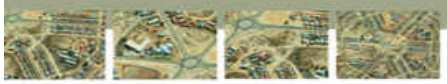

Geography ...
statistically speaking

"Statistics needs Geography; Geography needs Statistics"

Pali Lehohla

Statistician-General

AfricaGIS Conference, South Africa
31 October 2005

Statistical geography or geographical statistics is an expression of democracy

Statistics South Africa

Reflecting on a 23 year crusade

In 1982 November arriving in a homeland statistics office of Bophuthatswana from Lesotho, I was confronted with leading the planning and implementation of a census of the population for Bophuthatswana in 1985. The first question I had was where is the report of the 1980 Census. I was given an orange file and this was to become the guide for engaging in this next endeavour. This orange file had scanty maps with numbers in them. Associated with the map was a table containing the numbers on the map with more numbers per row and labelling of Stedelik and nie-Stedelik. Under Stedilik, I saw Mafikeng and a number corresponding to it and that number was 5600, and under nie-Stedelik was an aggregate figure in excess of possibly 40 000. For every Stedelik was some name and a corresponding number and for every nie-Stedelik was just a big number. I then asked my colleagues why are settlements in Bophuthatswana having the same name although they appear to be belonging to different geographic areas. Coming from Lesotho where villages each had a name and reported as such in a census report of Lesotho of both 1966 and 1976 and having been to Mafikeng and having noticed these similarities that villages had names I wondered why all these villages have the same name called “nie-Stedelik” in Bophuthatswana. I was informed that nie-Stedeliek means non-urban or rural, and I interpreted this non-urban or rural as “that which does not count.” This revelation constituted what was to become a personal mission and a crusade that I pursued to ensure that that which did not count counts and should be seen to count. A placename agenda was born in me, a fusion of statistics and geography of geography and statistics had revealed itself and presented possibilities I would have taken for granted. A crusade had begun. This crusade I drove against all odds in 1985 Census, the 1991 Census and I firmly put and led this agenda in the 1996 Census, which was the first inclusive census in South Africa. By 2001 Census a placename layer was now a-for-granted issue. A convergence of statistics and geography had occurred in the corridors of Statistics South Africa. We had moved from cartography to geography. An enumeration area was no longer a unit of logistics for census enumeration but had now been a critical expression of development challenges as it became a basis for analysis and dissemination. In that way geography was a central piece of revolutionarising statistical management, ensuring an end-to-end production to dissemination system. I therefore feel personally honoured that this august gathering has granted me this rare opportunity of reflecting on this journey of 23 years and sharing with you how I perceive the benefits that arise out of deliberately crafting and marshalling the relationship between geography and statistics.

Quality Statistics need geography

Good statistics requires good frames. The System of Statistics, comprising of National Accounts that overarches economic and social statistics, is based on the quality of information on frames such as the Business Frame and the Geographic Frame. These are governed by quality methodologies and standards. Statistics South Africa (Stats SA) has over the past few years, elevated the importance of the Geographic Frame, Geographic Methods and

Geographic Standards, in essence, the inclusion of *Geographical Knowledge* in the production and dissemination of statistics, similar to the Business Frame. Thus formalising this strategic positioning of geo-information in the System of Statistics.

