UNITED
NATIONS

E

Economic and Social
Council

Distr.
GENERAL

CES/AC.71/1999/5
13 January 1999

ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Meeting on the Management of Statistical Information Technology
(Geneva, Switzerland, 15-17 February 1999)

Topic (iv): The technologies and methods in the implementation of the 2000 round of Censuses

THE 2000 ROUND OF CENSUSES - A REVIEW OF MAJOR ISSUES

Submitted by the Central Statistics Office, Ireland[1]

I. BACKGROUND

1. A joint ECE-EUROSTAT work session on Population and Housing Censuses took place in Dublin during 9-11 November 1998. The meeting consisted of three sessions devoted to the following major themes pertaining to the 2000 round of censuses[2], namely:

- collection issues e.g. consultation with users, questionnaire design and content, and geography;
- processing issues e.g. scanning and recognition, automatic coding and edit and imputation; and
- dissemination issues e.g. responding to users' needs, the balance between paper reports and outputs using electronic media.

2. The timing of the work session was particularly appropriate as it allowed a valuable exchange of views to take place on the current state of play in the member states. Many countries had either recently carried out

[1] Prepared by Aidan Punch.

[2] Not all countries will have to concern themselves with every aspect of the themes dealt with at the Dublin meeting. For instance, countries using a register-based approach to census taking, while having to face challenges specific to registers, will not have to contend with fieldwork related issues.

pilot tests or were in the final stages of completing them. This allowed experiences to be shared on up-to-date developments in the run-up to the 2000 round of censuses.

3. The attendance at the meeting consisted of about 80 persons representing over 40 countries. The majority of those in attendance were census managers/practitioners.
Ten invited papers were presented, in each case the author being allowed about 30 minutes for delivery. Each of the invited papers was followed by a formal response prepared in advance by selected discussants and a general discussion open to the floor. Time was also allowed for brief presentations and discussion of a range of supporting papers[3].

4. The purpose of the present paper is to highlight the main issues raised which need to be addressed during the course of the 2000 round of censuses, emphasising where relevant those topics in which the use of information technology plays a major role. As the paper is written from the perspective of a census manager it reflects the preoccupations of a user of IT services rather than a provider of such services.

5. In addition the paper draws on the up-to-date experiences of a number of countries as reported in the proceedings of a recent International Seminar on Census Methodology, held under the auspices of the Office for National Statistics of the United Kingdom, which the author attended in Portsmouth, England on 29 April – 1 May 1998.

II. LAYOUT OF THE PAPER

6. The paper commences with a consideration of some of the strategic issues confronting national statistical institutes (NSIs) as they prepare for the 2000 round of censuses. The next section deals with census collection topics. This is followed by a consideration of data processing issues including the implications of a wholesale changeover from key to disk data capture techniques to scanning and imaging. The final section deals with the dissemination phase.

III. STRATEGIC CONSIDERATIONS

7. In terms of scale and cost, a Census of Population and Housing represents the single most important operation which most statistical offices undertake. The conduct of the census puts a public spotlight on the statistical office while the results are not just of major public interest; they also provide the bedrock for many other statistical operations. Because of cost and scale factors, censuses provide many opportunities for statistical offices, which otherwise would not be available to them, such as investing in computer systems which will benefit the organisation in its entirety. However, on the debit side, there are also associated risks to be managed.

8. Given the timing involved (at the end of one millennium and at the beginning of another) the 2000 round of censuses will come to be seen as a watershed in the long history of census taking. The pace of technological change, more than at any time in the past, will have enormous implications for how we collect, process and disseminate census information. This consideration is independent of the relative size, capability and stage of development of statistical offices. Therefore, it behoves all of us to carefully assess our level of preparedness before deciding on the course of action on which we will

---

[3] All of the papers for the Dublin work session are contained in the ECE Statistical Division's WWW site, at the following URL:
http://www.unece.org/stats/documents/1998.11.census.htm

embark[4].

