

## Identifying the challenges of creating an optimal dissemination geography for census

Helene Verhoef<sup>1,2</sup> & Amanda van Eeden<sup>3</sup>

<sup>1</sup>MPhil Urban and Regional Science student: Centre for Regional and Urban Innovation and Statistical Exploration (CRUISE), Department of Geography and Environmental Studies, Stellenbosch University, Stellenbosch, <sup>2</sup>Currently employed: Manager Geographic Frames, Statistics South Africa (Stats SA), Pretoria, South Africa <sup>3</sup>Centre for Regional and Urban Innovation and Statistical Exploration (CRUISE), Department of Geography and Environmental Studies, Stellenbosch University, Stellenbosch \*corresponding author, [avaneeden@sun.ac.za](mailto:avaneeden@sun.ac.za)

DOI: <http://dx.doi.org/10.4314/sajg.v4i1.4>

### Abstract

*The importance of census data in government and private-sector planning cannot be underestimated. However, the geographic level at which it is made available for different users, is a highly debateable issue. It is crucial that census data is disseminated in such a way that it satisfies most user needs as far as possible, to ensure that there is optimum use of the information and that maximum value for money is provided. In the past, Statistics South Africa disseminated data at the same geographic level created for data collection. This causes problems for data users and calls for the creation of a separate output geography rather than using the original collection geography.*

*The research was done on two levels: first, an overview of output geographies, as well as examples of developed and successfully used tools to generate these areas within a geographical information system. Some of these could be used in the South African milieu. Secondly the paper discuss aspects such as the population size variation of EAs, in order to inform the criteria for the creation of the ideal small area (SA) layer to satisfy the majority of user needs. Lastly the paper describes briefly the challenges faced to create the 2011 output geography. The results indicate a strong resemblance between the two EA population size patterns of 2001 and 2011, influenced by the EA demarcation rules. The challenges identified in the process of creating the SAL as a census output geography need to be taken into consideration to enable a more useful and user-friendly output.*

### 1. Introduction and background

A census is the most extensive and, arguably, the most expensive source of socio-economic data for any country. The census data is essential for the government, private sector, non-governmental organisations (NGOs), national intelligence, emergency agencies, research facilities and planners because it is unique in its coverage, (in area as well as in number and the detail of variables) and it is important to receive the data at the smallest possible geographical level for micro-planning. Each

one of these census data users has different data needs and applications, and they each make use of the data on different geographic scales (Robertson, 1969; Torrieri, 2007; Young et al., 2009; Dugmore et al., 2011, Martin et al., 2013). It is therefore crucial that census data is disseminated in such a way that it satisfies most user needs as far as possible to ensure maximum use of the information and maximum value for money (UN Statistical Division, 2010; Beard *et al.*, 2011).

Although it is common for countries to disseminate census information on administrative levels such as provincial, regional, divisional or municipal, few countries acknowledge that high-level statistics are not ideal for localised planning and monitoring. The universal problem seems to be the creation of relevant information for smaller areas while simultaneously adhering to confidentiality limits (Martin, 1998a; 1998b).

Enumeration areas, census tracks, enumeration districts, and mesh blocks are all units of convenience for different countries, created to manage and execute enumeration during the census period. All have certain specifications aiming at the best possible coverage and logistics management. These specifications are enumerator-orientated and not necessarily output-orientated. Another factor that plays a role in the usability of geographical entities is that census data is usually aggregated into various dissemination geographies which also differ in size.

The South African census takes place within an administrative frame, i.e. the provincial; district and municipal boundaries are taken into consideration for the demarcation of the collection as well as output geographies. Data is currently disseminated on two categories of administrative areas namely standard geography levels and non-standard geography levels. In the case of the standard geography levels the census statistics are disseminated from national and provincial levels down to small areas (SAs) (Statistics South Africa 2007). However, for the areas known as non-standard geographies such as the magisterial districts, health areas and electoral wards, where these boundaries follow service specific guidelines, special dissemination criteria were developed. These boundaries are not taken into consideration for the demarcation of enumeration areas.

Administrative areas as standard geography levels are currently the most used geographic areas for census data dissemination, but are usually not suitable for small-scale, localised planning, as much of the detailed information gets lost in generalisations, especially if the area is not homogeneous (Robertson, 1969; Paez and Scott, 2004; Cockings et al., 2013; Martin et al. 2013). As in 2001, the Census 2011 statistics are not released to external users at the enumeration areas (EA) level due to confidentiality issues. Consequently, the smallest area for which census statistics are readily available is at the sub-place (SP) level unless an effort is made to create separate output areas that are smaller than the place name areas but equal or bigger than an EA, such as the small area layer (SAL). Figure 1 indicates the difference in the size of output areas between the SPs and SALs.

