Paper 356-2013

SCAD: Development of statistical information systems for the provision of census data

Gregory Pole, Statistics Centre – Abu Dhabi

ABSTRACT

The Statistics Centre – Abu Dhabi (SCAD) was founded in 2008 and seeks to join the world's leading statistical organisations in statistical collection, production, and dissemination. In October 2011, SCAD conducted its first census of population and households. In addition to using innovative enumeration technologies (e.g. iPads), SCAD is also advancing the development of inventive and flexible tools for accessing rich census data. This is a positive shift towards greater access to public data in the Emirate. The tools SCAD has developed for the 2011 Census use SAS as a foundation and include: on-line Thematic Mapping, on-line Community Tables, and on-line Table Builder. These tools will be released to the Abu Dhabi government and public in 2012 and 2013 as web–based applications.

INTRODUCTION

ABU DHABI

Abu Dhabi is the federal capital of the United Arab Emirates (UAE) and the largest of the seven emirates. Geographically, Abu Dhabi lies on the borders with the Kingdom of Saudi Arabia and the Sultanate of Oman and the Arabian Gulf (see Figure 1). Over the past 40 years, Abu Dhabi has experienced significant population growth and economic development [1].

To manage the growth and prosperity of the Emirate, the Government of Abu Dhabi has developed a long-term plan called 'Abu Dhabi Vision 2030' [2].   It sets guidelines for key infrastructure, tourism and financial targets for the next 20 years. To support Abu Dhabi Government's Vision 2030, and the general policy agenda, an official agency that could provide statistics for decision making and policy setting was seen as necessary [3].

SCAD

Statistics Centre – Abu Dhabi (SCAD) was established in accordance with Law #7 for the year 2008 [4] as an autonomous body which enjoys financial, legal and administrative independence as the main authorised body concerned with official statistics in the Emirate of Abu Dhabi. SCAD is responsible for the collection, classification, storage, analysis and dissemination of official statistics covering social, demographic, economic, environmental and cultural indicators.

As a young statistical office, SCAD is in the fortunate position of being able to review and implement best practices from international bodies and leading National Statistical Organisations (NSOs). SCAD is aiming to be a world leader in innovative and efficient methods for data collection, analysis and dissemination. SCAD uses the UNECE Generic Statistical Business Process Model (GSBPM) as an underlying framework for its overall statistical system, but is eager to explore and implement new and cost-effective ways of operationalising these processes.

SCAD recognises its responsibility in improving the quality, timeliness, fitness for purpose and accessibility of statistics produced in the Emirate and is building tools and strategies to meet these responsibilities. A major step in achieving this vision was the organisation of a population census for the Emirate.

ABU DHABI CENSUS 2011

The first official census of Abu Dhabi was conducted in December 1975, as component of the larger UAE Census. At that time, the population of the Emirate was recorded at 211,812 persons. Federal censuses were subsequently conducted in 1980, 1985, 1995, 2001 and 2005. In 2005, the census population was counted at 1,399,484 persons.

The 2005 census outputs were largely limited to hardcopy and PDF tables covering broad census indicators and limited geographically to the Emirate and its three main regions (Abu Dhabi City, Al Ain and Al Gharbia). Census users were restricted to pre-structured data tables for research or decision making. Ad-hoc census data requests were available, but were delimited to particular areas of government.

The Abu Dhabi Population Census 2011 was part of an integrated census programme to produce accurate base information for the Emirate of Abu Dhabi. The first phase of the programme, Frame Update 1 (FUP1), recorded all buildings in Abu Dhabi. The second phase, Frame Update 2 (FUP2), described the key characteristics of all business establishments and all building units used for housing. The final phase, Abu Dhabi Census 2011, provided accurate information on the number of people and their demographic, social and household characteristics.

The Abu Dhabi Census collected information about everyone present and/or usually resident in the Emirate on the night of 3 October 2011. The data collected in the Census will provide population indicators which will give government agencies and other users a firm basis for development and monitoring. The project will also create the base frame for sample designs for future social and household surveys.

The fieldwork for the Census was conducted over five weeks, starting from 4 October 2011. Administrative Information about people living in labour camps, restricted-access sites, islands and other special dwellings and places, was provided through October and November.

Data was progressively edited and coded using centralized teams during the November 2011 to March 2012 period. It was then analysed (macro analysis), prior to the preparation of final data outputs in mid-2012.

In addition to using innovative enumeration technologies, such as iPads, during the census, SCAD has also been advancing the development of inventive and flexible tools for disseminating the rich census data.

The primary aim of the 2011 Abu Dhabi Census output planning was to make large amounts of census data available to users via online statistical tools. The plan specifies the development of web-based tools to support all users of statistics (novice through to experienced) to access and create customized outputs. This is a positive shift towards greater access to detailed public data in the Emirate and sets a benchmark for future statistical releases.