9. Many studies exist which examine how information systems (IS) are adopted by organisations in general in the attainment of business goals. Woolfe (1992) puts forward the following typical four-stage paradigm for the progressive adoption of technology:

- Stage 1: *Functional automation* in which IS automates the existing workflow within functional departments.
- Stage 2: *Cross-functional integration* in which IS integrates workflow across several business functions.
- Stage 3: *Process automation* in which IS enables the work of an entire cross-functional business process to be performed as a single entity.
- Stage 4: *Process transformation* in which IS enables business processes to be fundamentally re-designed and thereby conducted in ways that were previously impractical or impossible.

10. In the Woolfe classification stages 1 and 2 merely involve IS being used to automate business as usual while stages 3 and 4 involve the re-engineering of the organisation's business processes. Only minor organisational changes are required to carry out stages 1 and 2. The move to stages 3 and 4 however usually requires substantial changes in organisational structure, responsibilities and procedures. Clearly the major benefits are to be derived in trying to attain stages 3 and 4. However, the higher the stage being striven for the greater will be the challenge and the risk to the organisation.

11. The decision whether to adopt an incremental or "big-bang" approach to the introduction of increased technology in some or all phases of the census operation depends on a number of factors. The principal one is the frequency of census taking in the country. The approach followed in countries such as Australia, Canada and New Zealand tends to be incremental mainly because of the five yearly frequency of the censuses in those countries. On the other hand, countries such as the United Kingdom and the United States of America, with decennial censuses, tend to adopt more of a "big-bang" approach. A number of reasons can be put forward for this:

- The longer interval between censuses may render some of the approaches taken in previous censuses obsolete;
- With the passage of time the institutional memory diminishes and there may be little choice but to devise new methods of working;
- The longer lead-in time in planning afforded by a ten-year cycle allows more testing of new technology to take place.

12. The difference between the two approaches does not therefore necessarily imply that any one organisation is more risk averse than the other. Rather it is more a reflection of the realities facing organisations in trying to manage a large operation with the minimum of risk.

13. The scale of the challenge and unavoidable risk is also heightened where the IT environment may have been at a low level previously and where a decision is taken to catch up as quickly as possible. The transition from one level to another may simply be too much to take on board in one iteration. However, the temptation to have so-called leading edge tools at the disposal of the statistical office (especially as the census may be the only opportunity of getting access to such tools) may prove irresistible even though the associated complexity stemming from the technological environment

[4] In many cases this strategic assessment has been carried out some time ago. However, in some countries such considerations are still being addressed.

may pose a large challenge and risk to the census operation.

14. In many cases, even with adequate piloting, the census may turn out therefore to be the testing bed for much that is new in technology. This is somewhat ironic given that the risks involved are higher in a census than in most other statistical operations because of scale factors. Nolan (1998) cites the level of complexity introduced into the census process as one of the areas of major risk. This complexity may be as a result of striving for greater efficiencies and higher quality outputs. However, unless carefully managed, what may transpire may be the exact opposite of what was intended.

15. Another area of major strategic concern is the extent to which some or all of the processes involved in the census may have to be outsourced. This may arise either because the organisation does not itself have the necessary expertise in-house or alternatively because the cost of outsourcing may be cheaper than the homegrown option. With outsourcing comes an added layer of risk which has to be carefully managed.

16. The sums of money involved in outsourcing will normally be of such a magnitude that strict tendering procedures will have to be observed. As well as observing the strictures that accompany tendering, organisations will be very anxious to ensure that “all the angles are covered“. However, even with the best insurance (e.g. penalty clauses) it would be erroneous to assume that because an operation is outsourced the relevant risk is outsourced as well. The responsibility and accountability for the census operation and all of the attendant risk will always reside with the statistical office.

17. In the remainder of this paper no distinction is made between processes which are outsourced and those which the statistical office itself undertakes.

IV. COLLECTION ISSUES

IV.1 Consultation on topics

18. A necessary pre-cursor to any census is the need to consult the public about their requirements. This includes both the topics to be covered and the outputs to be produced. Such consultation is desirable both from the point of view of getting public support for the census and also ensuring that the topic coverage and output products continue to remain relevant, the latter especially in the context of a rapidly changing technological environment.

