Challenges in the Modernization of Statistical Production Process



Professor Paul Cheung
National University of Singapore

Two Critical Questions

- Are we happy with the current production process and the products?
  - Do we need to save money, increase productivity, improve information value
- How can we improve this?
  - improving quality, timeliness, marketing, usefulness, credibility, relevance, impact

Review: Statistical Production Process (I)

Measurement Framework

- Defining the Domain
  - eg: Tourism: expenditure on product/services
- International Standards
  - International Manual on Tourism Statistics
- Metadata: Definitions and Explanations

Review: Statistical Production Process (II)

Data Collection Instruments and Processes

- How to collect?
  - questionnaire, administrative sources, active or passive sensors, unmanned devices
- What is the process?
  - pre-arranged interviews, random choice, internet collection, sms, fax, automated devices

Review: Statistical Production Process (III)

Data Editing, Analysis and Archival

- Editing: manual or automatic, level of precision
- Analysis: pre-determined or data-mining
- Archival: protocol for storage and retrieval

Review: Statistical Production Process (IV)

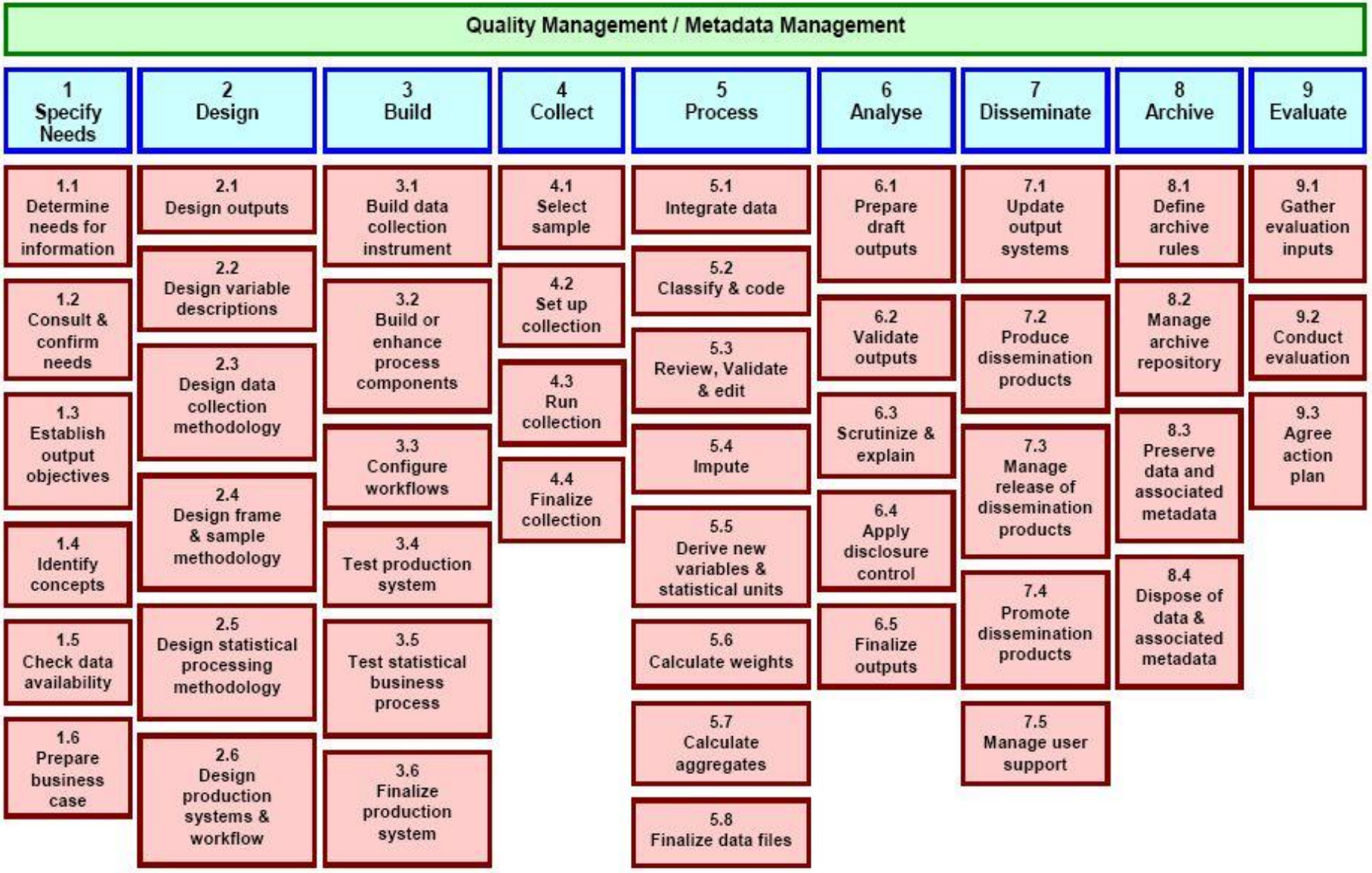
Output Production

- Types of output: hard copy, soft copy, internet
- Media Mode: press briefing, interview, circular
- Interpretation and Explanation

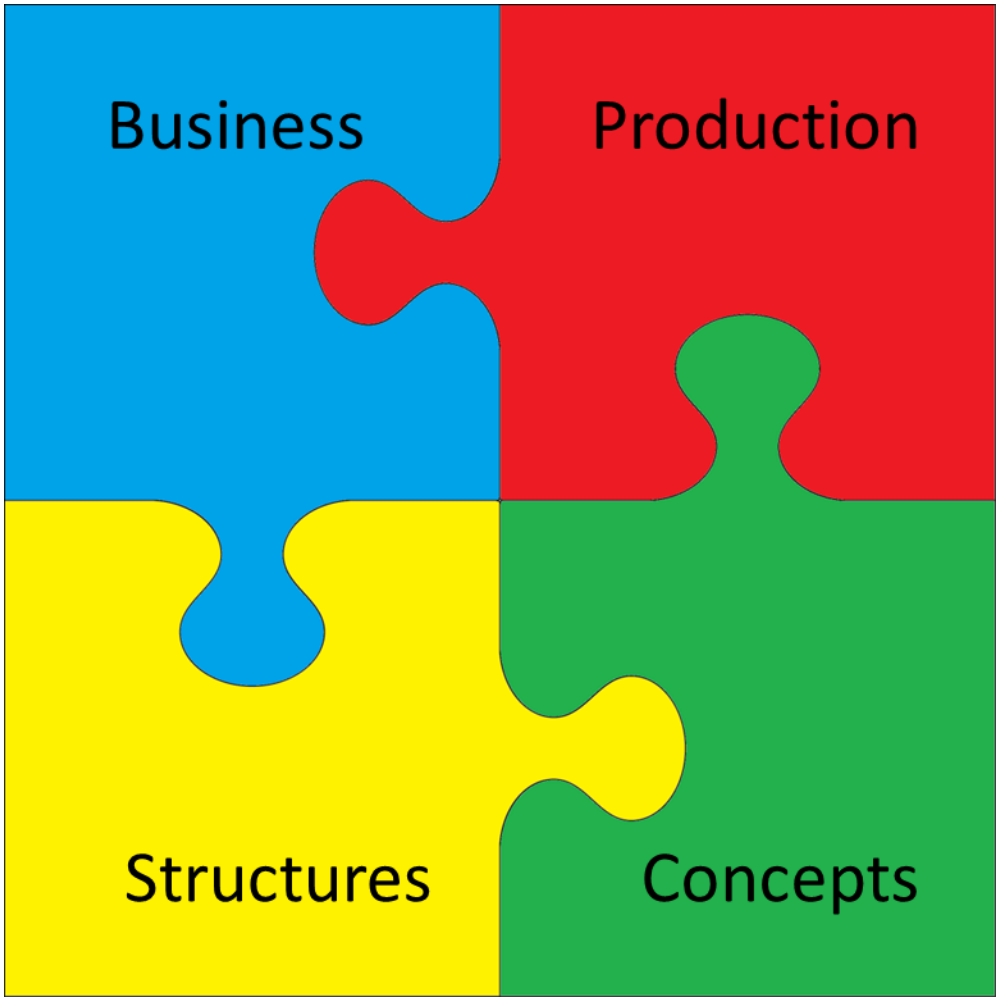
Generic Framework on Statistical Production Process

- Global statistical community desires a generic framework to review statistical production process
- International agreed modules on each of the production process with technical specifications
- Conference of European Statisticians leading this work. Two models proposed.