THE ABU DHABI CENSUS 2011 OUTPUT PLAN

PREREQUISITES

Before embarking on user consultation and international reviews of census tools, SCAD identified some prerequisite features to be incorporated into the tools. These included:

EASE OF USE/ACCESS

The tools should not require training, nor detailed instructions for use. They should be intuitive and easy to understand by a range of user types. Public users should not need to register, nor log-in to the tools.

SPATIAL REPRESENTATION

The Abu Dhabi government is very spatially integrated and use of geographic information systems is common place. The census data outputs therefore needed to incorporate some form of spatial representation.

SENSE OF COMMUNITY

Given Abu Dhabi's significant infrastructure and residential development, the ability to know more about local communities through census data was seen as a priority. Therefore the tools need to be able to tell a story about a self-defined community.

EXTRACT AND TAKE-AWAY

As the Census 2011 data would be made available electronically for the first time, SCAD wanted users to able to easily extract and take away the data for further analysis.

CONFIDENTIALITY

While a key driver for developing on-line census tools was to disseminate detailed data, applying suitable confidentiality controls for protecting the identification of individuals, was paramount. This papers section on 'Confidentiality' outlines the methodology employed by SCAD for the Abu Dhabi Census 2011 outputs.

USER SKILL LEVELS

SCAD recognised up-front, that different user types with different skill levels would be utilizing the tools. Any development suite would need have a layered approach and consider novice users ("tourists") through to experienced users ("harvesters/miners") [5].

METADATA

As metadata about the Census variables was going to be captured electronically and centrally in another project, the re-use and linking of these metadata into the tools was practical. Additionally, the metadata would be contextually linked [5], so as not to overload the user with related information.

PUBLISHING STANDARDS

It was necessary that the outputs (tables, text, maps, etc.) generated by the tools were not only consistent in presentation, but they also conformed to SCAD's internal Statistical Output Guidelines[6], which are largely based on the UNECE making data meaningful guidelines [7].

LANGUAGE ACCESSIBILITY

Arabic is the official language of the UAE. However, most government services are provided in both Arabic and English. The tools needed to cater for both language groups.

USER REQUIREMENTS

EXTERNAL USERS

As limited data was available from Census 2005, finding experienced external users was difficult. The approach employed was to review requests for census related data from SCAD's Client Relationship Management system (CRM). All requests for data pertaining to population, households and buildings were extracted and reviewed. The analysis showed that users were often interested in general demographic information about a single small area (e.g. "population statistics on Liwa"), or they were seeking very specific information at the Emirate or regional level (e.g. "educational status of citizens by age group living in Abu Dhabi").

GOVERNMENT USERS

SCAD has a number of key agencies with which it has signed Service Level Agreements. These agreements specify the types of data that both agencies would like to share. That is, what administrative

data SCAD would like to receive form agency X; and what agency X would like SCAD to provide. A review of these agency data requirements helped inform both the functionality and type of census data to be included in the output tools. Additionally, a number of meetings were held with strategic agencies and their thoughts on appropriate tools and data were captured.

INTERNAL USERS

Feedback and experiences from SCAD's professional staff has helped shape and design the tools. Local Emirati staff, with experiences in previous censuses and knowledge of local practices, were fundamental in ensuring the design met with local ideals. For example, decisions relating to the minimum output geography and the validity of output variables were guided by local knowledge.

INTERNATIONAL REVIEW

The census outputs planning included a review of international on-line census tools. A recent UN Statistical Commission paper [8], which reviewed the 2010 World Population and Housing Census Programme, states that "only 17 countries (16 per cent) use interactive online databases" as "their primary method of census data dissemination".

Many of the on-line tools connected to these databases were reviewed. In some respects, the Abu Dhabi Census 2011 tools represent an amalgamation of the best features found in these census applications from around the world. The tools reviewed include: Statistics New Zealand's Table Builder, ABS's Table Builder, OESR's Community Profiles and Thematic Maps, OECD stat, and UN Stats Census Info.

DEVELOPMENT

TIMEFRAME

The development period for the Abu Dhabi Census 2011 output tools was very short when compared with similar international developments. The contract agreement with SCAD's technology partner (SAS) was ratified in January 2011. The planned development period for the census tools was completion in mid-2012. As this was a new arrangement between SAS and SCAD, the agreed direction was to pilot test some of the tool concepts on other datasets as 'quick wins' for government-to-government data sharing. A thematic map application using Census Frame Update 2 data and a Table Builder application using Foreign Trade data were initially developed. These tools were a 'proof of concept' that the census applications could be successfully built (whilst meeting the prerequisite census criteria) and at the same time the project delivered detailed statistical data to key agencies within Abu Dhabi government.

The 'quick wins' project was largely completed at the same time as the actual census was being conducted (October 2011). This gave the development team nine months to re-engineer the Thematic Maps and Table Builder to suit the more complex census data, and to develop 'from scratch' a new Community Tables application.

SAS PARTNER

SCAD's selected statistical software vendor, SAS, is a vested partner in this project. SAS is recognized as an analytics provider of choice for many statistical offices globally. The Census 2011 Outputs project relied on SAS's data dissemination capabilities. The Abu Dhabi Census 2011 tools are the first instance of SAS being used as the primary platform for dissemination of official census data in the Middle-East.