19. In terms of topics to be covered, the scope for additional subjects is sometimes limited when account is taken of the need to keep the burden on respondents to a minimum and the necessity to cover the basic demographic and social questions. Some of the newer topics, which have come to the fore as a result of recent consultations, are income[5], disability, carers, and unpaid household work. Before being included in a census proper these newer questions have to be carefully piloted in a census test.

20. Income is likely to be the most problematic of these new questions because of its sensitivity and the possible negative impact it could have on response rates. Indeed, the impact of the inclusion of income in the census can only really be tested (on a sample basis) under actual census conditions. This is because the non-statutory nature of census tests does not allow the real effect of the inclusion of income to be assessed (see Hicks and Zelenbaba 1998 for a discussion of the Canadian census test carried out as part of the 1996 census).

---

[5] A question on income has been included in the censuses of a number of countries for some time. However, for others the 2000 round will be the first occasion for such a question to be included.

21. From time to time it is necessary to amend the wording of an existing question. This can arise when the existing wording is no longer relevant. The more prevalent reason for re-formulating questions in recent censuses has been to facilitate response and processing through the use of tick boxes. An example in the case of Ireland is the question on Irish language, which was asked unchanged over a long period in censuses[6]. Following consultation with the main users, it was decided to change the previous version of the question, which involved a write-in response, to a tick-box layout. While the results were in accordance with those derived from the piloting of the question, they did give rise to a discontinuity between 1996 and previous censuses. However, the assessment was that the time had come to introduce a new version of the question even if this gave rise to a break in continuity in a long-standing series.

22. Similar situations arise where new classifications are being introduced (e.g. occupation, socio-economic group), sometimes to facilitate automatic coding. The need to strike a balance between long-term comparability and relevance is one which has to be confronted as we move forward and it may not always be possible to maintain continuity with the past while ensuring that classifications retain their relevance.

IV.2 Questionnaire design

23. Once the questionnaire content has been decided, its layout has to be considered. These two facets are not independent, of course, as the layout of the questionnaire has to take account of the number of questions, their formulation as well as ease of response and processing.

24. The move towards scanning and recognition techniques, as the main means of data capture, places an even greater discipline than heretofore on questionnaire design. Critical issues are the use of colour, particularly dropout colours, the size of the tick boxes and the spaces between them. The necessary professional advice has to be sought at the questionnaire design stage to ensure that where scanning is used the relevant images are capable of being properly recognised. The impact of scanning on census processing is considered more fully in the section on processing.

25. A basic concern is whether the form is arranged in a matrix fashion (i.e. questions along the columns and persons along the rows, (or vice versa) or whether a number of pages is devoted to each person in the household (the so-called page(s) per person). Questions relating to dwelling characteristics have also to be included. Both versions are acceptable although there appears to be a preference in recent censuses for the page(s) per person layout. Even in this situation a choice has to be made between a booklet type questionnaire and separate questionnaires (normally double sided A3) for dwellings and persons. An advantage of this latter approach (which is followed in France and New Zealand for instance) is the avoidance of bulky 5 or 6 person booklets thereby saving on paper (and cost of printing). The major disadvantage is the need to ensure that the enumerator clearly codes the individual forms for a particular dwelling in such a way that they can be linked together for the purposes of family and household characteristics.

---

[6] In the wording of the question used in censuses between 1961 and 1991 respondents were asked to write, “Irish only”, “Irish and English”, “Read but cannot speak Irish” or leave blank as appropriate. In the version used in 1996 respondents were asked to tick a Yes/No box in response to the question “Can the person speak Irish”. In the case of those who indicated that they could speak Irish the frequency of speaking the language was also sought (i.e. daily, weekly, less frequently, never).

26. Where an individual form is used it may prove tempting to try to cover all the relevant questions on one double-sided A4 page. This constraint will undoubtedly curtail the number of questions to be asked or the layout of the questions to be included. The booklet layout on the other hand allows an uneven number of A4 pages (e.g. 3) to be included for each person. The requirement for a sufficient number of pages per person has become more relevant with the greater use of tick box questions, designed to ease the burden on respondents while at the same time simplifying processing.