Generic Statistical Business Production Model (GSBPM)



Generic Statistical Information Model (GSIM)



GSIM and GSBPM

Drivers for Modernization

- Rapid advancement in technology
  - internet, geospatial, video, speed, server space, sensors
- Changing attitudes of respondents and users
  - Respondents less cooperative,
  - Users demand timely, relevant data
- Evolving information value-chain
  - Need to compete and establish information value

A. Modernizing Data Collection

- Traditional surveys less emphasized. Too time consuming. Too slow;
- Multi-mode approach: internet, call center, administrative source, fax, sms, sensors;
- Active or passive data collection;
- Back-end system support and integrative environment important;

Mobile Phone Positioning Data for Tourism Statistics

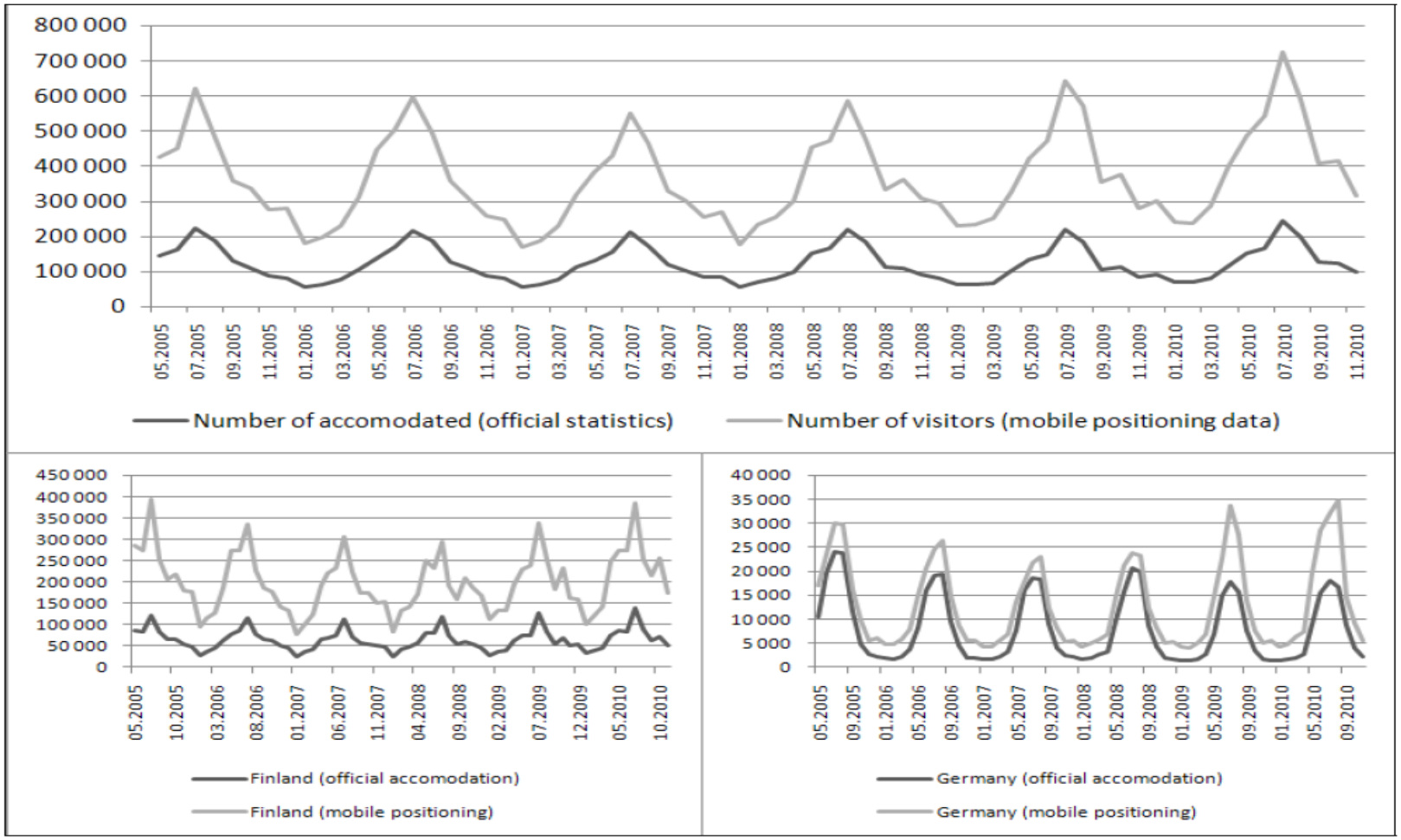
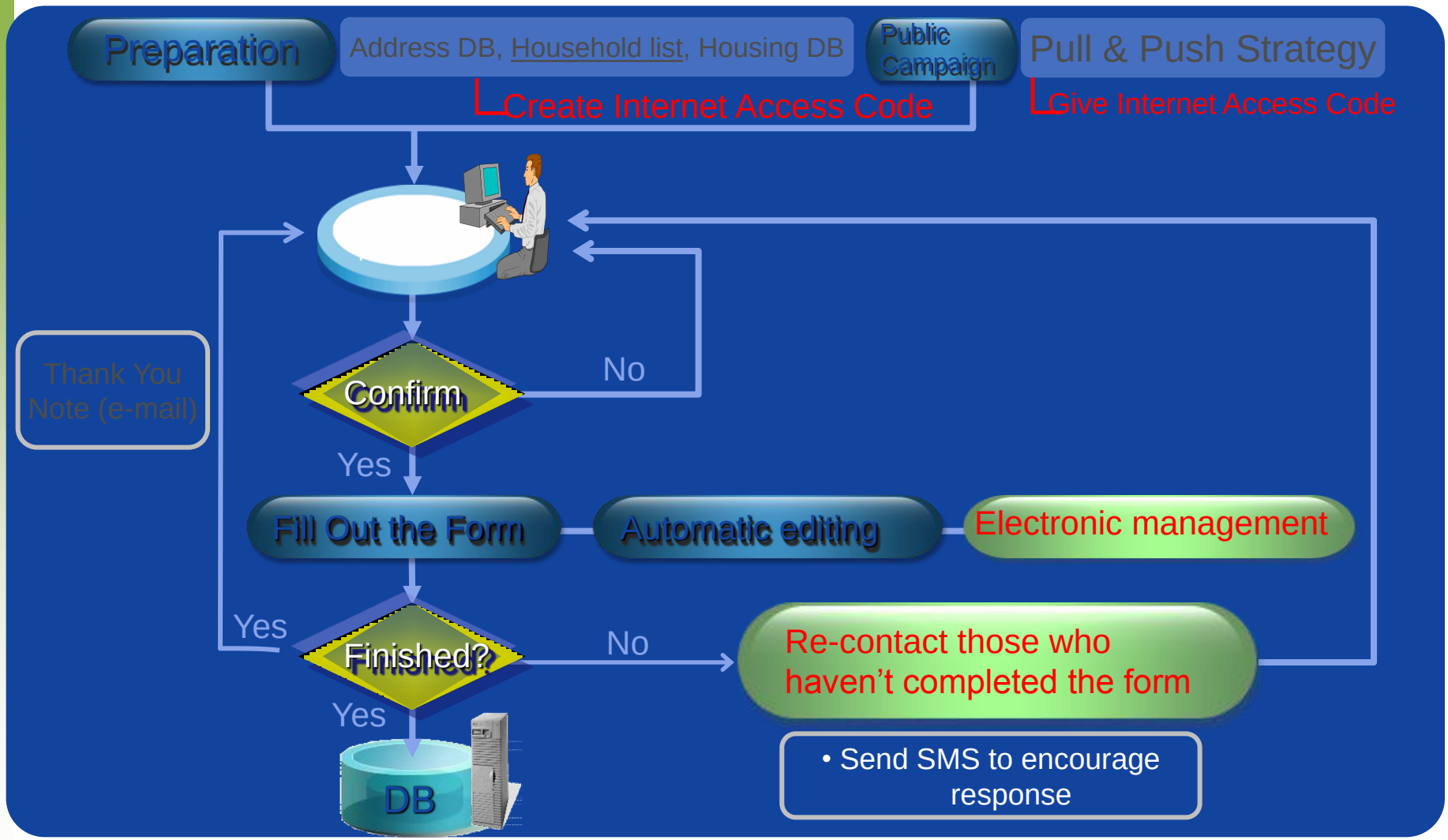


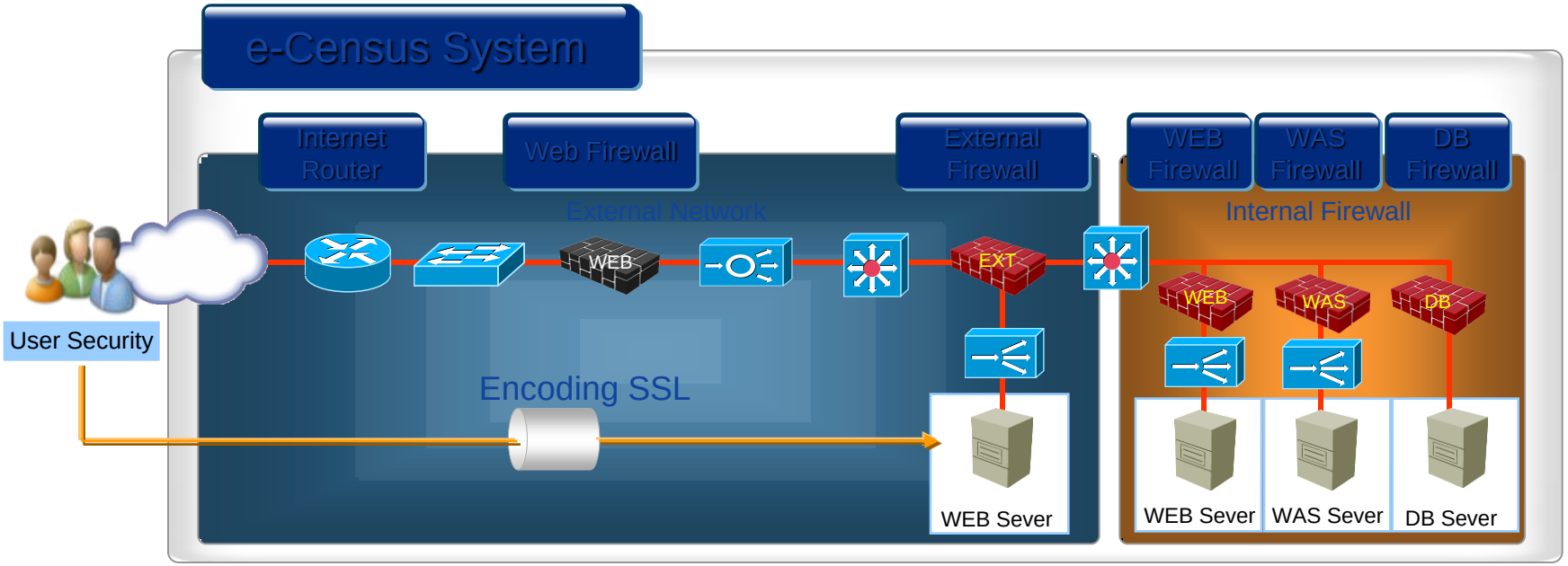
Figure 3. Comparison of official and passive mobile positioning based inbound tourism statistics for all tourists, Finns and Germans.