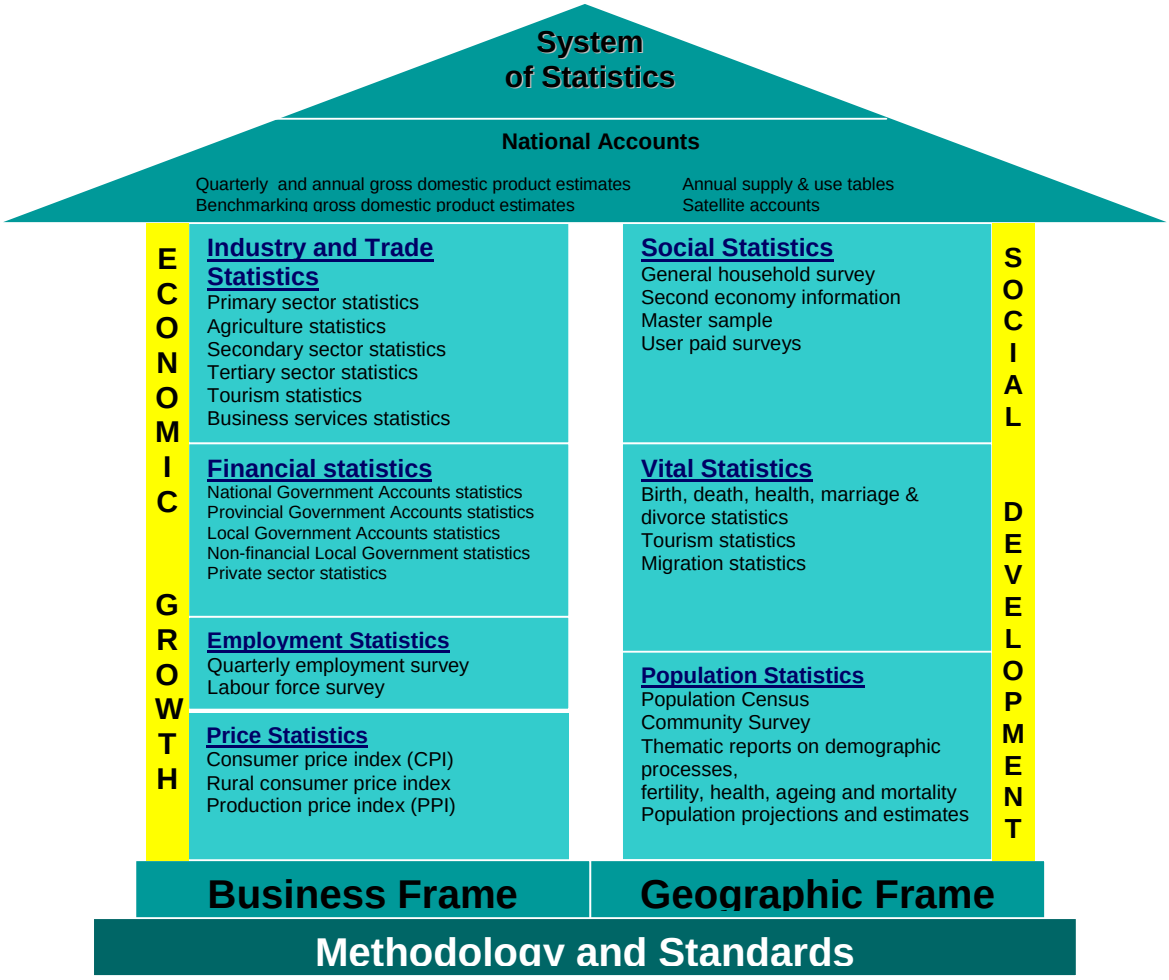

Figure 1: Geography foundation information for the System of Statistics

The importance of regional statistics for South Africa cannot be overemphasized. The gross geographic product is a statistical representation of economic attributes across space. Last Thursday night I got a call from the office of the Premier of the North West on the growth of the gross geographic product of the North West Province. Policy and decision-makers need to know *where* to focus on investment and development. These decisions have huge impacts on the economic and social futures and possibilities of the people of our country and indeed of any of any country.

Standardising Geographies for Statistics

Central to the change that occurred in Stats SA from 1996 was that we did not only recognise the importance of utilising standard geographies for collecting, disseminating and comparing data but we implemented a focused investment in geography because we knew that it constituted a relevance

glue. Initiated as standard *census geographies*, and expanding as the geographical framework within which various statistics can operate, the geographical frame maintains the relationships between geographical layers, and it enables proper geographical usage of statistical data. Elements of the frame supported by Stats SA are the combined responsibility of Stats SA (Enumeration Areas, Small Area Layer, Place names), the Municipal Demarcation Board (Electoral Wards and Local Municipalities) and Department of Provincial & Local Government (Provinces).