IV.3 Census coverage

27. The predominant definition of the population to be included in the census is that of the usually resident or *de-jure* population. Using this method persons have to be enumerated in their usual residence regardless of where they are on census night while all persons temporarily in the country at the time of the census are excluded. Clear operational rules have to be set out on the census form to cover instances where a person may have more than one residence. Particular difficulties may be caused in correctly accounting for school-going students who are away from home during term-time. The length of time a person may be in the country or intends to remain in the country will also call for clarification.

28. At the other end of the spectrum is the *de-facto* measure of the population. In terms of implementation, this is very simple; persons are enumerated where they spend census night. Therefore, usual residents who are absent from the State on census night are omitted while visitors are included. Family statistics will also be affected by the absence of persons elsewhere in the State on census night, unless of course an attempt is made *post-hoc* to locate these persons with their families.

29. A half-way house is implemented in New Zealand where in addition to using the *de-facto* measure a number of basic demographic questions are asked about usual residents who are absent on census night. These questions typically include name, sex, date of birth, family relationship and whether in or out of the country on census night. This enables a usually resident count to be provided while also maintaining family structure intact.

IV.4 Geography

30. For the purposes of the enumeration the country in question has to be sub-divided into mutually exclusive and exhaustive areas which are usually called enumeration areas or districts (EAs or EDs). The primary concern in the selection of these areas is that they represent manageable workloads for enumerators. Where possible regard should be had to administrative boundaries when EAs are being devised. In determining workloads account has to be taken of factors such as the amount of travel which may be necessary, the difficulty of enumeration (especially in inner city areas and in apartment blocks where access is restricted) and whether the populations are settled or transient. Particular problems arise in the enumeration of institutional or non-private households.

31. A balance has to be struck between determining the exact scope and content of EAs in sufficient time before the census and yet having the most up to date listings of buildings likely to be occupied around census time. Rural enumeration areas in which little new building has taken place therefore tend to be decided first while the last areas to be determined are those where there is building activity just before census time.

32. EAs normally represent the lowest level of geography distinguished in a census. However, because of changes from census to census due to the depopulation of some areas and the development of new population centres it is

difficult to maintain an unchanged set of EAs over time. The ability to geographically code households through the provision of the appropriate x,y co-ordinates or at the very least to code households at the level of grid squares will mean that output geography will be independent of the areas selected for enumeration. Where such is not the case then an attempt should be made to minimise the change in EAs from census to census.

33. Where they exist, national address databases exert a major influence on how censuses are organised. Where the address database is both up to date and comprehensive the possibility of mailing out questionnaires may be an option. However, because of the time necessary to prepare dwelling listings etc. the address database will have some level of obsolescence built into it. The alternative course of action being used in a number of countries is to provide each enumerator with a relevant address list for his/her enumeration district. This list would then form the basis for the delivery of questionnaires with care being taken to ensure that erroneous entries are corrected and that missing addresses are added during the course of the census fieldwork. The census geography will benefit greatly where the address database is already geo-coded at the dwelling level.

34. One of the greatest criticisms from enumerators in the past related to poor quality and out of date maps. This problem has been alleviated through advances made in electronic mapping which have greatly facilitated the provision of tailor made maps of appropriate scale for each enumerator.

35. On the output front, with the availability of affordable GIS software packages, census data are now typically bundled with such software thereby enhancing the usability of the data.

IV.5 Delivery and collection of forms

36. The delivery and collection methodology for census questionnaires has undergone critical reappraisal in a number of countries in the course of recent censuses. The long-standing approach was for enumerators to deliver blank census forms before census day and to collect the completed forms as soon as possible afterwards. This procedure is particularly labour intensive, especially the collection phase of the operation where repeated call-backs may add significantly to the workload.

37. Because of difficulties in recruiting enumerators, alternative delivery and collection methodologies have had to be explored. As mentioned earlier, where up-to-date national address registers exist these tend to be exploited during the delivery phase of the enumeration. Typically, enumerators are supplied with a list of addresses for their enumeration district. These addresses have to be validated and supplemented where appropriate during delivery.