Source: Mobile Telephones and Mobile Positioning data as source for statistics: Estonian Experiences, Ahas et. Al. (2011)

Korea Internet Census



Internet Census Security



User Security (Level 1)	Network Security (Level 2)	Web Service (Level 3)	Server Security (Level 4)	DB Security (Level 5)
PC Security	Access Control	Web Service Security	Security OS	Encoding Data
▪ PC keyboard ▪ Web browser ▪ No capture ▪ Encoding SSL	▪ Separating internal and external network ▪ Control irregular traffic	▪ Web Firewall	▪ Controlling accesses to server	▪ Encoding personal information

B: Modernizing User Engagement

- Recognized importance of consulting users extensively and systematically

- UK Example:
  - Measuring ‘national well-being’
  - 175 events, 2750 people, 34000 responses

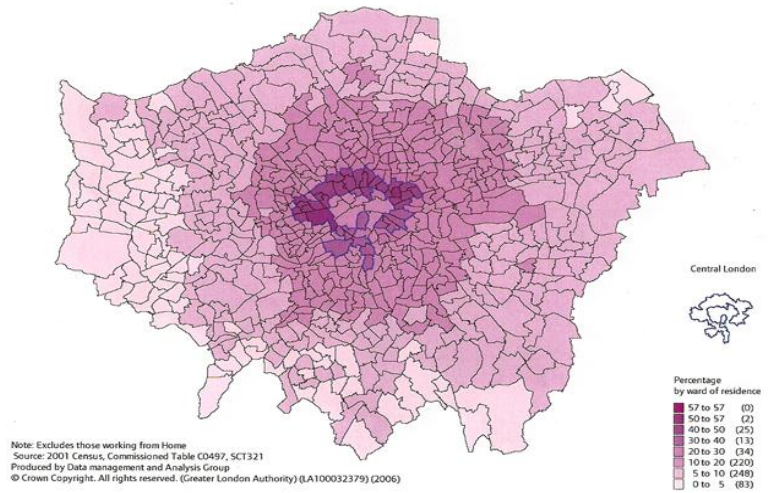
- Australia Example:
  - measuring ‘aspirations and progress’ :
  - advisory panel, social media, national forum, government workshops, international review

C: Modernizing Spatial Reference

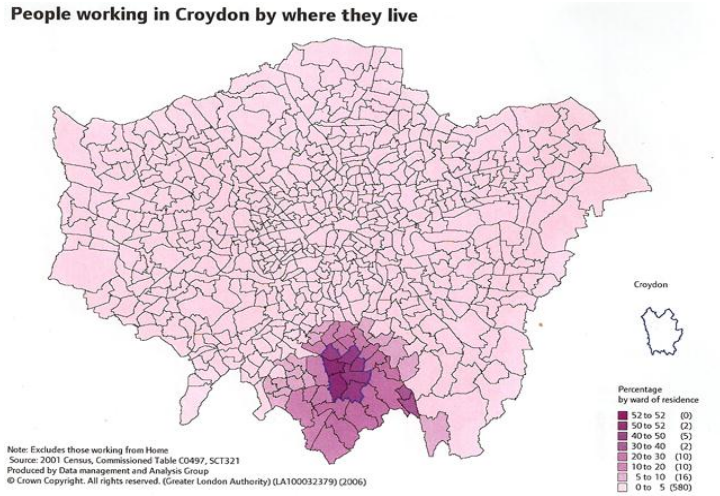
- Recognized importance of location-based information and pervasive use in devices
- Statisticians use ‘Metric’; geospatial information use ‘polygons’ as basic unit.
- Need a common language through geo-coding; All statistical data should be geo-coded
- UNSC considered ‘Statistical-spatial framework’ in 2013 session

Spatial Data Analysis

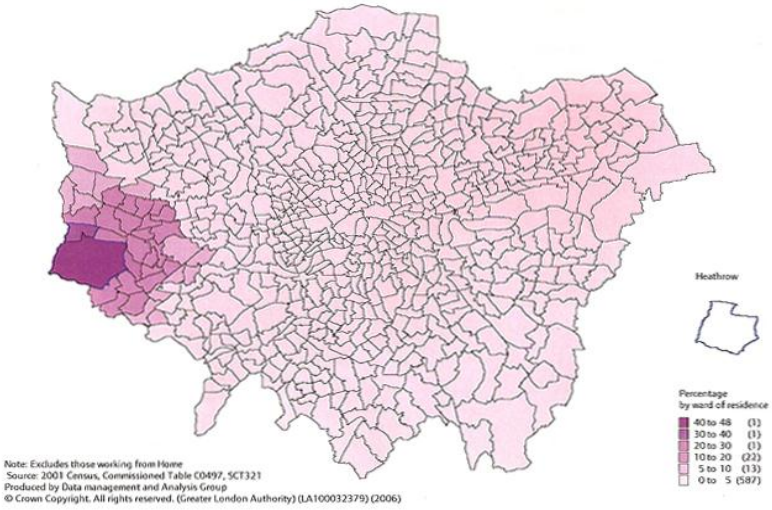
People working in Central London by where they live



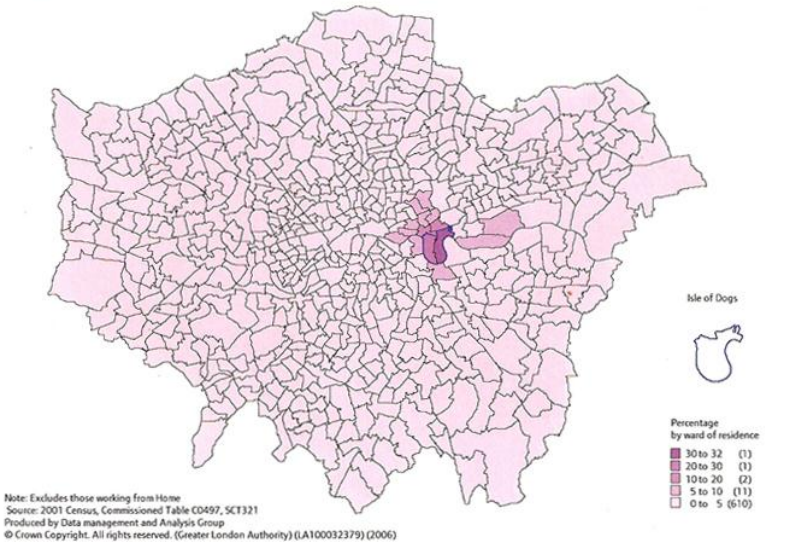
People working in Croydon by where they live



