Definition of National Census Geography

- Amor Laaribi
- UN-GGIM Secretariat
- UN Statistics Division
- New York

Positioning geospatial information to address global challenges

UN-GGIM | United Nations Initiative on
Global Geospatial Information Management

ggim.un.org

1

Overview

- ❑ National Census geography
- ❑ Criteria for EA delineation
- ❑ Coding scheme
- ❑ Geocoding
- ❑ Conclusion

National Census Geography

- ❑ Consists of an administrative hierarchy

- ❑ Hierarchy made up of **administrative areas**

- ❑ Normally includes statistical reporting units

Administrative Hierarchy

❑ Every country has its own specific administrative hierarchy

❑ Definition:

A system by which the country and each lower level set of administrative units (except the lowest) are subdivided to form the next lower level.

❑ Administrative areas for which census data will be reported

Example of Administrative Hierarchy:

❑ for the purposes of the census a country may have been divided into seven hierarchical levels in urban areas and six in rural areas
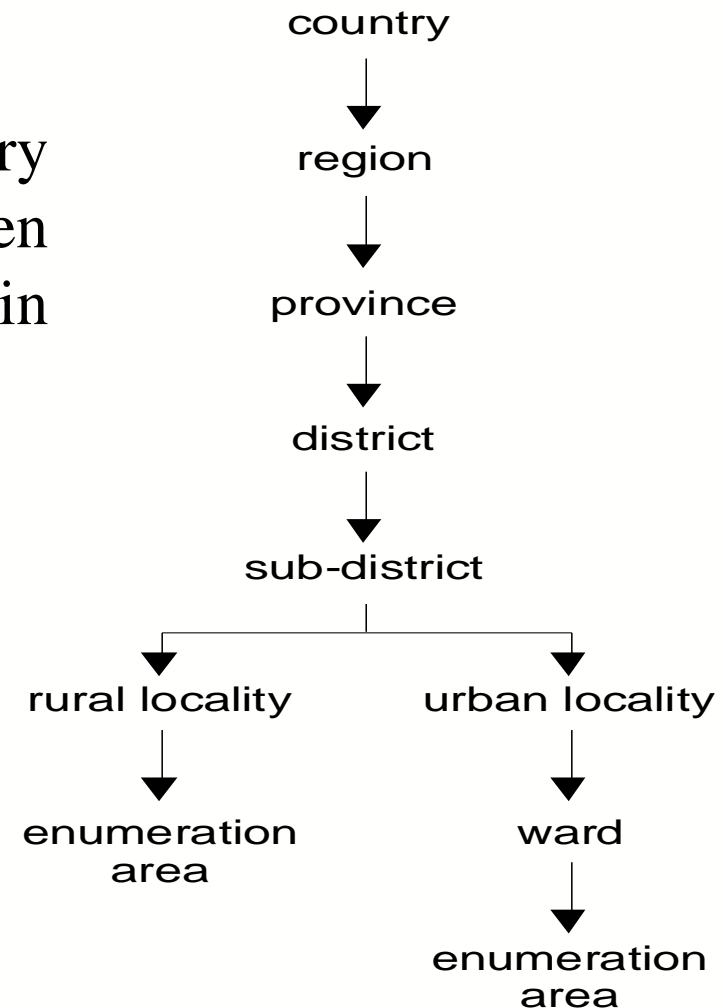
Generic census geographic hierarchy ➡

```
country
   ↓
region
   ↓
province
   ↓
district
   ↓
sub-district
   ↓
   ┌──────────────┴──────────────┐
   ↓                              ↓
rural locality              urban locality
   ↓                              ↓
enumeration                     ward
   area                          ↓
                            enumeration
                               area
```

Illustration of a Nested Admin. Hierarchy

❏ Relationships among all types of administrative and reporting unit boundaries are defined.