Throughout the development phases SAS and SCAD have worked closely together. Within SCAD, the Statistical Information Systems (SIS) team works with a team of SAS consultants with expertise in application and interface development.

APPLICATION ARCHITECTURE

The overall architecture is based on multi-tier applications - where user interfaces sit on top of J2EE application servers as the front-end, and census data resides in Oracle database tables as the back-end. SAS is used as the middleware which facilitates the secure data communication across the front-end and back-end layers through web services and OLAP cubes. See Figure 2 SAS Application Architecture.

CONFIDENTIALITY

THE METHODOLOGY

A SCAD working group was set-up to review and recommend the best approach for confidentiality. The groups' approach was to 'adopt and adapt' existing methods, drawing heavily on the data disclosure control work used by other NSOs such as Australia, New Zealand, Qatar, and the UK.

It was recognised that the proposed census tools would provide greater flexibility in the accessibility of data, but also added increased disclosure risk - particularly the Table Builder tool. The Table Builder tool allows the end user to select from a list of census variables and from a list of regions and then create a cross-tabulation of the result. This means that the user could potentially analyse census variables within small geographies.

The recommended and approved approach was a form of refined rounding - seeded random rounding to a base of 5. The intention of the approach is to 'hide' potentially disclosive values by either rounding them up or down to a 'base' figure. Additionally, the group recommended to add a 'key' (or 'seed') to reduce the possibility of disclosing a value through repeated analyses of the same cell, or through differencing by subsets.

Advantages

- Concept familiar and acceptable/understandable for users.
- Good level of protection.
- Extensively tested and used for census and other outputs.
- Raw data untouched.
- Seeding ensures cells are always consistent between analyses.
- Easy to implement using standard tools – SAS.

Disadvantages

- Post-tabular so requires treatment of every output (could result in delays in processing).
- Tables not additive.
- No protection for micro-data.

MINIMUM OUTPUT GEOGRAPHY

Another strategy employed to increase confidentiality was to define a minimum output geography based on an acceptable population and household threshold. A review of municipality boundaries (sectors) determined that some sectors had populations below a tolerable level for data dissemination. A threshold of ≥ 500 persons and ≥ 5 households was used to create new statistical sector geography. This merged non-compliant sectors with adjacent sectors until the threshold criteria were satisfied.

THE TOOLS

Following the review of user requirements and international tools, and the success of the 'quick wins' development - three tools were selected for dissemination of Abu Dhabi's Census 2001 results. These include: Census 2011 Thematic Maps, Census 2011 Community Tables, and Census 2011 Table Builder.   All three tools have been developed with dual language capability (English and Arabic).

ABU DHABI CENSUS 2011 THEMATIC MAPS

DESCRIPTION/FUNCTIONALITY

The Abu Dhabi Census 2011 on-line thematic mapping (or more precisely choropleth mapping) tool allows the user to quickly see census data represented geographically. This helps in recognising spatial variations and interrelationships of geographical distributions. 'Clusters' or 'hot spots' can be easily identified and users can undertake geospatial analysis of many demographic and social indicators at small area geographies. While standard annotated maps show where something is in a place, thematic maps tell a story about that place [9].

The tool includes standard mapping functions, such as: zoom levels, drag zoom, layer transparency, pan, export, and thematic colour palette. Detailed base maps, including satellite imagery, are provided through an Abu Dhabi government web mapping service (AD-SDI), as well as a Bing base map. Over 70 census variables, organised by theme, are available for selection.

The user steps for this tool are:

1. Select from a range of census variables (e.g. person 65 years and over)
2. Select a preferred geographic level (e.g. Statistical Sectors)
3. Generate the thematic map (selected data are banded into equal count quintiles)

See figures 3.1 to 3.4 for screen shot examples.

CHARACTERISTICS

Census variables: Seventy-one census indicators grouped under Demography, Education, Households, and Labour Force. The Persons data are based on place of usual residence (de jure).

Geography: Regions, Statistical Districts and Statistical Sectors will be provided to external public users. Statistical Sectors have population counts equal to or greater than 500 persons, and household counts equal to or greater than 5.

DEVELOPMENT TECHNOLOGY

To accommodate different user platforms (e.g. iOS, IE, Safari, Firefox, etc.) the thematic map application has been developed using both Flash and HTML5. This ensures that users with older browsers (e.g. pre-IE9) and users with newer browsers or mobile devices (e.g. iPad), can use the tool.

The Flash development uses the ESRI ArcGIS API for Flex in order to connect to the web mapping service. The API provides access to the map services using specific ActionScript components. The API is used to load the base map (terrain/street view) as well as providing basic navigation controls (e.g. zooming, area selection, etc).

The heat map (data map) overlaid on the base map is drawn using an ESRI GraphicsLayer. This layer allows drawing of any graphics (e.g. circle, polygons, etc.) and allows event handling (e.g. mouse click, mouse over, etc.). This layer is also used to render the different geography types (Statistical Sectors, Regions, etc).