People working in Heathrow by where they live

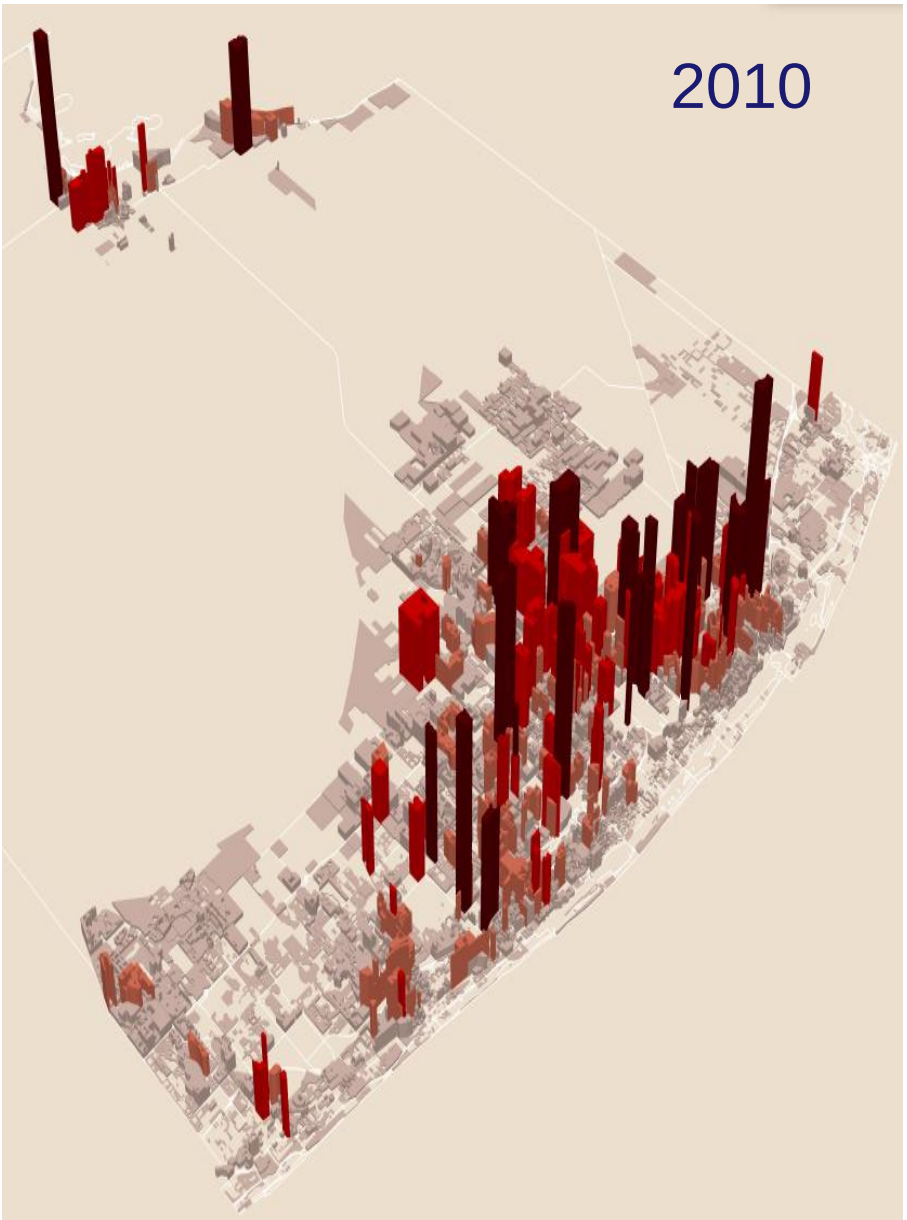


People working in the Isle of Dogs by where they live



Source: Alan Smith

3-D Sub Population Analysis



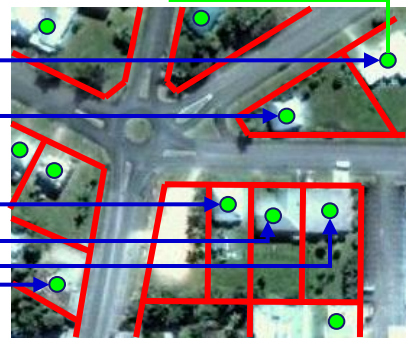
Location Information Framework

Aggregated to Local Government area or higher

Geocoded unit level data

25 Smith St = x,y: 35.5676, 135.6587

Aggregated to suburb or postcode



Location information at address level

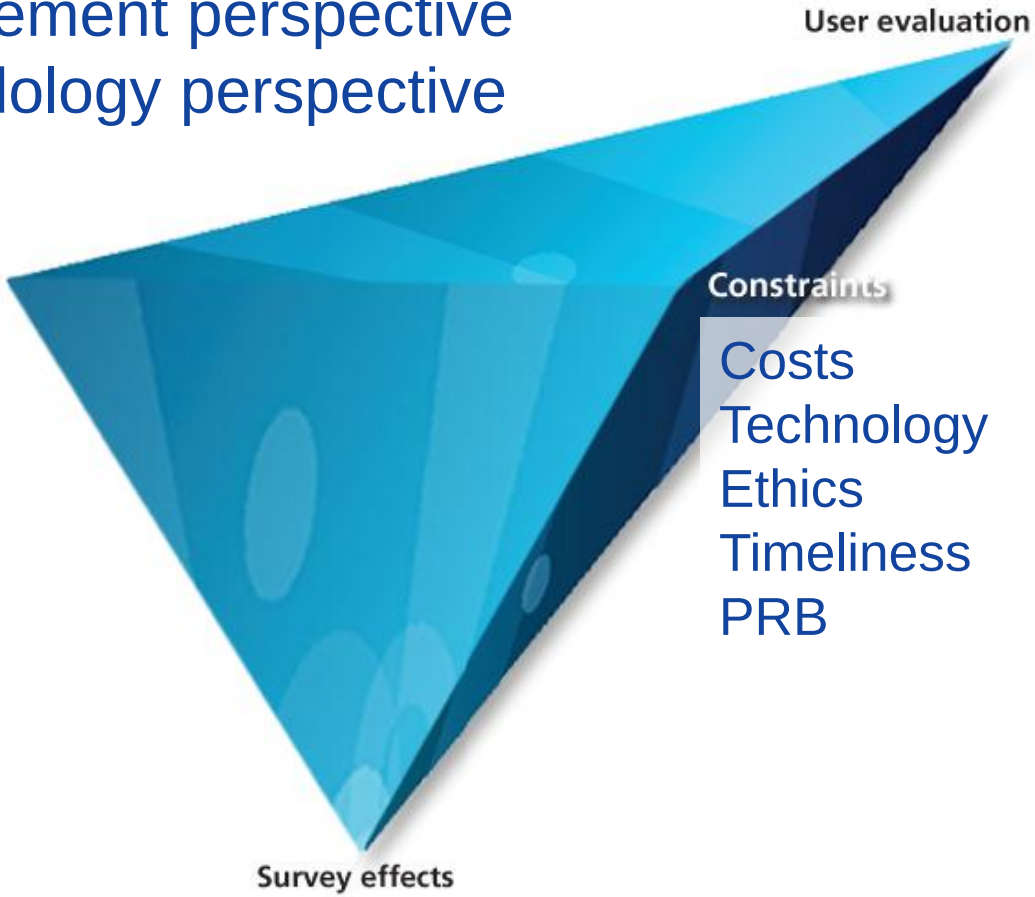
Analysis and aggregation across geographies

D. Modernizing Quality Assurance

- Quality assurance becomes more important especially after the Mexico/Greece case;
- European Commission puts strong emphasis with special mandate given to Eurostat;
- UNSC approved National Quality Assurance Framework with 19 dimensions in 2012 grouped under 4 headings: a) the statistical system, b) the institutional arrangement, c) managing statistical process, d) managing statistical output

The Quality Diamond

- Management perspective
- Methodology perspective



Relevance
Accuracy
Timeliness
Punctuality
Accessibility
Clarity
Comparability
Coherence

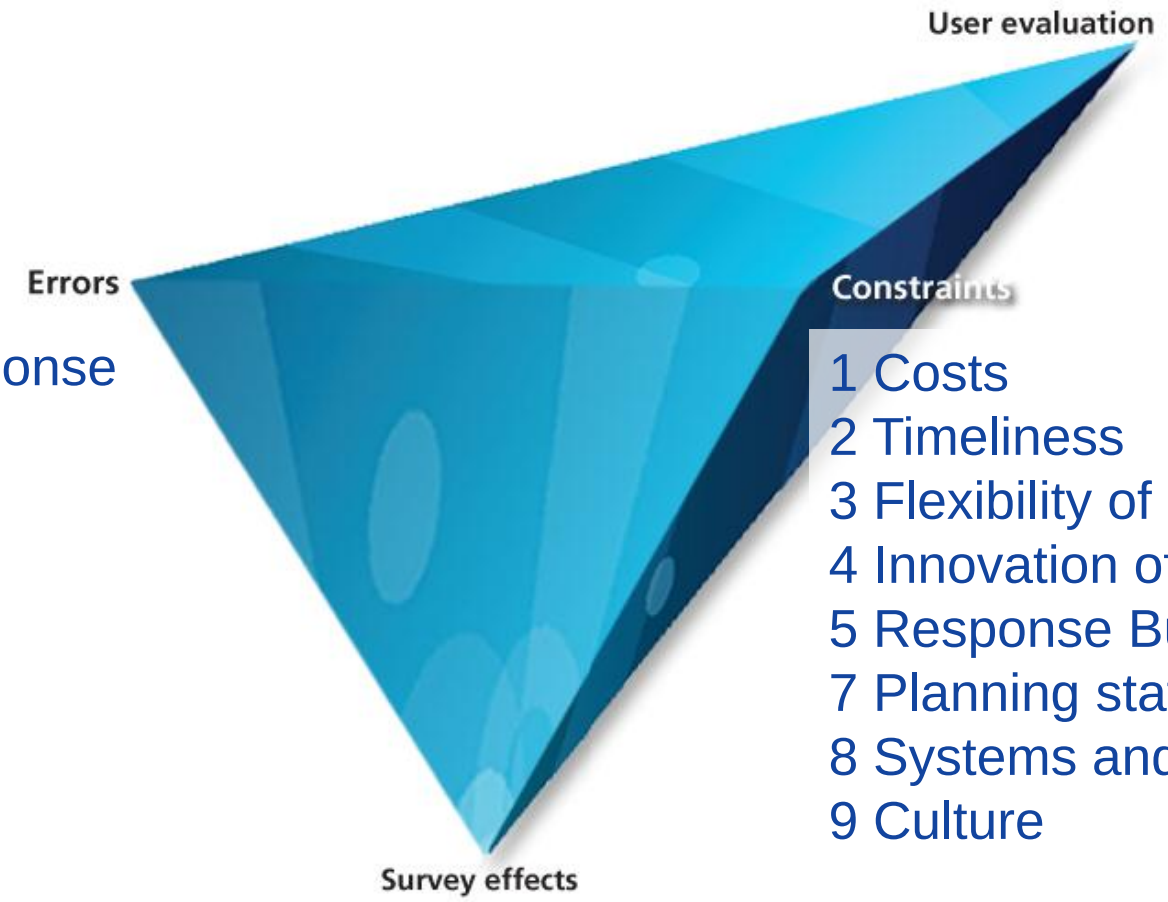
(Eurostat Quality Dimensions, 2011)