The ideal output geography should ensure that the majority of users can apply it for their data needs and this is currently not the case. The areas for which data is available are too large in terms of the total population as well as in physical size. The data also becomes too generalised to identify those specific portions of the community in need of certain services (Robertson, 1969; Vickers et al.,

2005; Martin, 2006). The same argument holds for research purposes, compounded with the added problem of sampling frames which need smaller homogeneous areas in order to be fully representative. The ideal output geography should therefore have entities with a physical area and population size as small as possible, as compact in shape as possible, as homogeneous in characteristics as possible, and fall within specified administrative boundaries.

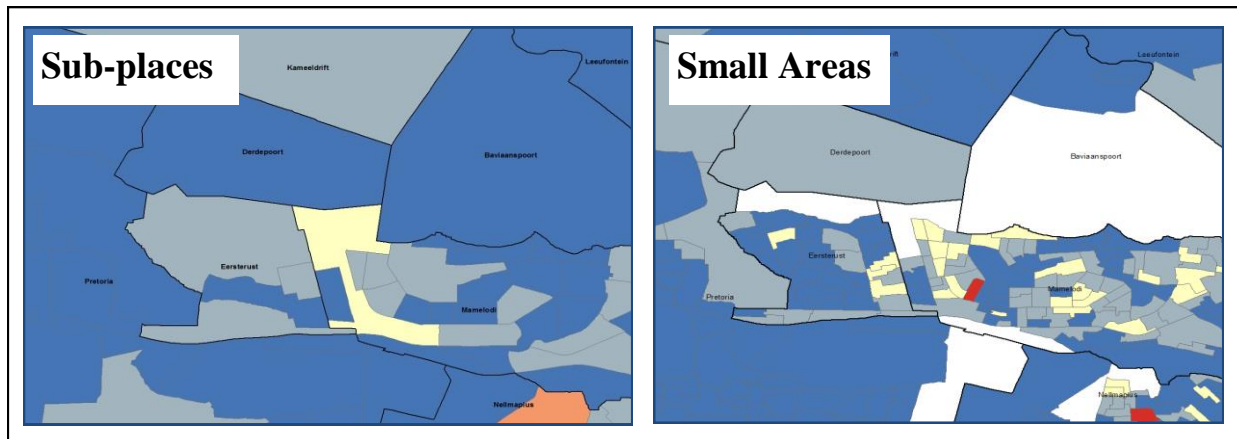


Figure 1. Difference in output area sizes for sub-place and small areas.

It is for the reasons above that the paper will aim to investigate challenges in creating the optimum output geography for census data in South Africa. In order to achieve this, the following objectives will be addressed, firstly to provide a brief overview on output geographies. Secondly the paper discuss aspects such as the population size variation of EAs, in order to inform the criteria for the creation of the ideal small area (SA) layer to satisfy the majority of user needs. Lastly the paper describes briefly the challenges faced to create the 2011 output geography and proposes further research.

## **2. Overview of output geographies**

Accessibility of information on the output geography for different countries tends to be limited to developed countries or, if accessible in developing and non-English-speaking countries, is available only at a very high level of geography. An effort was made to investigate documentation regarding census geographies in as many countries as possible (Table 1). These documents were scrutinised to get an idea of the general trends of dissemination products and tools which could prove useful in the creation of a more suitable output area for South Africa for Census 2011.

Most of the countries are disseminating census data at administrative areas of some kind. The efforts of Canada, the United Kingdom and New Zealand to optimise geographies, especially for dissemination purposes, warranted closer investigation – because of the similarity of their administrative and census systems to South Africa.

Table 1: Countries researched

Continent	Country
North America	Canada
	United States of America
South America	Brazil
Europe	United Kingdom
	Euro Stat Countries
Asia	Turkey
	India
Africa	Nigeria
	Kenya
	Zambia
	Namibia
	Botswana
	South Africa
Australia area	Australia
	New Zealand

## 2.1 Canada

Until 1996 Canada used EAs as primary collection areas as well as basic dissemination areas (DA) (Puderer, 2001). In 2001, Statistics Canada managed to create separate collection and output geographies by using the ‘block program’. The ‘block program’ was started by the creation of a national digital cartographic base for all areas to facilitate the automation of the delineation of dissemination areas. The programme georeferenced all dwellings to specific blocks, the polygons formed by the intersections of streets. The collection geography and output geography’s design can then differ, as blocks can be aggregated in various ways to suit different purposes.

The aim of the design criteria for the DAs was to increase temporal stability, reduce area suppression and get more uniformity, and to use intuitive boundaries to achieve compactness and homogeneity (Puderer, 2001). Not all of these criteria could always be adhered to simultaneously, and some trade-off conditions were implemented, for example the DA will respect census subdivisions and census tract boundaries.