The Place name geography was first developed and presented with Census 1996 data and this improved in Census 2001. People in South Africa can for the first time ever obtain information at a local level that is meaningful to them, that is, a community, village or suburb, where they come from. Based on a collection of enumeration areas, Stats SA's placename layer is the only layer that attempts to delineate an area, other existing databases records placenames as points. Stats SA will continue to correct the placename database with collaborating partners, since localised information is important for local area upliftment and nation building. The placename geography is the common geographic integrator for disparate datasets, especially with respect to the wealth of *untapped* information contained in administrative data, when combined with official statistics and analysed from a spatial perspective. To enable effective data integration on a placename level, standards will need to be maintained by all collectors of such data.

The next revolution: a dwelling frame

The next revolution is expanding the frame to a micro level of geography where each and every dwelling in South Africa is addressed. This new revolution should not take 23 years to realise. By standardising and maintaining address information already existing in formal areas of the country and assigning addresses to areas of the country deprived previously of formal addresses we will ensure that the Stedelik and nie-Stedilik divide is once and for all removed. The project has begun in earnest with the aim to complete coverage as we prepare for Census 2011. This detailed level of geography can re-define or correct all layers of geography in the frame, making elements of the frame more geographically reliable, for the production of reliable statistics.

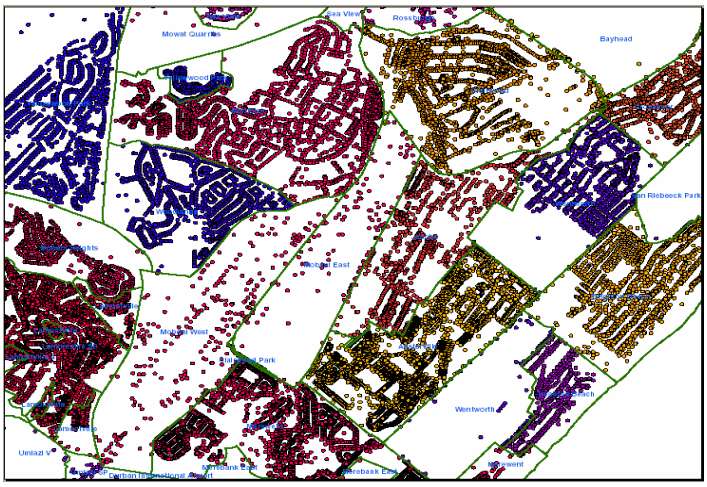

Figure 2: Correcting the place name frame with more detail geographies from the dwelling frame

Statistics needs Geography …

Statistics on its own certainly does not show the complete picture. As an example, the table below shows the number of businesses per Standard Industrial Classification code for a small town in KwaZulu-Natal. The table shows the number of active, start-up (birth) and closed (deceased) businesses in the town. Analysing the table at a glance, one can gather that most businesses are active and start-ups (birth)

SIC code	Active		Birth		Ceased		Deceased		Total
	Total	%	Total	%	Total	%	Total	%	
No classification	199	23	629	74	3	0	16	2	847
1	52	48	5	5	15	14	37	34	109
2	12	40	4	13	5	17	9	30	30
3	82	49	20	12	24	14	41	25	167
4	5	100	0	0	0	0	0	0	5
5	52	47	11	10	8	7	40	36	111
6	146	41	22	6	35	10	153	43	356
7	20	40	14	28	5	10	11	22	50
8	224	55	65	16	31	8	88	22	408
9	54	53	21	21	5	5	22	22	102
None	71	36	126	64	0	0		0	197
Grand Total	917	38	917	38	131	5	417	18	2382

Table 1: Most businesses in this small town are active and start-ups

The data can be presented graphically

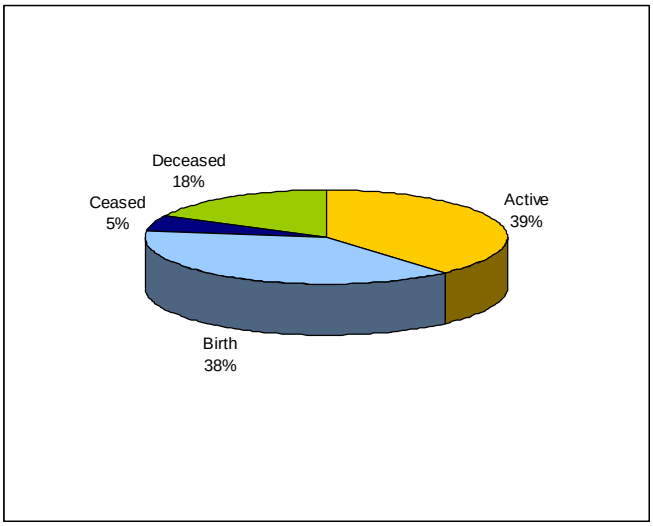

Figure 3: An equivalent assessment from a graph

However a map presentation gives more information. The map below contains the total number of active businesses (linked to addresses).