Specifications
Coverage
Sampling
Nonresponse
Measurement
Processing

Costs
Technology
Ethics
Timeliness
PRB

Question order effects
Mode effects
Cross-cultural effects

The Quality Diamond: The Management Perspective



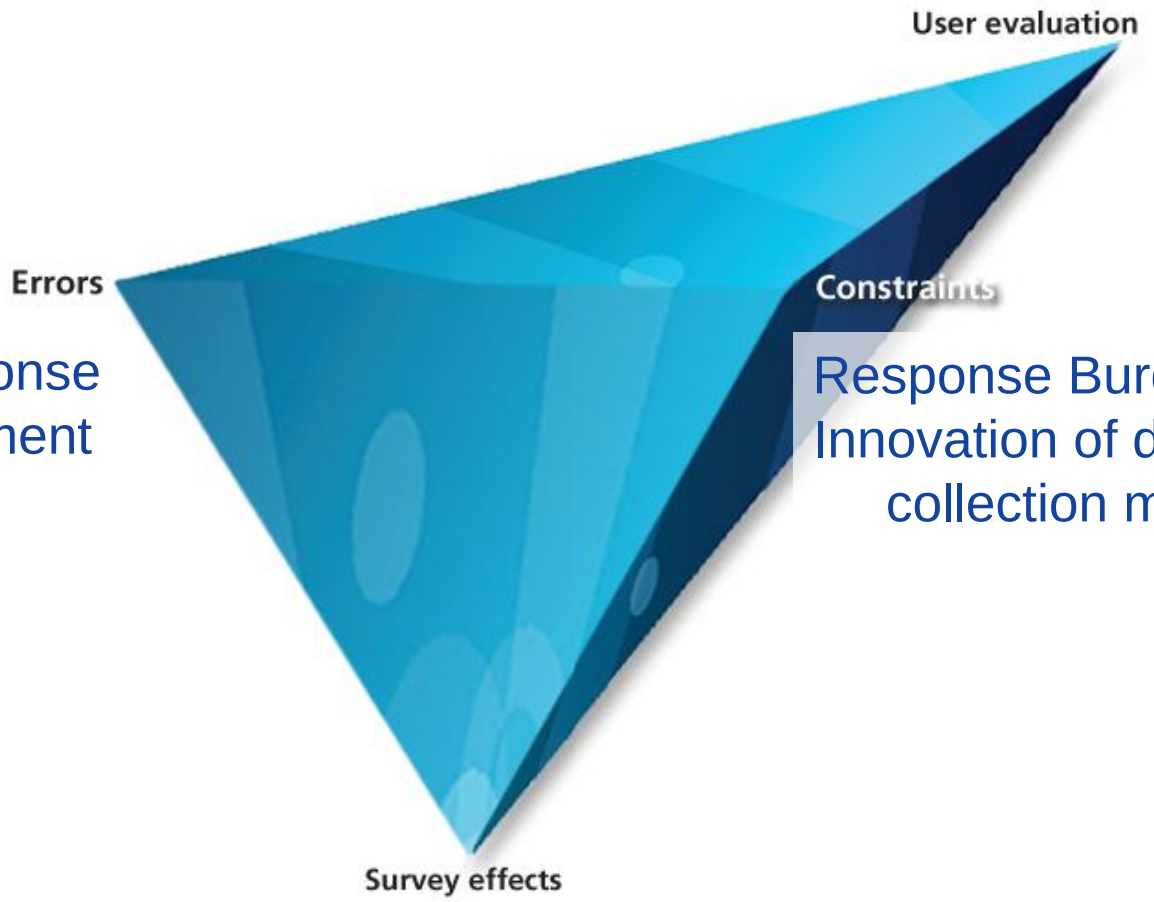
6 Managing customer demands and expectations

Non-response

1 Costs
2 Timeliness
3 Flexibility of process planning
4 Innovation of process
5 Response Burden
7 Planning staff
8 Systems and tools
9 Culture

Mixed-mode effects
Data Collection Strategy effects

The Quality Diamond: The Methodology Perspective



Accuracy

Non-response
Measurement
Coverage
Sampling

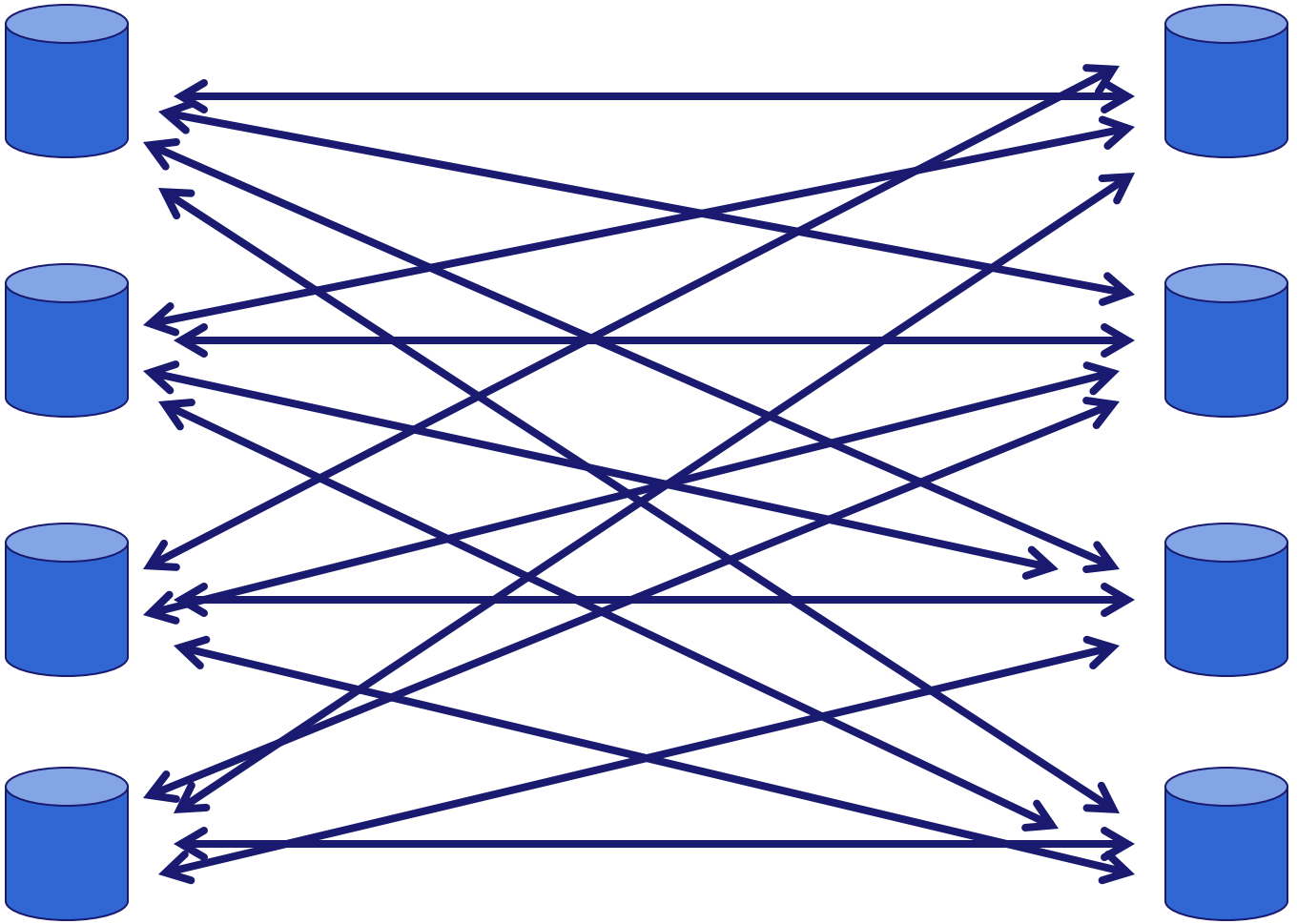
Response Burden
Innovation of data collection methods