❏ hierarchical levels may have actual administrative roles

❏ Other units may have statistical roles alone

Provinces

Districts

Localities
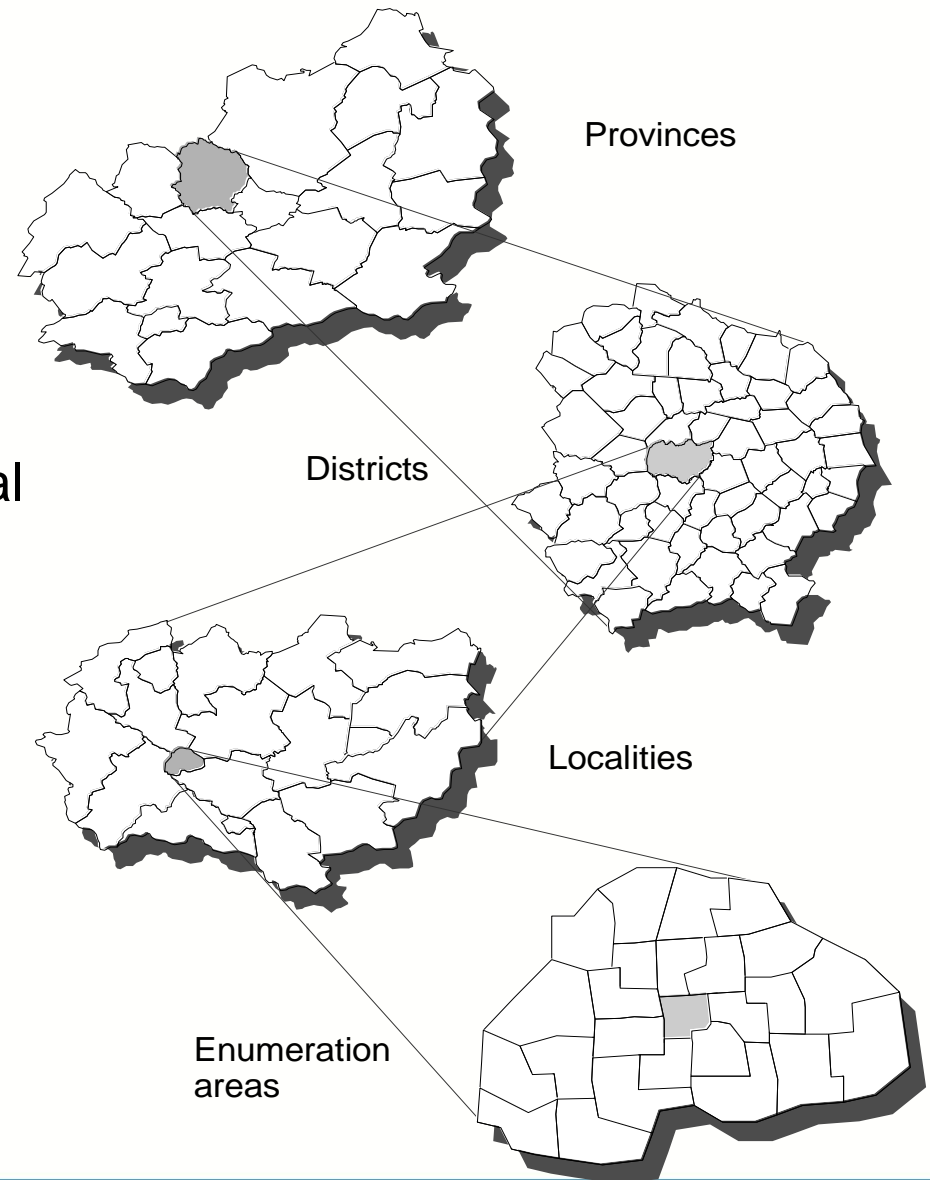
Enumeration areas

Enumeration Area (EA)

❑ EAs are the operational geographic units for the collection of census data

❑ Defined early in the census process.

❑ Some facilitate census data collection while others pertain to the usefulness of EAs in producing output products
  ▪ i.e., the relationship between data collection and tabulation units.

❑ Delineation of EAs is similar whether manual or digital cartographic techniques are used.

Criteria for delineation of EAs

❑ **Critical Factors of EAs:**

- ▪ Be mutually exclusive and exhaustive;

- ▪ Have easily identifiable boundaries on the ground;

- ▪ Be consistent with the administrative hierarchy;

- ▪ Be of approximately equally sized population;

- ▪ Be small enough and accessible to be covered by an enumerator;

- ▪ Be large enough to guarantee data privacy;

- ▪ Be useful for other types of data collection activities.

