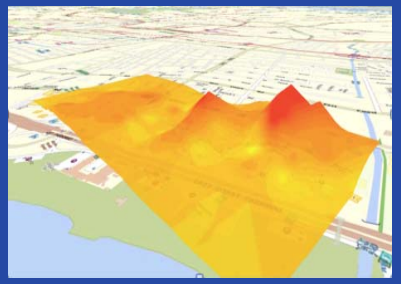
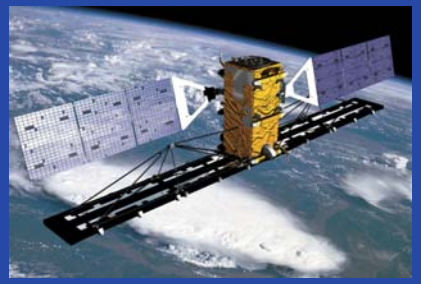
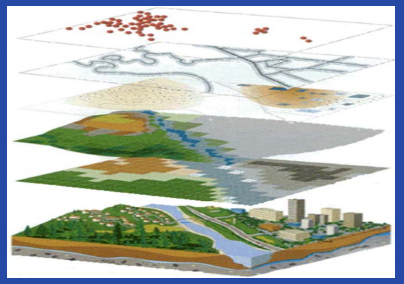
Using Location Information For Better Planning and Decision Support: Integrating Big Data, Official Statistics, Geo-information

Professor Paul Cheung
Professor of Social Policy and Analytics, NUS,
Chair, International Steering Committee for Global Mapping,
Former Director, United Nations Statistics Division

Rapid Information for Public Policy

- Increasing demand for information in public policy;
- Users want quick data and rapid assessment. Traditional sources (surveys) less emphasized. Too time consuming. Too slow;
- Statistical community adjusting – using multi-mode approach: internet, call center, fax, sms, sensors;
- Increasing use of administrative data;
- Realizing back-end system integration important;

Data Fusion as a New Challenge

- Data Fusion as a new Data Paradigm

- "Data fusion occurs when data from different sources are brought into contact and new facts emerge. Individually, each data source may have a specific limited purpose. Their combination, however, may uncover new meanings"

- "Policy attention should focus on Actual Uses of Big Data, and less on its collection and analysis"

1 May, 2014. White House

Promoting Global Data Fusion

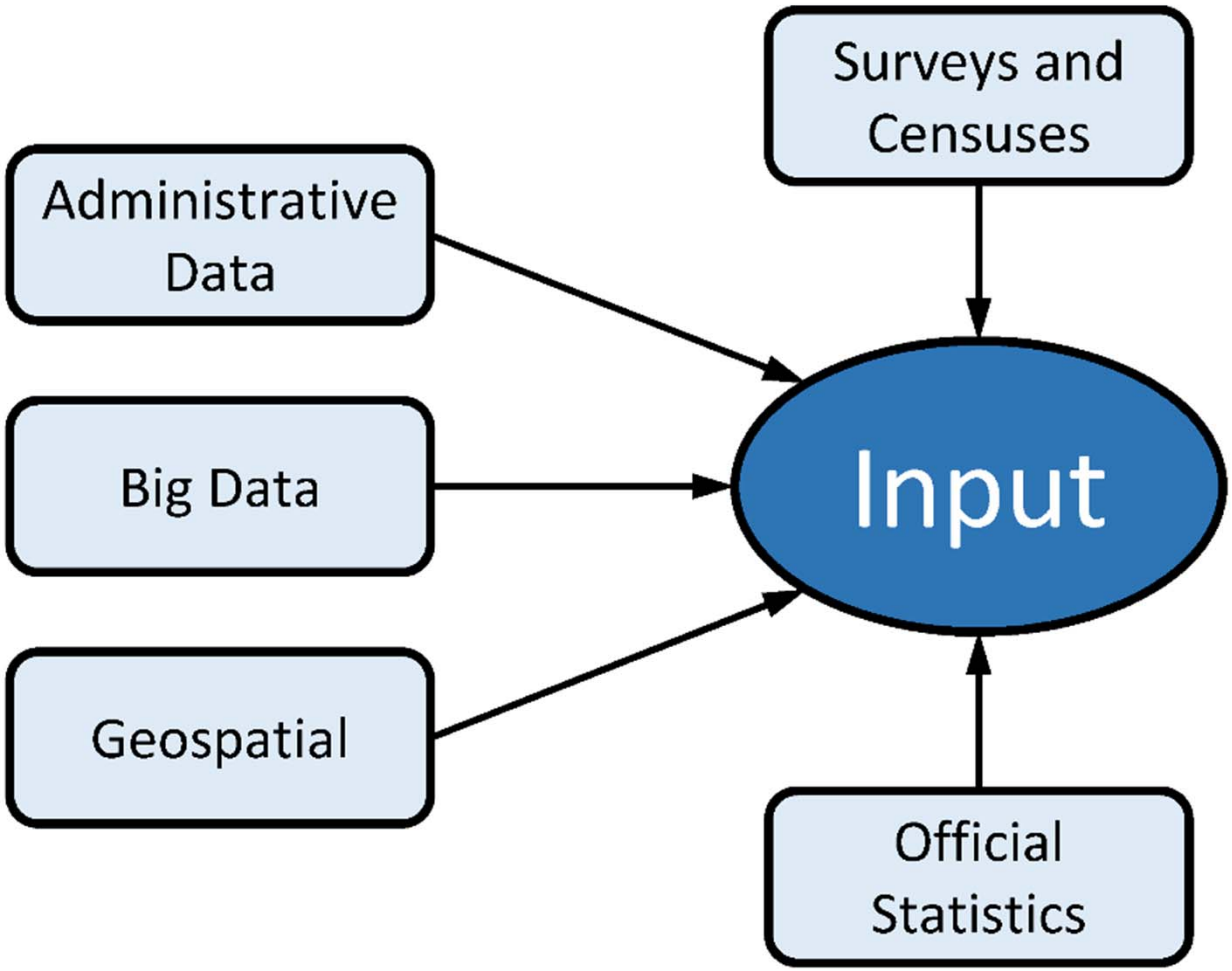
- UNSD has been taking the lead in this initiative;
- 2009 – Initiation of the UN-GGIM process;
- 2011 seminar in Korea explored integration of official statistics and geospatial information;
- 2013 UN Statistical Commission endorsed 'Statistical Spatial Framework' – development continuing;
- 2014 UN Statistical Commission discussed integration of Big Data and Official Statistics;
- Data Fusion process continuing, and will bring in new sources in the future.

Integrating Geo-Information, Official Statistics, and Big Data

- Three communities operating with different analytical schemes and data structures, with minimal overlap;
- Distinct culture, languages and practices;
- Comfortable as distinct professional communities –
  - Geospatial Community – mapping, imageries;
  - Data Scientists Community – big data analytics;
  - Official Statistics Community – structured indicators;
- But now compelled by emerging trends to look for the common ground.

What is the Common Ground? How to get there??

Integrating Multiple Data Sources for Decision Support



A New Era: Developing Location Information

- Rapid technological advancement overcoming security concerns, with innovative breakthroughs;
- Changing business model for the use, access and provisions of Geospatial (and related) Information;
- New business model gives rise to competition: eg: Google competes with government mapping agency;
- Data integration based on spatial framework gaining rapid pace (Location Information)

Place Policy needs Location Information

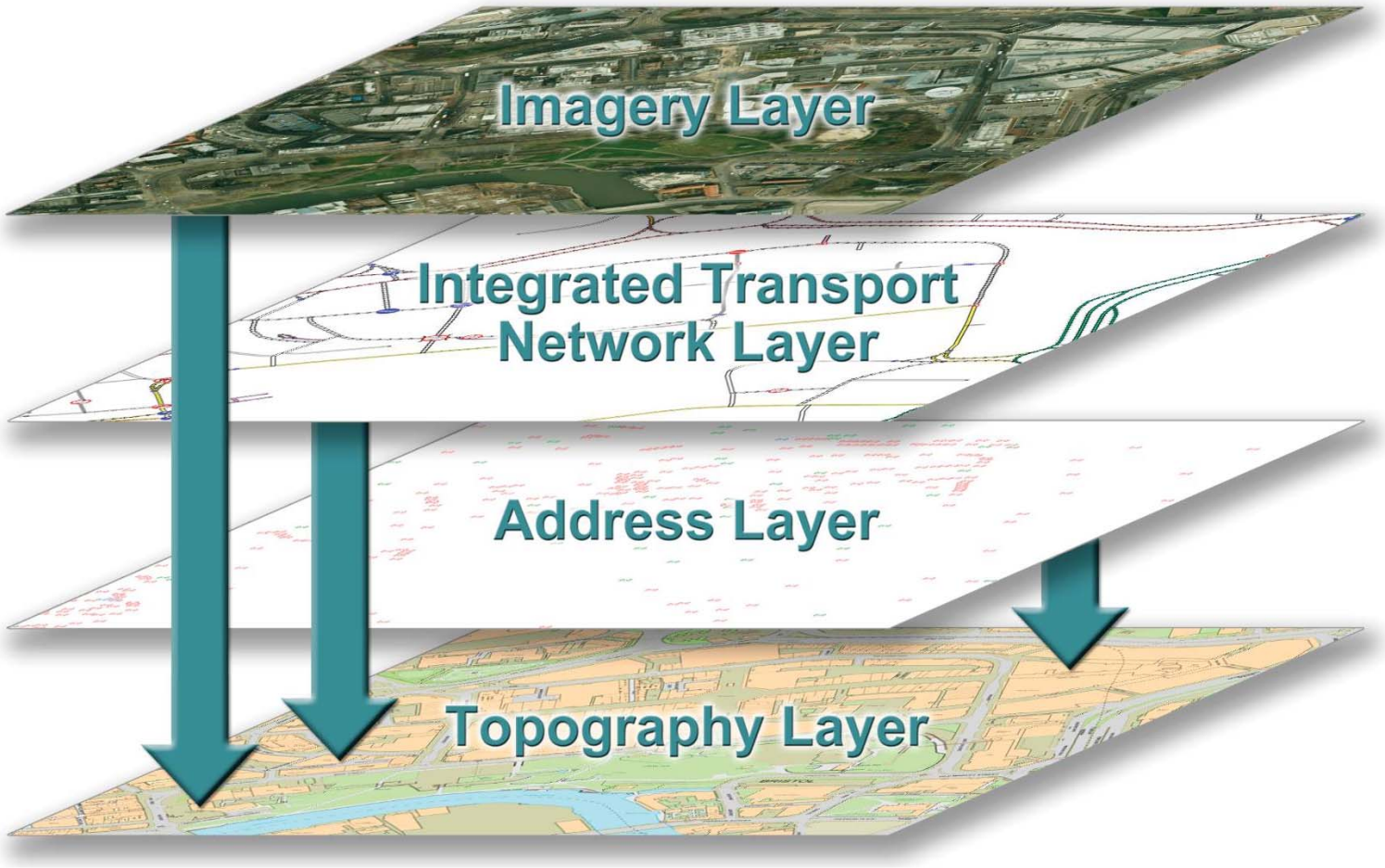
"Place policies target the prosperity, equity, sustainability, and livability of places – how well or how poorly they function and how they change over time"

