

Advancement in Statistical Processes from Census 2005 to Census 2011

Mr. Ghanem Khalfan Ghanem Al Mehairbi
Statistics Centre - Abu Dhabi (SCAD),
Abu Dhabi, United Arab Emirates, email: gkalmehairbi@scad.ae

Abstract

A population census is aimed at producing the demographic, social and economic characteristics of a given country/region. Statistics generated from census projects are used as baseline information for government entities in formulating policies and decision making processes. A census is conducted every 5 years for Abu Dhabi emirate. The census 2005 project initiated the process of presenting market driven statistics in a user friendly manner. The “then” technical advancements like PDAs for collecting building geo-coordinates, utilizing statistical systems and reports generating tools (e.g.: Oracle Discoverer, Oracle Reports Builder) were used by Abu Dhabi’s official statisticians. However, the role of technology in Census 2005 was limited to internal processing and the output was primarily paper base. In keeping with the SCAD’s mission and vision to provide relevant and reliable official statistics serving policy makers, the business community and the public; SCAD continuously strives for accuracy and consistency in collecting, processing, storing and disseminating official statistics. The census 2011 project has highlighted use of advanced IT and Geographical Information Systems (GIS) technologies across multiple business processes. Survey project planning and field work processes were simplified using spatially enabled web services and geo-enabled iPads. Spatial and non-spatial data were collected from “custodian” owners through active involvement of Abu Dhabi Systems and Information Centre (ADSIC) instead of reproducing and duplicating the data within SCAD. Use of statistical applications (SAS) resulted in simplifying statistical analysis through weighting, aggregation, imputation methods and presenting better and reliable statistic results. Online census output tools has been designed and developed for representing statistics through thematic maps and reports. Producing spatially enabled statistics is now possible through online census output tool. Statistical data is now accessible to government entities, business community and public. The data is customized for each audience using user rights policy, IT security policy and confidentiality rules.

Key Words: Dissemination, Statistical applications, Spatially enabled statistics, confidentiality, Census

1. Introduction

Statistics Centre - Abu Dhabi (SCAD) was established in accordance with Law No. (7) for the year 2008, with a view to develop and organize statistical work in the Emirate of Abu Dhabi in particular and the United Arab Emirates in general. The centre therefore has an independent legal personality and financial autonomy; having full legal capacity to act in a manner that is consistent with the emirate’s orientation towards sustainable development and strategic plans. It operates under the supervision and with full support of the Executive Council of the Emirate of Abu Dhabi.

SCAD has four main functions:

- Develop and organize a statistical system for Abu Dhabi.
- Contribute to the UAE’s national statistical system.
- Provide official statistics related to the conditions of the Abu Dhabi society.
- Support decision makers in Abu Dhabi

The statistical tools involved in the census projects are primarily used during the data collection, data analysis and data dissemination stages of the statistical process. This paper aims to address the progress made through effective use of technological advancements in statistical applications for Abu Dhabi census projects.

The paper begins with a brief detail on the technologies and methodologies used in Census 2005 in comparison to those used in Census 2011. The benefits achieved in Census 2011 are highlighted at the end of this paper.

2. Census 2005

2.1 Data Collection

Census 2005 involved collection of information through use of PDAs while the geographical location details were collected using handheld GPS devices. Owing to the technological limitations at that time, the XY accuracy was in the range of ± 5 meters (m). The positional accuracy was degraded inside of high rise buildings and by close proximity to high tension electrical towers.

During this period in Abu Dhabi there was no coordination among government departments regarding spatial data. Multiple government Entities were generating their own spatial datasets to serve their individual business needs resulting in duplication or creating redundant datasets. As a result the census team encountered problems in identifying the “true” custodians of the datasets along with specifying the metadata associated with the spatial layer.