Mixed-mode effects
Questionnaire effects
Data Collection Strategy effects

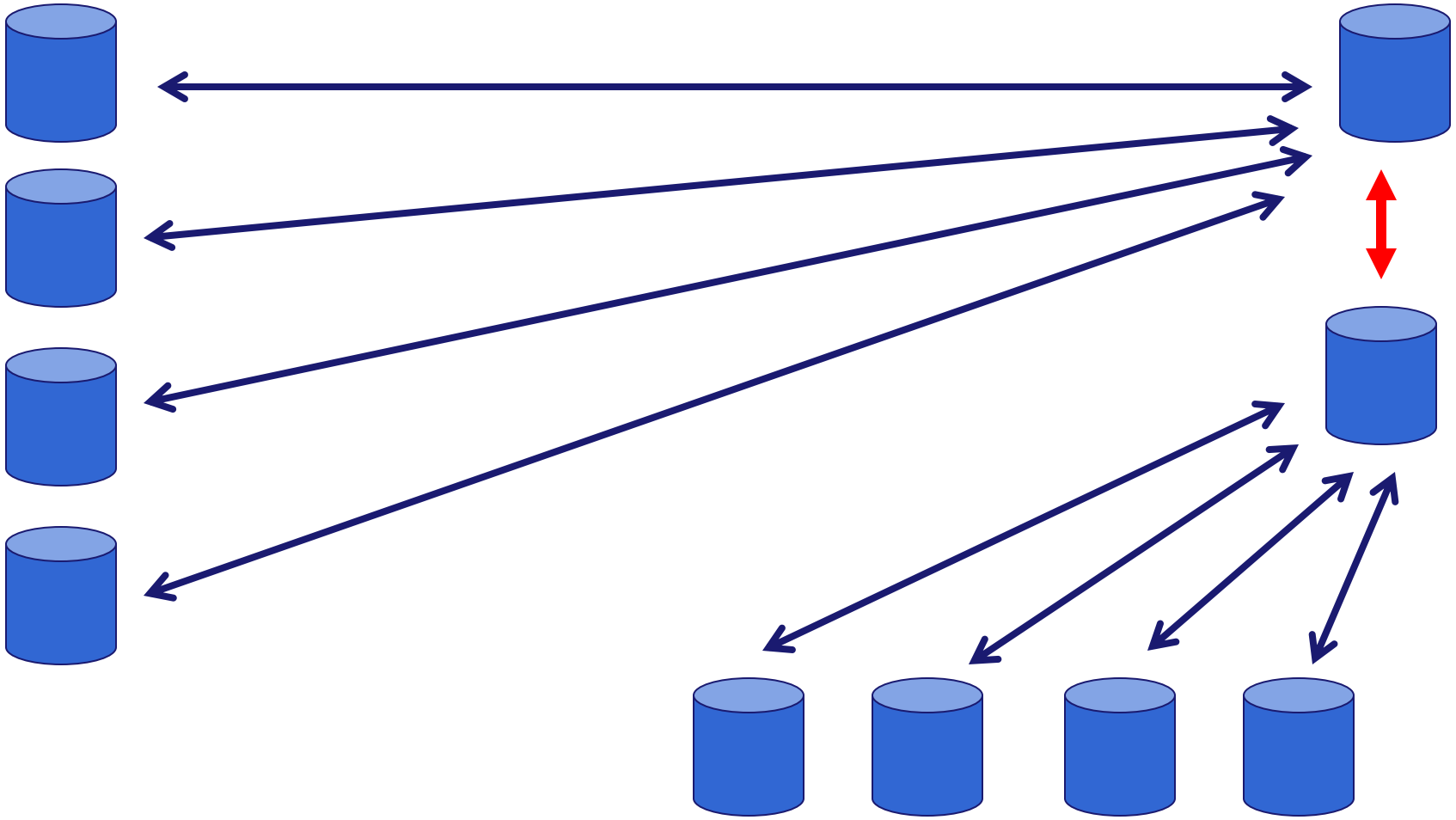
E. Modernizing Data Exchange

- Modern statistical system relies heavily on data exchange; no more ‘silo’
- How to transfer data effectively?
- SDMX has become global standard for data exchange, led by 7 agencies
- UN adopts SDMX as global exchange standard