Office of Management and Budget 2009

Place policy needs location information to measure successful outcomes.

OMB proposes standardized performance measures and data collection

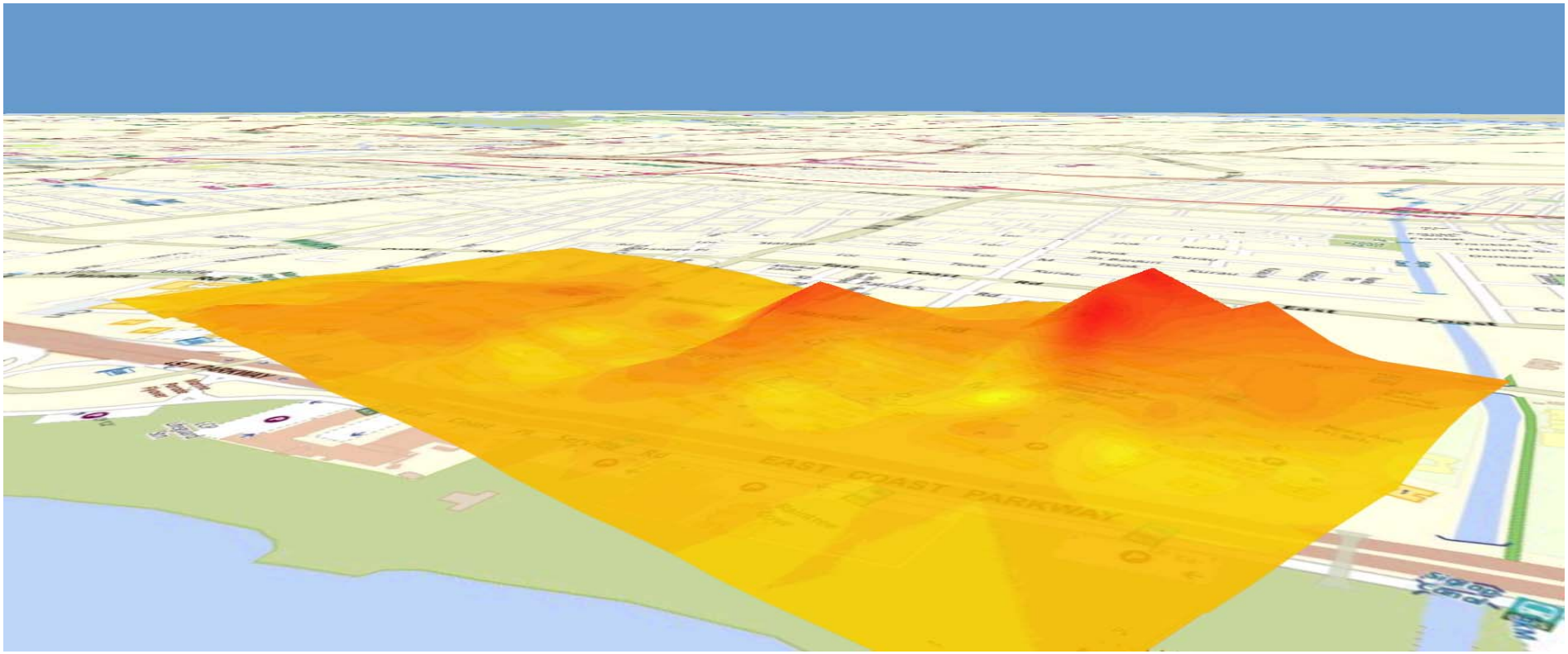
Mapping layers; Connecting Information



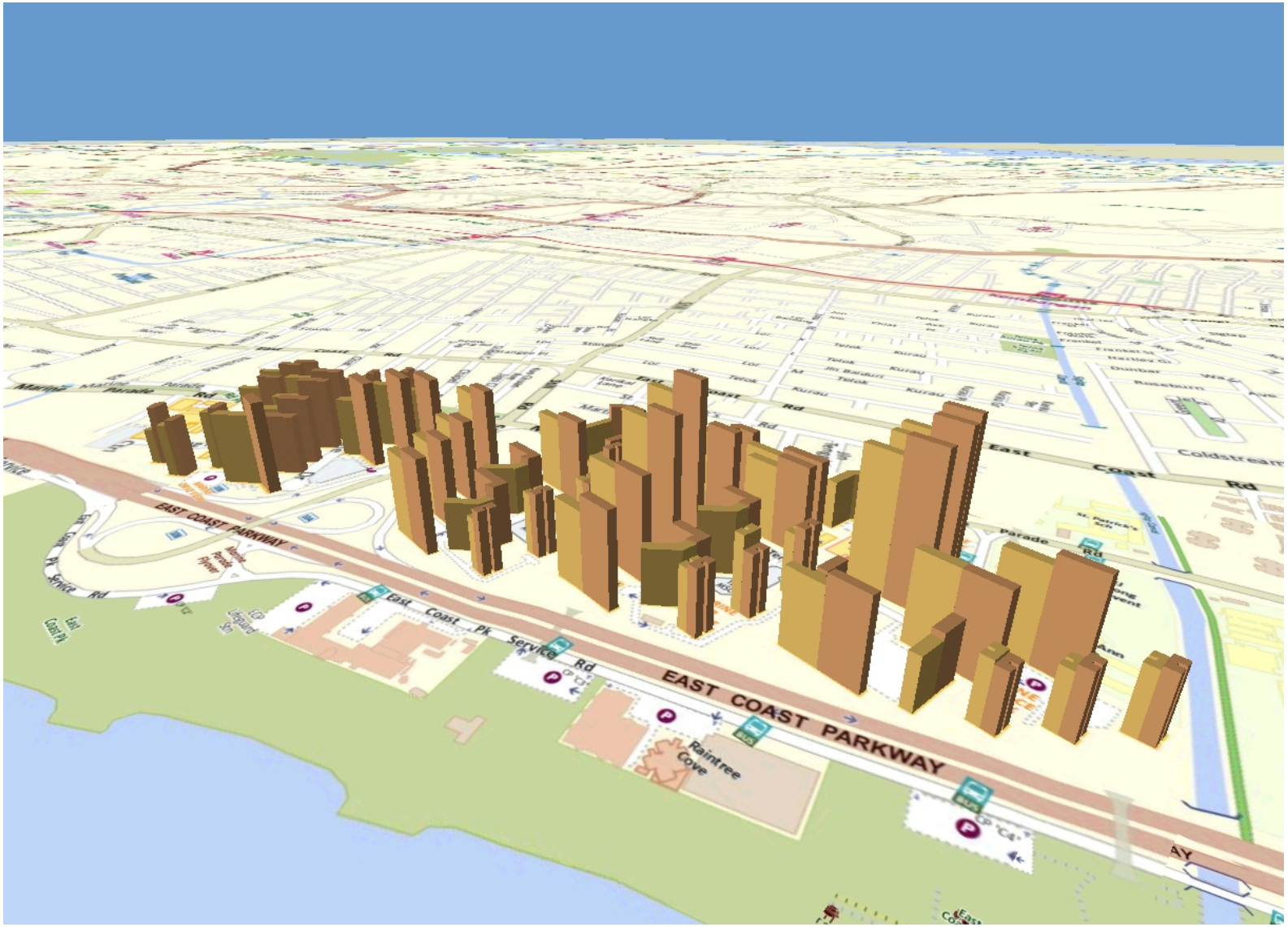
Source: Ordnance Survey International

Using Location Information