Wholesale trade and retail trade, repair, etc., hotels and restaurants, is grouped more or less south of the town. Most of these are mainly located in the south side, implying that the CBD is most properly located in the south side of town. Local planners can use such information for zoning, whilst entrepreneurs can determine the best location for their business, adding information such as transport routes can also maximise use.

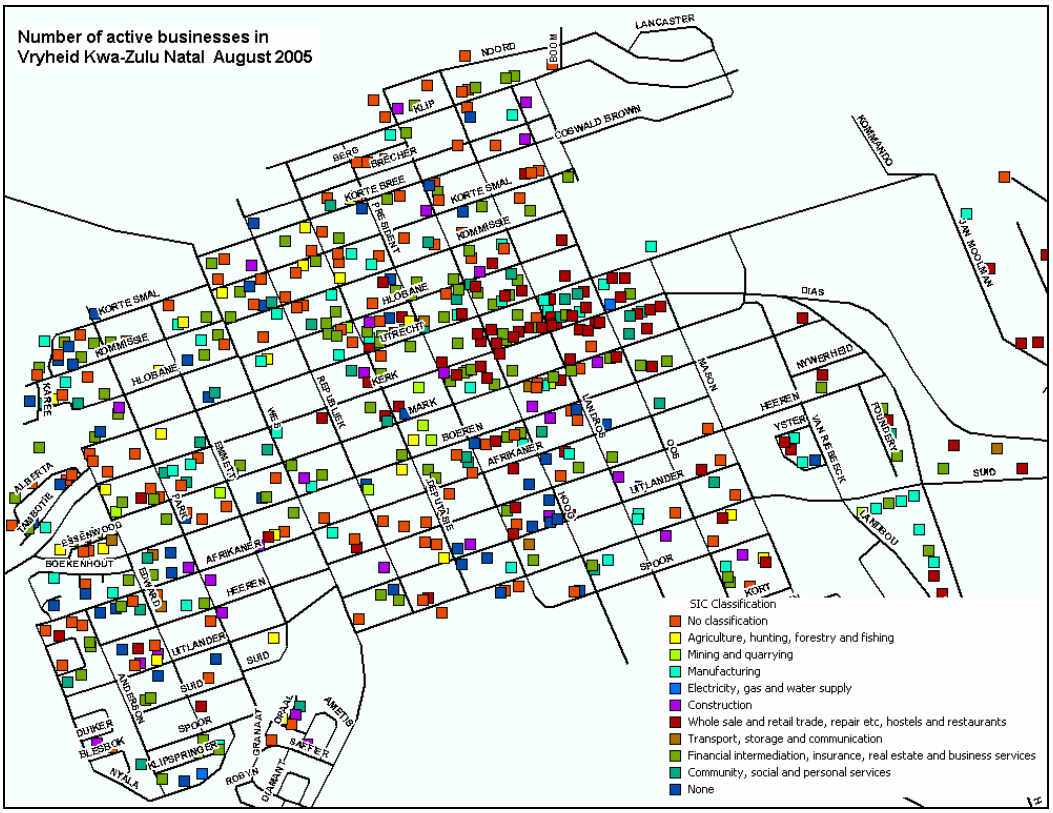

Figure 4: Getting the spatial benefits.

Geography needs Statistics …

Geography on its own misses the detail required to target interest. Although topographical maps are selective representations of the real world and imagery is a true reflection of the same world they do not reflect the socio economic attributes of that world. Their application vests in orientation, locality determination and feature - and pattern recognition. Adding statistics to geographical maps gives a sense of why one area is more important than the other and captures interest.

