Application of GIS in the 2011 Australian Census of Population and Housing and Plans for 2016

Alister D Nairn
Director of Geography, Australian Bureau of Statistics, Canberra, Australia
alister.nairn@abs.gov.au

Abstract

Conducting a Census is inherently a spatial undertaking and Geographical Information Systems (GIS) provide an essential capability in their planning, operations and information dissemination. A number of innovations were introduced for the 2011 Census, including using GIS and new sources of spatial data to improve the design of enumeration areas and output geographies, utilising geocoding to geographically locate dwellings, and the additional use of online mapping in the dissemination of census information for the wide range of geographies required by users of Census information. Planning for the 2016 Census is well underway and the uses of additional GIS capabilities are being tested. These capabilities include the use of GPS and mobile mapping applications to improve enumeration, and the development and use of an address register to facilitate a mail out for large parts of the country. Network analysis is also being tested to improve allocation of the field workforce. This presentation will cover the key areas where GIS has been utilised in 2011 and the additional capabilities being considered for use in 2016.

Key Words: Spatial data, Geographical Information Systems

1. Introduction

The Australian Bureau of Statistics (ABS) has recently completed the 2011 Census of Population and Housing. Geographic Information Systems (GIS) and geospatial data were utilised across all phases of the project from the design of collection and output areas, management reporting of progress with eCensus take up, data processing and geocoding, and ultimately in the information dissemination systems.

The effective use of GIS relies heavily on the availability of geospatial base-map data at the national level, and at a scale of sufficient resolution to identify individual parcels of land. This paper discusses the data sources used and the geospatial software utilised, together with the specific systems that have been developed.

Additional functionality is being developed for use in the 2016 Census, particularly associated with the move to developing an address register to mail out instructions to access the eCensus form to around sixty percent of the population. Network analysis has been tested and implemented in some parts of ABS' business areas to improve allocation of the field workforce and the use of GPS and mobile mapping applications are also being investigated to assist with enumeration.

2. Loading and storage of GIS data

Source datasets

ABS receives data quarterly from our supplier, PSMA Australia, for use in mapping and boundary design. PSMA Australia is responsible for aggregating topographic information sourced from Australia's state and territory governments. The PSMA

datasets that ABS uses include: Administrative Boundaries, Transport and Topography, Cadastre and a Geocoded National Address File (G-NAF).

Software used in load process and storage
Safe Software's FME (Feature Manipulation Engine), Python and Oracle SQL Developer are used to load data to the Master Spatial Database (MSD), an Oracle 11g database. The data is then available for use in MapInfo and ArcGIS and other corporate applications.

SQL Developer and SPARX System's Enterprise Architect are used to maintain the Master Spatial Database.

All the spatial data is stored within the Master Spatial Database. However, the base-map data from PSMA is stored separately to the ABS statistical boundaries. These boundaries are described in the Australian Statistical Geography Standard (ASGS), Australian Bureau of Statistics (2010).

Main processes / checks undertaken
The data received from PSMA are a set of highly normalised tables. These tables are connected together to form flat tables that contain the attributes required by the users. The flat tables are accessed by users through a set of views. Views are used to maintain a consistent naming convention and table structure. The views can also be quickly pointed to for each new version of the data with minimal disruption to users.

During the load process some geometry in the data is modified, such as joining parcels, or road segments. When the data is supplied (quarterly), quality checking is undertaken to identify any changes in the data model.

The data model
The introduction of the Australian Statistical Geography Standard (ASGS), has required the redesign of the data model for storing the boundary definitions and coding classifications. This database forms the single point of truth for the national statistical geography for other statistical processes within the ABS.

The philosophy of the ASGS is that all boundaries can be aggregated from a basic building block, the Mesh Block. The ASGS data model accommodates the 24 different boundaries and coding definitions, and also maintains the relationships between them.

The requirement for a "system agnostic" database
The MSD is a point of truth for the ASGS and PSMA data, and must be able to integrate with other statistical processing systems. For example, the ABS Address Register will have a dependency on the storage and maintenance of the G-NAF which is held within the MSD.

To enable this integration it is important that the database maintains a high degree of interoperability. To do this the data models involved have been designed to minimise the impact of proprietary data structures. For example, Esri software requires the use of proprietary data structures to support versioning and topology operations. These structures are presently implemented in separate systems and the inclusion of these processes within the MSD is an active area for research and development.

3. Design of geographic output areas

The smallest geographic area, from which all other statistical areas in the ASGS are built is the Mesh Block. A Mesh Block is designed to have either a zero population (eg

areas of water, parkland, industrial areas) or in populated areas contain 30 – 60 residential dwellings. There are 347,627 Mesh Blocks that entirely cover Australia.

Mesh Blocks were initially designed in 2005 and are now maintained using the 'Mesh Block Maintenance System' which was programmed in MapBasic and is a MapInfo Desktop Professional application. It is a complete system that allows the operator to identify changes such as areas with population growth and topographical alignment issues, and then undertake the necessary edits and checks.

