The UN–GGIM challenge to develop integrated core datasets

Advantages and disadvantages of grid–based, population and administrative geography approaches

Tim Trainor, U.S. Census Bureau
Session 3: Progress on the UN EG-ISGI work program items.

Build Integrated Core Datasets to meet the UN GGIM challenge

**DESIGN**

National Statistical Agencies Survey Lifecycles
- Data collection, Tabulation, Dissemination, and Analysis programs

**EVALUATION**

Standards and Best Practices
- Management, accountability, monitoring

**PRODUCTION**

Statistical data
- Object attributes in time

Geospatial data
- Object features in space

**DISSEMINATION**

Use in Case Studies
- Create narratives and research networks

Integrated Datasets
- Produce for all levels of public authority from local to global

**ANALYSIS**

Public Access to Information
- Decision-makers use integrated core datasets for global sustainable initiatives

Tasks

- Identify technical, institutional, and information policy issues for 2020 round of censuses (*Status: in progress*).

- Expand on a bibliography of research and a terminology document on international geocoding practices (*Status: In review*).

- Evaluate further the pros and cons of the administrative approach and the grid approach for geocoding at UN-GGIM conferences (*Status: ongoing with the goal of a Best Practices document*).
PROS and CONS of Administrative Areas

PROS:
• Spatial accuracy of data
• Field verification
• Imagery verification
• Geocoding / address verification
• Authoritative sources
• Local government involvement
• Local knowledge
• Nesting relationship w/other geographic areas
• Cadastral boundaries
• Data thresholds
• Separate land & water area
• Response rates
• Response options
• Response quality
• Sample frame
• Controls on disclosure

CONS:
• Comparability
• Boundary changes
• Traditional census data collection is becoming more infrequent for countries
• Use of non-visible boundaries
• Number of different geographic areas
• Cartographic considerations / generalization of boundaries
• Varying participation
• Irregular sizes
• Irregular shapes
• Variable density measures
• High costs to maintain the data
• Legal variation
• Regional variation
• Topographic variation
• Insufficient understanding of micro characteristics inside macro-scale units
• Data integration is difficult
PROS and CONS of Grid-based Statistical Areas

**PROS:**
- Global and local scope—fully scalable
- Uniform scale conducive to cross-border studies
- Comparability; better suited for Spatial Data Infrastructures (SDI)
- More attention to problem-oriented science
- Can locate people in space with more precision
- Good territorial framework for sampling
- Can aggregate to different kinds of territorial units
- Ready to use with GIS analysis
- Easily generated from point-based georeferenced data
- Able to see clusters
- Easy and cost-efficient to collect
- Micro-scale analysis using flexible size grids
- Data integration is possible with newer data sources, (i.e. ground-based, imagery, internet)
- Stable over time; time-series not affected by admin unit changes
- Independent from traditional data collection procedures
- Widely used in science and practice

**CONS:**
- Disclosure control /cell size
- Grid cell sizes in rural areas
- When merging datasets, there is a need to change from one coordinate system to another before the data compilation into grids
- European terrestrial reference system (ETRS80) is based on Lambert Azimuthal Equal Area coordinate reference system with fixed projection center; different projections may be needed in other parts of the world
- Coding systems [scale intervals vs quadtree solutions
- Due to high data volume, errors are difficult to find and correct
- Various grids may be adopted within regions or countries
- Areas with dynamic or transient population fluctuations pose numerous complications for regional analysis
- Spatial and temporal cross-validation models using multiple sources of geographic, physiographic, and socio-economic data in conjunction with imagery analysis is necessary
UN GGIM is Looking forward

- Create a international geospatial statistical framework.
- Publish best practices in linking geospatial information to statistics through the research and efforts of the EG-ISGI.
- Assess the value of small geographic administrative areas vs statistical grids, or the combination of the two approaches.
- Accomplish goals and targets for local, regional, and global sustainable development initiatives through global geospatial information management, data monitoring, and accountability.
Grid-based data (global to local) vs Administrative

10,000 km window/100 km grids (Global scale)

1,000 km window/10 km grids (International regions)

100 km window/1 km grids (National regions)

10 km window/100m grids (Urban Districts)

1 km window/10 m grids (Urban neighborhoods)

100 m window/1 m grids (Urban blocks)

United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov
The solution is perhaps not an either/or. Can we combine administrative areas with grids to create an international geospatial framework or infrastructure?

Case studies at the U.S. Census Bureau: the representation of population and agricultural systems.

- 1992 *Agricultural Atlas of the United States*
- Haiti Demobase, 2010
- *Population Distribution of the World* (data provided to LandScan/Oak Ridge National Lab)
U.S. Census Bureau International Map Viewer

2015 population data for countries and territories
U.S. Census Bureau
International Map Viewer

Data layers now online:
• Total population
• Growth rate
• Life expectancy
• Infant mortality rate

Future: More variables

www.census.gov/population/international/data/intlmapviewer.html
References:

Backer, Lars H. “The Geostat Hypotheses; In search of a high resolution foundation for Geostatistics, an interactive response to operational user needs in a time of transition. European Forum for Geography and Statistics (EFGS).