38. However, it is the collection phase of the enumeration which places the greatest strain on the field operation. The growth of one-person households as well as the greater emphasis on security and privacy by tenants has resulted in mail-back finding favour with the census authorities of a number of countries. While for some countries (e.g. UK) the approach is being implemented for the first time in the 2000 round of censuses, in others (e.g. Canada) it has been in operation since 1971. Despite the fact that mail-back was considered a feasible option in Australia it has been decided to adhere to the traditional collection system, which because it worked well did not need to be changed (see Skinner, Struik, Davis and Lamb 1998).

39. A major advantage of the traditional collection system is the direct ownership and accountability engendered by making enumerators responsible for all aspects of the enumeration in their allocated area. At the conclusion of

the collection stage, completed census forms can be arranged systematically by street, enumeration district etc., thereby ensuring localised control over the coverage and comprehensiveness of the enumeration. Following extraction by enumerators a quick preliminary headcount of the population can also be easily organised. However, the greatest advantage of the direct collection of completed census forms by enumerators is the opportunity it affords for the scrutiny of the information on the doorstep[7]. Once the enumeration has concluded the scope for such checking is greatly diminished[8] and resort has to be had to statistical techniques for imputation etc.

40. One of the shortcomings of the direct collection methodology is the perceived difficulty with confidentiality where completed census forms are being collected by enumerators who may be known in the neighbourhood. Because it will be necessary to recruit enumerators with knowledge of the local area this is a price which has to be paid. The alternative is a complete mail-back approach.

41. Where mail-back is being organised a decision has to be taken whether this will be done on a centralised basis or to regional collection centres. The use of the latter will allow information to be fed back more rapidly to the field to facilitate the follow-up of non-respondents. Workloads will be more manageable while corrective action can be taken swiftly where particular local problems arise. Because of the volumes of mailed back questionnaires, particularly during the period immediately after census day, and given that field follow up has to take place in a relatively short space of time, automatic systems of recording returns are essential.

42. The major plus of the use of mail-back is that processing can begin immediately a batch of forms arrives in the central (or decentralised) processing centre without having to wait until all forms are properly sequenced and bundled. The saving may be as much as two months, which in the context of tight production time-schedules is not insignificant.

V. PROCESSING THE CENSUS

V.1 Data capture

43. The predominant method of data capture used in censuses up to and including the 1990 round was to key the information directly to disk. This approach was both costly and time-consuming. For example, in the 1996 Irish census it took 120 staff trained on keyboard skills 3 months to capture (i.e. key and verify) 1.1m forms at an average rate of 6,500 key depressions per hour.

44. As reported in Suharto 1996, just over a quarter of the 178 countries, which carried out a census as part of the 1990 round, made use of scanning and optical mark reading (OMR) techniques. Some countries used recognition of marks only (i.e. tick boxes) while others recognised marks and numeric data. Only a few attempted the recognition of alphabetic characters. Therefore, at an overall level the use of imaging, scanning and recognition is relatively recent and is being used for the first time by quite a few countries as part of the 2000 round of censuses.

---

[7] Opinion is divided on this. For instance, the Japanese experience supports the contention (see Ishida 1998) while the pilot test carried out in Northern Ireland offers a refutation (see Caven et al., 1998).

[8] Although see Hicks and Zelenbaba 1998 for the use of telephone follow-up of failed edit questionnaires. In the same paper it is admitted that telephone follow-up is becoming more difficult with the passage of time.

45. It is widely accepted that recognition systems are more accurate and cost effective today than they were say, 10 years ago and that by the year 2000 entering data into computer systems will no longer be the bottleneck that it once was (see Woods 1995). Notwithstanding the progress made in computer technology over the past 20 years the rate of adoption of recognition systems appears to have been slower than expected. However, this is confidently expected to change in the coming years.

46. Statistics New Zealand was one of the first statistical offices to use imaging and scanning. From their perspective (see Statistics New Zealand 1996) the main benefits were:

- Publication time-lag reduced by 3-4 months
- Cost savings in data capture
- Reduction in paper handling and storage
- Easier access to responses on forms
- Quality can be systematically controlled
- Processing does not need to be carried out in a specific order
- Reduction in staff to be recruited and trained.