In order to adhere to confidentiality concerns certain measures were implemented. Geographic areas with a population count of less than 40 persons had their characteristic data removed in so-called ‘area suppression’. A minimum population of 500 persons has been stipulated for a DA. Population and dwelling counts will be released by block but with no characteristic data, and the lowest level with population and dwelling data will be the DA. (Puderer, 2001)

## 2.2 United Kingdom

The various Census Offices of the United Kingdom invested in an extensive volume of research and user consultations to introduce a number of major innovations for the 2001 census, and these were refined for 2011. This strategy for disseminating the results was regarded as revolutionary by Leventhal (2003: 1). He was of the opinion that it will significantly change the way users will use census data. Enumeration districts (ED), the area covered by an enumerator (the same entity as the EA in South Africa), were designed for fieldwork purposes during censuses. According to Leventhal (2003) their variation in population size as well as composition render them less than ideal as a base for analysing data. The small output areas or dissemination geography for the UK are known as Output Areas (OAs). They are designed in such a way that each contains around 125 households – populations are to be as homogeneous as possible in tenure and dwelling type, and areas have regular shapes and follow 'natural' boundaries where possible. The exception is Scotland where the size is around 50 households (Scotland's Census, 2013). Furthermore, these OA boundaries are nested within the administrative area hierarchy, i.e. civil parishes/communities, wards and local authority districts.

The various design procedures used for the UK 2001 and 2011 output geography are described and discussed in detail by Martin (1998a), Leventhal (2003) and Cockings *et al.* (2011). The difference in the previous procedures (a), versus the current split in collection and output geography (b), are illustrated in Figure 2. The use of a geographic information system (GIS) and automated processes are central to the development of the output areas, as indicated at stage (b).

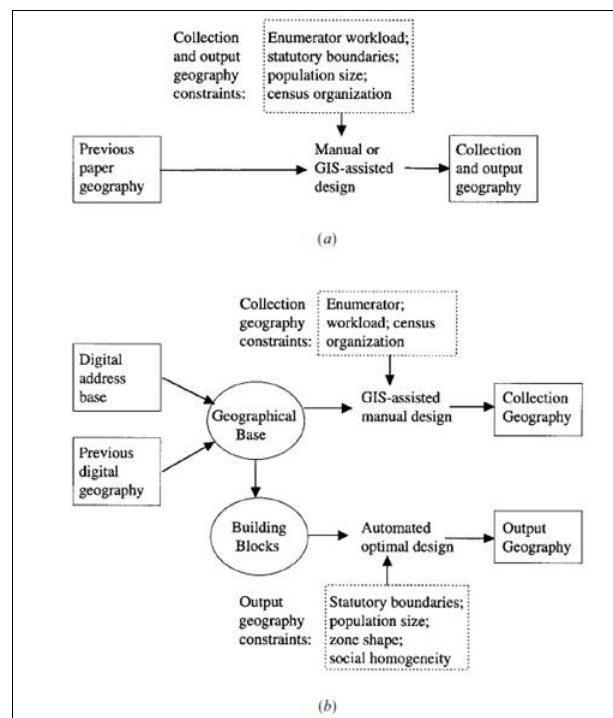


Figure 2. United Kingdom's census output design procedures - (a) at stage three and (b) stage four.

Source: Martin (1998a:677)

The benefits of automated procedures are that they reduce the subjectivity of manual procedures, which are often reliant on people’s local knowledge or intuition. They ensure the application of more systematic and objective methodologies and efficient standards. The AZTool, developed by Prof. David Martin and Dr Samantha Cockings of the University of Southampton, was successfully used to produce the output geographies for the 2001 census for England and Wales. This publicly available tool was used again for zone maintenance in 2011.

According to Cockings et al. (2011), the characteristics which initially underpinned the design of the zoning system (AZTool) in 2001, (e.g. population size or homogeneity of a kind) tended to change over time. It was proposed that rather than keep inappropriate areas or redo everything, the existing system could be modified only by splitting and merging the areas that do not fit the new specifications. Since England and Wales had already developed output zones (OA) in 2001, the 2011 census output geography was merely a case of maintaining the existing zoning system using an automated zone-design technique.

The constraints and criteria employed in the maintenance procedures were as follows: population and household thresholds were set for the different geographic levels as specified in Figure 3; the target population (number of households) was 125 for OAs, 600 for lower-layer super output areas (LSOAs) and 3000 for middle-layer super output areas (MSOAs); homogeneity was measured using intra-area correlation scores for accommodation type and tenure; compactness was monitored by calculating perimeter<sup>2</sup>/area; the minimum boundary length was set at 10% of the total perimeter of the shared boundaries. The last constraint was regional in nature; lower-level output geographies must align to respect higher-level boundaries. Any zones which still had problems after all constraints had been relaxed were left as they were.