Many examples: Municipal and Urban Planning and Services



Current and Projected 65+, 2012-2022



Projected Elderly, 2022

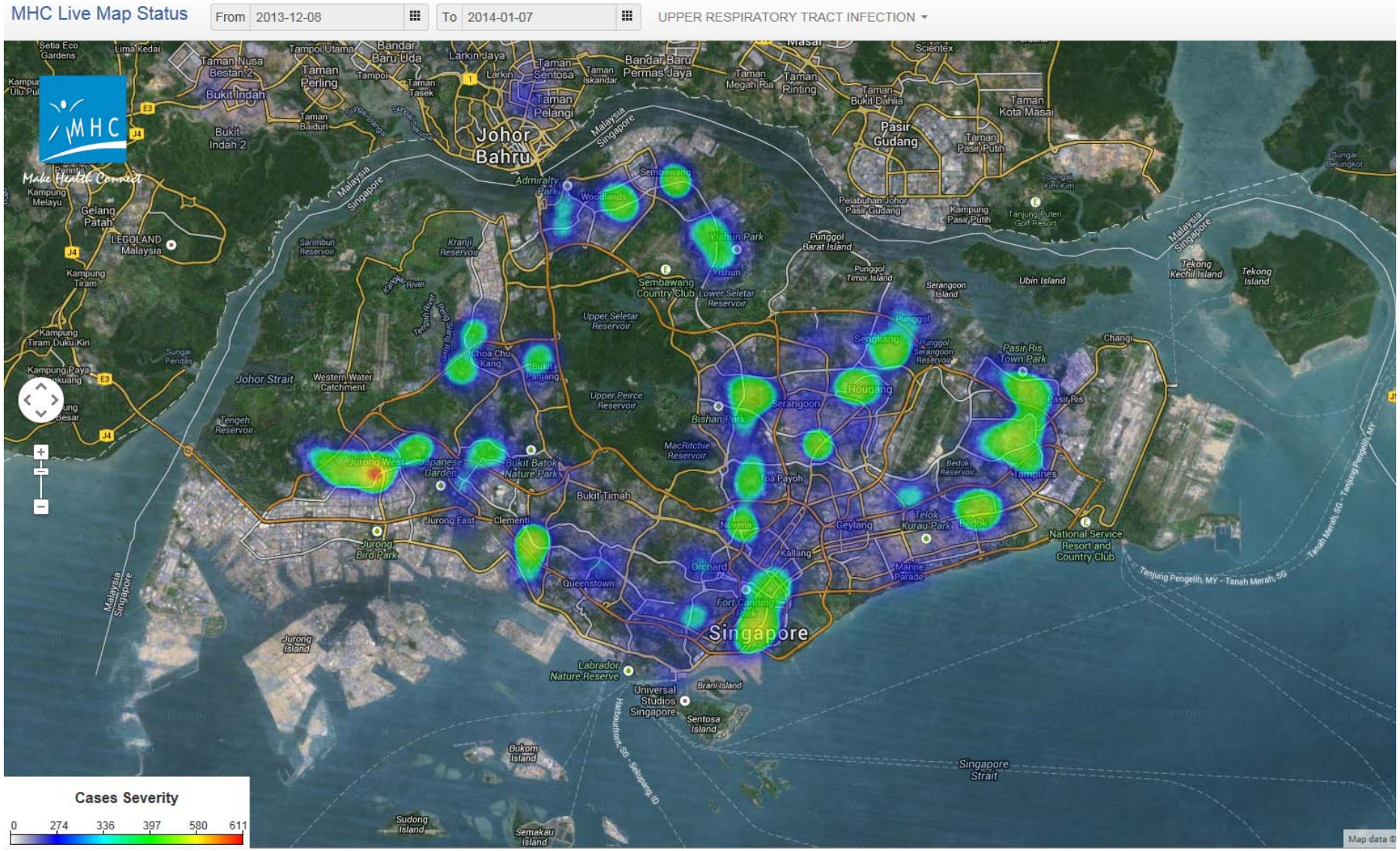
Elderly Residents (65+) in the Marine Parade Constituency (2012-2022)

Total Residents aged 65 and above*: 4851

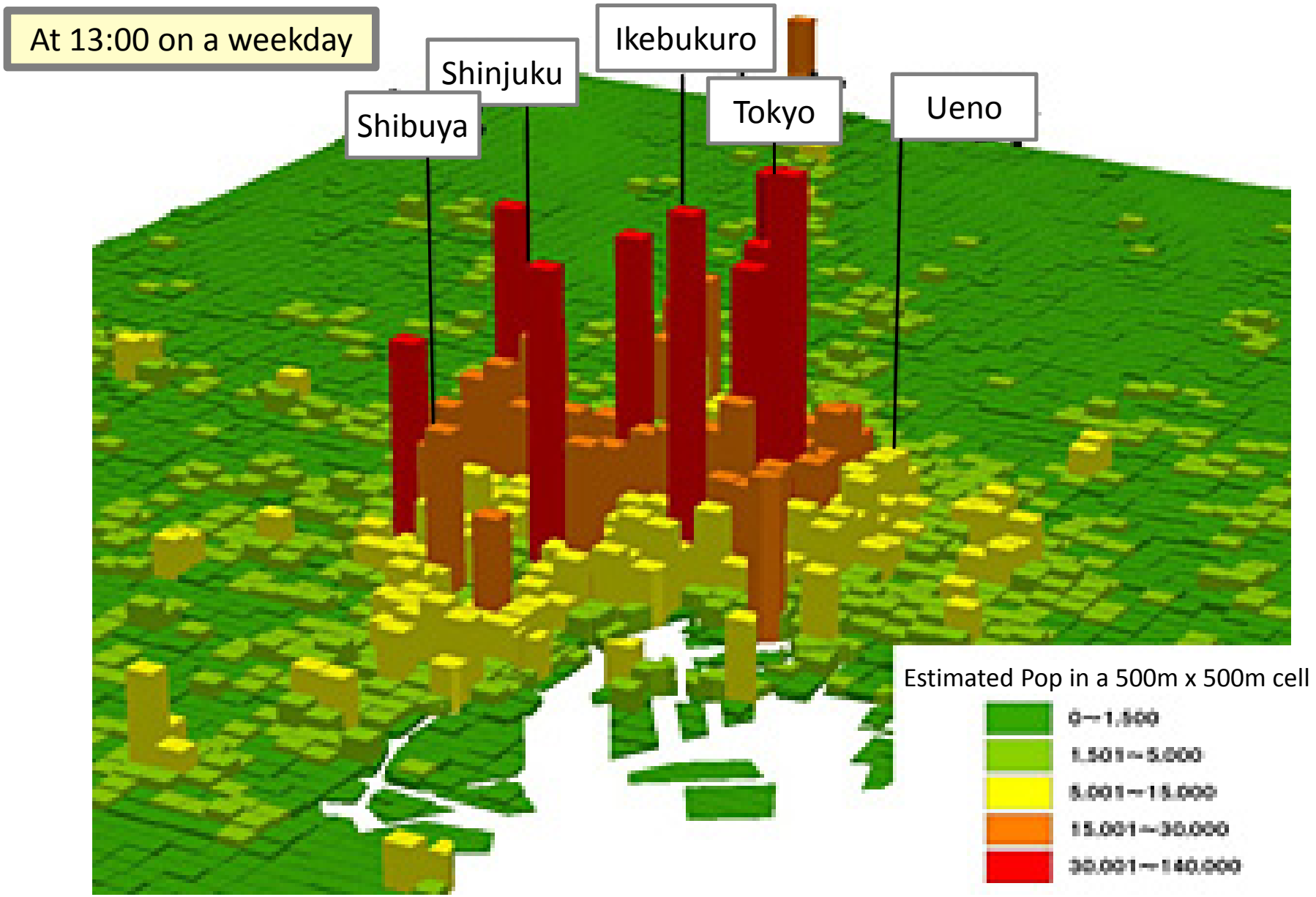
*Residents refer to Singapore Citizens & Permanent Residents "alive" as at end 2012