47. Using the long-standing key to disk method of data capture, the main processing stages of coding, keying and editing were of necessity both rigid and highly sequential. A criticism of the method (see Blum and Ben-Moshe 1998) was that it was not possible to determine from the computer file exactly what was on the questionnaires[9]. By way of contrast, in the new technological environment, the order of the processing is not constrained by the need to handle paper questionnaires. This means in particular that, where mail-back is being used, questionnaires can be imaged and scanned as they arrive in the scanning centre, thereby saving up to 6-8 weeks in the publication time schedule[10]. Capturing on a free-flow basis implies that clerically compiled preliminary results are no longer an option. However, this may be a small price to pay if the definitive results can be made available on a significantly more timely basis than was possible heretofore.

48. As elaborated in Blum and Ben-Moshe, the implementation of imaging and scanning imposes its own set of disciplines. Careful consideration has to be given to the texture and thickness of the paper used in the questionnaires. The background colours and the colours used for printed text and characters to be optically read also need to be selected with due care and attention. Where the questionnaire is in booklet form, the guillotining should be done away from the scanning equipment to avoid dust particles impairing the quality of the scanned images.

49. Clearly, from the point of view of scanning, closed tick-box questions are to be preferred to written responses. The accuracy levels achieved in scanning numerical data have been reported widely (see for example Blum and Ben-Moshe, and Ishida). What is less clear is the extent to which written character strings will be scanned in the coming round of censuses. It is no co-incidence that these strings occur in what have been recognised as the most difficult questions to process on the census form (e.g. description of occupation, employer’s name and address). While from a scanning perspective, capitalised responses in constrained boxes may be preferred, the difficulty which this causes for the respondent should not be lightly ignored. There may,

---

[9] This is not entirely true in the case of the 1996 census held in Ireland in which all fields, other than qualifications and employer’s business and address, were keyed. In particular the actual descriptions entered for country of birth and occupation were keyed and automatically coded.

[10] This approach requires that every page of every questionnaire contains relevant identification variables.

therefore, be a need for some level of keying from image depending on the scanning throughput achieved for alphabetical characters at the pre-determined accuracy level specified.

V.2 Automatic coding

50. Because of the volumes involved, particularly for the larger countries, and given the difficulty of coding responses to questions such as industry branch and occupation, a number of statistical offices in the past confined the coding operation to a sample of the responses received. In some cases a long form was administered to a sample of respondents while in others all respondents were asked to complete the same form and a sample of questionnaires were coded for the “difficult” questions. Both approaches were unsatisfactory from the point of view of providing small area data. The possibility of recognising written responses coupled with automatic coding techniques now affords the opportunity of 100 per cent processing for all variables.

51. Recent experience in the use of automatic coding has been impressive. In the case of the 1996 Irish census the approach was used successfully for coding occupation (see Keogh 1998). All occupation strings were keyed. The occupation classification used was a derivative of the UK Standard Occupation Classification (SOC) comprising 346 unit groups. Just over 60 per cent of the 2.1 million responses to the occupation question were accounted for by as little as 2.5 per cent of the unique descriptions returned. It was decided to code these 14,500 descriptions using expert coders. This meant that 60 per cent of the file was coded using direct string matching. Of the remaining unique strings, 43 per cent were coded using the Precision Data[11] software package operating in batch mode. By re-coding a 1 in 40 random sample of these the accuracy level was assessed as 85 per cent. The final 20 per cent of the file was coded with Precision Data operating in computer assisted mode. The overall accuracy level was assessed at 95 per cent, well in excess of the accuracy level achieved using clerical coding.

52. Where the volumes are larger (e.g. for countries such as US and UK), the objective would be to code as many as possible of the descriptions using either direct matching or automatic coding operating in batch. Therefore, in order to minimise the volume of codes for computer assisted coding using operators it may be necessary to trade off on the accuracy level to be achieved. This will entail striking a balance between staff costs and time available on the one hand and users’ expectations on the other.

V.3 Edit and Imputation

53. There is a concern to ensure that the data we publish are plausible and internally consistent. This implies that corrections may have to be made to the information collected. At the most basic level the response to a particular question must be within the range of allowable values. Also there needs to be consistency between the responses to different questions in respect of the same record and between different records within the same household or family.