Geography <sup>a</sup>	Population thresholds <sup>b</sup>		Household thresholds <sup>c</sup>	
	lower	upper <sup>d</sup>	lower	upper <sup>d</sup>
OA	100	625	40	250
LSO	1 000	3 000	400	1 200
MSOA	5 000	15 000	2 000	6 000

<sup>a</sup> OA—output area; LSOA—lower-layer super output area; MSOA—middle-layer super output area.  
<sup>b</sup> Population thresholds = household thresholds × 2.5 (equating approximately to average household size).  
<sup>c</sup> Household threshold values from Mitchell and Ralphs (2007, table 1.1, page 4).  
<sup>d</sup> No upper thresholds published in 2001 for OAs or LSOAs. Values are given by 2001 OAPS target mean × 2 (as in Ralphs and Mitchell, 2006). MSOAs did have a published upper threshold of 4000 households, but here = 6000 households (as in Mitchell and Ralphs, 2007) to be consistent with ratios used at other levels.

Figure 3. Threshold population for zones in United Kingdom. *Source: Cockings et al., (2011:2408)*



### 2.3 New Zealand

New Zealand also developed zones to optimise their geographies for data reporting (Ralphs, M & Ang L, 2009). Output areas for New Zealand were created by using a modified AZTool. According to Statistics NZ, the algorithm deployed by the tool is Openshaw's zone design algorithm (Openshaw, 1977 and 1978) which addresses scale effects and geographical partitioning (i.e. the modifiable areal unit problem, MAUP). They started from a feasible initial solution and swapped zones in and out until they were satisfied with the result. Their geographical hierarchy and population size, for which data was provided, is illustrated in Figure 4.

Classification name	Number of spatial units in New Zealand	Mean 2006 population
Meshblocks	41,362	87
Area units	1,909	2,100
Territorial authorities	72	55,100
Regional councils	16	251,000

Figure 4. Standard geographical hierarchy for New Zealand 2006. *Source: Statistics Ralphs and Ang,(2009:1).*

The problem with the smallest units, the meshblocks, was that they were designed primarily for data collection rather than output, and resulted in areas with wide-ranging population sizes, contributing to problems with confidentiality in census data (Ralphs and Ang, 2009). Some of these boundaries also crosscut significant patterns of local socio-economic variation on the ground. They aimed to standardise the output zones by population size but also took compactness of zone shape into consideration. Some of the criteria included maximising social homogeneity, ensuring that output zones exceeded the specified population-size threshold, and that output zones were nested within the larger territorial-authority geography.

The countries investigated all aim to supply their census data at the ideal output geography. The main aspect is that the smallest area's boundaries are made up of street blocks or zones, using natural boundaries such as streets or rivers for example. These blocks or zones can then be aggregated if needed for confidentiality purposes as well as to accommodate areas where social homogeneity is too low. The UK and New Zealand used an automated tool to generate the ideal areas based on set criteria.

### 3. Current South African census geography

The structure and nature of the 2011 collection and possible output geographies will be described along with some of the problems encountered, as well as a critique on the 2001 attempt to create small areas for dissemination purposes.

The 2011 South African census collection and dissemination geography is organised in a nested hierarchical model (Figure 5). The collection geographies range from the national to provincial and district municipality (DC) level, as well as metros, local municipalities (LM), MP areas, SP areas and EAs. The upper echelon within the collection and dissemination hierarchy is based on the official boundaries from local municipalities and up to metros, district municipalities and provinces (Statistics South Africa, 2007). Aggregated census data is disseminated to users at all these geographic levels discussed above, except on the EA level. The only layers that are not both collection and dissemination areas are the small area layer (SAL), designed specifically for dissemination only and the EAs, designed for only data collection. The SAs are the lowest level for dissemination. The dwelling frame informs the EA size as well as place names and is available as a spatial data set with descriptive attributes related to the points. No census data is linked to the dwelling frame.