Projected Total Residents aged 65 and above (by 2022): 5635

Real Time: Upper Respiratory Tract Infection Map

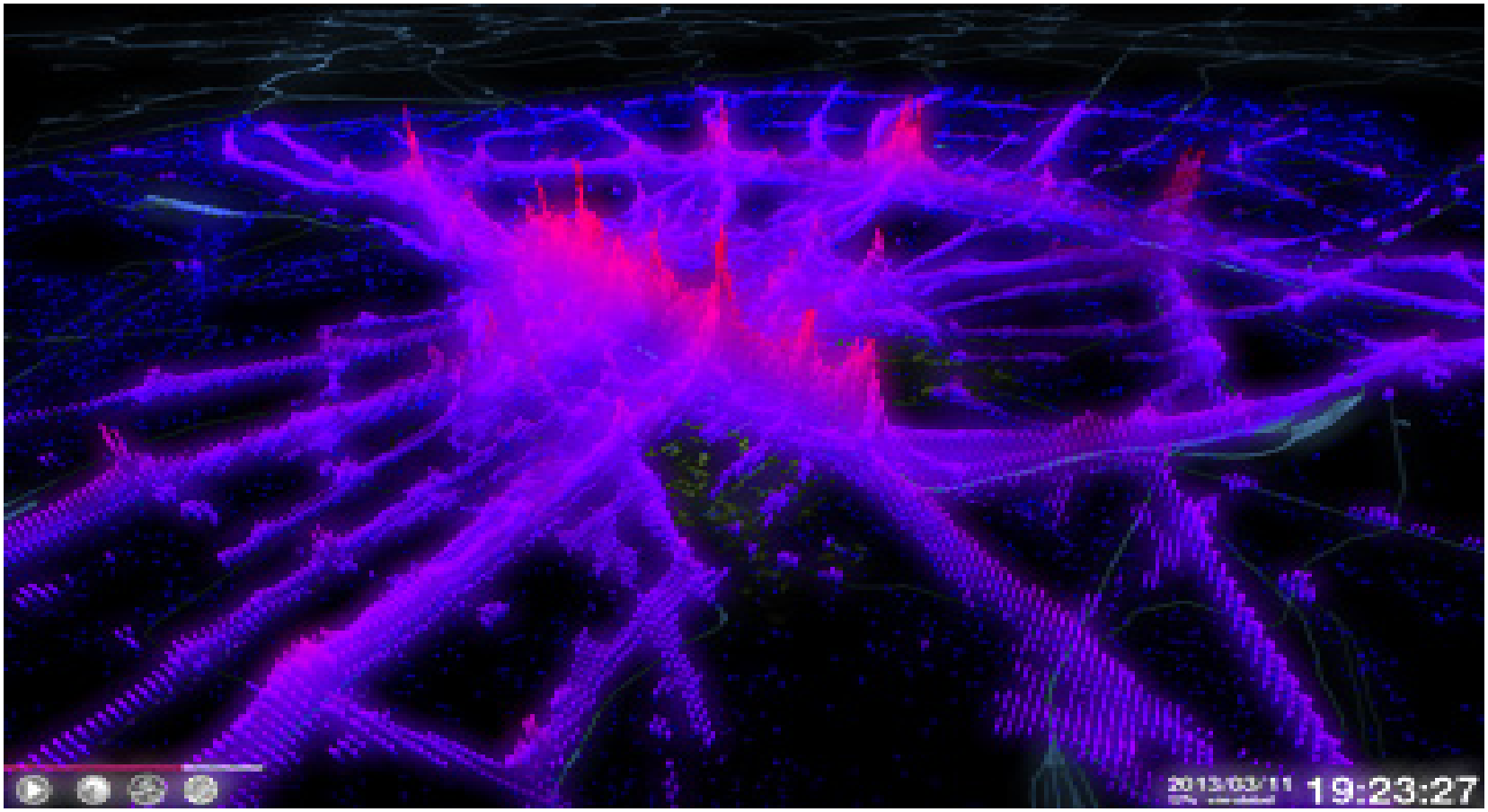


Sample image of Mobile Geospatial Statistics:
Population Distribution of Central Tokyo (23 Wards)



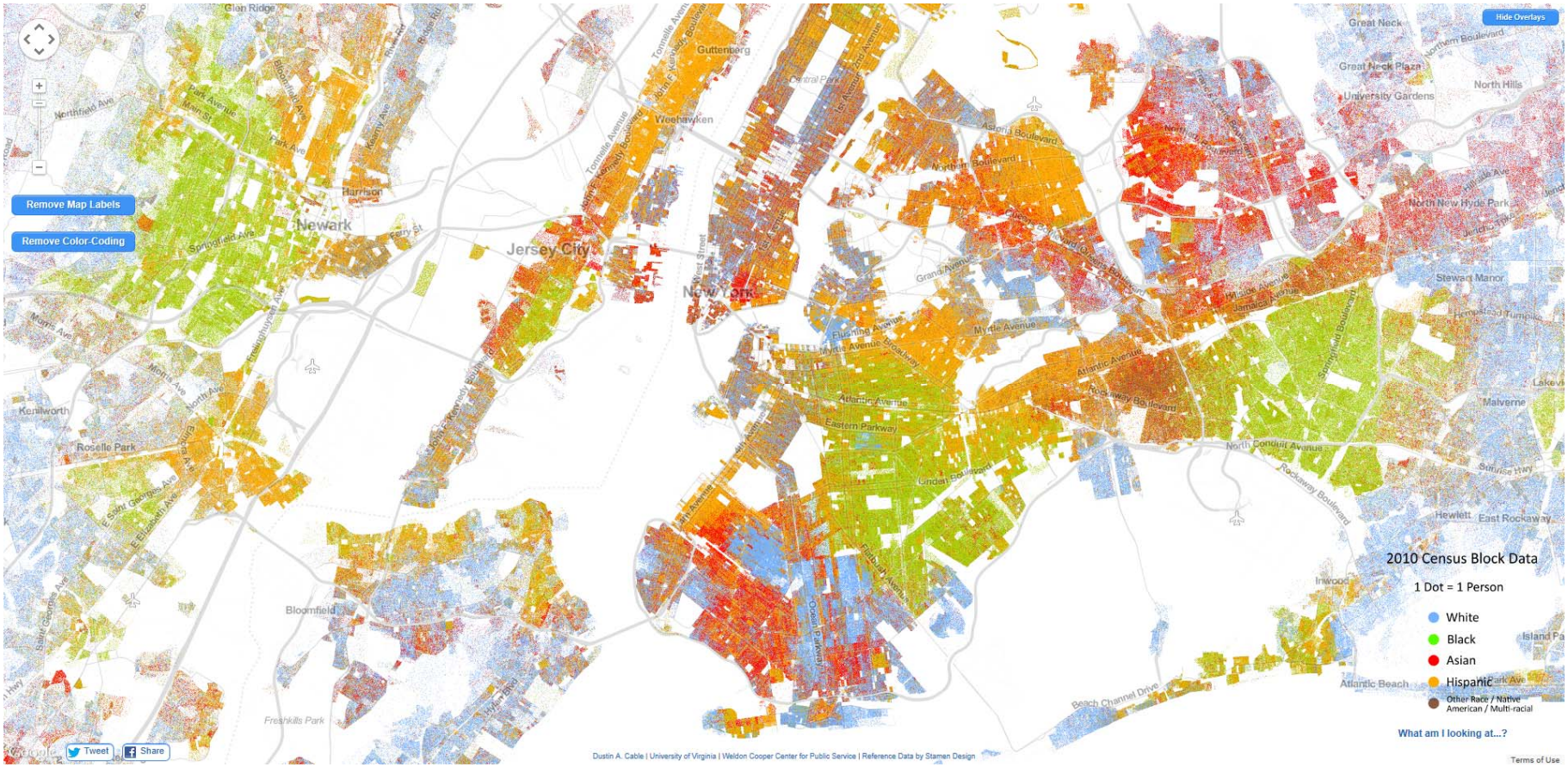
Taken from NTT docomo website

Traffic Pattern Location Information



Tokyo Traffic Pattern- Congestion Heatmap, derived from mobile phone data, March, 2013, 7:23 pm

