

The Australian 2006 Census and the Internet

Summary

1. The role of the Australian Census of Population and Housing is to accurately measure the characteristics of all people in Australian territories on census night. The census aims to provide high quality data for small geographic areas and for small population groups.
2. This paper briefly outlines the processes that were put in place to provide an Internet option to the Australian public at the 2006 Census, the experience with the solution and the range of issues and risks that needed to be managed.

Drivers

3. In Australia, there is a government policy that citizens should be able to undertake their business with government by using the internet. These expectations and the legal framework for the use of internet for government purposes are set out in the *Electronic Transaction Act 2001*.
4. The ability to shop, bank and complete a tax form on-line has engendered a community expectation that modern organizations should be providing internet facilities.
5. As well, the use of the Internet provided new opportunities to reach geographic areas and population groups that were difficult to enumerate using traditional census methodology that depend on face to face contact with a Census enumerator. These include people in secure apartment buildings, mobile population and people in very remote areas (an example in Australia is people in the Australian Antarctic bases) and the disabled population who have, in the past, had to rely on others to complete their census form. .
6. Provision of an internet form provides greater convenience for the public completing the census form. With the Australian "drop-off/pick-up methodology, people need to be at home to hand their form to the collector. With the internet they can complete and return the form at their own convenience.
7. The internet also can improve the perception of privacy of Census information. While the ABS goes to some trouble in ensuring that census enumerators do not work in their immediate neighborhood, this is not always possible. The internet provides another option for the return of the census form for people who have concerns about this matter.

Outline of the Australian e Census

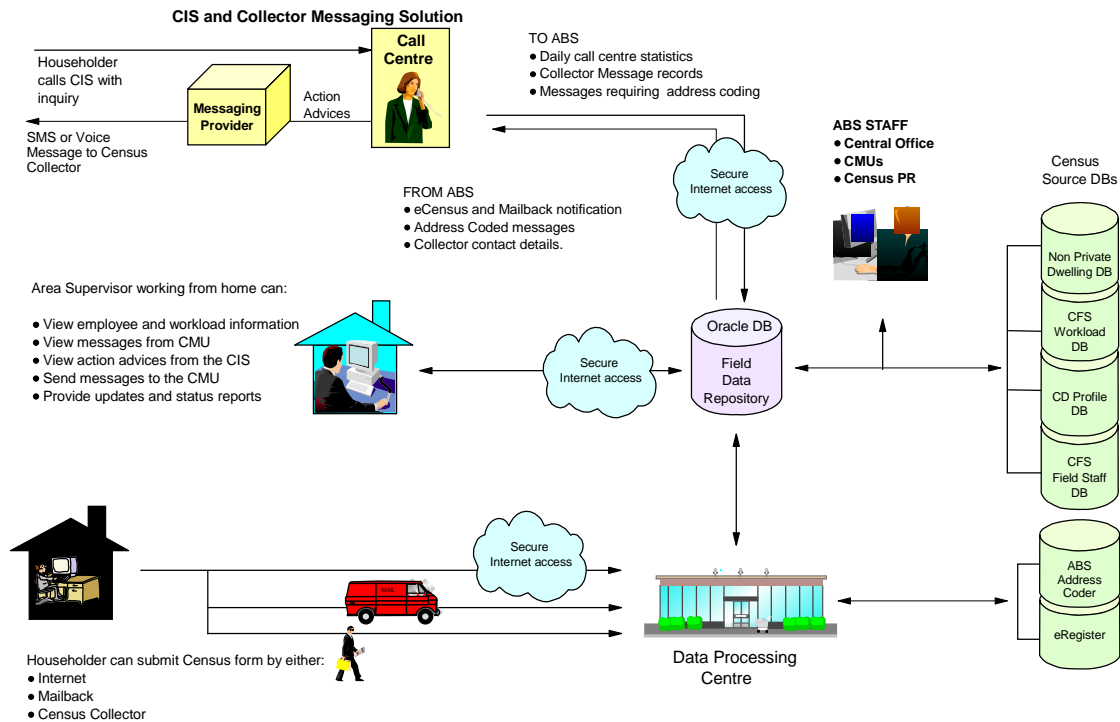
8. For the 2006 Census, Australia conducted a drop-off/pick-up census on a “de facto” basis, that is, for the purposes of the census count, people were counted where they were on census night not where they usually lived (usual residence data are derived from a usual residence question on the census form). A census enumerator delivered a census form to each address within a designated area and returned in the three weeks after census night to collect a form. The householder was encouraged to complete the census form on census night itself with reference to all people staying in the household that night.