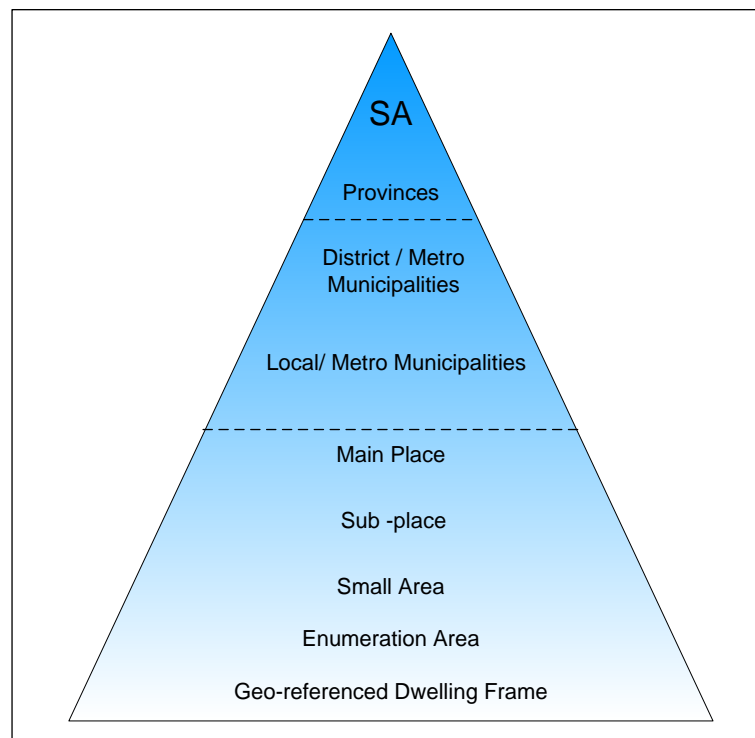


Figure 5. Nested hierarchy for the South African census of 2011

The first attempt to create an output layer, known as the SAL, was done without much consultation and research and the areas created are therefore not necessarily optimal output areas (Grobbelaar, 2005). The chief criterion was simply a minimum population of 500 and, if not, adjacent EAs of the same geographic type, had to be merged to adhere to the threshold population



size of 500. The geography type of 2001 is a classification of EAs that distinguish between the dominant land management types i.e. urban formal, urban informal, farms and traditional areas. Avenell et al. (2009) analysed the SAL in the process of conducting research on deprivation in South Africa. Since a minimum population of 500 was required, 50.4% of EAs are identical to the SAs, and the remaining EAs were merged in various combinations to comply with the population requirement. Problems such as fragmented EAs and therefore SAs were detected: the non-contiguous geographic structure created a problem when merging different EAs, for example, isolated villages surrounded by open space (Figure 6). So-called ‘island EAs’, such as small villages comprising only one EA surrounded by another EA, (mostly vacant) also created problems with merging – especially if more than one ‘island’ occurred within the same larger surrounding EA.

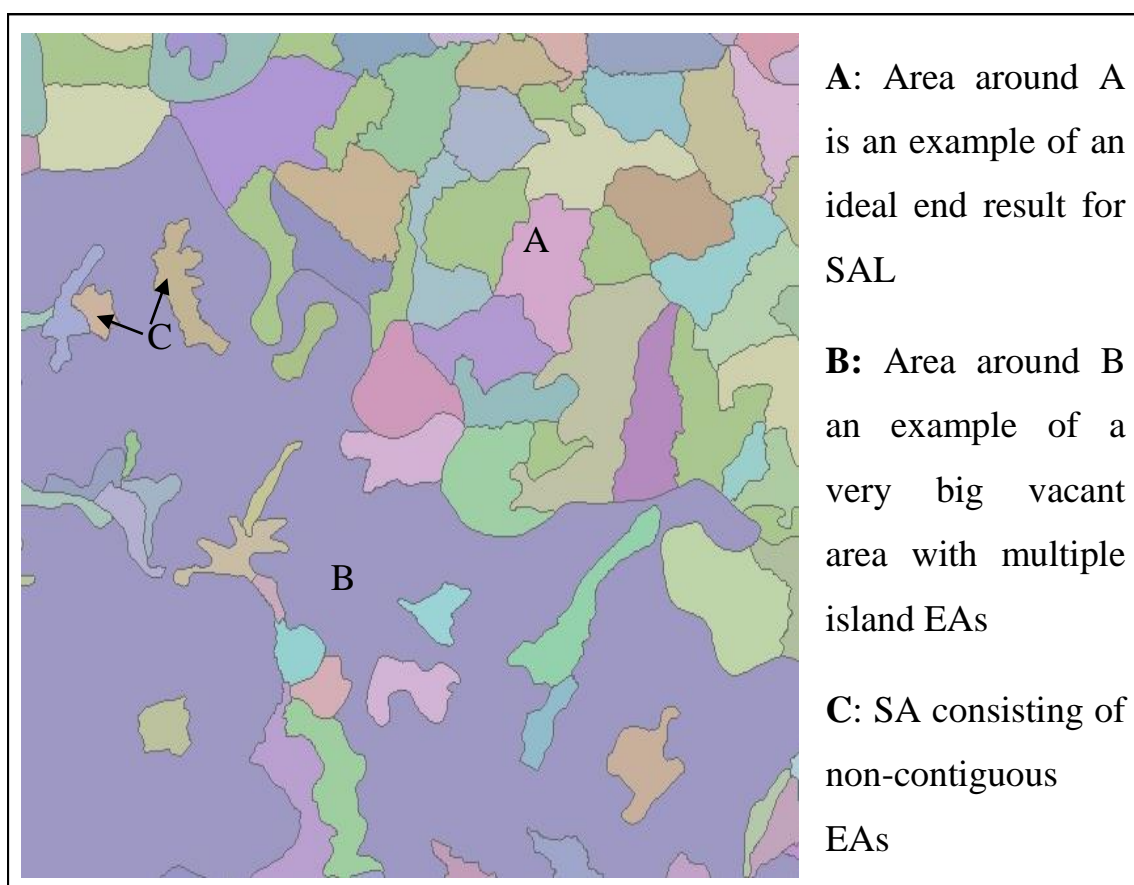


Figure 6. Small Areas built with EAs inherit the EA demarcation problems.