The HTML5 development uses Canvas elements to draw graphics (e.g. circle, polygons, etc.). It also supports event handling (e.g. mouse clicks, overlay, etc.), similar to the Flash based GraphicsLayer. The application uses

Dojo and jQuery as the underlying JavaScript frameworks. Both provide core methods to register user events and to render user interface components.

SAS 9.2 is used as middleware which facilitates the secure data communication across the front-end and back-end through web services. Flash and HTML development generate and consume SOAP messages based on user actions (e.g. change of geographic level or census variable).

ABU DHABI CENSUS 2011 COMMUNITY TABLES

DESCRIPTION/FUNCTIONALITY

The concept of the Abu Dhabi Census 2011 Community Tables is to provide an instant statistical snapshot of a user-defined community. A statistical office can never know all the geographical areas of importance of its clients, so empowering users to define their own 'community of interest' was an important aspect of the tool. The tool allows the user to select one or more geographic areas from either a mapping interface, or a hierarchical folder structure. The aggregated area can be labelled with a user-defined name representing the new community.

The user can select from 24pre-defined census tables to form the basis of the output report. These tables were designed by SCAD's Population and Social Statistics department and represent the tables that best describe a community and reflect local UAE requirements.

The final community report can be provided in PDF, Excel, or HTML formats. The output report structure includes: a title page, table of contents, regions selected, census tables, and explanatory notes (metadata).

The user steps for this tool are:

1. Select community of interest (e.g. four Statistical Sectors in Khalifa City)
2. Select tables of interest (e.g. Buildings by Building Type, or Population by Education Attainment, Citizenship and Gender)
3. Provide a user defined name, select output type (PDF, Excel, or HTML), and generate report.

See figures 4.1 to 4.6 for screen shot examples.

CHARACTERISTICS

Census variables: Twenty-four census tables grouped under Dwellings, Demography, Education, Families and Households, and Labour Force. The Persons data are based on place of usual residence (de jure).

Geography: Regions and Statistical Districts will be provided to external public users. Statistical Sectors will be provided to government users. Statistical Sectors have population counts equal to or greater than 500 persons, and household counts equal to or greater than five.

DEVELOPMENT TECHNOLOGY

The Community Tables tool is a multi-tier application with the user interface sitting on top of J2EE application server (front-end) and all confidential data residing in back-end Oracle tables. SAS is used as the middleware which facilitates the secure data communication across front-end and back-end through web services.

Community Tables is a thin client application which is built on HTML 4.1 and CSS 2.0 specification, which will allow users to generate reports in pre-defined file formats and save them locally. The application is hosted on Oracle's Weblogic application server. As J2EE is a platform independent web framework, the Community Tables can be extended to any platform operating system without any changes in the application. The Community Tables user interface is built using J2EE Java Server Pages (JSPs) which enables the creation of dynamic

HTML content, by controlling the HTML generation through java infused inside JSP. Javascript is used by the application for client side features, to provide a complete web 2.0 experience.

Similar to the Census Thematic Maps application, the Community Tables region select mapping feature has been developed using both Flash and HTML5 for compatibility and cross-platform purposes. The application business layer handles all requests from the client and performs necessary action. The business layer implemented as J2EE servlets along with core java classes creates the necessary objects and stores them in session to be used by the JSPs. The business objects are the java beans which hold the complex data in ready to be used object structure.

Dataset design for Community Tables is based on SAS cubes created from underlying Oracle census tables. The tool uses multi-dimensional data cubes for pre-aggregation of data and fast processing of queries.

ABU DHABI CENSUS 2011 TABLE BUILDER

DESCRIPTION/FUNCTIONALITY

The purpose of the Abu Dhabi Census 2011 Table Builder tool is to provide clients with the highest degree of freedom in selecting and combining census data. The Table Builder application is an online tool which allows users to create their own census output tables by accessing variables (using hierarchical tree folders) contained in pre-defined census data sets and by modifying the table structure to suit their own requirements.

The user steps for this tool are:

1. Select census variables of interest (e.g. Occupation, Usual residence one year ago)
2. Design table structure (e.g. move variables between columns and rows)
3. Retrieve census data (database extraction)

This tool produces results that are fully confidentailised and includes a seeding algorithm that ensures that the same result is produced each time the same variables are selected - even when sub-sets of the results are produced.

The table builder also allows for easy export (Excel) of the results and offers some elementary data analysis options (e.g. row and column percentages).

See figures 5.1 to 5.4 for screen shot examples.

CHARACTERISTICS

Census variables: Over 100 census variables are available, not including the many hundreds of geographic locations which can also be selected. The variables are available by dimensions (data sets). These are: Persons, Households, Units, and Buildings. The Persons data are based on place of usual residence (de jure).

Geography: Emirate, Regions and Statistical Districts will be provided to external public users. Statistical Sectors will be provided to government users. Statistical Sectors have population counts equal to or greater than 500 persons, and household counts equal to or greater than 5.