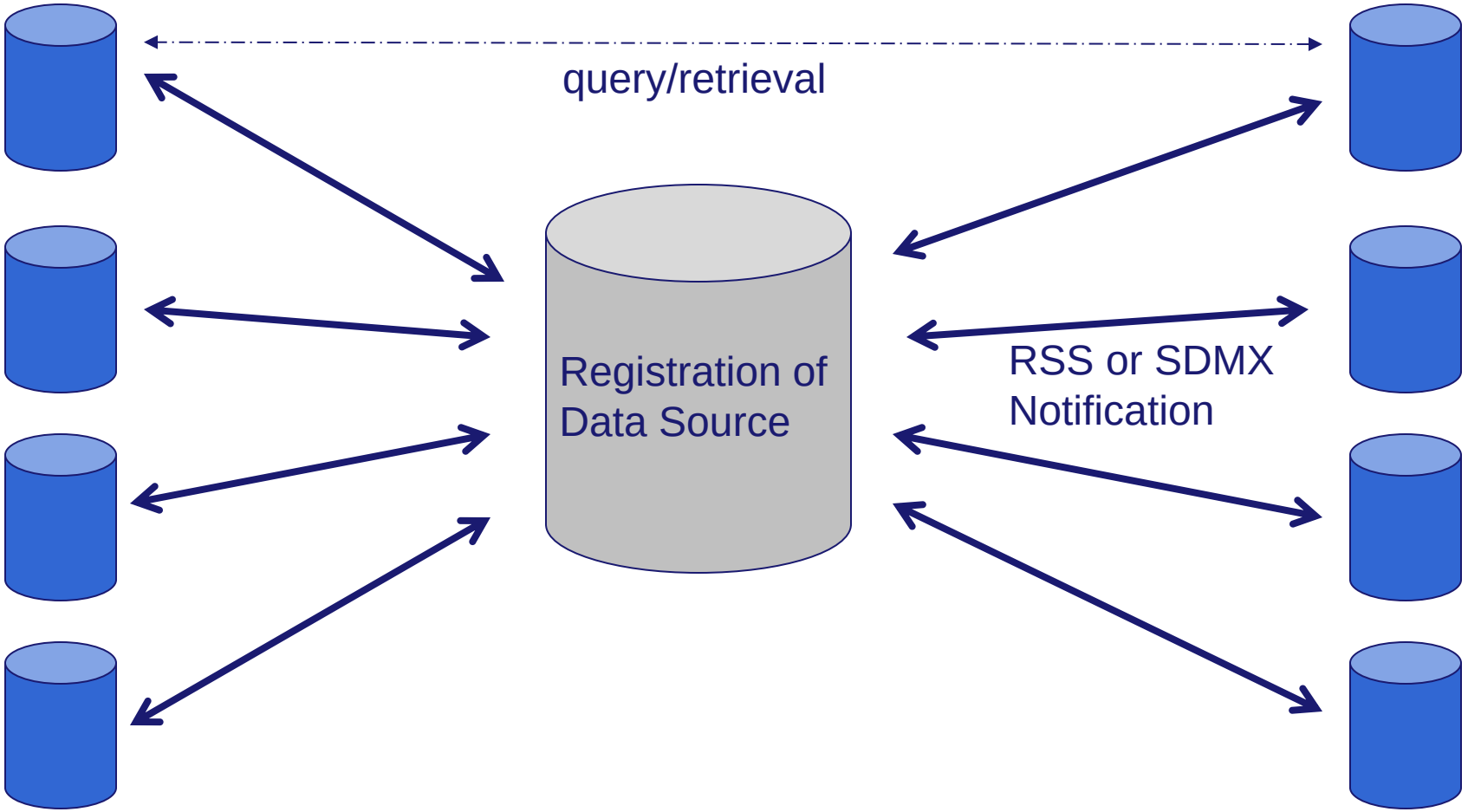
Bilateral Exchange



Gateway Exchange



Data/Metadata-Sharing Exchange



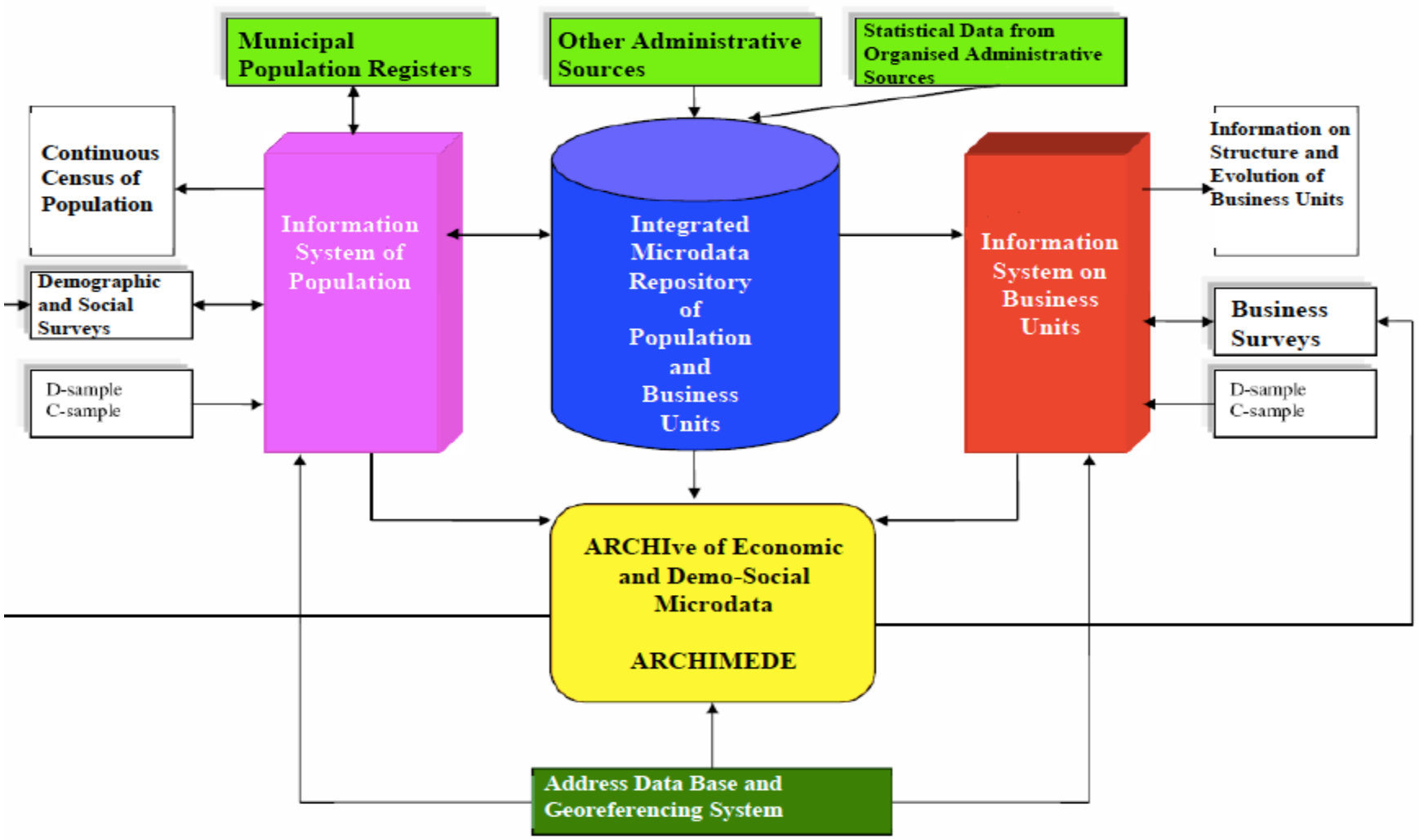
Data/Metadata Reporters

Data/Metadata Consumers

F. Modernizing Archival and Retrieval

- Data can be used anytime, anywhere.
- Data Archival : Principle of Immutability.
  - meta data important, data catalog
- Data retrieval : new standards emerging in health care or financial industries. Indexing and search capabilities
- TheWorld Bank has developed special tool kits

Italy Data Archival System

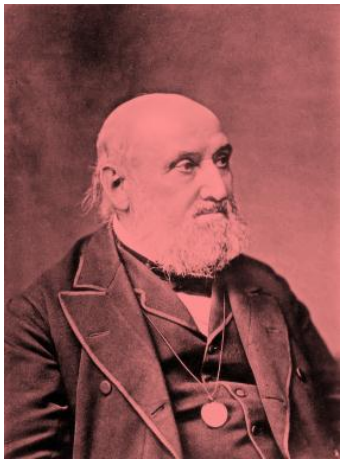


G. Modernizing Dissemination

- Data dissemination has undergone the most changes with the Open Data era
- Much more will happen in the future as technology rapidly changes
- But resistance is still there. More political than technological.

The Old Rule

William Farr (1807-1883)

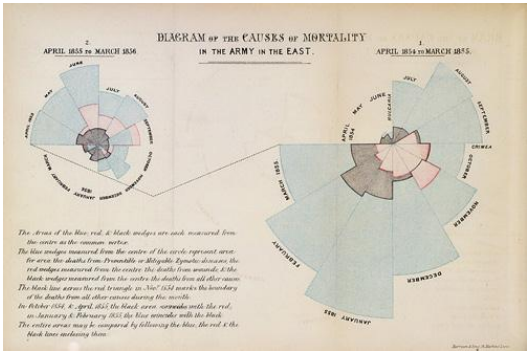


"You complain that your report would be dry [without the graphics]…

…THE DRYER THE BETTER.

Statistics should be the *dryest* of all reading"

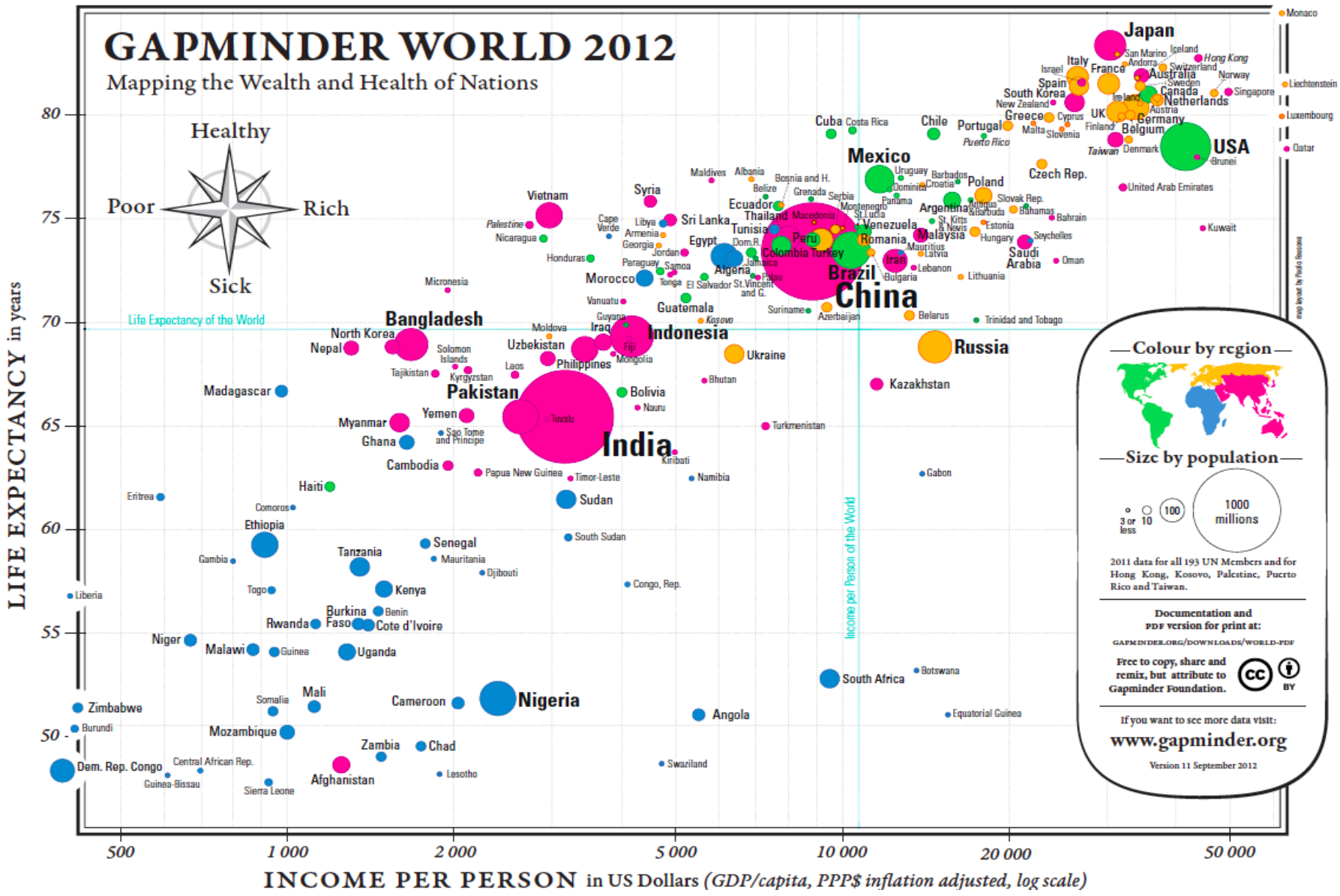
letter to Florence Nightingale, 1861



Modernizing Dissemination

- Open data initiative, access to microdata
- Transiting from print version to multiple-platforms: web, mobile, more interactive
- Catering to users at different level – user-defined data needs
- Forums for discussion/inputs
- From information providers to knowledge builders
- Visualization
- Location – specific data dissemination

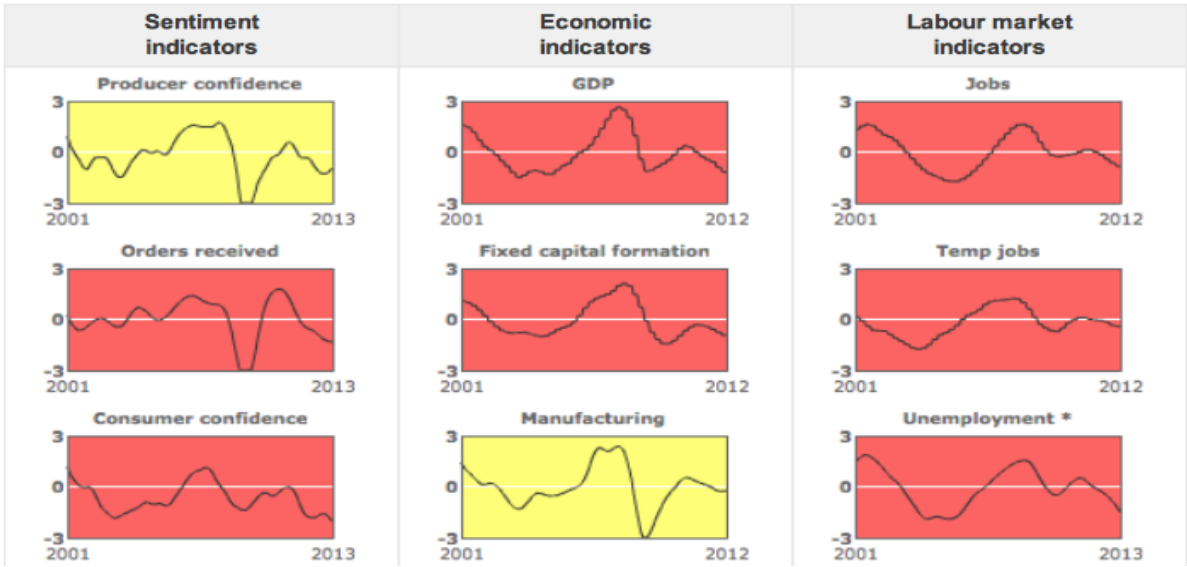
Data Visualization



Themes

- Agriculture
- Construction and housing
- Education
- Enterprises
- Financial and business services
- Government and politics
- Health and welfare
- Income and spending
- International trade
- Labour and social security
- Leisure and culture
- Macroeconomics
- Manufacturing and energy
- Nature and environment