2.2 Data Analysis

Oracle Report Builder was used to generate final statistics. In 2005 a small number of administrative records were collected from sites with restricted access such as oil company facilities, these records were merged into the PDA data. The data was then be loaded into the Oracle database and analyzed by IT staff with guidance from statisticians. The validation, coding, imputation rules were discussed and decided by the statisticians along with the definition of derived variables. The implementation was carried out by IT database specialists. Census 2005 did not have a data dictionary, metadata or GIS maps that would have helped in standardizing the process.

SCAD also used Oracle Discoverer as a Business Intelligence tool for ad-hoc requests. However, there were limitations in entertaining specific customer defined queries or requests for producing business driven statistics.

2.3 Data Dissemination

All statistics during Census 2005 were generated in tabular formats and limited to the specified list of themes from the government. There was no data confidentiality rules applied and outputs could not be customized for different user types. The final results were printed on hard copies and/or delivered in PDF formats through external storage media to the government entities. Final statistics data was not available via the internet this meant that the public and external users did not have access to the results. The only spatial data dissemination was done by another government entity, Abu Dhabi Systems and Information Centre (ADSIC). They added some of Census 2005 information in their GIS portal for Abu Dhabi Government. No other spatial data was disseminated such as thematic maps or a statistical atlas.

جدول 1- إجمالي المباني حسب شكل المبني والمنطقة
Table 1- Buildings by type and district

District	المجموع Total	جزر الإمارة Emirate Islands	المنطقة الغربية Western Region	منطقة العين AL Ain			منطقة أبوظبي Abu Dhabi			المنطقة
				المجموع Total	ريف Rural	حضر Urban	المجموع Total	ريف Rural	حضر Urban	
Palace	215	42	21	84	15	69	68	19	49	قصر
Multi-storey building	4271	2	118	519	2	517	3632	179	3453	مبنى متكرر
Two-storey building	3614	44	339	2229	206	2023	1002	227	775	مبنى من طابقين
One-storey building	26771	817	7753	13599	8144	5455	4602	3171	1431	مبنى من طابق
Villa	22345	67	1910	9565	738	8827	10803	457	10346	فيلا
Low-cost house	31008	595	2734	14245	7056	7189	13434	12747	687	بيت شعبي
Shed	13124	67	1727	8065	6460	1605	3265	880	2385	شبره
Caravan	6296	338	2456	2137	933	1204	1365	1016	349	كرقان
Shack	5651	27	1039	4292	3389	903	293	293	0	عشة، خيمة، صندقة
Others	4174	80	593	1889	581	1308	1612	651	961	أخرى
Total	117469	2079	18690	56624	27524	29100	40076	19640	20436	المجموع

Population, Housing and Establishments Census 2005 1 تعداد السكان والمنشآت 2005 - إمارة أبوظبي

Figure 1: Output table from Census 2005

3. Census 2011

3.1 Data Collection

The establishment of Abu Dhabi Spatial Data Infrastructure (ADSDI) through Abu Dhabi Systems and Information Centre (ADSIC) resulted in provision of a centralized repository of data (spatial and non-spatial) as well as eliminating the need to build redundant/duplicate datasets. It also ensured provision of the best spatial data quality for SCAD from ADSDI custodians.

Technological advancements in IT and spatial domains enabled SCAD to use a standardized and defined data structure template for its business processes as well as simplifying the data exchange processes between SCAD and ADSDI community through effective usage of web services and map services.

A comprehensive database comprised of spatial references and relevant attribute information was developed within SCAD through the use of GIS applications and IT tools. Bespoke data collection software was developed using spatially enabled web services and geo-enabled iPads. Oracle BI dashboards were used to monitor the progress of the field work providing daily reports and real-time monitoring.

The use of administrative records was expanded to ease respondent and enumerator burden. All large labor camps (housing more than 100 persons) were asked to send administrative records via a fixed template as an excel sheet in place of having enumerators interview each person.