A management layer is first downloaded with all the Mesh Blocks, together with the layers of topographic information required to make informed design decisions. The layers include the latest cadastral parcels, address points, roads, rivers, Statistical Areas Level 1 (SA1s), points of interest and administrative boundaries, to name a few.

If Mesh Blocks require editing they are amended using maintenance tools within the system. Changes to Mesh Blocks consist of alignment, nudges and deletions; however change will primarily be undertaken by splitting the Mesh Block. The maintenance system allows the operator to split a Mesh Block, retire the old Mesh Block code and generate new and unique Mesh Block codes. The system also has functionality to topologically validate the boundaries.

Once the Mesh Blocks are finalised, GIS systems are also used to build the other areas defined in the ASGS.

4. Design of census collection areas

The collection area design system was developed using MapBasic for MapInfo Desktop Professional to produce Census Collection Areas (CLWs) that met specific criteria using groups of whole Mesh Blocks. These criteria varied depending on whether the areas were urban or rural.

The system was based on designers initially downloading a management area. This included downloading Mesh Blocks, topographic layers and ABS derived Special Dwellings and Indigenous Locations layers. Designers would then select single or multiple Mesh Blocks and activate a “display workload estimates” function. This enabled the designer to see if the selection of Mesh Blocks met the estimated dwellings criteria, which in urban areas was 300 to 550 dwellings, and in rural areas was between 80 and 300 dwellings. If the criteria were met, then a CLW would be created. This process was repeated until every Mesh Block in the management area had been assigned to a CLW.

Once the management area was completed, the CLWs were reviewed by the relevant ABS Regional Office. After review by the Regional Office, the management areas underwent a quality assurance process prior to the printing of the collector maps.

5. Production of maps for collectors and field managers

A map production process was developed that automatically took the Census collection area boundary files, determined an appropriate scale for printing, added the topographic base information and applied an appropriate ruleset for cartographic presentation. The placement of text was automated using the Maplex software (now incorporated into Esri ArcGIS), and then the map base and text were combined into a PDF file. The software used included MapInfo, Maplex and Adobe Acrobat Professional and Distiller software.

Overall the process of map design and printing took approximately 6 months and involved the printing of approximately 60,000 maps consisting of Collector maps, Area Supervisor workload maps, District Manager maps and Quality Management Area maps. Google KML files were also produced for each collection area and loaded onto the ABS website. This enabled the areas to be viewed by field staff and supervisors using Google Earth.

For 2016 the system is being modified and simplified to use Esri ArcGIS software.

6. Management reporting of progress with enumeration

In 2011, experimental processes and systems were developed to automatically extract the eCensus take-up rate and pre-emptively make it available to all ABS staff. The goals of this experiment were to communicate eCensus management information to all interested ABS staff and to make use of interesting and impressive data visualisation methods and technology. The information was displayed on large screens at prominent places within the ABS building.

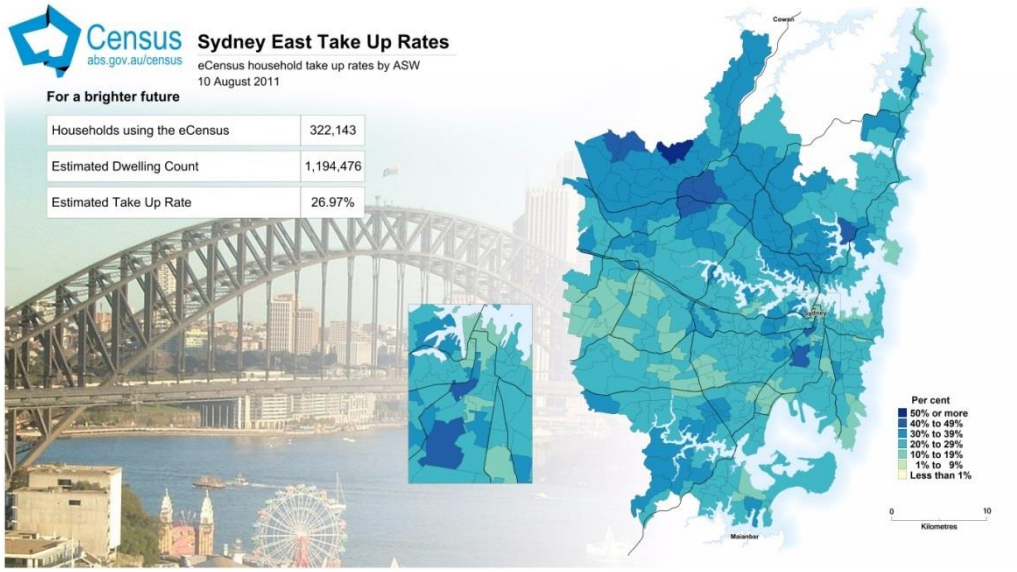

Figure 2: Map of eCensus take up rates

7. Geocoding of census data (from forms and e-census)

The 2011 Census was the first year that the collection areas were not the same as the output areas. This necessitated the accurate geocoding of all the address information supplied on the Census forms and introduced additional challenges to ensure the coding was accurate and completed in time. Table 1 provides approximate numbers of the addresses that required geocoding.

