS in the EU [Online] Available: `http://ec.europa.eu/eurostat/web/nuts`

[6] Portal Site of Official Statistics of Japan [Online] Available: `http://www.e-stat.go.jp/SG1/estat/eStatTopPortalE.do`

[7] j STAT MAP [Online] Available: `https://jstatmap.e-stat.go.jp/gis/nstac/`

[8] Open libraries to compute World Grid Square codes and their reference positions from World Grid Square codes have been released from `http://www.fttsus.jp/worldgrids/en/library`

[9] A list of grid square codes for 2nd level administrative areas in Japan, Statistics Bureau, Ministry of Internal Affairs and Communications, `http://www.stat.go.jp/data/mesh/m_itiran.htm`

[10] Research Institute for World Grid Squares, `http://www.fttsus.jp/worldgrids/`

[11] C.D. Elvidge, P. Cinzano, D.R. Pettit, J. Arvesen, P. Sutton, C. Small, R. Namani, T. Longcore, C. Rich, J. Safran, J. Weeks, and S. Ebener, “The Nightsat mission concept”, International Journal of Remote Sensing, Vol. 28, No. 12 (2007) pp. 2645–2670.

[12] B. Klemens, A. Coppola, M. Sbron, “Estimating Local Poverty Measures Using Satellite Images”, World Bank Group, Policy Research Working Paper, 7329, June 2015.

[13] C. Mellander, J. Lobo, K. Stolarick, Z. Matheson, “Night-Time Light Data: A Good Proxy Measure for Economic Activity?”, PLoS ONE, Vol. 10 No. 10 (2015) e0139779.

[14] T. Hengl , J.M. de Jesus, R.A. MacMillan, N.H. Batjes, G.B.M. Heuvelink, et al., “SoilGrids1km — Global Soil Information Based on Automated Mapping”, PLoS ONE Vol. 9, No. 8 (2014) e105992.

[15] MESHSTATS, `https://www.meshstats.xyz/`

[16] Global environmental layers, `http://worldgrids.org/`

[17] Open Street Map, `http://www.openstreetmap.org/`

[18] LOD for government statistics, e-Stat, Statistics Bureau in Japan, Ministry of Internal Affairs and Communications, `http://data.e-stat.go.jp/lodw/`

[19] A.-H. Sato, T. Watanabe, “Measuring Activities and Values of Industrial Clusters based on Job Opportunity Data Collected from an Internet Japanese Job Matching Site”, 2016 IEEE International Conference on Big Data (Big Data), 5-8 Dec. 2016, 2016, pp. 2199–2208

[20] A.-H. Sato, C. Shimizu, T. Mizuno, T. Ohnishi, T. Watanabe, “Relationship between job opportunities and economic environments measured from data in internet job searching sites”, Procedia Computer Science, Volume 60 (2015) pp. 1255–1262.

[21] A.-H. Sato, “Microdata analysis of the accommodation survey in Japanese tourism statistics”, 2015 IEEE International Conference on Big Data (Big Data), Oct. 29 2015-Nov. 1, 2015, pp. 2700–2708.

[22] A.-H. Sato, H. Sawai, “Geographical risk assessment from tsunami run-up events based on socioeconomic-environmental data and its application to Japanese air transportation”, Procedia CIRP, Vol. 19 (2014) pp. 27–32.

[23] Japan Seismic Hazard Information Station, National Research Institute for Earth Science and Disaster Resilience, http://www.j-shis.bosai.go.jp/

[24] Anticipated inundation zone data in 2012, Ministry of Land, Infrastructure, Transport, and Tourism in Japan, http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-A31.html