9. For 2006 Census the householder was also given, in addition to the paper census form, a sealed envelope containing a unique 12 digit eCensus Number. If the person chose to use the Internet rather than return a paper form they could log on to the census web site and with a combination of the Census Form Number contained on the paper census form and the eCensus Number in order to gain access to the eCensus form.

10. The eCensus form was an online rather than a downloadable form. ABS examined the experience of other organizations in Australia and overseas that had used downloadable forms. The conclusion was, given the relative simplicity of the census form, an online form would be much more user friendly than a downloadable form.

11. Once the person has completed and submitted the eCensus form, a SMS (short messaging service) message is generated to the cell phone of the relevant census enumerator informing the enumerator that an eCensus form has been lodged and that the enumerator need not visit the house again.

12. This messaging system was integrated into an overall field communication system that also dealt with other census related inquiries from the public and the flow of management information between the census field staff and ABS (Australian Bureau of Statistics) management. An overview of this process is outlined in the diagram below:



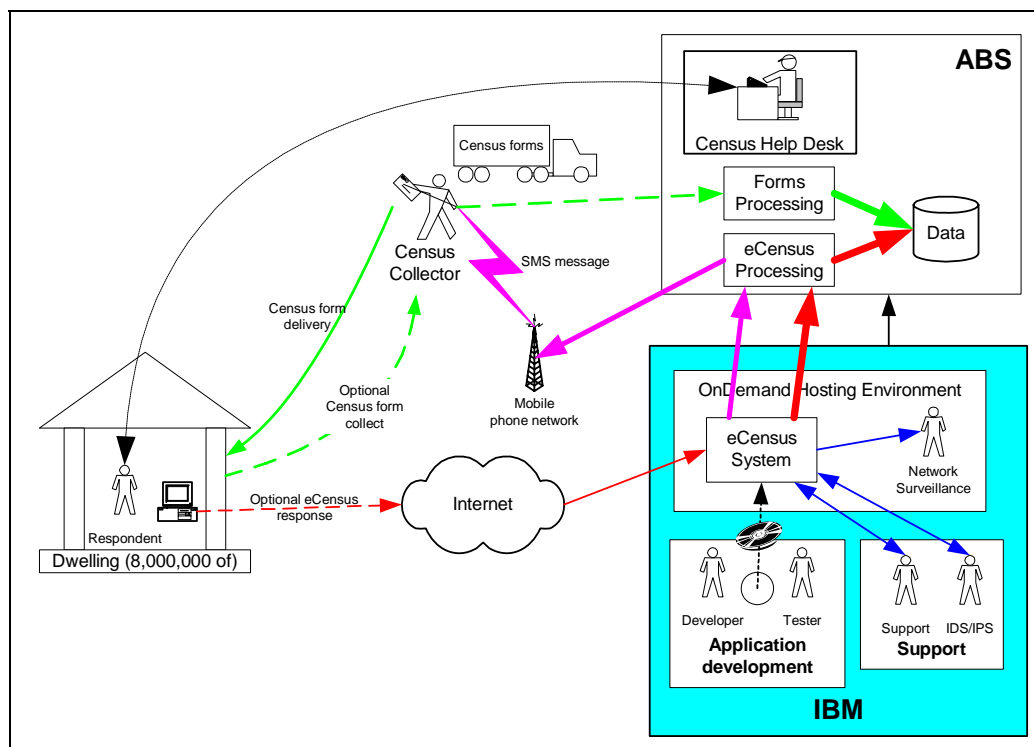
13. Individuals within households and people in institutional dwellings were also able to complete an eCensus form even if other members of the household or institution completed a paper form. Systems were also in place to monitor and handle duplicate forms from the one household (eg where there is both an eCensus form and a paper census form for the same household and/or people).

