**USE OF IT IN THE
AUSTRALIAN CENSUS OF POPULATION AND HOUSING**

Briefing for the Expert Group Meeting on
Effective Use of IT in Population Censuses
Bangkok 10-12 December 2007

Background

Australia's fifteenth national Census of Population and Housing was conducted on 8 August 2006, and the next will be conducted in August 2011.

2   The 2006 Census was notable for the amount and range of technology use, and the amount of innovation that occurred. This paper provides details of technology use in 2006, looks forward to innovations planned for 2011, and indicates ABS policy regarding experiences in sharing information with other countries.

IT Infrastructure for Census 2006

3   The Australian Census of Population and Housing is supported by a wide range of IT infrastructure which covers the field operations, data capture, processing and output phases of the Census. This infrastructure has been developed and extended over several Census cycles. The applications and services used are a mix of in-house developments and vendor supplied systems.

4   With a five year Census cycle, ABS re-uses large amounts of IT infrastructure from one Census cycle to the next. Each Census, some IT facilities are re-developed and some are created to support new business processes or to take advantage of new technologies which can provide process efficiencies.

5   In 2006, the new IT facilities were used in the following areas:

- on-line recruitment
- web-services for integration with the Census Inquiry Service
- web-based communication for Area Supervisors
- SMS based communication for Collectors
- eCensus
- a new workload management system to provide flexibility and efficiency in data processing
- Oracle data store for processing systems
- meta-data driven Edits library
- new methodology for data imputations
- web-based output products (QuickStats, MapStats, Census Tables, Table Builder, CDATA Online)

Geography

6   Commercial GIS software (MapInfo Professional) is used for delineation of the boundaries that are defined as part of the Australian Standard Geographic Classification. Collection District maps for collectors are also prepared using MapInfo together with in-house developed Map Basic programs. Maplex by ESRI is utilised for placement of text on the maps. Geographic information is stored in Oracle spatial and loaded from source data using SQL scripts and SAFE FME

7   Google Earth has also recently been used to provide access to imagery to provide a "reality check" against the mapping data ABS receives in vector format via mapping and land administration agencies.

Field Operations

8   For Area Supervisors and Collectors, applications for employment in 2006 could be submitted via a web-based facility as well as by phone and paper channels.

- The web-based channel was developed in-house using Lotus Domino.
- The phone channel was provided by a third party (ie outsourced) and application data was integrated in-house with data from the web-based and paper channels.
- The paper channel was captured in-house using ICR and imaging technology.

9   Area Supervisors were provided with a web-based system to obtain and provide information to operations staff. This included information on collectors, requests to the Census Inquiry Service, eCensus and mailback returns. In turn, Area Supervisors provided updates on workloads characteristics, form deliveries and collection activity. This facility was developed in-house using ASP.Net for the web-site and C#.Net for web-services to connect to an in-house Oracle data store.

10   The Census Field System provided processes to engage and pay Field staff, that is, Area Supervisors and Collectors. The Census Field System also provided processes to manage and allocate workloads for Field staff. This system is a set of Lotus Notes applications developed in-house for the 2001 Census and re-used in 2006.

11   All inquiries to the Census Inquiry Service which resulted in an action by a collector such as deliver an extra form, collect a form, not at home on Census night etc - were notified to the collectors via SMS (text messages). This mechanism was also used to notify collectors of the receipt of an eCensus form or a mailed back paper form.

12   ABS developed C#.Net web services in-house for Address Coding services and for SMS transmission to collectors. These web services operated between ABS and the third party vendor who operated the (outsourced) Census Inquiry Service.

Data Capture

13   Data Capture was conducted via two channels - the paper form and the eCensus. The paper channel was collected and returned to the Data Processing Centre for capture using ICR and imaging technology. The ICR and imaging functionality used is the Intelligent Forms Processing (IFP) application from IBM. This facility was operated in-house by the ABS in both 2001 and 2006. A similar application is used by the ABS for business based collections.

14   Data from the Census paper questionnaires was captured using 12 Inotec Scamax 510 document scanners. A total of 65.5 million pages were scanned over 71 working days. The images from the scanners were transferred to 18 form processing PCs where the IFP application processed the images using ICR technology. The resultant data and associated images were then loaded into Oracle and image data stores.

15   The eCensus facility was developed, hosted and operated by IBM Australia. The application was based on the Websphere application platform and hosted in an IBM data centre on dedicated P-Series servers. Both a Java script and HTML facility were provided to cater for accessibility requirements. All respondent data was encrypted from the respondent to the ABS and all points in between. Only ABS held the key for decryption. The eCensus facility was developed by IBM in partnership with the ABS with ABS retaining responsibility for functionality and approval for security processes. Data was delivered to the ABS every 8 hours, with notifications of completed forms delivered each hour. These notifications were used to generate the SMS (text) messages to collectors described above.

16   eCensus data was provided to the ABS in an encrypted format which matched the format of the ICR data from the paper channel. Images were generated for eCensus data. eCensus and paper channel data were then processed using the same facilities.