Type of Address	**2011**
Dwelling Address	9,300,000
Usual Address	1,000,000
1 year ago address	2,000,000
5 year ago address	9,000,000
Workplace address	10,000,000

Table 1: The approximate number of addresses requiring geocoding for 2011

Various techniques were used in the geocoding process, but at the core of these processes was the G-NAF file which contained official addresses for Australia together with a geographic coordinate. Address matching software was used which parsed the addresses and matched them to the G-NAF record. Where a match was not possible, manual coding was undertaken using interactive GIS based coders and other techniques, involving inspecting the collector record books and the marked up collector maps. In some cases Google Earth and Street View were also used as an additional source of intelligence.

8. Use of GIS in the census dissemination system

Data from the Census is widely used across business, government and communities to better understand the social characteristics of Australia and inform decision making. Over the last decade a range of advancements in the dissemination of data have been made with the emergence of geospatial tools changing the way we communicate data.

The ABS is committed to making Census data widely available and giving users the freedom to analyse and view data in any way they want. Census data is inherently geographic and giving users the ability to seamlessly navigate on a website through the use of geographic tools to select, or customise, areas of interest is critical to ensuring users can discover the data that meets their needs. The ABS census website includes a range of facilities that enable users to 'search' geographically. Users can select an area, aggregate areas to create custom shapes and search at an address level to discover the full range of pre-aggregated areas in which that address is located.

Increasingly users want the ability to view data thematically on maps and create custom shapes of geographies to tailor the data to meet their unique needs. The ABS uses a platform called TableBuilder to support these dissemination needs. TableBuilder is a web based platform that uses SpaceTime Research's SuperWEB and Esri ArcGIS Server to enable users to interact, customise and extract data in a range of formats. The contextual topographic base map is now accessed via a web service. TableBuilder gives users ultimate flexibility to interact with Census data by combining data items to create tabular outputs and then viewing these outputs thematically on a map interface. These tools are freely available on the ABS website and have opened up the use of Census data to a new audience.

9. Additional GIS capabilities being developed for 2016

Preparation is well underway for the 2016 Census. Although final decisions are still to be made, a new Census process based on a mixed mode of delivery is envisaged. The process will utilise an address register and mail out of Census materials for the more urbanised areas of the country, while traditional manual delivery will occur in the more difficult to enumerate areas. This change in methodology will also drive other changes in the use of GIS for the 2016 Census which are outlined below.

Development of an address register

An address register is being developed to support the mail out of materials to around sixty percent of all households. The address register will be based on G-NAF, which contains the most comprehensive list of official addresses in Australia. One significant issue, however, is that G-NAF contains the addresses of all properties, and not just the residential dwellings required to be contacted in a Census. One of the challenges in developing an address register suitable for supporting the Census, and in the longer term other household surveys, will be to identify a way of maintaining a land use code against the address. This code will need to differentiate between dwelling addresses

and other types of non-dwelling addresses, such as vacant sites, industrial sites, etc. Applying this land use classification, and ensuring the overall currency and accuracy of the register, will require a large body of quality assurance work. This is likely to entail field canvassing of many of the addresses, prior to using the register to construct a "Census Frame" for the 2016 Census.

Use of network analysis to improve allocation of field workforce

Esri's Network Analyst together with the NAVTEQ networked road dataset for Australia has been successfully utilised to calculate road distances between a matrix of addresses. The output from this is currently used in an optimisation process for the allocation of survey interviewers to interviewees. Based on the success of this trial, other applications are being developed and tested, including:

- developing a process for ordering addresses within an area into a logical list for field staff to "Check List". The list is sorted into order that minimises walking / driving times.
- optimising the creation of adaptive workloads as a basis for field checking non-responding dwellings during the Census enumeration process.
- optimising workload area design to minimise travel outside a workload area to complete enumeration.

Use of GPS and mobile mapping applications to improve enumeration

Providing enumerators with mobile electronic devices (tablets and or phones) has the potential to create significant advantages, such as:

- providing maps to assist with locating the correct dwellings.
- undertaking data validation and coding processes in the field.
- linking a statistical instrument (Census or survey form) to a location in the field at the point of entry using a GPS enabled mobile device.
- reducing costs associated with scanning and editing forms.
- facilitating the more efficient communication with and management of field staff.

Some testing of devices has already taken place which has identified many issues associated their use. The issues being evaluated include data security, reliability of devices, costs and occupational health and safety, to name a few.

10. Conclusions

Effective use of GIS within and organisation such as the Australian Bureau of Statistics relies heavily on the access to good base data, strong internal management of that data, flexible software tools and most importantly staff with geospatial capability to meet the changing needs of the organisation. The application of GIS used in the 2011 Census described in this paper successfully contributed to all aspects of the Census, and additional capabilities are being developed for 2016.

References

Australian Bureau of Statistics (2010) "*Australian Statistical Geography Standard*", http://www/abs.gov.au/geography, accessed April 2013.