Modelling Racial Diversity in New York



Building geo-portal
gathering data and utilizing them

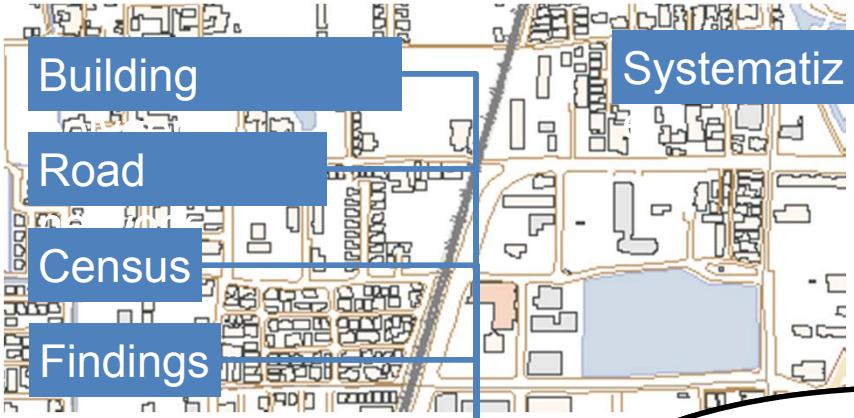
Static

Building

Road

Census

Findings



Systematiz

Dynamic

Seismic Intensity

Tsunami Height

Satellite Imagery

Probe/sensor data

Collaborate

Urban Resilience
Geo-Portal Online

Information for

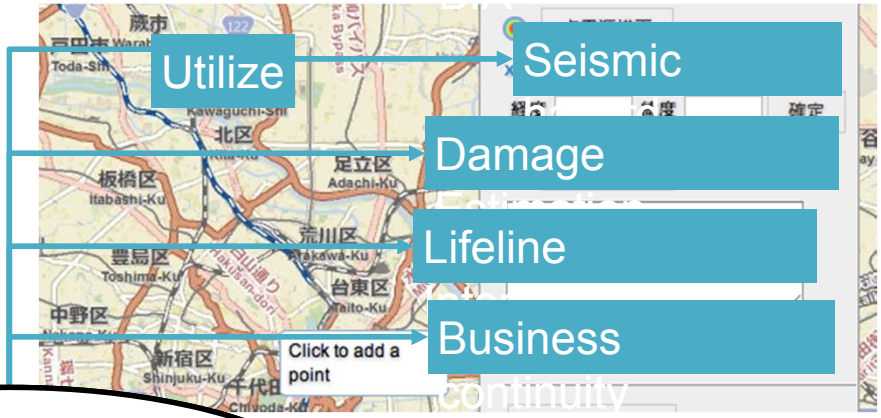
Utilize

Seismic

Damage

Lifeline

Business



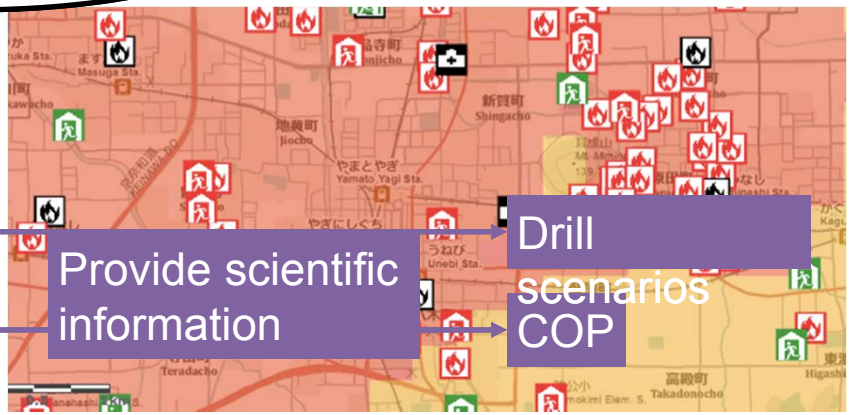
continuity

Information for

Provide scientific information

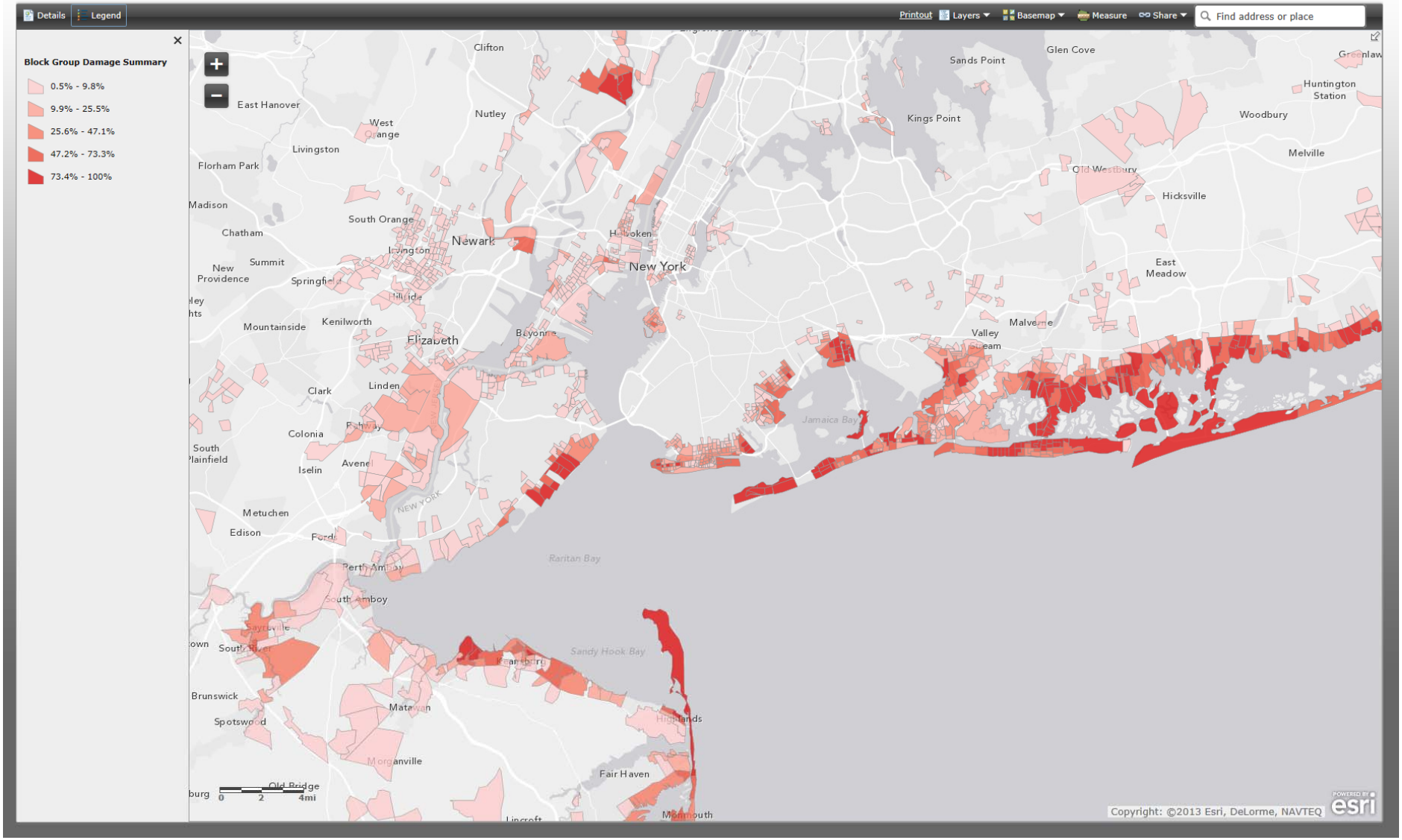
Drill scenarios

COP

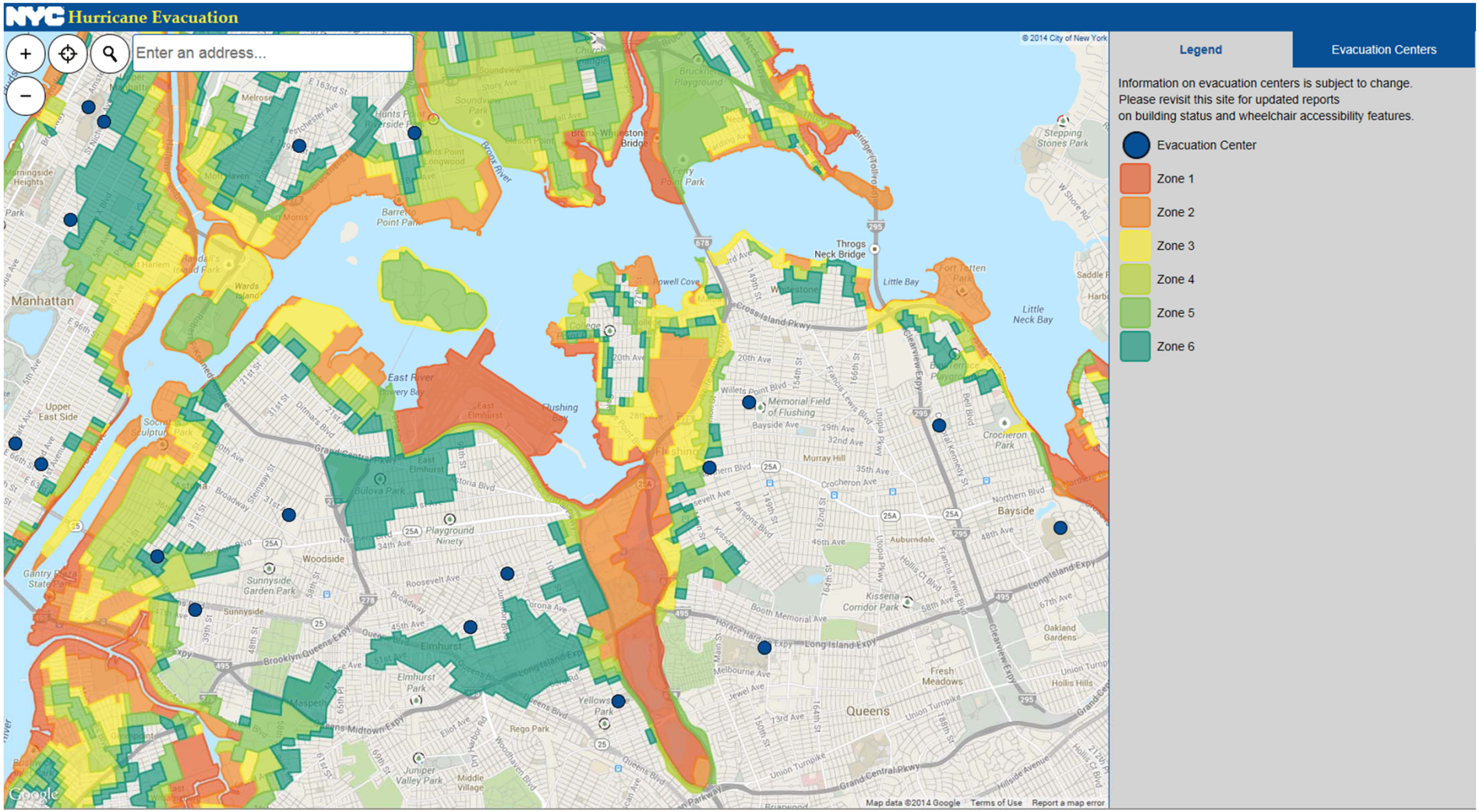


Hurricane Sandy Damage Estimates

Sandy Damage Estimates by Block Group



Hurricane Evacuation Zones (New York City)



Issues in Integrating Official Statistics and Geo-information

A. Treating Location as Basic Unit of Observation;

B. Full Integration of NSDI and NSDS:

NSDI: National Spatial Data Infrastructure

NSDS: National Strategy for Development of Statistics

C. Governance and Institutional arrangement

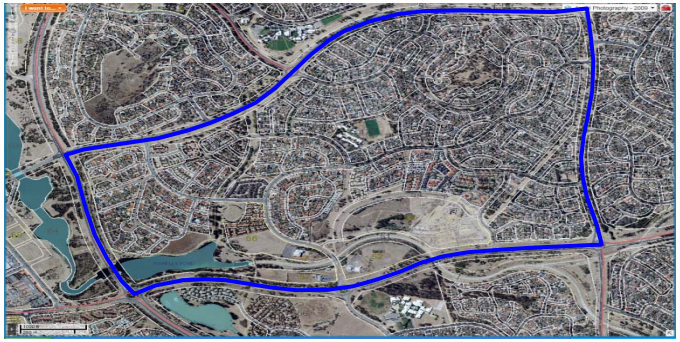
D. Evolving new National Information Infrastructure (NII)