14. In addition to standard Internet security (the same as used by banks etc), Internet transmissions were encrypted.

15. The basic business processes and systems described above were devised by ABS staff and tested in a number of small field test using basic infrastructure developed by the ABS in-house. This was to ensure that any solution could be integrated seamlessly into the overall census processes. .

16. For the census itself, the ABS outsourced major components of the overall solution. These included the development and deployment of the Internet application as well as the Internet infrastructure. As well, the Census Inquiry Service and elements of the field communication system were also outsourced. The major rationale behind outsourcing was the requirement to provide for a high peak load capacity on and around census night for the eCensus and the need to have ready access to the communications networks. ABS lacked both the infrastructure and the expertise in these aspects of the operation.

17. The diagram below outlines the elements of the eCensus undertaken by the ABS and those undertaken by IBM.



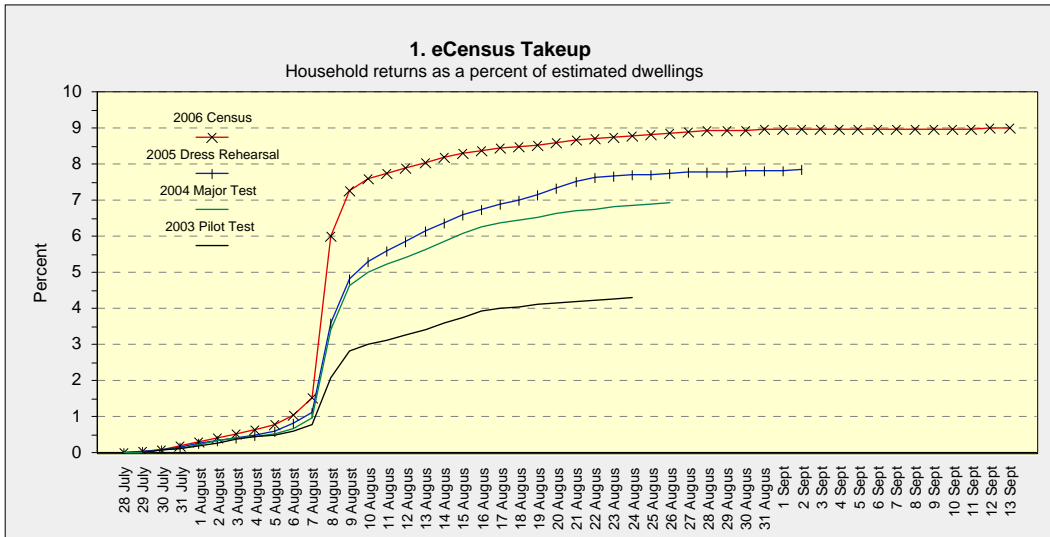
18. For 2006 Census, enumerators were advised by SMS messaging to their cell phone whether a census form has been received either by Internet or by mail. Field managers were also able to organize and undertake intensive follow-up for those areas with large numbers of forms not received.

Performance of eCensus

19. The 2006 eCensus system was opened to the public just after 8pm on 27 July, with enumerators due to commence delivery of forms on 28 July. The first eCensus respondent submitted their online form at 20:29 on 27 July.