Processing Facilities

17   Processing systems have all been developed in-house (with the exception of an address matching algorithm from IBM). These systems include automated and manual coding, editing, imputation, validation and workload management processes together with management information systems. They have been developed over several Census cycles and so utilise a range of technologies including: C, C++, Centura, C#.Net, PL/SQL, XML, Lotus Notes and Oracle Discoverer. Facilities include a meta-data driven edits definition system, snippet and form viewing facilities, computer assisted coding facilities and an index management facility. There are also a range of administration systems covering human resource functionality such as employment, pay, leave and training.

18   The Census Data Processing Centre (DPC) contained it's own local computer data centre providing a range of Windows 2003 servers supporting the file and print services, Lotus Notes servers etc. The primary processing server for data capture and processing was a Sun E4900 running Solaris 8 with 4 dual core CPUs at 1.4Ghz each. It had an additional capacity-on-demand board which provided a further 4 dual core CPUs which was activated during peak production. This connected to approximately 500 desktop personal computers used by Census processing staff.

Census Product Generation (Output)

19   Data validation and aggregation are undertaken using SAS and SuperCross. Data are loaded into the ABS Information Warehouse and prepared into output products including QuickStats (a summary of key Census statistics for a selected area), MapStats (users can create their own thematic map of an area of interest), Census Tables (a quick and easy way to select individual Census tables) and Census Profiles (a comprehensive picture of an area through a series of standard tables). (see www.abs.gov.au).

20   QuickStats is a set of Lotus Domino static documents. MapStats returns data in a dynamically generated thematic map. Census Tables extract data dynamically and present it as an Excel file for download. Census profiles provide a wide range of Census data for an area in an Excel file. These products were developed jointly by ABS and Space Time Research (STR). The technology includes Lotus Domino, TomCat, STR web components and ESRI mapping services (ArcIMS and ArcSDE).

21   Products still under development include CDATA Online (a comprehensive range of data complemented by powerful web mapping application software) and Table Builder (a SuperWeb product which allows users to construct their own tables via an interactive web interface, using a database containing the 2006 Census Unit Record File). These are joint developments between ABS and STR.

Lessons Learned

22   In-house ABS IT services have been primarily deployed to those activities which are core to the ABS and to the Census program. Where there are solutions or expertise available from the broader IT industry, ABS has contracted third parties to provide these services and worked closely with the selected vendors.

23   Third party applications were used in 2006 on the following applications:

- Census Inquiry Service (call centre)
- Phone based collector applications
- eCensus
- Address Matching Algorithms
- Performance, Volume and Security Testing

24   For eCensus, the ABS lacked the experience, the expertise and the resources to develop and operate a one-off, large scale, secure, web-based collection system. The IT industry was able to provide this expertise and to work closely with ABS to develop the facility. This work included consideration of usability and accessibility requirements.

25   ABS also lacked the experience and ability to load test the system to the scale of estimated usage on Census night. Again, the IT industry has experts in this field and ABS employed their services to load test the eCensus, the on-line collector recruitment facility, the Area Supervisor web-based application and the Call Centre capability to manage calls and web-service activity.

26   Each load test uncovered issues which, had they not been found and addressed, would have caused significant systems failure when operating at peak capacity.

27   The third significant testing aspect required for eCensus was review and testing of system and data security. ABS undertook in-house security assessment of the application and infrastructure designs and implementations as well as operational procedures. ABS also engaged multiple third parties to undertake independent security reviews and vulnerability tests of the eCensus to ensure that the ABS could be confident of fully protecting the privacy and accuracy of respondent data.

28   Adoption of imaging, ICR and automatic coding technologies have resulted in significant productivity gains and better quality outcomes. However, for these technologies to be successful it's important that the following issues are carefully considered during planning for the Census:

- the design of the Census questionnaire must take into account the requirements for ICR while at the same time ensuring that the design does not make it more difficult for householders to complete,
- comprehensive testing of ICR is required and recognition software engines need to be tuned to maximise the recognition of characters while taking into account, and reducing the effect of, substitution errors (e.g. when a 4 could be misregonised as a 9),
- significant resources need to be devoted to building and testing indexes to support automatic coding. Comprehensive indexes that reflect current "real world" responses and society usage of different terms will ensure high automatic coding rates and eliminate false positive matches (i.e. when a response is automatically matched to an incorrect code), and
- indexes need to be continually updated and tuned during processing to maximise the benefits from this technology

Strategic Directions for Census 2011 and beyond

29   A strategic goal for Census 2011 and later cycles is to improve coverage of the Census as the attitudes and culture of the Australian public change over time and obtaining full participation in the Census becomes more difficult A key method of supporting this goal is to enable dynamic and efficient deployment of Field staff during the Delivery and Collection operational phase.

30   Census Collectors are allocated an area in which to deliver and collect forms and this allocation is quite static. If we are able to manage the workload of collectors more dynamically, resources could be deployed as required. For example, where householders choose to complete their Census form on-line, collectors will not need to return to those households to collect a paper form. If an area has a high level of use of the eCensus channel, the collector may have additional capacity and could be deployed to assist in another area.