In a study done by Avenell et al., (2009) they had a problem with the fact that the SAL consist of EAs which had already been aggregated, that the minimum population was 500, and that the merging of the EAs was based only on population size and geographic proximity – they preferred neighbourhood and population type for their study. Big differences in population densities and types also gave problems, creating stranded single-EA datazones with sub-minimum populations surrounded by vast areas of merged EAs with low-density populations. Another very valid problem

noted was that the SAL geographies inherited geographic problems associated with the EAs, such as multipart polygons. These EAs are not contiguous and consequently the SALs have a non-contiguous structure as well, which is problematic for data zone creation. The only advantage they mentioned in using the SAL was the limited census data which is publicly available without interpolation at this level, since census data was not released on EA level, and would have been derived from higher geographic levels such as sub-place name areas. Only the spatial boundaries for EAs were in the public domain.

Census data is also aggregated to other user-determined geographical areas, also known as non-standard geographies such as wards, police districts, education, and health areas. This is done on a 'best-fit principle' since the building blocks are EAs, the boundaries of which do not necessarily coincide with these areas, and therefore not within a true nested hierarchical order (Ralphs, 2011).

#### **4. Confidentiality issues: Standards and Rounding**

Various documents from Census and Statistical agencies report that confidentiality and statistical disclosure control remain issues at all levels of geography, but more so at the lower levels (Puderer, 2001; Leventhal, 2003; Statistics South Africa, 2010). Confidentiality or non-disclosure rules are in place to protect individual respondent identities and characteristics. Area suppression is frequently used to remove all characteristics data for geographic areas below the specified population size. Most countries have confidentiality rules and policies to protect the identity of respondents (Duke-Williams, and Rees, 1998; Leventhal, 2003; Statistics Canada, 2011).

With regards to the issue of confidentiality, disclosure and dissemination of data in South Africa, Statistics South Africa (1999: 20) states that: '6) The results of the compilation and analysis of the statistical information collected in terms of this Act may not be published or disseminated in a manner which is likely to enable the identification of a specific individual, business or other organisation, unless that person, business or organisation has consented to the publication or dissemination in that manner.' Disclosure control is discussed in detail in the Generic Operational Manual for Social and Population Surveys (Statistics South Africa, 2010). According to the manual, the primary goal is to ensure that the data of the specific individual return can be inferred to within a narrow range. It is also stressed that it is necessary to protect all basic demographic characteristics information, whether it concerns something probable to be considered sensitive, such as income or not.

Guidelines for reducing the disclosure risk in frequency, as well as magnitude tables, are provided in the manual. They include, for example, 'cell suppression; changing the row and column definitions by collapsing categories or by regrouping or top coding the category values, perturbing data through the addition of noise to the micro-data or the addition of noise to the tabular data, such as rounding; and other procedures that make the micro-data file from which the tabulations are run safe from disclosure.' (Statistics South Africa, 2010: 67). It is stressed that it is critical to ensure that

all releases of public-use micro-data files are reviewed in detail before release. As it is impossible to define what measures to follow for all possible requests, the manual concludes that confidentiality protection should include some common sense that cannot be replaced by rules. Stats SA does not release micro-data for cells of less than 3 individuals, but usually aims for 5 up to main place level. It is regularly used as the rule or standard for data confidentialisation at that level. For ward-level data and higher, the value is increased to 10.

Although the countries reviewed in this research vary in their perception of the specifications for output geographies, or do not bother to have special separate dissemination areas at all, they all battle to supply relevant data to data users who have a problem to solve, for which the general census products are not satisfactory. The issue of confidentiality is universal but is managed by a variation of standards regarded as fit for each country's purpose.

## **5. Challenges with the creation of the 2011 SAL**

Only some of the size and shape problems encountered with the 2001 SAL could be resolved, since EAs were again used as building blocks. Ideally geo-referenced unit records should be used to generate an ideal output unit that will not be influenced by either collection geography or already aggregated data. At the time of the creation of the SAs for 2011, the DF with addresses was not nationally completed and the census address listings not digitally available; the implication therefore was that not all records could be linked (georeferenced) to a locality smaller than an EA.

The minimum and maximum population size was set for each EA type using the standard deviation. EAs having above the maximum population that were merged automatically had to be split or manually re-merged with another neighbouring EA with less population. EA polygons which needed to be merged manually had to be merged with polygons with a low population. EAs originally above the maximum population were not split and formed a SA on its own. If a SA has a too low population, confidentiality might be compromised, however, if it is too high, it would not serve its purpose of being a 'small' entity. The different EA types were treated differently because the aim was to create SAs which were as homogeneous as possible.