Take up rate

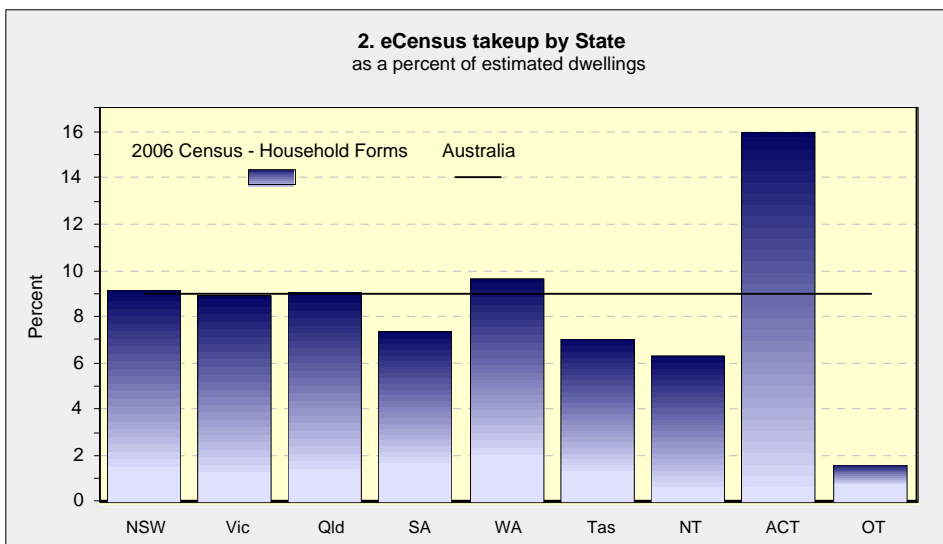
20. The number of Household forms submitted via the eCensus was **775,856**, representing take up by **9.0%** of the estimated dwellings in Australia. This take up is higher than the take up for the Dress Rehearsal. The final take up rate for the Dress Rehearsal was 7.9% of the 39,800 dwellings in the area.



Take up by State/Territory

21. The following graph shows the take up in each State and Territory. As can be seen, take up is significantly higher in the Australian Capital Territory (ACT) than other states. Of the more populous States, Western Australia (WA) is slightly above the national average, with South Australia (SA), Tasmania (Tas) and Northern Territory NT below the national average. The most populous states, New South Wales (NSW), Victoria (Vic) and Queensland (Qld) are about the national average.

22. This take up largely reflects internet connectivity in each State and Territory which is highest in the ACT (82%), and lowest in Tasmania and the Northern Territory (44% each). The lower take up in the Northern Territory will also be partly due to the fact that the eCensus was not offered in discrete Indigenous communities, which represent a significant proportion of the Northern Territory's population.