Hierarchical Data Structure : Location as Basic Unit of Observation

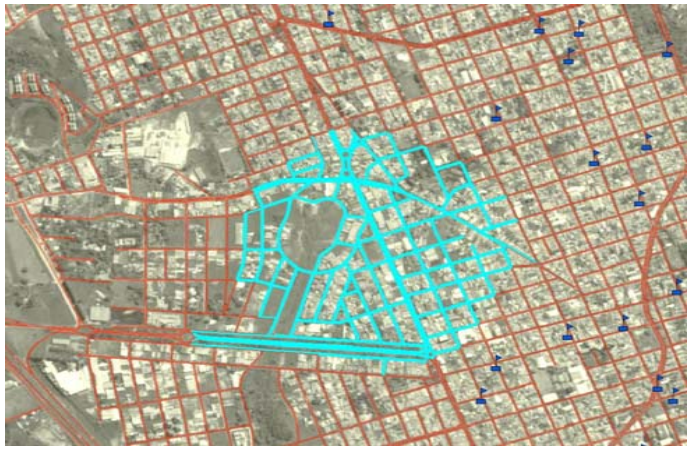
25 Smith St, Town Z x,y: 35.5676, 135.6587

Address / Geocode

Cadastral property parcels



Census Districts/Postal codes

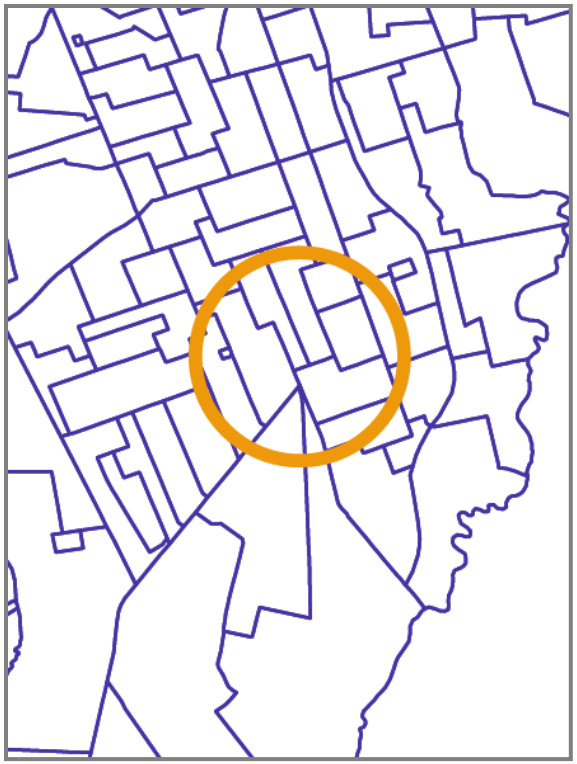


Block Face

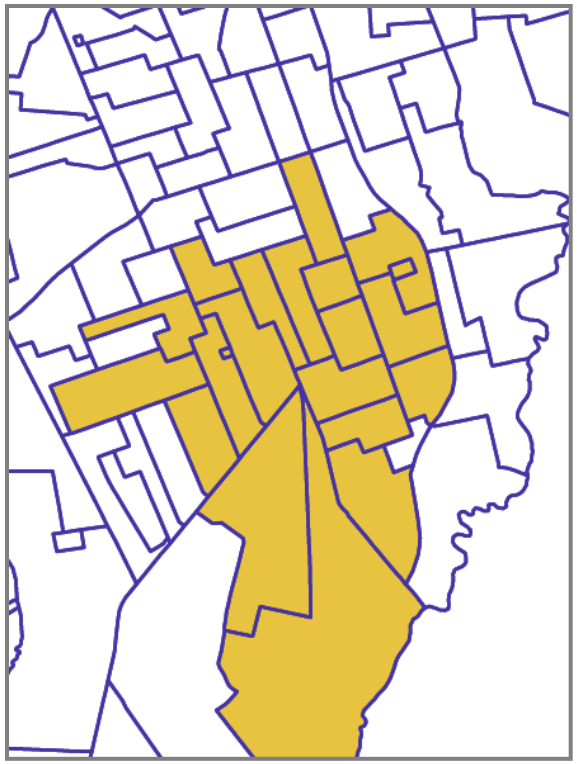
From Polygons to Points of Relevance (POR)

Users demand increasing precision.

What is the smallest spatial unit possible??



area of interest



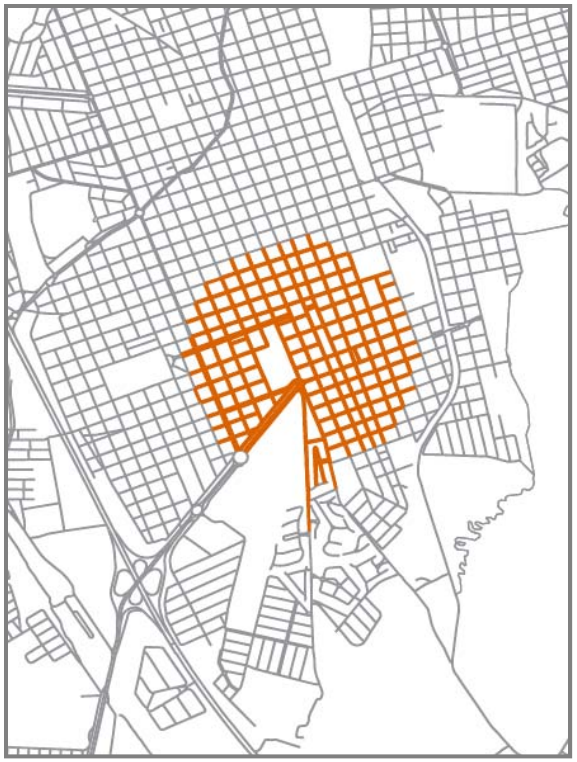
intersection result

Smaller Polygons, More Precise Data

Confidentiality the key constraint

But users demand (and will supply) POR data

area of interest



intersection result

From Polygons to Point-Based Information

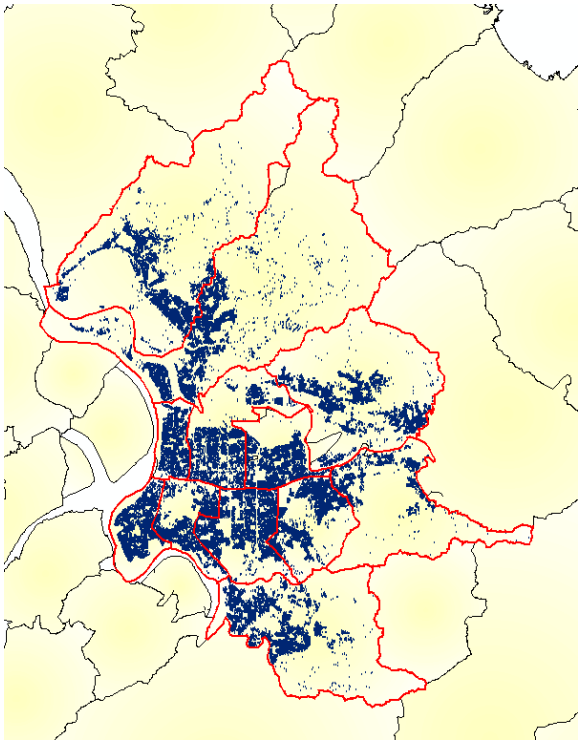
- Points likely to complement Polygons as the organizing framework for data integration, providing location-specific Information;
- The dynamic movement from Point to Point will pull out packets of Point-of-Relevance information on a string;
- Point-based information will be able to facilitate the convergence of information from multiple sources for a particular location;
- Points identified by Geocodes or Addresses.

Example from Chinese Taipei

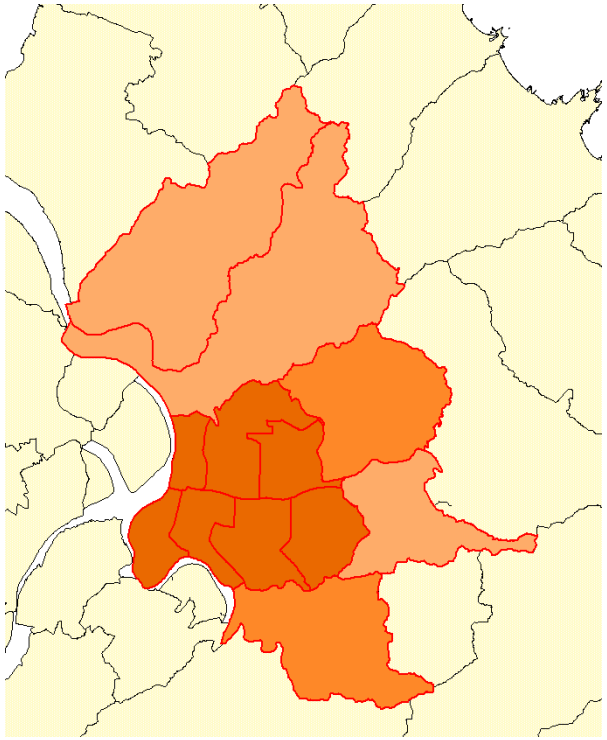
Problems

- Jurisdictional units is usually too large to provide detail information on local area of interest