Criteria and process for ground delineation of EAs (cont.)

❑ Size of EAs can be defined in 2 ways:

- **1. Population** (most important in terms of censuses)

  - ❑ The chosen population size varies from country to country and is determined based generally on pretest results.

  - ❑ For most practical purposes, the population size of an enumeration area will be in the low to mid hundreds.

- **2. Surface Area and Size**
  - Enumeration can proceed more quickly in towns and cities than in the countryside so accessibility and time require consideration

Criteria and process for ground delineation of EAs (cont.)

❑ **Pre-delineation of EA's requires:**

- Estimating the number of persons living in an area and their geographic distribution. (e.g. Count the housing units, determine the associated number of households and multiply by an average household size.)

  ❑ Commonly done from information from a recent survey or registration system

❑ **The number of housing units can be determined through**

- cartographic field work,
- cooperation with government officials,
- extrapolation from previous census results,
- aerial photographs or satellite imagery

Criteria and process for ground delineation of EAs (cont.)

❑ **In the Delineation of EA boundaries:**

- ▪ Boundaries need to be clearly observable and easily identifiable on the ground

- ▪ Enumerators need to be able to find the boundaries of the area for which they are responsible.

  - ❑ Natural features that can be used for this purpose are
    - ▪ Roads;
    - ▪ Railroads;
    - ▪ Lakes, creeks and rivers;
    - ▪ Fences, or any other feature that defines a sharp boundary.

Criteria and process for ground delineation of EAs (cont.)

❑ In some instances, it is unavoidable to use EA boundaries that are not clearly visible on the ground.

  ▪ In this case, an exact textual description and appropriate annotation on the EA maps is required.

    ☐ Examples are offset lines and extended lines.

UNSD-CELADE
Regional Workshop on
Census Cartography for

UN-GGIM

United Nations Initiative on
Global Geospatial Information Management

Positioning geospatial information to address global challenges

ggim.un.org

Criteria and process for ground delineation of EAs (cont.)

❑ Criteria in determining ideal EA size:

- ▪ the number of people one enumerator can count in the time period scheduled for data collection.
  - ☐ census pretest can determine the number of housing units (HUs) that an enumerator can cover per day.
  - ☐ Local officials can be called upon to provide small-area estimates, or the areas in question can be visited by NSO field personnel.

- ▪ In areas that have not experienced dramatic change, estimates can be adjusted from the previous census based on the time elapsed.

Delineation of supervisory (crew leader) areas

❑ Supervisory areas provide the means for crews of enumerators to be effectively managed

❑ The design of supervisory is usually straightforward

❑ Supervisory areas consist of groups of usually eight to twelve contiguous EAs which share some of the same characteristics as enumeration areas

❑ The EAs assigned to the same supervisory area must be compact to minimize travel times and of approximately equal size

❑ They should be included in the same field office area, which usually is defined according to administrative units

UNSD-CELADE
Regional Workshop on
Census Cartography for

UN-GGIM

United Nations Initiative on
Global Geospatial Information Management

Positioning geospatial information to address global challenges

ggim.un.org

Coding Scheme

❑ A unique numeric code needs to be assigned to each enumeration area or administrative unit.

❑ This is the numeric code that provides the link between the aggregated census data and the digital EA boundary database stored in a GIS.

❑ Ideally the coding scheme needs to be determined on a country by country basis.
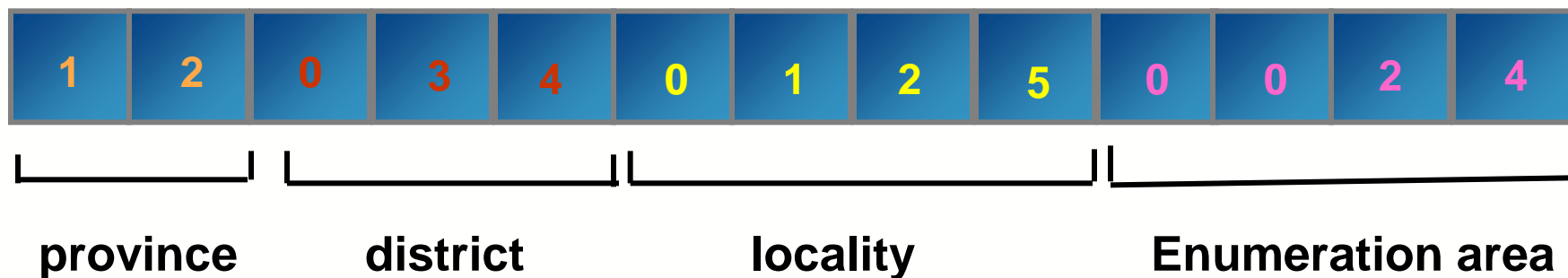
Coding Scheme (cont.)