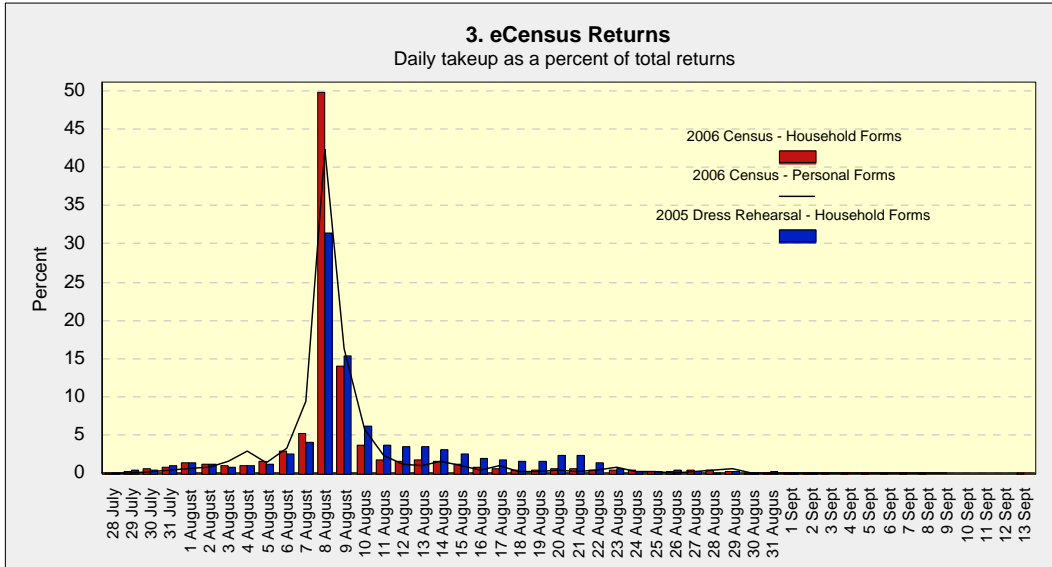
Submission pattern

23. The eCensus take up was characterised by a large spike in submissions on census night and a high level of use on the following day. Table 2 shows the distribution pattern for eCensus submissions across the different field phases, with comparison to the testing program. The most notable feature of this table is the proportion of forms submitted on census night, which at 40.5% was significantly higher than was seen in any of the field Tests (22-26%). There was also a lower proportion of submissions on the following day, and throughout the collection period.

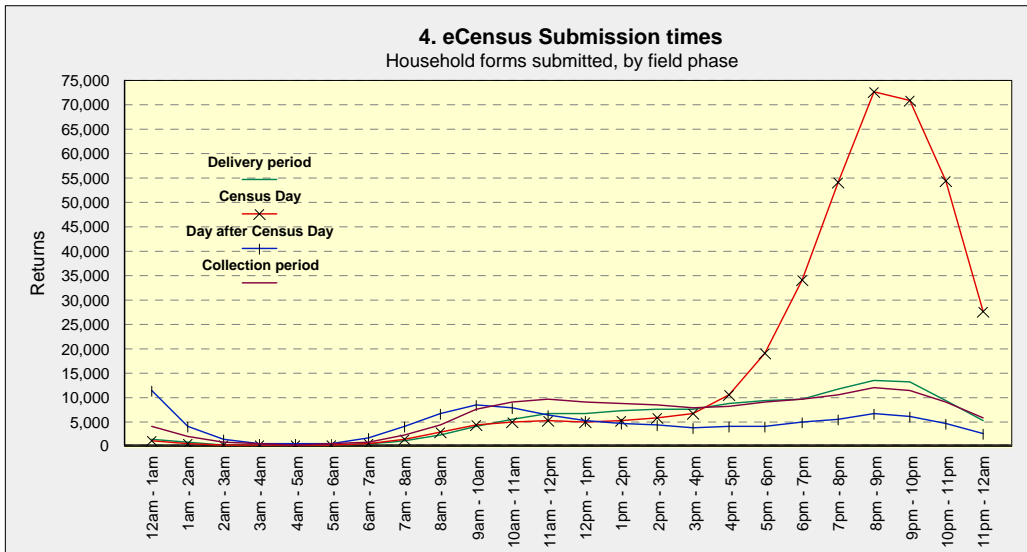
24. The different pattern seen in the census may be due to the awareness raising public relations campaign, which emphasises census night. It is also possible that the higher proportion of submissions seen in the collection phase of the Tests was affected by the type of staff employed as enumerators: it is possible that enumerators employed in Tests are of higher quality (due to the smaller number of positions to be filled) and are therefore more diligent in collection and follow-up.

Phase	2006 Census	2005 Dress Rehearsal	2004 Major Test	2003 Pilot Test
Delivery period	16.9	14.6	14.0	18.6
Census Day, prior to 6pm	9.3	6.9	9.4	7.2
Census Day, after 6pm	40.4	24.6	25.8	22.6
Day after Census Day	14.0	15.4	17.7	17.2
Collection period	19.3	38.6	33.2	34.4
Total	100.0	100.0	100.0	100.0
Estimated dwellings	8,617,000	39,800	20,600	6,450

25. The following graph shows the daily distribution of eCensus submissions. Comparative data from the Dress Rehearsal are also shown. Apart from the expected census night spike and strong activity in the 'shoulder' period (especially the following day), this graph also shows that a small proportion of people complete their forms on the Tuesday prior to census night (1 August), as well the last weekend on which enumerators were active in the field.