SAs were created within sub-places, which meant they also fell within main place areas and local municipalities. An attempt was made to merge EA polygons where needed in such a way that they belong to the same EA type and geography type to ensure homogeneity in terms of land use. EAs are classified firstly as being either urban, farm or traditional (geography type) and then as any one of the following EA types: formal residential, informal residential, traditional residential, farms, parks and recreation, collective living quarters, industrial, small holdings (agricultural holdings) commercial and vacant. Vacant EAs and other EAs with low populations (0-10) were excluded in the creation of the spatial output areas. To avoid having a non-contiguous SA layer in densely populated areas, EAs with zero population (such as parks or open areas, e.g., servitudes) were merged with the nearest EA irrespective of type, except with formal or informal residential areas, as

it will influence the population densities of small areas. It was not always possible to adhere to the requirement that EA polygons in need of merging must belong to the same EA type and geography type. In such cases the resulting SA is of a mixed type.

Only adjacent EA polygons were merged to adhere to the population threshold requirement, and an attempt was made to avoid the creation of multipart polygons as far as possible. Single-point contiguity was not allowed and such EAs had to be re-merged with a neighbour with a shared boundary or multipart polygon. In addition, an attempt was made not to create SAs straddling both sides of a major road, for practical reasons, should the SAs be used as fieldwork entities.

The implications of these different challenges on creating an optimal output geography is firstly that it require intensive manual work that are usually impacted by the individual's subjectiveness rather than the objectiveness you get with set rules in an automated environment. It is also very time consuming. Secondly, due to EAs that were used as building blocks, the SAs inherited the design specifications for enumeration which are mostly not resulting in ideal small areas or output areas mostly because of the different sizes of the different EA types.

## **6. Concluding remarks and recommendations**

Through an investigation into the output geographies and processes used to disseminate census data in other parts of the world it became apparent that Statistics South Africa can test and assess their processes against some of these ideas. Canada, New Zealand and the UK have examples of ideal small area output geographies generated by using roads or postal areas to form 'building blocks' from which both the collection and output geographies could be created. The current lowest geographies used by Statistics South Africa consist of two separate layers, namely the SAs, designed specifically to be used as a dissemination layer and another, the EA layer to be used for collection only.

The initial broad guideline (2001) of 500 households in the SAL does not satisfy user needs and it is not conducive to an optimum output geography for the South African Census. The attempt made for the 2011 to adjust the size and characteristics of SAs to provide for a better, more user friendly output geography. This was a first attempt in getting closer to the ideal output geography of entities with a physical area and population size as small as possible, as compact in shape as possible, as homogeneous in characteristics as possible, and fall within specified administrative boundaries.

Statistics South Africa is currently conducting a test exercise to create 'blocks' using existing road networks. The updated address listings from Census 2011 fieldwork, will be the link to the blocks, and census data could then be aggregated to this street-block level geography. The blocks with data could be used in an automated process using the AZTool for the generation of output geography areas as well as various other administrative or planning areas such as voting districts and wards, police areas, etc.

The DF was updated with the captured address listings from Census 2011 fieldwork, and will be used to link the census questionnaires to the spatial locality, such as the address associated with a structure or land parcel. A unique identifier for each structure, the map-reference number that was generated for census enumeration, is the link between the captured address listings, the census questionnaire and the geo-referenced point on the accompanying EA map. Analysis will be conducted to establish the coverage of usable physical addresses per EA before attempting to link the questionnaires. The initial aim is to at least attempt the creation of the building blocks with associated data for the Metros and secondary cities that have generally good coverage of digital address points.

The benefit would be that all other areas that used these standardised building blocks as input will have more accurate aggregated census data, and the process of generating any spatial entities could be automated using standardised implementation criteria. Manual demarcation should then be a limited exercise saving time and effort and the overall quality improved. With the creation of smaller 'building blocks' it would be possible to address individual user needs without major impact on staff requirements for once-off products.

## **7. References**

- Avenell, D, Noble, M & Wright, G 2009, *South African datazones: A technical report about the development of a new statistical geography for the analysis of deprivation in South Africa at a small area level*, CASASP. Working Paper No. 8, Oxford: Centre for the Analysis of South African Social Policy, University of Oxford.
- Beard, C, DeGennaro, M & Thompson, W 2011, 'Cartographic support for the 2010 Decennial Census of the United States', *Cartography and Geographic Information Science*, vol. 38, no. 3, pp. 320-325.
- Cockings, S, Harfoot, A, Martin, DJ & Hornby, D 2011, 'Maintaining existing zoning systems using automated zone-design techniques: methods for creating the 2011 Census output geographies for England and Wales', *Environment and Planning A*, vol. 43, pp. 2399 – 2418.
- Cockings, S, Harfoot, A, Martin, DJ & Hornby, D 2013, 'Getting the foundations right: spatial building blocks for official population statistics', *Environment and Planning A*, vol. 45, pp. 1403 – 1420.
- Duke-Williams, O & Rees, P 1998, 'Can Census Offices publish statistics for more than one small area geography? An analysis of the differencing problem in statistical disclosure', *International Journal of Geographical Information Science*, vol. 12, no. 6, pp. 579 - 605.
- Dugmore, K, Furness, P, Leventhal, B & Moy, C 2011. 'Beyond the 2011 Census in the United Kingdom: With an international perspective', *Journal of Market Research*, vol. 53, no. 5, pp. 619 – 650.