DEVELOPMENT TECHNOLOGY

Similar to the Census Community Tables, the Table Builder application has been designed in a modular and multi-tier architecture with the front-end being a J2EE web application hosted in a Weblogic environment. The middle tier consists of SOAP Web services hosted in SAS environment, which interacts with Oracle database in the back-end. All the data exchanges between front-end and SAS are in the form of XML.

J2EE web application forms the presentation layer of the application. The User interface is developed with J2EE based Java Server Pages (JSPs) and the Business layer in core Java. The web service client creates an

abstraction for the presentation layer and manages all the communication from presentation to SAS middle tier. XML Parser module converts necessary information in the XMLs to usable value objects. XML parsers are based on JAXB which binds the data/information to the value objects. Dynamic Table creator will work with the presentation (JSPs) and generate the table structure from the value objects created by the XML Parser component.

The SAS middle tier consists of SAS web services exposed as SOAP end points from SAS stored processes. Web services collect the information about dynamic table creation and hands it over to the stored processes running in background. These stored processes extract the information from OLAP cubes and sends the response back to the front-end. This layer is a complete SAS environment which enables interaction between the user interface and Oracle database in the back-end in an independent manner.

CONCLUSION

SCAD recognises its responsibility in expanding and improving the quality, timeliness, and accessibility of statistics produced in the Emirate of Abu Dhabi. The suite of on-line tools, developed by SCAD using SAS to disseminate the rich Census 2011 data, is evidence of SCAD's commitment to this cause. This project has raised expectations of the types of output SCAD can produce and has set a positive benchmark for other future statistical outputs.

REFERENCES

[1] The Government of Abu Dhabi, Abu Dhabi Development Statistics 1960 - 2010, Statistics Centre - Abu Dhabi (SCAD), 2012

[2] The Government of Abu Dhabi, The Abu Dhabi Economic Vision 2030, Abu Dhabi Council for Economic Development, 2009

[3] The Government of Abu Dhabi, Strategic Plan of Statistics Centre - Abu Dhabi (2009 to 2013), Statistics Centre - Abu Dhabi (SCAD), 2009

[4] The Government of Abu Dhabi, Law No. (7) 2008 Regarding Statistics Centre – Abu Dhabi, Statistics Centre - Abu Dhabi (SCAD), 2008

[5] Siu-Ming Tam and Regina Kraayenbrink, Innovations in Official Statistics - Background paper presented to the United National Statistic Division seminar New York, Friday 20 February 2009, Australian Bureau of Statistics, 2009

[6] Gregory Pole, SCAD Statistical Output Guidelines, Statistics Centre – Abu Dhabi (SCAD), 2012

[7] United Nations Economic Commission for Europe,Making Data Meaningful Part 2 A Guide to Presenting Statistics, United Nations, 2012

[8] United States Census Bureau, Economic and Social Council Statistical Commission, Forty-third session - Report of the United States of America on the 2010 World Programme on Population and Housing Censuses, United Nations E/CN.3/2012/2, 2012

[9] Xin Liang, Yan Liu and Liming Zhu, A Component-based Approach to Developing Thematic Mashups, School of Computer Science and Engineering, University of New South Wales, Australia, National ICT Australia Ltd., NSW, Australia, 2015 [sic].