26. As noted previously, the time of day data shows a large peak in submissions in the evening of census night. The graph below shows that during the delivery period, there is increasing use throughout the day, leading to a peak in the early evening. census day itself shows consistent use throughout the day, with a large peak starting to build from 4pm onwards, tapering off after 11pm.



Note that the green 'Delivery period' and purple 'Collection period' lines represent an aggregate period, while the red 'Census Day' and blue 'Day after Census Day' lines each represent a single day's activity

Data quality

27. Evaluation of the 2004 Major Test revealed that people who used the completed the eCensus had a different set of characteristics from those that completed a paper form. They:

- were, on average, four years younger than paper form respondents;

- lived in larger households;
- more mobile: eCensus users were more likely to have a usual residence one and five years ago that was different to usual residence on Census Night;
- were more likely to be studying and participating in the labour force;
- had higher income levels;
- had higher levels of both school and non-school qualifications;
- were less likely to be providing unpaid care; and
- were more likely to have broadband access.

28. Data quality indicators from both the 2004 Census Major Test and the 2005 Census Dress Rehearsal revealed that eCensus forms had a lower item non-response rate than the paper forms and also had more questions completed (including questions often overlooked in paper forms such as “other” boxes with write in questions). While this may indicate some modal affect, it also may simply reflect the higher socio economic status and motivation of those who decided to use the eCensus.

Critical Success Factors

29. Some of the critical success factors are listed below:

1. Capacity of final solution.

30. The final eCensus solution used for the 2006 Census was sized to deal with up to 25% of the population submitting their census form online over the census period and with 20% doing so on census night. Extensive volume testing and systems tuning was undertaken prior to the census. Volume testing was also undertaken by an independent consultant to verify the figures being supplied by IBM. Volume testing covered both the ability of the systems and infrastructure to handle both sustained load as well as spikes.

31. During census operations, the eCensus system was closely monitored for any unexpected surge in usage. The capacity of the system was never really put to the test – with peak load on census night reaching only 15% of capacity.

32. There were contingency plans in case the capacity was exceeded. The first of these prevented any further people logging on to the system once a certain load was reached. This was set at a level that ensured that people already logged on could continue to complete their forms without any significant delay. Those people denied access would receive a message that the system was unavailable and to log on later. Public relations messages were also ready to inform the public of any capacity constraint on the eCensus site and that they

may need to postpone logging on. These contingency plans were never required to be implemented.

2. Perceived and actual security of the eCensus system and the internet

33. The eCensus security solution was compliant with the Australian Government Protective Security Manual (PSM) and government IT security guidelines contained in ASCI 33. It utilised proven security technologies such as multiple levels of firewalls, intrusion detection and protection devices and application servers built on hardened operating systems. The solution made extensive use of strong encryption to ensure the confidentiality of respondent data. Systems management policies were put in place to ensure that IBM employees had only tightly controlled access to eCensus infrastructure components.

34. In order to ensure that IBM staff were not able to access eCensus data all submitted and unsubmitted data were encrypted prior to being saved. The data were encrypted using a dedicated hardware device initialised with a private key by ABS staff. IBM did not have access to the private key. The encryption used is rated as secure until 2015 (that is using conceivable processing resources it would take at least until 2015 to break the encryption). Application code reviews and technical auditing were in place to ensure that IBM did not copy eCensus data in any form. At the conclusion of eCensus ABS security staff oversaw the secure deletion of eCensus data from all IBM systems. This deletion was carried out in accordance with Australian Government standards for the deletion of security classified data. Additional encryption techniques were used to protect the data while in transit over the internet or within IBM's network.

35. ABS IT Security section and the Australian Department of Defence were involved in the solution design. It was reviewed at numerous stages by an external security consultant. Prior to the system going live the Australian Department of Defence were engaged to conduct vulnerability assessments of the solution. External security consultants conducted detailed application code reviews and carried out active penetration testing.

3. The capacity of Australian ISPs (Internet service providers) to handle increased levels of Internet traffic on and around Census night.