- Gotway, CA & Young, LJ 2002, 'Combining Incompatible Spatial Data', *Journal of the American Statistical Association*, vol. 97, no. 458, pp. 632 – 650.
- Grobbelaar, N 2005, The development of a Small Area Spatial Layer to serve as the most detailed geographical entity for the dissemination of Census 2001 data. *Africa GIS Conference*, Pretoria, October 2005.
- Leventhal, B 2003, 'Developments in outputs from the 2001 Census', *International Journal of Market Research*, vol. 45, no.1, pp. 3 - 19.
- Manuel, T (Minister in the Presidency: National Planning) 2013, *The National Assembly budget vote debate on Statistics South Africa*. Speech held, 15 May, African National Congress Parliamentary Caucus, Pretoria, viewed 1 August 2013, <<http://www.anc.org.za/caucus/show.php?ID=3006>>
- Martin, DJ 1998a, 'Optimising census geography: the separation of collection and output geographies', *International Journal of Geographical Information Science*, vol. 12, pp. 673-685.
- Martin, DJ 1998b, '2001 Census output zones: from concept to prototype', *Population trends*, vol. 94, pp. 19-24.
- Martin, D 2006. Last of the censuses? The future of small area population data, Royal Geographical Society
- Martin, D, Cockings, S & Harfoot, A 2013. 'Development of a geographical framework for census workplace data', *Journal of the Royal Statistical Society*, vol. 176, no. 2, pp. 585-602.
- Morphet, C 1992, 'The Interpretation of Small Area Census Data', *Area*, vol. 24, no. 1, pp. 63-72.
- Openshaw, S 1977, 'A geographical solution to scale and aggregation problems in region-building, partitioning, and spatial modelling', *Transactions of the Institute of British Geographers*, vol. 2, no. 4, pp. 459–72.
- Openshaw, S 1978, 'An optimal zoning approach to the study of spatially aggregated data' in I Masser, and PJB Brown (eds.), *Spatial representation and spatial interaction*, Springer US, Boston.
- Paez, A & Scott, DM 2004. 'Spatial statistics for urban analysis: A review of techniques with examples', *GeoJournal*, vol. 61, pp 53-67.
- Puderer, H 2001, *Introducing the dissemination area for the 2001 Census: an update*, Working paper No. 2000-4, Ottawa: Statistics Canada.
- Ralphs, M 2011, *Exploring the performance of best fitting to provide ONS data for non standard geographical areas*, Newport: Office for National Statistics.
- Ralphs, M & Ang L 2009, *Optimised geographies for data reporting: zone design tools for Census output geographies*, Working paper No. 09-01, Wellington: Statistics New Zealand.

- Riva, M, Apparicio, P, Gauvin, L & Brodeur, J 2008, *Establishing the soundness of administrative spatial units for operationalising the active living potential of residential environments: an exemplar for designing optimal zones*, *International Journal of Health Geographics*, vol. 7, no. 43, viewed 20 March 2014, <<http://www.ij-healthgeographics.com/content/7/1/43>>.
- Robertson, IML 1969, 'The Census and Research: Ideals and Realities'. *Transactions of the Institute of British Geographers*, no. 48, pp. 173-187.
- Scotland's Census 2013, *Scotland's Census 2011: Outputs Prospectus*, Edinburgh: General Register Office for Scotland.
- Statistics Canada 2011, *Data quality and confidentiality standards and guidelines (Public)*, Ottawa: Statistics Canada.
- Statistics South Africa 1999, Statistics Act, Act 6 of 1999. *Government Gazette of South Africa*, vol. 406, no. 19957. Cape Town: Government Printer.
- Statistics South Africa 2007, *Using the 2001 Census: Approaches to analysing data*, Pretoria: Statistics South Africa.
- Statistics South Africa 2010, *Generic Operational Manual for Social and Population Surveys*, Pretoria: Statistics South Africa.
- Torrieri, NK 2007. 'America is changing, and so is the census' *The American Statistician*, vol. 61, no. 1, pp. 16 – 21.
- Vickers, D, Rees, P & Birkin, M 2005, *Creating the National Classification of Census Output Areas: Data, Methods and Results*, Working paper No. 05/2, School of Geography: University of Leeds.
- Young, C, Martin, D, & Skinner, C 2009. Geographically intelligent disclosure control for flexible aggregation of census data, *International Journal of Geographical Information Science*, vol. 23 no. 4, pp. 457-482,