31   Currently, direct communication with collectors is limited, as is real time management information about the collection activity - so it is difficult to tell where there may be additional capacity and where there may be areas in need of assistance. Training for Collectors is quite limited and direct communication with them is also limited. The current tools available to collectors are their paper Collector Record Book and their own mobile phone. ABS is continuing to review the use of hand held electronic devices to support the work of Census Collectors. In the first instance, this would most likely include real-time (or near real-time) electronic capture of Collector Record Book data. It could also extend to include electronic maps and GIS capability, dynamic task allocation, and facilitate direct communication with collectors.

32   During the testing cycles for Census 2011, ABS is undertaking trials of ePen technology. This technology utilises specialised paper and a pen which contains imaging technology along with standard pen and paper (ink) functionality. Data is still recorded in the paper record book, but as it is also imaged, it can be uploaded to the ABS near to the time of it's capture. This would enable more effective management of the Collector workforce and better information on coverage issues whilst staff are still in the Field and can take action to address the issues. The ePen option is also considerably cheaper than other hand held devices and would require less training to ensure its effective use.

33   In 2006, communication with Collectors was one way and not guaranteed to work. SMS messages have approximately a 7% failure to deliver rate. Whilst use of SMS was a good innovation for 2006, ABS is seeking to improve this communication. For 2011, it is expected that communication with collectors will primarily be via a web-based application - an extension of the facilities used for Area Supervisors in 2006. It is expected that use of SMS technology to communicate with collectors will diminish significantly.

34   A key mechanism to support dynamic allocation of Collector resources is to increase the level of adoption of eCensus by the Australian public. When households complete an on-line Census form, the Collector does not need to make a return visit to collect a paper form thus freeing up resources for re-deployment.   The ABS had a successful deployment of eCensus in 2006 with very limited advertising of the facility. The option to complete a form on-line was made known to the householder as the collector delivered their Census form. ABS achieved a 9.0% take up of eCensus in 2006 (776,000 forms). It is expected that the take-up can be increased to somewhere in the order of 30% for 2011 and higher again for 2016 with the gradual removal of the paper form channel.

Technical Assistance

35   The ABS provides capacity building in the core work of a national statistical agency (or members of decentralised statistical systems). The aim of this is to build the organisational capability of the agencies assisted through provision of training or advice. In providing capacity building assistance, the ABS wishes to establish good statistical policies and methodologies through transfer of ABS knowledge and practical skills.

36   The ABS very rarely undertakes statistical work program for other countries. Rather, the ABS aim is to help, guide, train and mentor local staff to undertake the work and to build capability for the future. Hence an important criterion which the ABS examines before responding to requests for assistance is the expected long-term benefit which will result to the statisticians in the recipient country from the help provided.

37   ABS technical assistance efforts can be delivered in the following ways:

- ABS staff undertaking statistical consultancies or conducting training courses in another country (generally short-term);
- ABS staff contributing to workshops or seminars arranged by international statistical organisations;
- ABS staff advising on planning for statistical programs;
- staff from other countries visiting the ABS for study tours or training;
- providing documentation on ABS practices; and
- contributing to development of resources, facilities and systems by international statistical organisations to provide assistance to countries in the region.

38   ABS is willing to consider proposals for assistance of the type listed above for countries in the Asia and the Pacific region. ABS is not funded and does not have the capacity to provide wide assistance, so the nature of the assistance sought, and the funding of ABS participation, are important considerations.

39   As a statistical agency, the ABS offers technical assistance, but does not undertake projects which require it to undertake significant levels of administrative process, such as the organisation of regional workshops. Note also that ABS responds to requests for statistics, but does not tender for technical assistance work.

40   A further important consideration for the ABS is the timing of requests for assistance. It is important that requests are received with sufficient time available for the assistance to be effective. In terms of Census assistance, it is also important that any assistance is not timed to coincide with the peak time for preparation and implementation of the Australian Census.

Conclusion

41   The ABS uses IT extensively in all phases of its Census processes. It has built up over time, an extensive range of IT applications, built partly in house, supplied partly by external vendors, and developed partly in collaboration with external suppliers. These have provided significant gains in efficiency and data quality.

42   ABS is willing to consider requests for assistance by Asia and Pacific NSOs in the ways indicated in paragraph 37 above and subject to the conditions indicated in paragraphs 35-40 above. Any requests for assistance should be addressed initially to:

> Director
> International Relations Unit
> Australian Bureau of Statistics
> Locked Bag 10
> BELCONNEN ACT 2616
> AUSTRALIA
>
> Email: anne.reader@abs.gov.au

44   Enquiries about any aspects of the content of this paper should be directed in the first instance to Denis Farrell, First Assistant Statistician, Population and Environment Statistics Division, Australian Bureau of Statistics at denis.farrell@abs.gov.au.

Australian Bureau of Statistics
December 2007