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Gregory Pole

Enterprise: Statistics Centre – Abu Dhabi

Address: P.O. Box: 6036 Abu Dhabi, UAE

Work Phone: +971.2.81.00.597

APPENDIX A: FIGURES

Figure 1: Abu Dhabi location

Figure 2: SAS application architecture
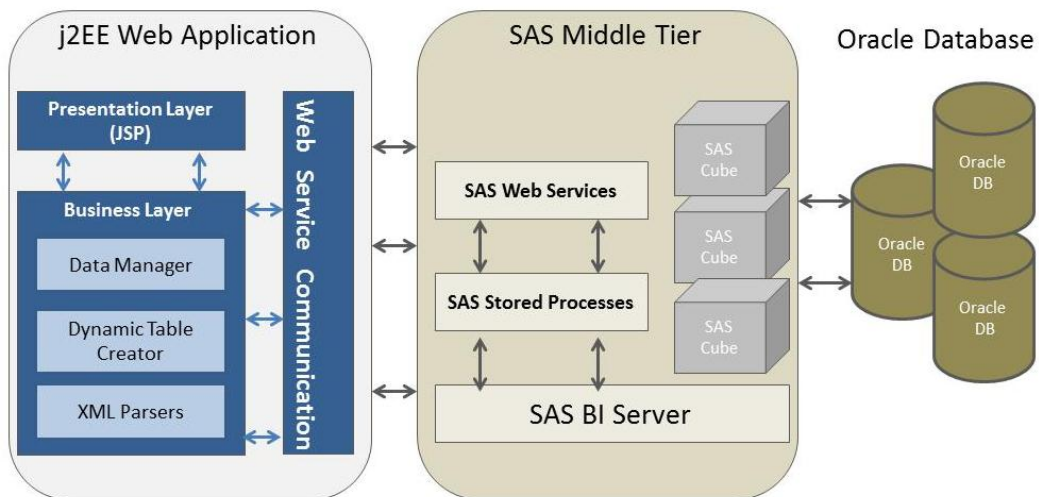
**Figure 3.1: Thematic Maps - data selection and base map**
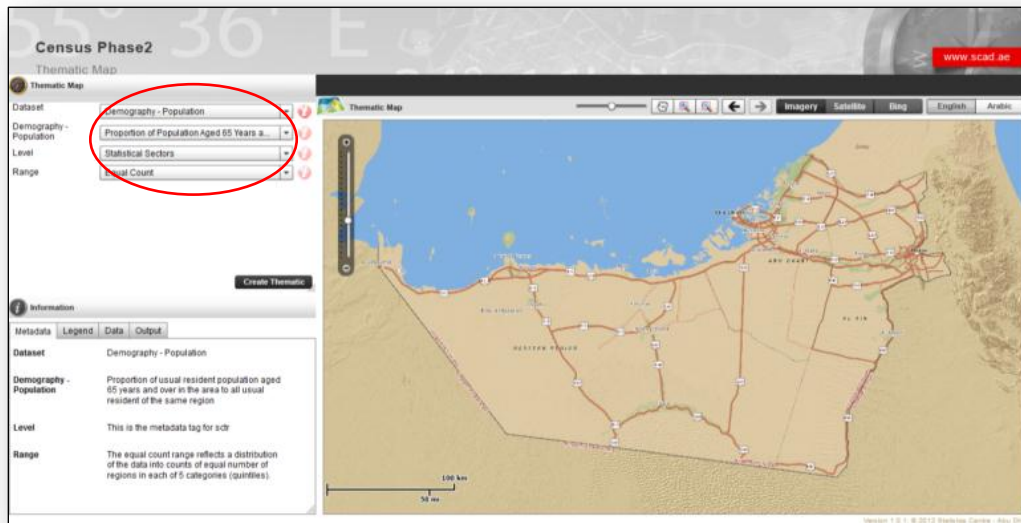