❑ However, the rules used to assign codes need to be unambiguous and should be designed in collaboration within the NSO, especially between the mapping/GIS unit and those managing the data core.

❑ The most important principles when designing a coding scheme are

  ▪ **flexibility**,

  ▪ **expandability**,

  ▪ **intuitiveness** and

  ▪ **compatibility** with other coding schemes in use in the country.

UNSD-CELADE
Regional Workshop on
Census Cartography for

UN-GGIM
United Nations Initiative on
Global Geospatial Information Management

Positioning geospatial information to address global challenges

ggim.un.org

Example of a generic enumeration area coding scheme

❏ Geographic units are numbered at each level of the administrative hierarchy—usually leaving gaps between the numbers to allow for future insertion of newly created zones at that level.

| 1 | 2 | 0 | 3 | 4 | 0 | 1 | 2 | 5 | 0 | 0 | 2 | 4 |

province   **district**   **locality**   **Enumeration area**

Example of US Coding Scheme

250131402013

Digits 1-2 = State code

Digits 3-5 = County Code

Digits 6-11 = Census Tract Code

Digit 12 = Blockgroup code

State > County

County

Tract > Block Group
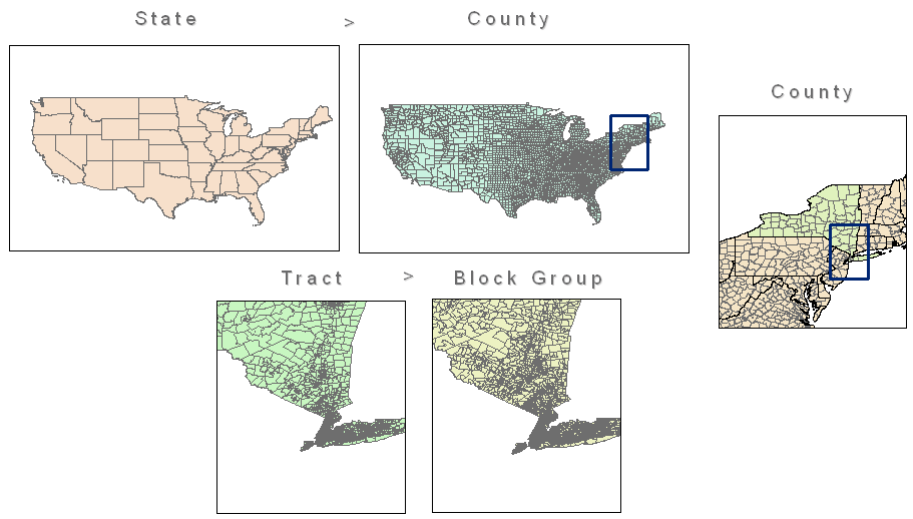
Coding scheme (cont.)

❑ In cases where administrative and reporting units are not hierarchical, special coding conventions need to be developed.

❑ Important to be completely consistent in defining and using the administrative unit identifiers, since they are the link between the GIS boundaries and the tabular census data.

❑ The census office should therefore maintain a master list of EA and administrative units and their respective codes and commit any changes made to the master list to the GIS and census databases.

Geocoding Concept

❏ The UN definition of "geocoding" is more broad, representing the connection between statistical observations and real-world locations expressed in terms of latitude and longitude or other locational attributes (in other words, geocoding is a way to ensure the data know where they are!)

❏ Other: Geocoding is the process of finding associated geographic coordinates (often expressed as latitude and longitude) from other geographic data, such as street addresses, or postal codes.

