

Data Infrastructure for Data Analytic Platforms



京都大学
KYOTO UNIVERSITY



Dr. Aki-Hiro Sato
(Kyoto University & JST PRESTO)
aki@i.kyoto-u.ac.jp

*The world's mount of data
is increasing year by year
(about 4.4 ZB/year in 2013 and about 44
ZB/year in 2020)*

*Human cognitive capacity
is finite
(300MB ~ 500 MB/life)*

*There is an enormous gap
between big data and
human processing capacity*

Data Infrastructure About 44ZB/year by 2020 (ZB = 10^{21} Bytes)

Data sources:

1. Internet e-commerce sites
2. Satellite imageries
3. Government statistics

Computational Resources:

1. HPC (Supercomputers)
2. Cloud computing

Standards:

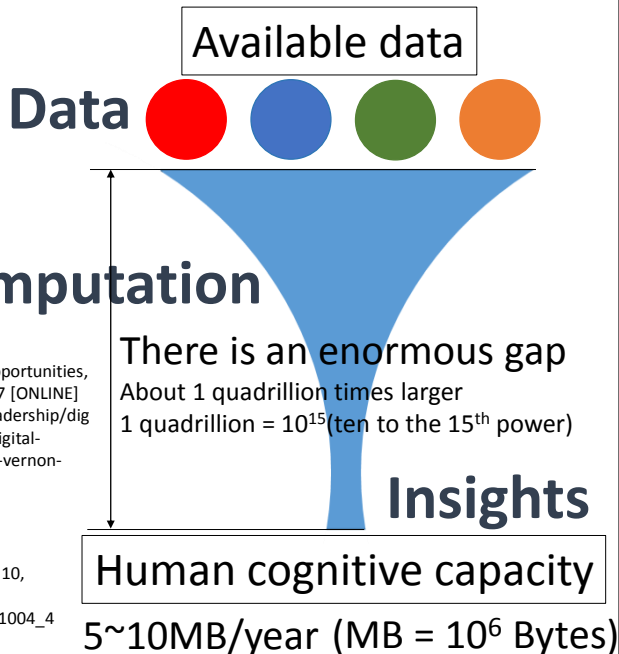
1. Statistical Standards
2. IT Standards

Domain knowledge:

Purposes:

The Digital Universe of Opportunities,
Accessed on 18 June, 2017 [ONLINE]
<https://www.emc.com/leadership/digital-universe/2014iview/digital-universe-of-opportunities-vernon-turner.htm>

T.K. Landauer, Cognitive Sci. 10,
477-493 (1986),
DOI:10.1207/s15516709cog1004_4

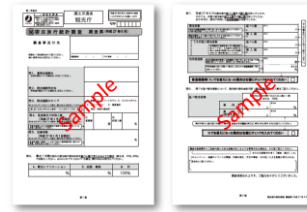
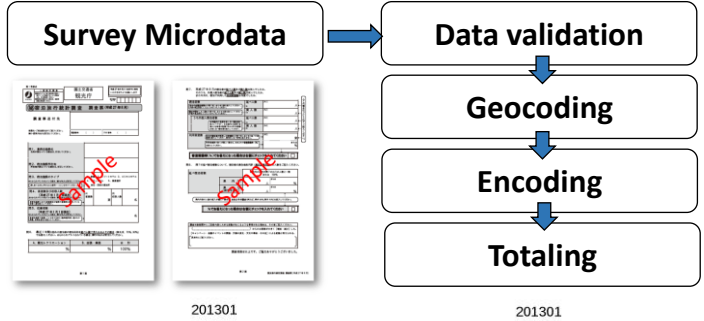


Accommodation Survey in Japanese Tourism Statistics



- The Accommodation Survey in Japanese Tourism Statistics is a monthly survey conducted by the Japan Tourism Agency of the Ministry of Land, Infrastructure, Transport and Tourism (MLIT).
- These data are collected from hotels, inns, and other private and public accommodations listed in the establishment frame database defined in Article 27 of the Statistics Act (Ministry of Internal Affairs and Communications).

How to create Grid Square Statistics (JIS X0410)

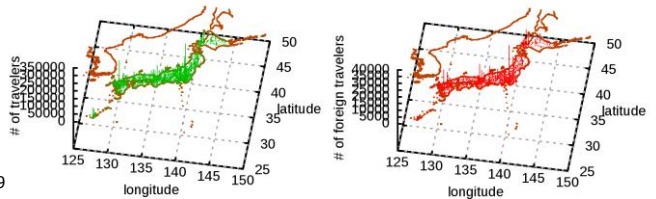


201301

201301

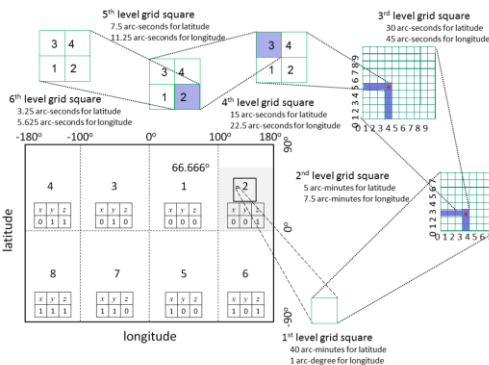
of travelers

of foreign travelers



A.-H. Sato, 2015 IEEE International Conference on Big Data (Big Data), Oct. 29 2015-Nov. 1, 2015, pp. 2700-2708, DOI: 10.1109/BigData.2015.7364070