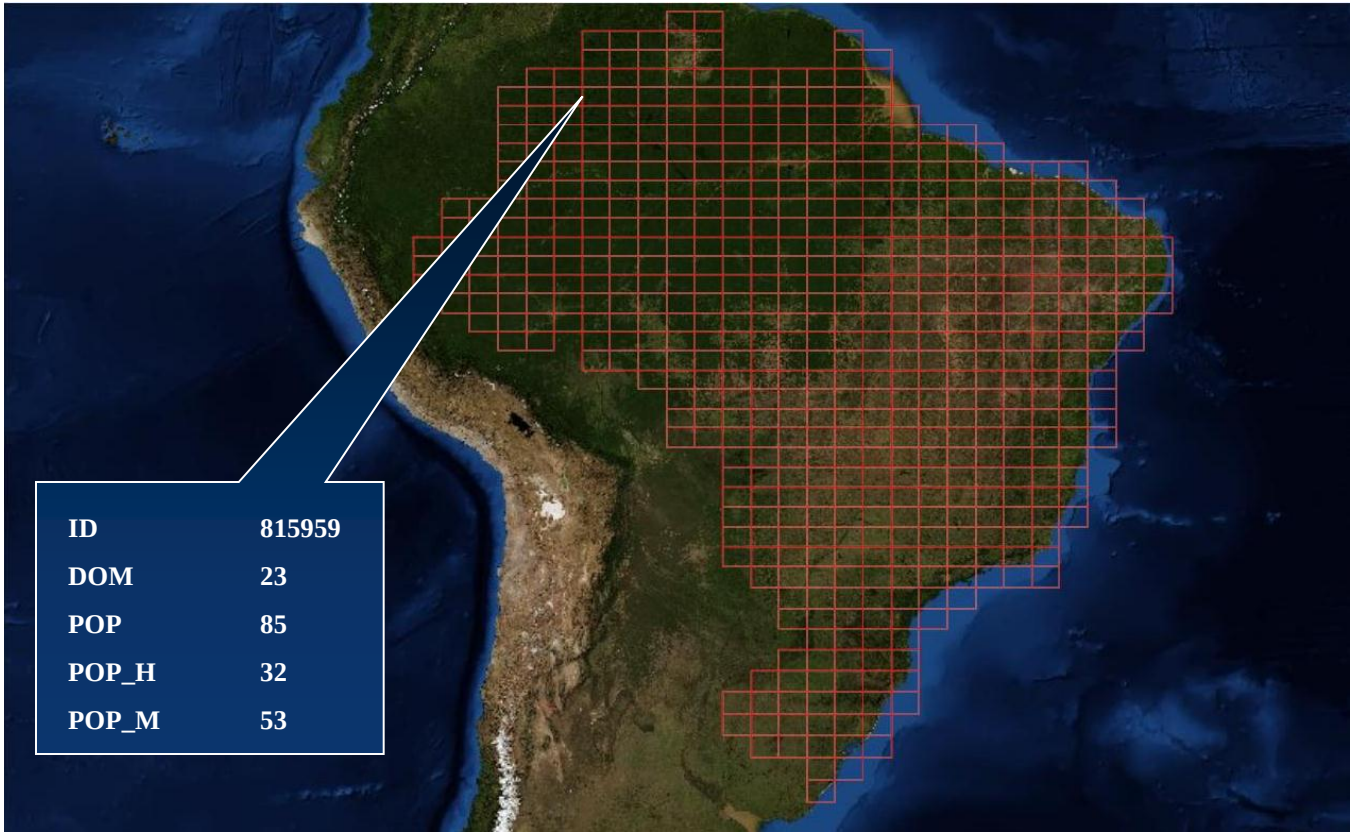
Business Cycle Dashboard February 2013



Spatial data Dissemination

Statistical Grids:

- hierarchical spatial structures formed by regular cells and used to make aggregated data available

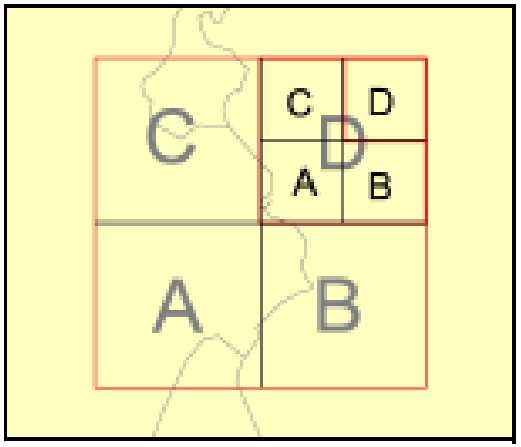


Source: IBGE

Statistical Grid as Basis

Statistical Grids:

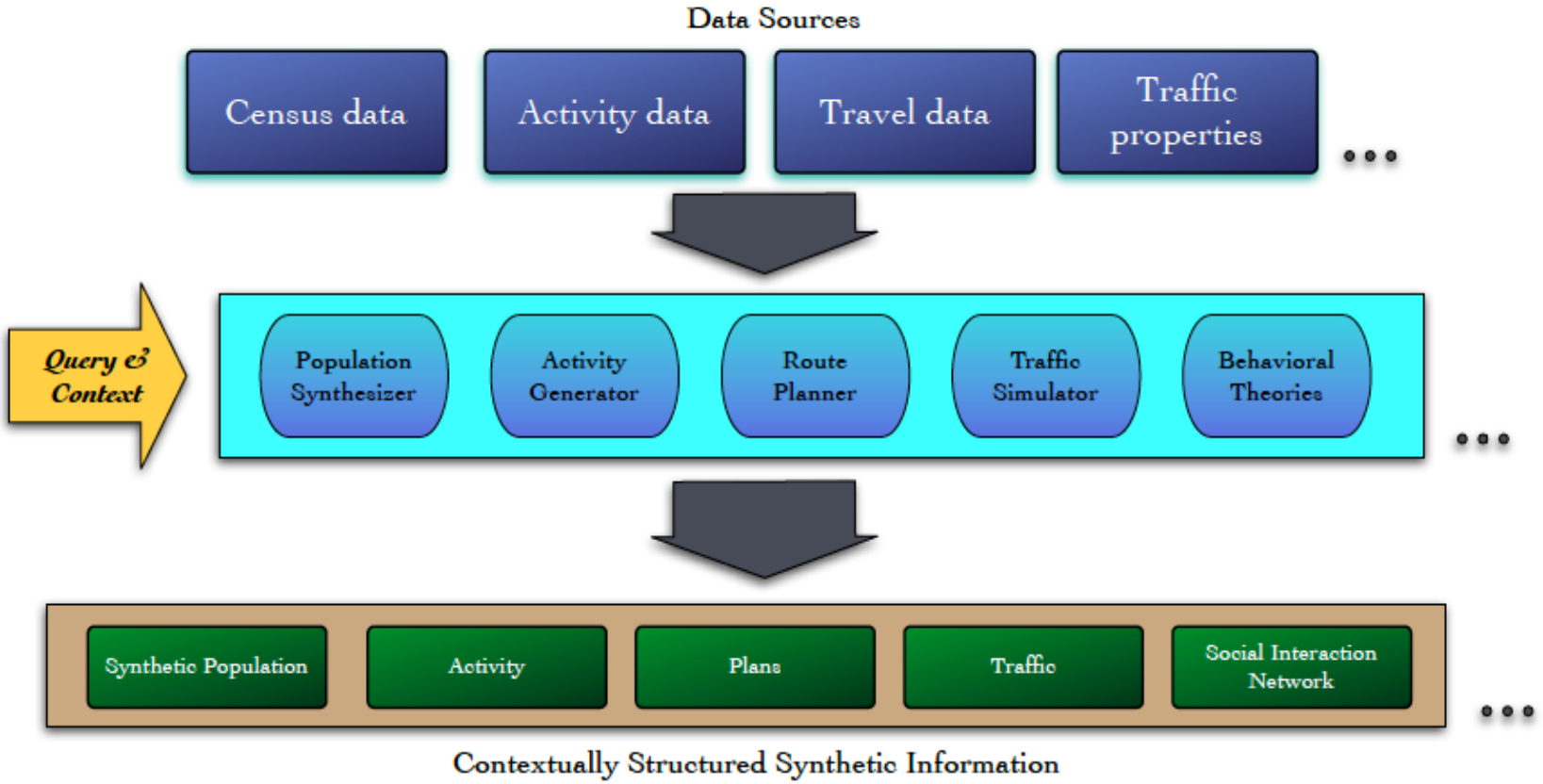
- independence from political-administrative boundaries
  → direct comparability
- no change over time
  → direct comparability
- regular distribution
  → computing efficiency
- hierarchical structure
  → allows multi-scale analyses
- easily handled with GIS tools
- vector or raster data structure



Source: IBGE

New Frontier: Dynamic Complex System Analysis

Figure 2: Schematic of the information integration process



Conclusions

- A new world of possibilities, with new product lines and improved productivity;
- Use of technology will intensify rapidly;
- Incorporate Data Analytics and Complex Systems in official statistics;
- Official statisticians must adopt new image;
- Be prepared for Rapid Future Changes.