Geocoding…

❑ Geocoding (Geospatially enabling unit records) involves taking location information for these statistical units (such as address) and linking this information to a location coordinate (i.e. x, y, z coordinates) and/or a small geographic area.

❑ The geocodes, the location coordinates and geographic areas codes, obtained from this process can be stored directly on the statistical unit record or linked in some way to the record.

❑ The concept of geographic coding or "geocoding" goes beyond what is conventionally understood by the GIS industry to mean address-matching. It covers a continuum of spatial scales: from individual housing units to enumeration are level up to higher administrative or national levels.

Main Geocoding Methods

❑ In September/October 2012, 52 National Statistical Organizations completed a "Survey of Linking Geospatial Information to Statistics" as part of a United Nations Statistical Commission (UNSC) programme review. The Australian Bureau of Statistics (ABS) used information from this survey to write a programme review titled, "Developing a statistical-spatial framework in national statistical systems".

❑ Results:

Main methods of Geocoding

❑ **Enumeration geography**

- 31% of responding countries listed enumeration geography as their main geocoding method.

- Use of enumeration geography as the main geocoding method was dominated by countries from Asia, Africa and non-European Union countries. This may reflect the developing nature of their data collection and/or geocoding capabilities.

- Using this geocoding method, enumeration geography is linked to the unit record at the point of collection, via the workload assignment for the enumerator that this geography represents.

- The main issue with using enumeration geography for geocoding is that it forces enumeration geography and dissemination geography to be tied together; where any dissemination geography must be built out of the enumeration geography, limiting the possible dissemination uses.

Methods of geocoding…

❑ **Use of national registers**

  ▪ Nearly one-quarter (23%) of responding countries nominated the use of national registers as their main geocoding method. Registers were predominately used in European Union countries and included address, building or dwelling registers.

  ▪ Unit record information (such as, name and address, or register identification numbers) are matched or linked to a formal or statistical register which also includes location coordinates or small geographic area codes.

❑ **Address coding**

  ▪ A further 9% of responding countries use address coding as their geocoding method.

  ▪ Address coding is a related but less formal approach to using national registers, where a range of geospatially enabled address databases are used to obtain geocodes.

Geocoding Methods…

❑ Other geocoding methods:

- 8% of all responding countries are also using **block, locality and community coding**, and 6% are opting for direct capture using **GPS.**

❑ Additional geocoding methods

- For those countries who listed the additional geocoding practices, **address geocoding** was the most common at 11%. These countries all used address coding in conjunction with national registers as their main method of geocoding and were mainly European Union countries. It is possible that these countries use address coding to compensate for gaps in national register data.

- Countries are also using block, locality and community coding and direct capture using GPS as additional methods – each 8% of all responding countries. It seems that block, locality and community coding is being used in a wide variety of contexts, and for differing reasons.

- While use of GPS data capture appears to be being used in specific contexts or as a supplement to registers or address coding indexes.

Table below lists the main geocoding methods in order of popularity.

Main Geocoding Methods	Approximate % using this method	Comments
Enumeration geography	31%	This is a traditional approach that takes the geospatial reference from the geography that the data is collected within, but forces enumeration geography and dissemination geography to be tied together.
National registers	23%	Strong in Europe.
Address coding	9%	The major countries using this method are Australia, New Zealand, USA, UAE, Poland and Portugal.
Block, locality and community coding	8%	Stronger in the Americas and Asia, and similar to enumeration geography but generally the areas are larger.
Direct capture (via GPS)	6%	This method is being used in a number of developing countries throughout the world, where these countries may be leapfrogging older methods.

Table below lists the additional geocoding methods.

Additional Geocoding Methods	Approximate % using this method	Comments
Address geocoding	11%	Mainly European countries.
Block, locality and community coding	8%	Stronger in the Americas and Oceania
Direct capture (via GPS)	6%	Stronger in the Americas.
National registers	6%	Stronger in the Americas.

Conclusion

❑ Administrative areas for which census data will be reported: One of the earliest decisions in Census planning

❑ A census geographic hierarchy

❑ A coding scheme: flexible and expandible

❑ Geocoding: link with GIS

Thank You!