Preliminary report for the Global Forum on the Integration of Statistical and Geospatial Information

Global geographic classification and geocoding practices

Executive summary

This preliminary report was compiled by a working group of the UN Expert Group on the Integration of Statistical and Geospatial Information following the first meeting of this Expert Group. The objective was to investigate the various practices and approaches used to determine and represent geographical units for statistical purposes.

The following key points are noted in this report:

- The use of frameworks is limited to National Spatial Data Infrastructures and the INSPIRE directive requirements in Europe, and a limited number of country specific statistical spatial frameworks.
- Enumeration geography is the main method used to geocode statistical unit record data (i.e. data relating to individual persons, households, dwellings, businesses or buildings); however, this method is a very traditional method and can limit the usefulness of the data release by tying enumeration and dissemination geographies together.
- National registers are the next most popular method of geocoding, followed by the related address coding. These methods provide highly accurate and flexible geocodes by producing location coordinates and/or small area geographic codes.
- Direct capture using GPS, or similar technologies, in the field is being increasingly used, especially in developing countries, which seem to be leapfrogging older techniques.
- The majority of countries use regional government administrative boundaries as the primary geography to disseminate statistics. These meet key client needs but are subject to change, which can affect time series comparisons.
- Other geography types being used alongside these administrative geographies include: enumeration geographies – linked to the geocoding approach mentioned above; function-based statistical geographies – to define urban, rural and remote areas; postal geographies; and grid based geographies – growing in popularity in Europe as means of providing small area geography.

Background

In September/October 2012, 52 National Statistical Organisations 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". In 2013, the UNSC and the United Nations - Committee of Experts on Global Geospatial Information Management (UN-GGIM) both considered this programme review and accepted its recommendations to develop better linkages between geospatial and statistical communities, and develop a global statistical-geospatial framework based on the ABS Statistical Spatial Framework. Both UN bodies agreed to establish a UN Experts Group and conduct an international conference to pursue these aims.

The first meeting of the UN Expert Group on the Integration of Statistical and Geospatial Information was held from 30 October – 1 November 2013. At this first meeting, one of the key issues discussed was the various practices and approaches used to determine and represent geographical units for statistical purposes. Therefore, to assist with understanding this issue the Expert Group included the following item on its work program:

> *Develop a global consultation questionnaire on (a) geographical classifications and practices and (b) geocoding practices.*
>
> - *Build on the existing questionnaires from the global geospatial information management process to ensure coordination rather than duplication.*
> - *In time to present findings to the fourth session of the Committee of Experts (UN-GGIM) and the international conference.*

After reviewing the original responses to the 2012 survey it was decided that sufficient information was already provided to complete the consultation without having to resort to any duplication of effort with regards to gathering further information. The review of the 2012 survey information has been completed and this paper provides analysis of the responses to the 2012 survey relevant to geocoding practices, data linking and integration practices, and geographic classification practices. In addition to the 2012 survey information, limited follow up was conducted with specific countries who had already responded to obtain extra information particularly with regard to geocoding. Some additional responses and updated information was also obtained at the International Workshop on the Integration of Statistical and Geospatial Information, Beijing, 9-12 June 2014. This brought the total responding countries to 64.

Analysis of Questionnaire Responses

Introduction

The analysis below describes:

1. the geospatial frameworks that are in use;
2. geocoding practices that are used in geocoding statistical information;
3. And the types of geography that are used to disseminate statistics.

It should be noted that the statistics quoted in this report are only representative of the 64 responding countries, and cannot necessarily be taken to be representative for all countries or the countries within a region. That said, the information will provide a reasonable indication of the practices employed across the world.

Geospatial Frameworks

Where countries provided information about their geospatial frameworks, the main responses given were "Infrastructure for Spatial Information in the European Community (INSPIRE)" and "National Spatial Data Infrastructure (NSDI)". Responses indicating that INSPIRE was their geospatial framework were all from European Union (EU) countries, reflecting the role of INSPIRE within the EU. NSDI responses were more widely distributed across countries reflecting the more general implementation of NSDI's globally, particularly within developed countries.

This shows that until now the emphasis with regards to geospatial framework has been on meeting the requirements of geospatial data users and that there is a need for further development of frameworks that incorporate statistical uses.

Only Australia, with the Statistical Spatial Framework, and Mexico, with the National Geostatistical Framework, use frameworks that integrate both geospatial and statistical data uses.

Geocoding practices

For the purpose of this report, geocoding practices are those processes that are used to geospatially enable statistical unit record data (i.e. data relating to individual persons, households, dwellings, businesses or buildings). 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. This process is generally described as geocoding. 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.