Figure 3.2: Thematic Maps - output layer and functional tools
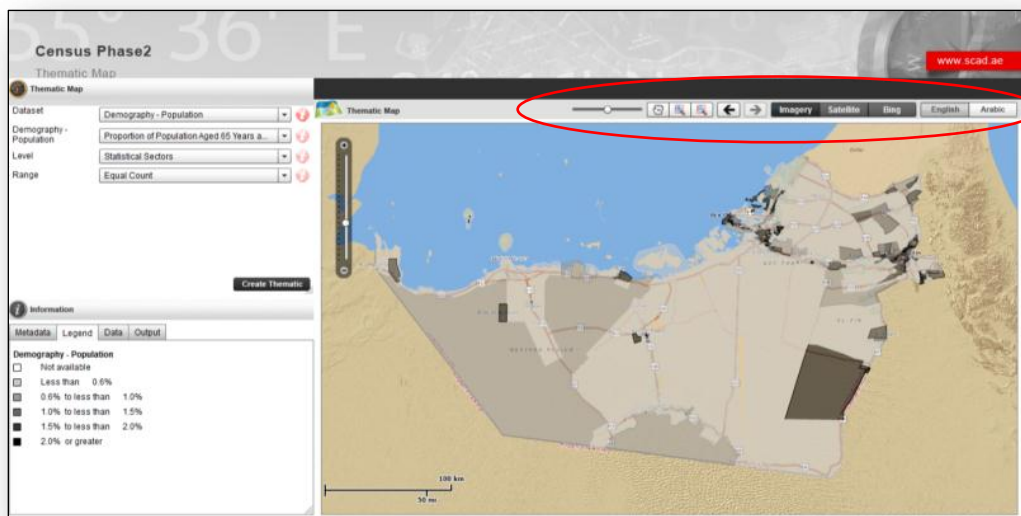
Figure 3.3: Thematic Maps - output layer zoomed
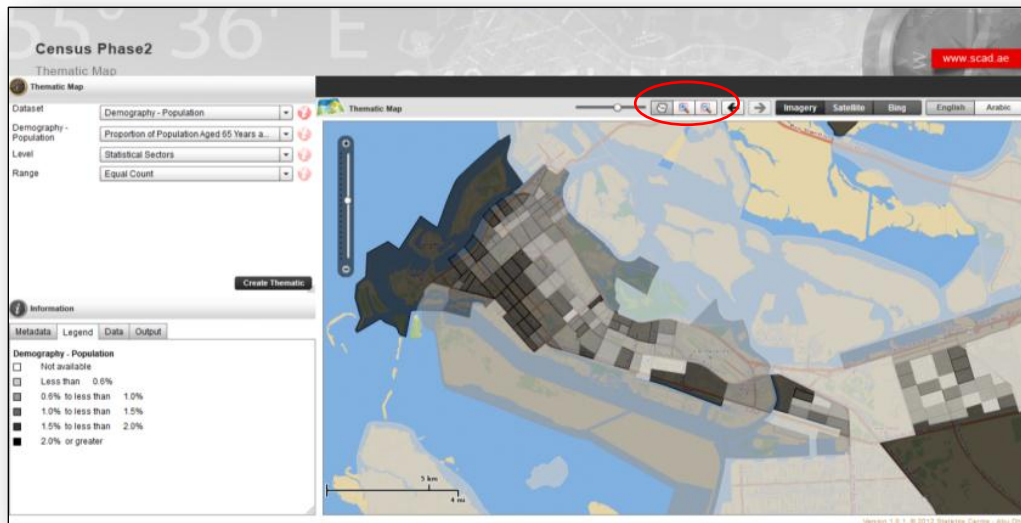
Figure 3.4: Thematic Maps –output layer with info box
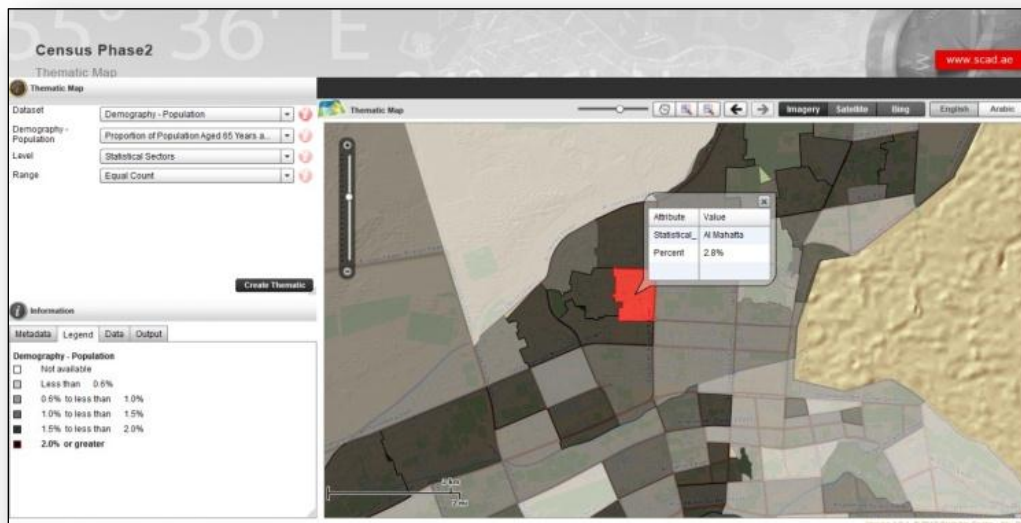
Figure 4.1: Community Tables - region selection (Step 1)
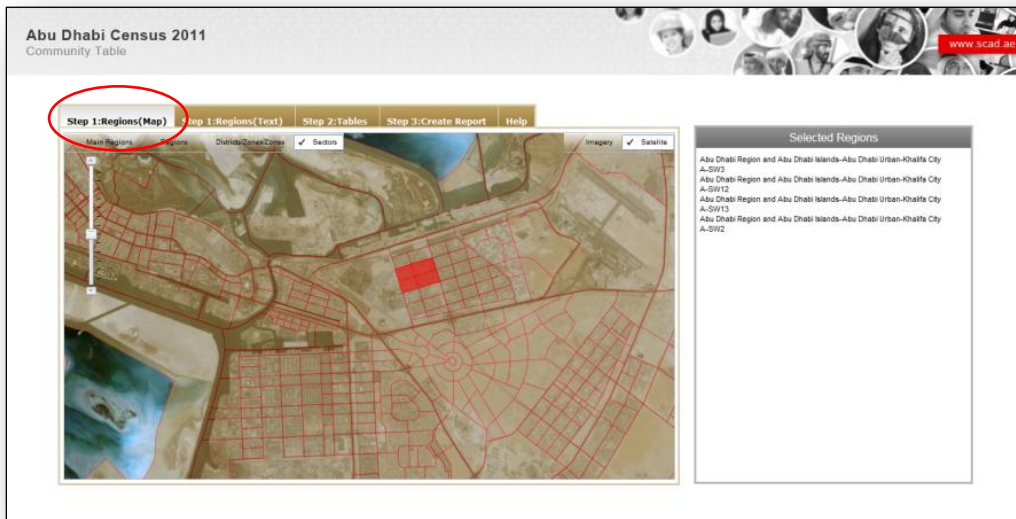
Figure 4.2: Community Tables – table selection (Step 2)
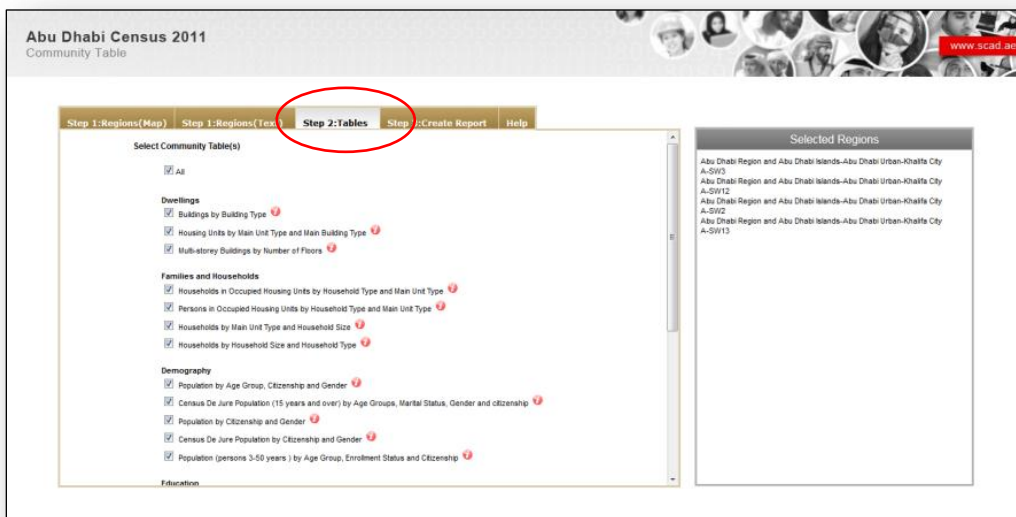
Figure 4.3: Community Tables – output selection (Step 3)