54. Devising edit rules can be a complicated process. The more detailed the rules the greater will be the intervention in the data set with consequential cost and timeliness implications. There is also a danger of imposing the statistician’s view of the world on the data set. Statistical data editing has

[11] Developed by Inference Group of Australia.

been the subject of specialist study for quite some time[12]. A multiplicity of different editing techniques have been reported on. Some of these are being assessed by statistical offices for use in the 2000 round of censuses (see for example Vickers and Yar 1998 and Valente and Massimini 1998).

55. While there appears to be agreement on the principles involved in data editing (i.e. minimum change to data and operational efficiency) there is not unanimity in how the adhere to these principles.

56. Another area where the practice tends to vary from country to country is the use of imputation, especially for missing values. While most countries impute for missing values of the basic demographic variables, for the remaining variables there is a divergence in the practices used. Some countries, among them Australia and New Zealand, explicitly show “not stated” as a category in published tables[13] while others impute for all missing values. The practice adopted may well depend on the feedback received from users.

VI. DISSEMINATION ISSUES

57. The major challenge facing statistical organisations for the next round of censuses is the timely provision of accurate results. Without exception, users are calling for the results of censuses to be made available more quickly and in a more flexible fashion than heretofore. The dissemination area is therefore the one where gains in technology will be most conspicuous to the public.

58. Some of the goals being set by statistical offices for their censuses are extremely ambitious. To quote a few:

- The Australian Bureau of Statistics is planning a significant improvement in timeliness over 1996 while maintaining data accuracy at the high levels achieved in that census (Skinner, Struik and Nauenburg 1998)[14].
- Statistics New Zealand have set as their goal to have most of the standard tables available within 12 months of census day (Nolan 1998).
- The goal of the US Bureau of the Census for 2000 is more timely delivery and more access to a wide variety of users (Bounpane 1998).

59. A major question exercising the minds of census planners at present is the appropriate mix of outputs to meet the needs of users, both existing and emerging. A feature of past censuses was the large volume of printed output generated. Statisticians tried to cater for users’ needs by compiling and printing as many as possible of the tables they thought users would require. However, because of the difficulty involved in table layouts and pagination for printed reports this concentration on printed output merely added to the delays in making census results available.

60. In order to respond adequately to the calls from users for more timely data, greater use will have to made of technological means of dissemination than has been the case in the past. This is not to say that printed reports will not continue to have a role to play in disseminating census results. Rather, the balance will have to switch from the printed medium to electronic

---

[12] For example, a UN Work Session on statistical data editing was held in Cork, Ireland in October 1994.

[13] Ireland also shows “not stated” explicitly. The extent of non-response varies by question. Typical values in 1996 were: industry 2.3%, occupation 3.2%, level of education 3.3%, Irish language 5.5%.

[14] By any yardstick 1996 was already a major success as final results for 80 per cent of topics were produced inside 11 months with the remainder inside 15 months.

and other forms of publication. Such printed reports as will be made available will tend to be less detailed than before. Maintaining printed reports, albeit on a more limited basis than for past censuses, should adequately cater for the needs of those requiring historical volumes to put beside the volumes from previous censuses. It should also cater for the user as yet unacquainted with modern publication methods.

61. Knowing exactly when to consult users in a rapidly changing technological environment has been raised by Vukovich (1998); too early and technological advances may have overtaken the plans; too late and there may not be sufficient time to put the necessary preparations in place. The provision of data on diskette and CDROM will undoubtedly continue to have a major role to play. However, it is generally recognised that the Internet will provide the greatest scope for innovation in disseminating the results of the 2000 round of censuses[15].

62. As a specific session in this conference is being devoted to the impact of the Internet on statistical dissemination it is not proposed to dwell on the issue here. However, the exciting proposals of the US Bureau of the Census should not go unremarked. These include pre-determined table shells which can be populated according to the user's choice of sub-population and geography and an internet query system capable of supplying real time special data requests. Given the transferability of technology, many statistical organisations will have more than a passing interest in being able to replicate such developments.