World Grid Square Codes



Definition of the 3rd level world grid square code

$$[(b-2x)latitude \times 60 + 40] = p$$

$$a = [(1-2x)latitude \times 60 + 40 - p] \times 40$$

$$[a + 5] = q$$

$$b = (a + 5 - q) \times 5$$

$$[b \times 60 + 30] = r$$

$$c = (b \times 60 + 30 - r) \times 30$$

$$[(1-2y)longitude - 100z] = u$$

$$f = (1-2y)longitude - 100z - u$$

$$[f \times 60 + 7.5] = v$$

$$g = (f \times 60 + 7.5 - v) \times 7.5$$

$$[g \times 60 + 45] = w$$

$$h = (g \times 60 + 45 - w) \times 45$$

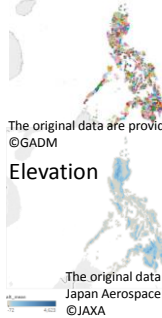
$$o = 2^2 \cdot x + 2 \cdot y + z + 1$$

$$\begin{matrix} o00p0qrvr & (p < 10, u < 10) \\ o0p0qrvr & (10 \leq p < 100, u < 100) \\ o0p0qrvr & (p \geq 100, u < 10) \\ o00p0qrvr & (p < 10, u \geq 10) \\ o0p0qrvr & (10 \leq p < 100, u \geq 10) \\ o0p0qrvr & (p \geq 100, u \geq 10) \end{matrix}$$

Research Institute for World Grid Squares

- Documents
- Data
- Libraries
- Events

Administrative Area



The original data are provide from GADM. ©GADM

The original data are provide from Japan Aerospace Exploration Agency. ©JAXA

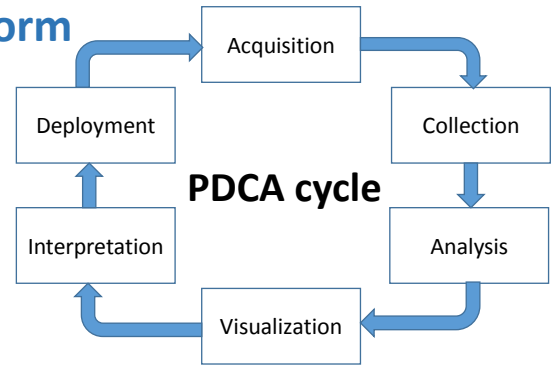
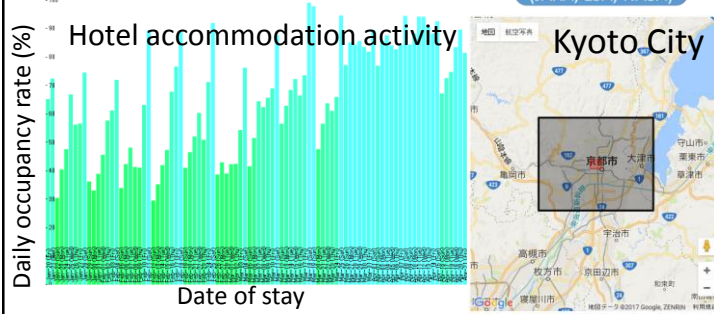
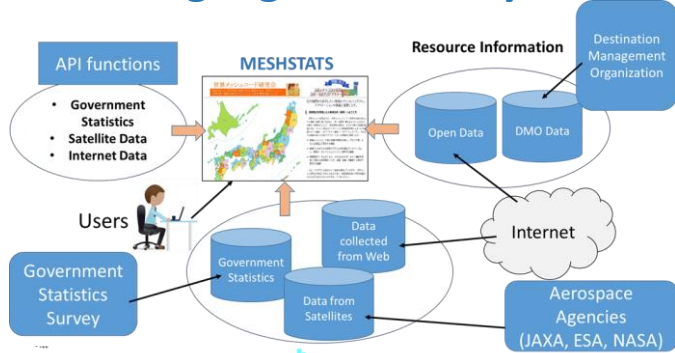
Level	Level1	Level2	Level3	Level4
Methcode	206225	20622546321	206225463212	2062254632122
NW	(35.3333333, 135.775)	(35.0833333, 135.75)	(35.0333333, 135.725)	(35.0033333, 135.7)
SE	(34.9666666, 136.025)	(34.9666666, 135.875)	(35.005, 135.7875)	(35.005, 135.7875)
East-to-West Span [km]	20.9180	11.3864	1.1408	1.1408
East-to-West Span [km]	51.6264	11.4110	1.1408	1.1408
North-to-South Span [km]	73.8775	9.2348	0.9235	0.9235
Area [km2]	8743.8949	105.2445	1.0524	1.0524

Level	Level1	Level2	Level3	Level4
Methcode	20622546321	206225463212	2062254632122	20622546321222
NW	(35.0291666, 135.775)	(35.0270603, 135.778125)	(35.02600416, 135.7796875)	(35.0250000, 135.78125)
SE	(35.0246666, 135.78125)	(35.025, 135.78125)	(35.025, 135.78125)	(35.025, 135.78125)
East-to-West Span [km]	0.5700	0.2052	0.1428	0.1428
East-to-West Span [km]	0.5704	0.2052	0.1428	0.1428
North-to-South Span [km]	0.4617	0.2309	0.1154	0.1154
Area [km2]	0.2634	0.0698	0.0165	0.0165

This work is supported by JST PRESTO Grant Number JPMJPR1504, Japan.

<http://www.fttsus.jp/worldgrids/>

Multi-language data analytics platform



<https://www.meshstats.xyz/meshstats/>