Figure 4.4: Community Tables – output report – title page

Figure 4.5: Community Tables – output report – tables (preliminary data)

**Figure 4.6: Community Tables – output report – tables (preliminary data)**

**Figure 5.1: Table Builder – variable selection**
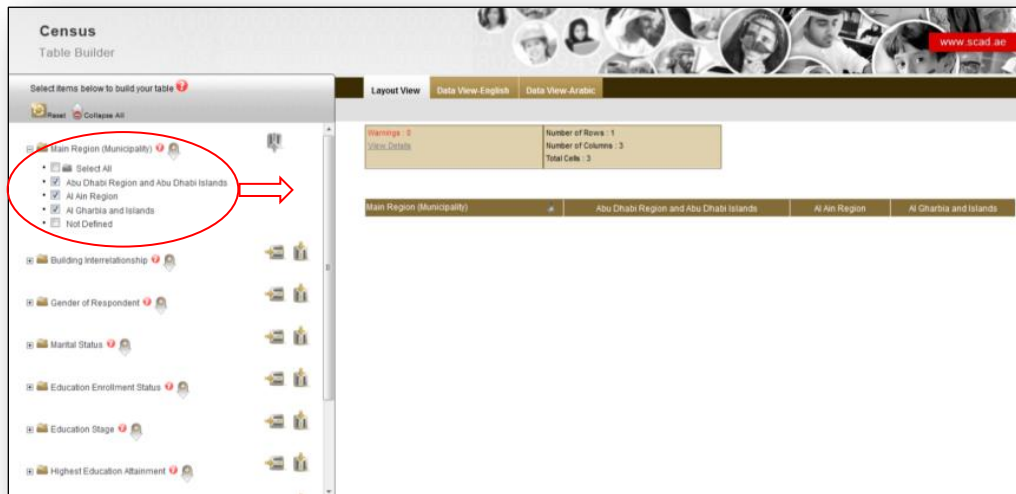
Figure 5.2: Table Builder – variable selection and table build
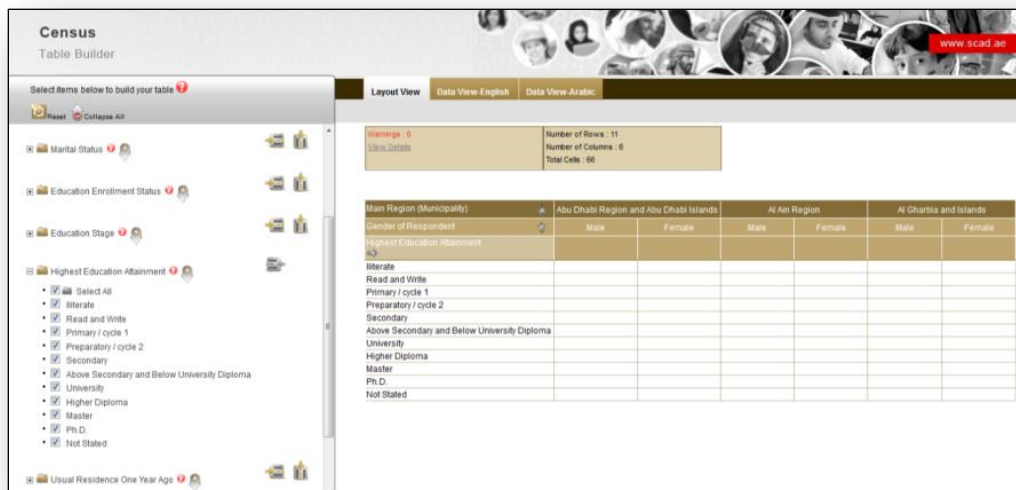
**Figure 5.3:Table Builder – data view – English (preliminary data)**

Figure 5.4: Table Builder – data view with percentages – Arabic (preliminary data)