Once the unit record data is geocoded it can then be used in geographic information systems (GIS), where geospatial operations can be performed on the data. These geospatial operations can occur during the statistical production process and in the creation of aggregate statistics for release.

<u>Main geocoding methods</u>

Close to one-third (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. It should be noted that a number of these countries appeared to be working towards improved geocoding abilities.

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. The design of

enumeration areas often conflicts with user requirements, and in some instances can result in additional design criteria for enumeration geography that reduces its effectiveness as a workload management tool.

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.

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.

Table 1 below lists the main geocoding methods in order of popularity.

Table 1 Main Geocoding Methods

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.

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 2 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)	8%	Stronger in the Americas.
National registers	6%	Stronger in the Americas.

Dissemination geography

Dissemination geographies are those areas that statistics are released for. Unit record data are allocated to these geographies based on the geocodes (or in some instances a proportional allocation across these geographies), and the data for these units are aggregated for each geographic area.

Main dissemination geography

For the majority of countries responding (84%), the primary geography for which they disseminate statistics is their existing regional government administrative boundaries. Disseminating statistics on these administrative boundaries directly meets the needs of a core set of key clients for statistical agencies: regional and local governments. However, these administrative boundaries may be subject to change and these changes will usually be driven by administrative requirements, rather than statistical considerations. These changes often cause complications for statistical time series comparisons. In some instances, these administrative geographies will not provide a small area geography that can be used for finer level geographic analysis or as building block for other geographies.

Table 3 below lists the main dissemination types in order of popularity.

Table 3 Main Dissemination Geography Types

Main Dissemination Geography Types	Approximate % using this type	Comments
Administrative geography – regional government	84%	Administrative geography based on regional government.
Enumeration/Census geography	6%	Countries such as Mexico, Swaziland, United Arab Emirates, and Oman use this type.
Statistical geography – functionally based with population criteria.	3%	Only used by Australia and New Zealand as the main dissemination geography (other countries reported these as additional dissemination geographies, see below).

Additional dissemination geographies

Most countries release statistics on more than one type of geography and the reported mix of geographies in use were very diverse. The majority of countries that nominated additional dissemination geography types listed enumeration geography (representing 28% of all countries). This reflects the important and fundamental role that enumeration geographies have played in census collection and dissemination activities, and the degree to which they have become incorporated into statistical analysis and decision making processes.

Functional based statistical geography (i.e. rural/urban/remote categorisation) was also common at 25%. The importance of rural and land use analysis is reflected in the common use of this type of geography.

The use of postal geography in 19% of responding countries perhaps reflects an attempt to satisfy the need for small area geography in these countries; though it may also reflect an attempt to supply data on a geography that many commercial organisations use to geospatially enable their own data from. The recent program in Europe of exploring the use of grids to meet small areas statistical needs is reflected in 17% of responding countries using this type of geography. Grids are also used in Columbia and Nigeria.

Table 4 Other Dissemination Geography Types

Other Dissemination Geography Types	Approximate % using this type	Comments
Enumeration/Census geography	28%	Used across many continents, most popular in Asia.
Statistical geography – functional based (rural/urban/remote)	25%	Used across many continents.
Administrative geography – postal	19%	Used across many continents.
Grid (raster) Cells	17%	Strong in Europe.
Electoral geography	16%	Used across many continents.
Statistical geography - population based	16%	Used across many continents.

Attachment - ***UN EG-ISGI consultation on geocoding practices, linking and integration practices and geographic classification practices – detailed results***

Please update or correct results in tables below using form included.