3.2 Data Analysis

SCAD implemented Generic Statistical Business Process Model (GSBPM) across all departments and different teams/departments were entrusted with responsibilities at SCAD. A detailed responsibility matrix was drawn up to govern access rights for each person to Census 2011 data. This resulted in refinement of processes as compared to Census 2005 and documentation of methodology.

Various statistical processes like donor imputation, deterministic imputation, aggregation, validation and variables derivation were further enhanced through use of SAS and thus had a direct positive impact on process efficiency. A lot of paper based processes used in Census 2005 were eliminated through use of these statistical processes. Data processing was handled by methodology team with guiding by Population and Demography section.

Disclosure controls were defined and applied for all output. As part of Census 2011, a statistical geography was defined and boundaries were drawn for statistical districts and statistical sectors. The minimum geography was the statistical sector which was defined as a collection of municipality sectors with a minimum of 500 persons and 5 households. Controlled random rounding methods were applied to all outputs.

3.3 Data Dissemination

Tabular data were refined along with defining database structure, metadata schema and standards. The metadata developed can support multi-linguistic details (Arabic and English).

Integration of Statistical Information Systems (SIS) data with GIS data was made possible through SAS BI applications and helped SCAD in producing thematic maps and a statistical atlas. Three output tools were developed to produce and present the outputs from Census 2011 project namely: thematic maps, community tables and table builder. The aim of developing these tools was to make rich census data more widely available/accessible to end users. All tools were developed to be accessible through most of internet browsers and operating system e.g., IE7, Samsung Galaxy and IPADs.

The tools were made available through the SCAD website and are currently accessible to government entities through SCAD data portal. Upon approval from Executive Council, the same tools shall be made available to the public internet users. Through Census 2011 output tools, the end user can generate standard or customized reports on their own instead of making requests through SCAD's customer support. All tools are integrated with census metadata to present variable names and their definitions in both Arabic and English.

Strict data confidentiality rules were applied as part of information security management system (ISMS) policy because of the highly sensitive nature of the data at SCAD. In all, 3 levels of confidentiality were applied for Census 2011 data:

- Level 1 – Data disseminated using a statistical geography comprised of districts and sectors.
- Level 2 – Each variable was evaluated and designated for release to users.
- Level 3 – All outputs randomly adjusted to a multiple of 5 “on the fly” to allow any query and ensure confidentiality.