63. Other issues of concern in the dissemination area, which are outside the scope of this paper, are charging policy, the role of data intermediaries, the provision of micro-data files to accredited users and protecting statistical confidentiality, regardless of output medium.

VII. CONCLUDING REMARKS

64. Censuses of Population and Housing are large and complex operations in which most of the processes involved benefit enormously from the application of technology. The present paper attempted to highlight the major issues arising in the forthcoming round of censuses, emphasising in particular those where it was considered information technology had the greatest role to play. While the paper is written at a general level it is hoped that the issues raised will help inform the ongoing dialogue between census managers and information technology specialists to ensure a successful outcome to the 2000 round of censuses.

---

[15] The ramifications of using the Internet in the census collection process have yet to be fully worked out. However, those conducting censuses around 2005/2006 will have to address the issue.

REFERENCES

Blum, Olivia and Eliahu Ben-Moshe, 1998. "When Image Met ASCII: Data Processing in the 1995 Census of Population and Housing in Israel", Joint ECE-EUROSTAT work session on Population and Housing Censuses, Dublin.

Bounpane, Peter, 1998. "Data Dissemination plans for the United States Census Bureau for Census 2000", Joint ECE-EUROSTAT work session on Population and Housing Censuses, Dublin.

Caven, Norman, Trevor Evans and Robert Beatty, 1998. "The Northern Ireland Census of Population and Housing: Looking Back and Looking Forward", *Journal of the Statistical and Social Inquiry Society of Ireland*, Vol. XXVII, Part V.

Hicks, Doug and Milena Zelenbaba, 1998. "Meeting Census data collection challenges (in Canada) as we move into the 21st century", Proceedings of 1998 International Seminar on Census Methodology organised by the ONS in Portsmouth, UK.

Ishida, Yashuo, 1998. "Planning Japan's 2000 Population Census", Joint ECE-EUROSTAT work session on Population and Housing Censuses, Dublin.

Keogh, Gerard, 1998. "Automatically Coding Occupation Descriptions from the 1996 Census of Population of Ireland", NTTS '98, International Seminar on New Techniques and Technologies for Statistics, Sorrento, Italy.

Nolan, Frank, 1998. "Planning for the Census of Population and Dwellings in New Zealand". Proceedings of 1998 International Seminar on Census Methodology organised by the ONS in Portsmouth, UK.

Skinner, Tim, John Struik, Anthony Davis, and Andrew Lamb, 1998. "Data Collection Methodology (in Australia)", Proceedings of 1998 International Seminar on Census Methodology organised by the ONS in Portsmouth, UK.

Skinner, Tim, John Struik and David Nauenburg, 1998. "Processing Methodology in Australia)", Proceedings of 1998 International Seminar on Census Methodology organised by the ONS in Portsmouth, UK.

Statistics New Zealand, 1997. *Imaging the 1996 New Zealand Census of Population and Dwellings*, December.

Suharto, S., 1996. "Innovative techniques used in the 1990 round of population and housing censuses". INNOTEC, The Hague, 22-26 April, as reported in Valente, Paolo and Massimini, Giorgio, 1998. "Processing the Italian Population and Housing Census Data", Proceedings of 1998 International Seminar on Census Methodology organised by the ONS in Portsmouth, UK.

Valente, Paolo and Giorgio Massimini, 1998. "Processing the Italian Population and Housing Census Data", Proceedings of 1998 International Seminar on Census Methodology organised by the ONS in Portsmouth, UK.

Vickers, Paul and Mohammed Yar, 1998. "The development of the donor edit and imputation systems for the UK 2001 Census", Proceedings of 1998 International Seminar on Census Methodology organised by the ONS in Portsmouth, UK.

Vukovich, Gabriella, 1998. "Dissemination issues in the 2001 Hungarian Census of Population and Housing Programme", Joint ECE-EUROSTAT work session on Population and Housing Censuses, Dublin.

Woods, John, 1995. *Recognition Technology for Data Entry – a Guide and*

Directory. Cimtech Limited, University of Hertfordshire Press.

Woolfe, R., 1992. *The Path to Strategic Alignment*, Amsterdam: CSC Index Indications, Vol. 9, No. 2.