36. An inability of Australian ISPs to handle high levels of Internet traffic generated by use of the eCensus could have severely impacted upon the take up rate of the eCensus. Should the services of a major ISP have been unavailable, a large number of people would have been unable to connect to the internet, and therefore unable to log on to the eCensus form. This may have resulted in respondents choosing to revert to a paper form, or not completing a census form at all.

37. A strategy for the notification of ISP's in the lead up to the 2006 Census was implemented, in an attempt to manage some of the issues arising from increased internet traffic generated by the eCensus. There was a problem at the start of the census operations with some smaller ISPs not including the eCensus

address in their indexes making it impossible for their customers to access the eCensus site.

4. Malicious attacks on the e Census web site

38. A denial of service attack would not only reduce the ability of people to lodge their census form on line but could also have had an adverse affect on the image of the ABS and of the census. ABS along with IBM implemented a range of measures to contain any such attacks. These include continuous monitoring to detect possible attacks with the options of locking out source ISPs for up to thirty minutes or provision of additional access routes to the census site. There were a few occasions where unusual levels of activity were detected from certain IP addresses that lead to these addresses being denied access.

39. There was also a contingency that, in cases where these attacks could not have been dealt with quickly, public relations messages would have firstly assured the public that their census information is secure and secondly provide information about alternatives such as delaying using the e Census system or using the paper census form. This contingency was not required.

5. Integration of eCensus into other census systems and procedures

40 The successful implementation of the e Census was reliant on the successful integration of the eCensus system into existing census systems and procedures. This was to both maintain the quality of the census and the cooperation and support of the public.

41. For census field systems, this included the successful operation of the Census Inquiry Service (that handled inquiries from the public about the census) and the census field communication systems, to ensure that every household and person was counted and that households that completed an eCensus form were not followed-up unnecessarily.

42. Changes were made to the census processing system to accept eCensus forms and integrate them into the processing streams. Extra checks were implemented for duplicated household and person records and a method has been implemented to determine which of the duplicated records was retained.

6. Ease of access and usability of the eCensus system

43. A key requirement for a successful eCensus was that the system should be readily accessible, easy to access and easy to use. All households were given an eCensus number that would allow them to access the eCensus system. The eCensus system was compatible with nearly all browsers – both commercial and open source and could be used via by both dial up and broadband connections. A great deal of attention was given to ensuring that the form itself was easy to use and extensive usability testing was undertaken. As well, a form that was compatible with the major screen readers used by the blind was also made available. In addition to help facilities built into the form itself, the ABS also provided a technical telephone help line through the Census Inquiry Service.

44. There was an opportunity on the eCensus form for users to provide comments. These comments were overwhelmingly positive, in particular about how easy the form was to use. There were a very small number of complaints from people who had ongoing difficulties with the eCensus system and it was not always possible to identify the source of the problem.

Towards 2011 Census

45. Technically, the eCensus worked extremely well and currently ABS envisages repeating the technical aspects of the solution – including the outsourcing of the infrastructure.

46. Decisions about the Census public relations campaign, publicity material and advertising had to be finalized well in advance of the completion of the eCensus system. A decision was taken not to actively promote eCensus in this material. For 2011 Census, the internet will be promoted as the prime means of taking part in the Census. This will be done as part of a move to a “mail-back” Census to replace “pick up” by Census enumerators.

47. Opportunities will be taken to reduce the need for paper forms (and therefore reduce costs). The current process of delivering paper forms to all households as well as offering eCensus increases costs of conducting the Census. One approach might be to encourage people to register in advance their willingness to use eCensus.

48. The 2006 eCensus form had limited editing – with only a few of the key demographic fields made mandatory. While the quality of 2006 eCensus forms appear to be much better than the paper forms, it may be possible to further improve the quality by more interactive editing while the form is being completed. This could range from undertaking some simple cross record checking (eg ensuring ages in the household are consistent with stated family relationships) to asking for more details about occupation or industry if the information provided was insufficient to accurately code. Any changes should not be at the cost of discouraging people from taking part in the Census.