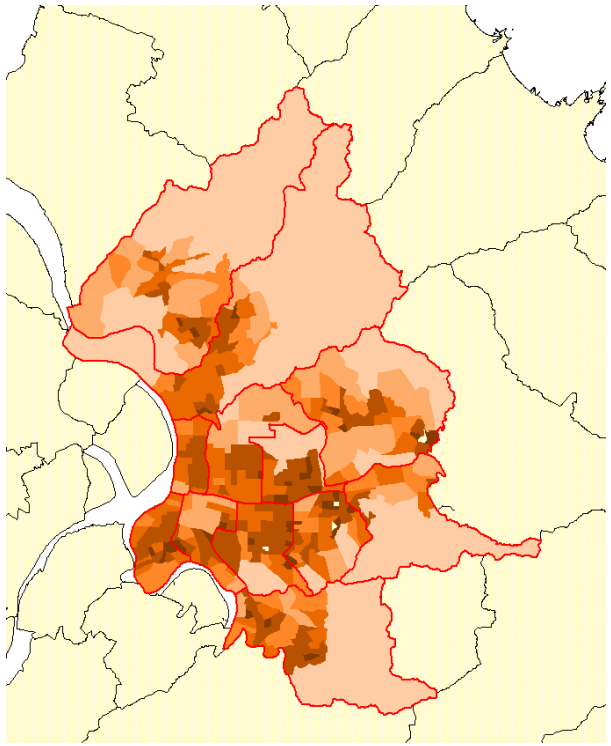
Original individual data

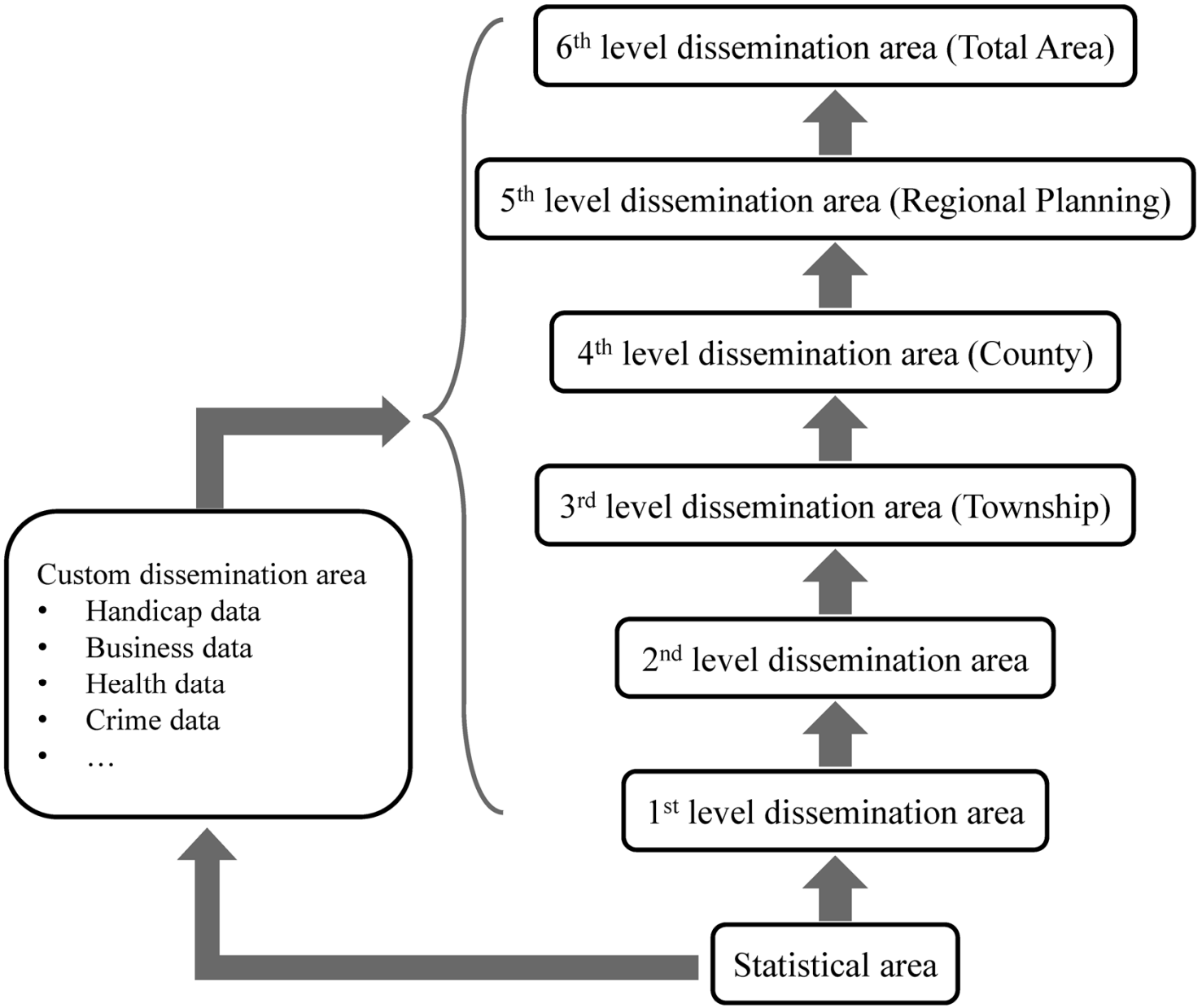


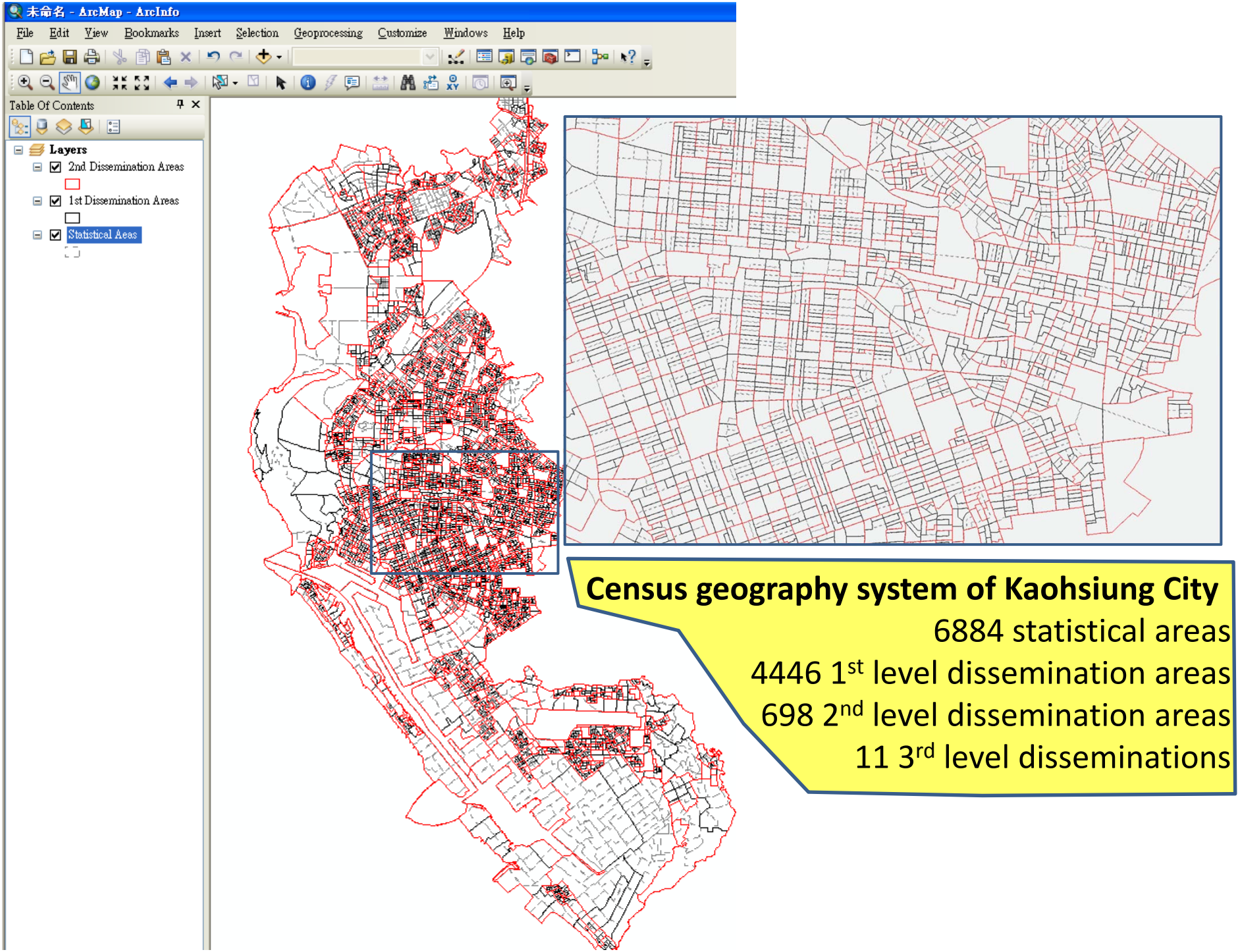
Aggregated by township level boundaries



Aggregated by village level boundaries




6th level dissemination area (Total Area)
5th level dissemination area (Regional Planning)
4th level dissemination area (County)
3rd level dissemination area (Township)
2nd level dissemination area
1st level dissemination area
Statistical area
Custom dissemination area
• Handicap data
• Business data
• Health data
• Crime data
• …