	Armenia	Austria	Australia	Bangladesh	Barbados	Belarus	Botswana	Cambodia	Cape Verde	Chile	Columbia	Croatia	Cyprus	Czech Republic
Dissemination Geography/Region Types														
Administrative Geography - regional government (sub national eg. state, provence)	MAIN	MAIN	OTHER	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN
Administrative Geography - service delivery (eg. Hospital regions, school zones, etc)			OTHER	OTHER							OTHER			
Administrative Geography - postal			OTHER								OTHER		OTHER	
Statistical Geography - population			MAIN	OTHER		OTHER		OTHER			OTHER			
Statistical Geography -				OTHER		OTHER		OTHER			OTHER			
Statistical Geography - functional (eg. rural/urban, remoteness, etc.)		OTHER	OTHER	OTHER		OTHER		OTHER						
Grid (Raster) Cells		OTHER									OTHER			
Electoral Geography		OTHER	OTHER	OTHER				OTHER						
Environmental Geography			OTHER	OTHER							OTHER			
Enumeration/Census Geography				OTHER				OTHER			OTHER	OTHER	OTHER	
Other (give details)														
Geocoding Method														
Direct capture in the field (e.g. GPS)										MAIN	OTHER			
Address Geocoding (matching reported address to an address index)		OTHER	MAIN											OTHER
National Register - widely used Address/Building/Property Register (matching records to a register using a unique identifier)		MAIN						OTHER	MAIN		OTHER			MAIN
Block/Locality/Community (coding a respondent to a small geographic			OTHER					OTHER		OTHER	OTHER			
Region (coding a respondent to a larger geographic region eg. town)						MAIN								
Enumeration Geography (geography based on an enumeration area, eg. Census tract)			OTHER	MAIN			MAIN	MAIN			MAIN			
Reported area - postal (respondent reported postal area)			OTHER											
Reported area - other (respondent reported geography e.g. administrative geography)								OTHER						
Other (give details)														

	Ecuador	Egypt	Estonia	Finland	Fiji	France	Germany	Ghana	Hungary	Indonesia	Iran	Israel	Italy	Japan
Dissemination Geography/Region Types														
Administrative Geography - regional government (sub national eg. state, provence)	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN
Administrative Geography - service delivery (eg. Hospital regions, school zones, etc)		OTHER												
Administrative Geography - postal				OTHER			OTHER							
Statistical Geography - population														
Statistical Geography -														
Statistical Geography - functional (eg. rural/urban, remoteness, etc.)	OTHER	OTHER			OTHER									
Grid (Raster) Cells			OTHER	OTHER		OTHER								
Electoral Geography		OTHER		OTHER								OTHER		
Environmental Geography												OTHER		
Enumeration/Census Geography		OTHER						OTHER						OTHER
Other (give details)														
Geocoding Method														
Direct capture in the field (e.g. GPS)					MAIN									
Address Geocoding (matching reported address to an address index)			OTHER	OTHER		OTHER								
National Register - widely used Address/Building/Property Register (matching records to a register using a unique identifier)			MAIN	MAIN		MAIN			MAIN			MAIN		
Block/Locality/Community (coding a respondent to a small geographic	MAIN										MAIN			
Region (coding a respondent to a larger geographic region eg. town)														
Enumeration Geography (geography based on an enumeration area, eg. Census tract)	OTHER							MAIN		MAIN			MAIN	MAIN
Reported area - postal (respondent reported postal area)														
Reported area - other (respondent reported geography e.g. administrative geography)														
Other (give details)														

	Jordan	Lithuania	Kenya	Kuwait	Madagascar	Malaysia	Mexico	Moldova	Mongolia	Morocco	Mozambique	Namibia	Netherlands	New Zealand
Dissemination Geography/Region Types														
Administrative Geography - regional government (sub national eg. state, provence)	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN	OTHER	MAIN	MAIN	MAIN	MAIN	MAIN	MAIN	OTHER
Administrative Geography - service delivery (eg. Hospital regions, school zones, etc)							OTHER		OTHER	OTHER				
Administrative Geography - postal							OTHER		OTHER		OTHER			
Statistical Geography - population						OTHER	OTHER		OTHER	OTHER				MAIN
Statistical Geography -							OTHER		OTHER	OTHER				
Statistical Geography - functional (eg. rural/urban, remoteness, etc.)			OTHER		OTHER				OTHER	OTHER				OTHER
Grid (Raster) Cells													OTHER	
Electoral Geography			OTHER											
Environmental Geography									OTHER					
Enumeration/Census Geography							MAIN	OTHER		OTHER		OTHER		
Other (give details)														
Geocoding Method														
Direct capture in the field (e.g. GPS)							OTHER		OTHER	MAIN				
Address Geocoding (matching reported address to an address index)							OTHER		OTHER					MAIN
National Register - widely used Address/Building/Property Register (matching records to a register using a unique identifier)		MAIN		MAIN					MAIN				MAIN	
Block/Locality/Community (coding a respondent to a small geographic								OTHER						
Region (coding a respondent to a larger geographic region eg. town)														
Enumeration Geography (geography based on an enumeration area, eg. Census tract)			MAIN			MAIN	MAIN	MAIN		MAIN				
Reported area - postal (respondent reported postal area)							OTHER							
Reported area - other (respondent reported geography e.g. administrative geography)							OTHER							
Other (give details)														