Figure 2: Census 2011 output tools

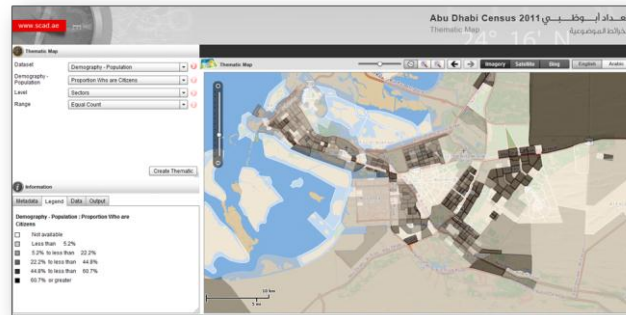


Figure 3: Census 2011 output tool - thematic map

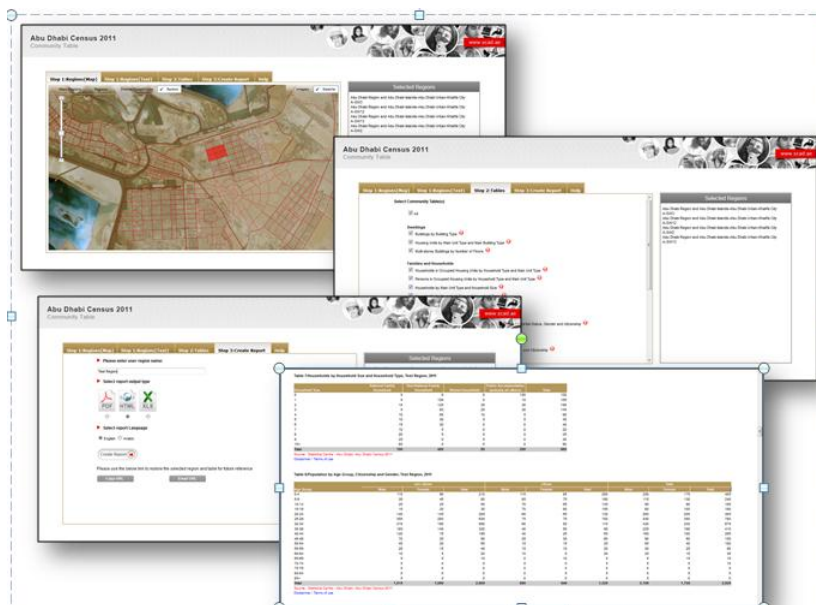


Figure 4: Census 2011 output tool – community tables

Main Region	Abu Dhabi Region and Abu Dhabi Islands		Ajman Region		Ajman and Islands		Row Total
	Male	Female	Male	Female	Male	Female	
Total	16,985	16,985	50,525	16,760	13,125	2,500	135,630
Illiterate	35,720	16,985	50,525	16,760	13,125	2,500	135,630
Read and Write	61,565	27,305	47,820	19,870	9,910	2,180	142,555
Primary Cycle 1	31,540	28,660	21,500	28,165	14,200	4,045	216,255
Preparatory / cycle 2	53,405	28,005	30,725	17,990	5,940	2,585	141,645
Secondary	108,770	68,530	48,895	32,990	16,905	4,710	279,210
Above Secondary and below University	18,360	28,410	12,895	9,860	4,365	1,375	61,565
Diploma	98,695	73,250	27,765	25,540	8,000	4,175	227,425
University	19,940	6,030	2,365	1,825	875	325	22,855
Master	16,690	9,380	3,905	1,875	690	300	21,895
Doctorate	3,185	1,185	1,340	435	110	45	6,300
Not Stated	13,460	1,555	1,495	580	2,625	310	24,725
Total	520,445	295,600	285,100	153,025	78,735	22,545	1,355,445

Figure 5: Census 2011 Output Tool - Table Builder

4. Results

The following marked improvements were seen in Census 2011 project as compared to Census 2005 project:

- a) Simplification in data collection processes from other government entities through use of web and map services.
- b) Usage of spatial data from “Custodians” instead of producing duplicate/redundant datasets.
- c) Standardization of data structure, building data dictionary and multi-linguistic metadata standards to serve SCAD’s needs
- d) Enhancement and process improvement through use of SAS BI application.
- e) Dissemination tools developed for Census 2011 are initiative tools and produce results highly relevant to user needs.
- f) Generation of thematic maps output tool and statistical atlas as output for Census 2011 in addition to tabular data. The thematic maps enabled SCAD to add value and present a different dimension to the online statistics generated. It also helps to easily recognize spatial ‘clusters’ or ‘hot spots’ and tell a story about that place.
- g) Implementing online strict data confidentiality and ISMS policies to safeguard the sensitive nature of the data.
- h) Users can design their own census output tables by accessing variables and by modifying the table structure to suit their own requirements. Additionally, all the statistics for Census 2011 are being made available to the end user on the website as against the Census 2005 where only government entities were provided with the details.
- i) Metadata of Census 2011 is integrated in all Census 2011 output tools in addition to a data dictionary report in both Arabic and English.
- j) Statisticians performed coding in SAS instead of IT staff.

5. Conclusion

As stated previously, one of SCAD’s important functions is to support decision makers in Abu Dhabi. As part of this, SCAD provides official, reliable and timely statistics to various government entities. It is imperative that SCAD must choose to define and adopt the best statistical approach and presentation in order to enhance the decision making processes. With Census 2016, it is envisioned that SCAD will further improve upon the methodology adopted for Census 2011 and make optimum use of the advancement in technologies, new trends applicable to SCAD’s business needs.