Census geography system of Kaohsiung City

6884 statistical areas

4446 1st level dissemination areas

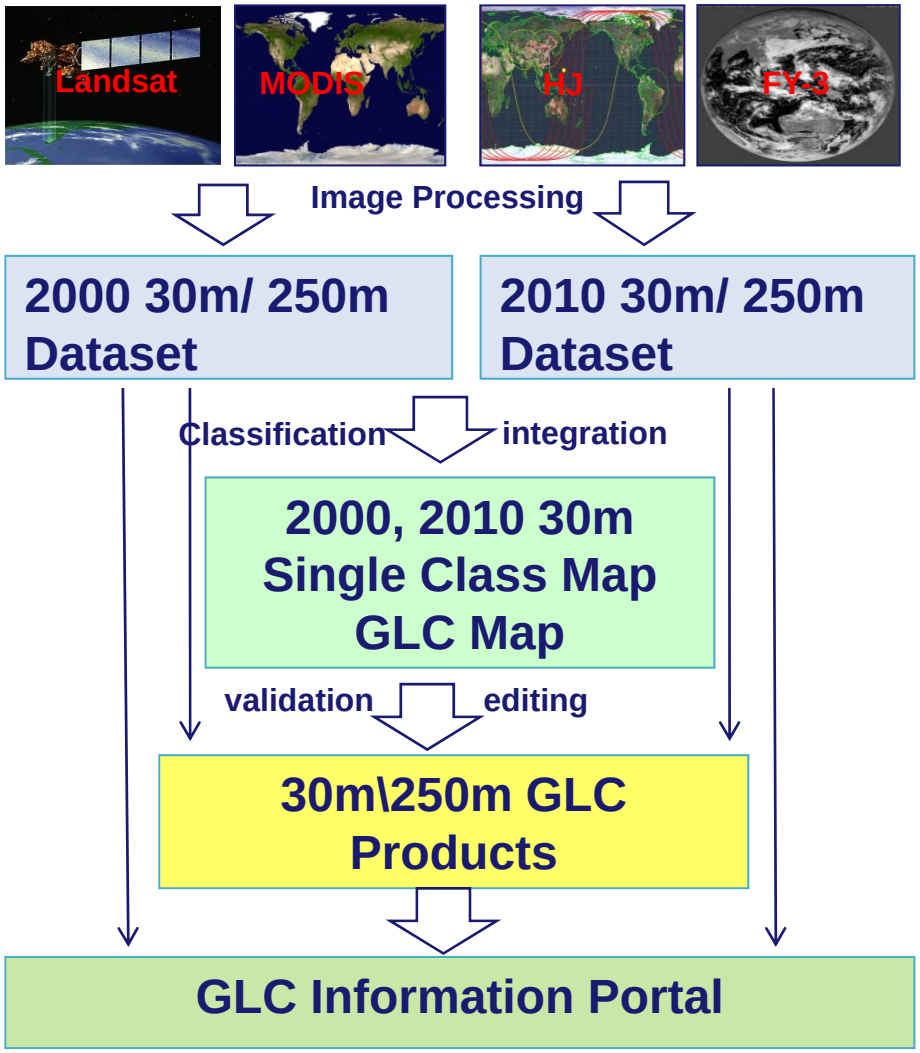
698 2nd level dissemination areas

11 3rd level disseminations

Building Location-Based Data Structure

- No consistent Geocode to link statistical data to Location;
- Many countries working on National Address Management Framework to define an unique geocode data structure;
- Urgently need location-based data management practices with multiple databases linked through geocode;
- Statistical-Spatial Metadata Interoperability, Integrating SDMX/DDI (statistics) with ISO-19115;
- Need enabling policies and protocols.

Global Location Information: China's Global Land Cover Mapping

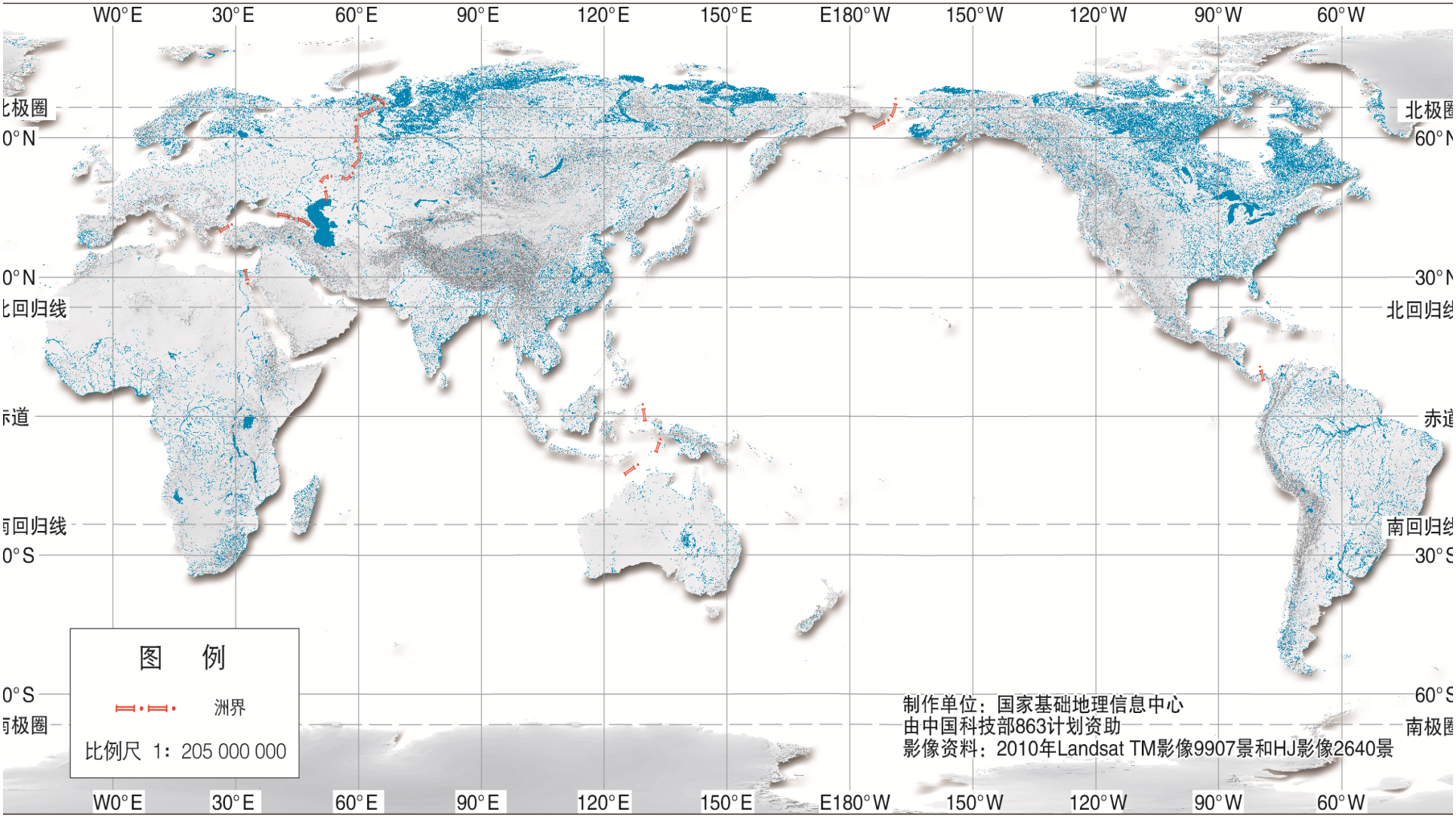


- **Mapping land cover of the whole globe at 30 m and for two baseline years (2000 and 2010)**

- **The first time in the world**

From Prof Chen Jun, NASG

Global Land Surface Water



From Prof Chen Jun, NASG

Assessment of Land Surface Water Areas

	Area of Surface Water (in 10,000 sq km)		Change (in 10,000 sq km)	% change
	2000	2010		
Asia	125.00	124.28	-0.72	-0.58%
Europe	32.58	31.59	-0.99	-3.12%
Africa	28.61	27.19	-1.42	-5.22%
North America	153.23	153.02	-0.21	-0.14%
South America	28.79	26.78	-2.01	-7.49%

From Prof Chen Jun, NASG

Coordinating NSDI and NSDS

- NSDI and NSDS at present two independent processes;
- NSDI sets norms for sharing spatial information;

  - Specifying the technology, policies, criteria, standards and people necessary to promote geospatial data sharing

- NSDS sets strategy for statistical development;

  - No reference to location information in NSDS

- Greater cooperation allow full implementation of Statistical Geospatial Framework (S. Africa as example)

Governance and Institutional Arrangement

- Mapping agency and Statistical agency under different jurisdictions and with different mandates;
- Coordination needs to be further enhanced;
- Mapping agency primary responsibility is the base map and the geographic and cadastral layers;
- Statistical agency adds layers of information from their pool of official statistics, but also produce census maps;
- Coordination and collaboration based on mutual respect is important

The Road Ahead: Developing a National Information Infrastructure (NII)

- Need a National Information Infrastructure (NII) to coordinate Data Fusion (meta data and layers);
- Need a core data system linking all agencies and new tools in information integration;
- Need new thinking on how to use information;
- Need a new national governance arrangement to manage the process of gathering and using information;
- Need cross-border coordination as well such as borderlands information systems or global thematic mapping.

Developing a NII through CDO Leadership

- Governments and enterprises have established 'Chief Data Officers' (CDOs) to have the business responsibility to capture and exploit data for decision-making purposes;
- CDO's role will become more important in future, as decisions are based not on single source, but through data fusion and aggregation;
- CDO manages the location of central database and the coordination of information layers to enhance efficiency;
- The new NII will be critical for national competitiveness and productivity gain.

THANK YOU

paul.cheung@nus.edu.sg