	Nigeria	Oman	Palestine	Phillipines	Poland	Portugal	Republic of Moldova	Romania	Serbia	Seychelles	Singapore	Slovakia	Slovenia
Dissemination Geography/Region Types													
Administrative Geography - regional government (sub national eg. state, provence)	MAIN		MAIN	MAIN	MAIN	MAIN			MAIN	MAIN	OTHER	MAIN	MAIN
Administrative Geography - service delivery (eg. Hospital regions, school zones, etc)	OTHER												
Administrative Geography - postal	OTHER										MAIN		
Statistical Geography - population	OTHER				OTHER						OTHER		
Statistical Geography -											OTHER		
Statistical Geography - functional (eg. rural/urban, remoteness, etc.)	OTHER					OTHER							
Grid (Raster) Cells	OTHER				OTHER	OTHER							OTHER
Electoral Geography	OTHER									OTHER	OTHER		
Environmental Geography	OTHER												
Enumeration/Census Geography	OTHER	MAIN	OTHER	OTHER	OTHER						OTHER	OTHER	
Other (give details)													
Geocoding Method													
Direct capture in the field (e.g. GPS)	OTHER	MAIN			OTHER								
Address Geocoding (matching reported address to an address index)					MAIN	MAIN							
National Register - widely used Address/Building/Property Register (matching records to a register using a unique identifier)	OTHER				OTHER						MAIN		MAIN
Block/Locality/Community (coding a respondent to a small geographic													
Region (coding a respondent to a larger geographic region eg. town)	OTHER												
Enumeration Geography (geography based on an enumeration area, eg. Census tract)	MAIN			MAIN						MAIN			
Reported area - postal (respondent reported postal area)													
Reported area - other (respondent reported geography e.g. administrative geography)													
Other (give details)													

	Sri Lanka	Swaziland	Sweden	Switzerland	Tunisia	Turkey	Ukraine	UAE	US
Dissemination Geography/Region Types									
Administrative Geography - regional government (sub national eg. state, provence)	MAIN	OTHER	MAIN	OTHER	MAIN	MAIN	MAIN	MAIN	MAIN
Administrative Geography - service delivery (eg. Hospital regions, school zones, etc)									
Administrative Geography - postal			OTHER						OTHER
Statistical Geography - population									
Statistical Geography -									
Statistical Geography - functional (eg. rural/urban, remoteness, etc.)									OTHER
Grid (Raster) Cells				MAIN		OTHER			
Electoral Geography									
Environmental Geography									
Enumeration/Census Geography	OTHER	MAIN			OTHER			OTHER	
Other (give details)									
Geocoding Method									
Direct capture in the field (e.g. GPS)									
Address Geocoding (matching reported address to an address index)								MAIN	MAIN
National Register - widely used Address/Building/Property Register (matching records to a register using a unique identifier)			MAIN	MAIN					
Block/Locality/Community (coding a respondent to a small geographic					MAIN				
Region (coding a respondent to a larger geographic region eg. town)									
Enumeration Geography (geography based on an enumeration area, eg. Census tract)	MAIN				OTHER	MAIN	MAIN		
Reported area - postal (respondent reported postal area)									
Reported area - other (respondent reported geography e.g. administrative geography)									
Other (give details)									

UN EG-ISGI consultation on geocoding practices, linking and integration practices and geographic classification practices - QUESTIONNAIRE

Please provide your contact details:

Country:

Name:

Organisation:

Email address:

What geospatial information do you attach to your unit record data?	MAIN Please tick one	OTHER Tick all used
Administrative Geography - regional government (sub national eg. state, provence)		
Administrative Geography - service delivery (eg. Hospital regions, school zones, etc)		
Administrative Geography - postal		
Statistical Geography - population based		
Statistical Geography - socioeconomic		
Statistical Geography - functional (eg. rural/urban, remoteness, etc.)		
Grid (Raster) Cells		
Electoral Geography		
Environmental Geography		
Enumeration/Census Geography		
Other (give details)		

What approach do you use to geocode your unit record data?	MAIN Please tick one	OTHER Tick all used
Direct capture in the field (e.g. GPS)		
Address Geocoding (matching reported address to an address index)		
National Register - widely used Address/Building/Property Register (matching records to a register using a unique identifier)		
Block/Locality/Community (coding a respondent to a small geographic region)		
Region (coding a respondent to a larger geographic region eg. town)		
Enumeration Geography (geography based on an enumeration area, eg. Census tract)		
Reported area - postal (respondent reported postal area)		
Reported area - other (respondent reported geography e.g. administrative geography)		
Other (give details)		

Please forward to: Geography@abs.gov.au