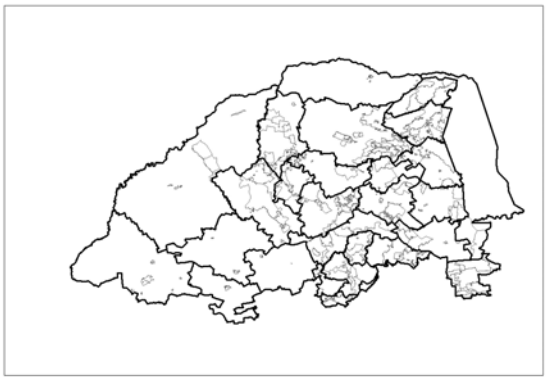
Figure 5: Delineated municipalities of Limpopo.

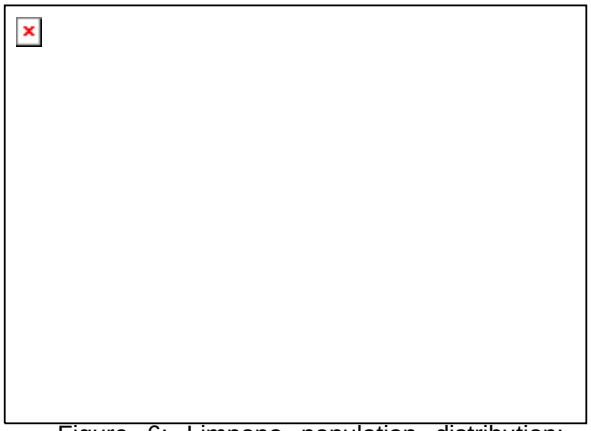
Figure 6: Limpopo population distribution: high and low population areas.

A fixed or static geography is ideal for statistical purposes especially for temporal analysis to benchmark change. However, static geographies do not reflect the dynamic nature of demography and the actual situation on the ground. On the other side of the coin, geographies are dependent on the profiles of communities that they enclose. Without statistics on the social and economic situation of communities, proper demarcation of geographies to be used to serve these communities fail. Thus, the inter-dependency of up-to-date geographies used for the collection of up-to-date statistics and the statistics needed to update geographies to keep it up-to-date, feeds off each other in a continuous cycle to ensure optimal data collection on an optimal spatial data frame.

Figure 7: An area in 2000 delineated as a single area.

Figure 8: The same area in 2005 will require several delineations.

Challenges that need to be addressed for synergies

It is evident that geography is the key integrator for disparate data. It is also evident that there is much more value if all information can be associated with each other for a complete picture. Because of the importance of statistical geography or geo-statistics cannot be left to chance. In South Africa this is one of the biggest challenges. In 1996 geography was driven by the need to run a census, subsequently, the need for elections drove geography. In many countries these have been prime movers of geography as a consequence standards get largely driven by census and or election needs and any other need becomes incidental. It is therefore important to locate geo-statistical awareness at the heart of what we all do because all of us are collectors or custodians of information collected across time and space. Awareness that information collected, besides the fact that data collection is an expensive and time consuming exercise, must be collected in a structured manner that enables statistical integrity. This requires establishing and maintaining standard geographies. The Municipal Demarcation Board (MDB) has already started this i.e. the move towards standard municipalities and electoral wards. A similar endeavour is required for other geographies like placenames and addresses. Basic data collection instruments like forms and questionnaires will benefit from geographic standards.

Having more detail geographies such as locations of each and every dwelling in the country, or Stats SA's enumeration areas consisting of more or less 120 dwellings, and its associated attribute information, is what we all require for planning. The user community wants enumeration area data! The aspect of data confidentiality is important to a statistical agency and gives statistical agencies credibility with its respondents. The development of the small area spatial layer (SASL) was in direct response to the statistical User community needs for a more detailed level of census data dissemination. If not for the statistics associated with the EA spatial layer, which formed the foundation for the SASL, the balance between optimising confidentiality and refining a more detailed level of dissemination would not have been possible.

Conclusion

Statistics need Geography and Geography needs Statistics to complete the picture and target interest. Both Statistics and Geography cannot be regarded as separate entities since the combination is fundamental as we move forward.