Statistics New Zealand
TATAURANGA AOTEAROA

Social and population statistics architecture for New Zealand

Christine Bycroft

newzealand.govt.nz

This information paper is one of several that look at how to meet future information needs for social and population statistics. Statistics New Zealand is not committed to adopt any findings, methodologies, or recommendations outlined in the papers.

All papers are available on the future approach to social and population statistics page on the Statistics NZ website (www.stats.govt.nz).

Citation
Bycroft, C (2011). *Social and population statistics architecture for New Zealand.* Wellington: Statistics New Zealand

ISBN 978-0-478-35384-6 (online)

Published in July 2011 by
Statistics New Zealand
Tatauranga Aotearoa
Wellington, New Zealand

Contact
Statistics New Zealand Information Centre: info@stats.govt.nz
christine.bycroft@stats.govt.nz
Phone toll-free 0508 525 525
Phone international +64 4 931 4610
www.stats.govt.nz

Abstract

This paper sets out a proposed future direction for social and population statistics in New Zealand across the Official Statistics System.

The social and population statistics architecture provides a high-level view of the primary data sources and the ways in which they can be combined to meet information needs in the future. Information needs change over time, and we can expect continued demand for more detailed, complex, and comprehensive data. These demands must be met within constraints imposed by funding, by respondent burden, and by privacy concerns.

This proposed architecture is a continuation and development of the existing model of a Census of Population and Dwellings and household sample surveys supported by administrative data. There is no paradigm shift; rather it is an evolving system.

Each of the primary sources has limitations. The census reaches the whole population but is infrequent and must be simple enough for self-completion. Sample surveys can provide depth, but not detailed breakdowns for subgroups. Asking respondents directly produces high quality information, but imposes costs on both the survey agency and the respondent's time. Administrative data incurs no new respondent burden, but is limited to information related to the specific administrative purpose.

The solution is two-fold: to be as efficient and effective as possible in conducting the census and household survey programme; and to combine administrative and survey sources using statistical techniques that make up for the relative weaknesses of each source through the strengths of the other.

Rather than each survey being treated as a stand-alone survey, sample surveys will be managed as a coordinated and flexible Integrated Household Survey programme across the main survey vehicles. Administrative, survey, and census data will become more intertwined through using statistical modelling techniques and increased data linkage. A strong emphasis is placed on making more use of administrative data.

The main alternative option to this model is to replace the census with a register-based system, where many administrative datasets are linked to a reference population register. This register-based system is not seen as a viable solution in the medium term because no population register currently exists in New Zealand. While some countries successfully operate a population register, the public and political acceptability of establishing such a register in New Zealand is unknown. Any proposal to move towards this approach in the long-term would require changes to cross-government infrastructure, substantial investment, and a long development time. The proposed model does not rule out being able to progress to a register-based statistical system should this be considered in the long term and, in any case will advance our understanding for future decisions.

Contents

List of figures

Figures by chapter

1 Introduction

Defining social and population statistics

Social and population statistics are statistics about people. Social statistics, along with economic and environmental statistics, form the three major groupings of the Official Statistics System. While there is considerable overlap among the interests of these groups, it makes sense to consider them separately from a structural perspective because the units about which information is collected are different. The units of interest for social statistics – individual people, families, and households – are all based on the person, while the economic statistics unit model is based on businesses, and environmental statistics relate to physical entities (eg land, water, energy).

Programme of Official Social Statistics

Development of a systematic programme of social statistics was seen to have lagged behind that for economic statistics, and this was recognised by establishing the Programme of Official Social Statistics (POSS) in 2003. POSS was provided with 10-year funding from 2005/06, to establish a coherent system of official social and population statistics across the government sector. The POSS work programme includes consolidating existing surveys into a managed programme, introducing new surveys to fill information gaps (eg the General Social Survey) and exploiting other data sources such as administrative databases.[1]

Statistics architecture

POSS also recognises the need to develop a long-term view of the overall design of the data sources that contribute to meeting the information needs for social and population statistics. We do this through developing a social statistics 'architecture', following the success of a vision for economic statistics architecture that was used to guide the development of economic statistics (McKenzie, 2008). A statistical architecture is a tool to help structure thinking, and guide forward-looking and cohesive development of a statistical system to meet specified goals, within constraints (Cornish, 2007).

This paper sets out a high-level view of a social statistics architecture for the future. It is deliberately directed towards a fairly simple overview, as this is seen as being more likely to be understood by a wide variety of stakeholders, and allows for flexibility in working out the details of how it might be implemented. The architecture also sets limits. It is as important to say what will not be done, as for what is included. One aim is to avoid constantly re-visiting basic questions, which is both costly and impedes progress.

The architecture is designed to meet the key enduring information needs for social and population statistics within constraints imposed by the system, and is aligned with government goals, with the vision for the whole Official Statistics System, and with Statistics NZ's strategic direction. It provides an authoritative direction that can be used for planning work programmes and long-term asset management.

Successful implementation of this architecture is closely dependent upon aspects of the statistical infrastructure that remain outside the scope of this paper. This infrastructure includes data collection and modes of questionnaire design, standard methodologies and tools, IT systems, access, and dissemination.

The main information needs and the main constraints are given in section 2. We then describe the present situation, before setting out the future social statistics architecture. Alternative options are briefly outlined, and the paper concludes with a short discussion.

---

[1] The other element of the work programme is improving analytical capability, disseminating information, and access to data.

2 Information needs and constraints

POSS aims to meet the key information needs of government and the community. Official social and population statistics serve a rich variety of uses. All the statistics contribute in some way towards allowing New Zealanders to understand their society and are a vital component of good government. They provide objective statistical information for use in evidence-based policy making, and support the democratic process by providing independent information for the voting public to assess government performance.

These uses can be broadly categorised as:

- allowing New Zealanders to understand their society
- monitoring social outcomes
- allocating government funding
- improving understanding of causal relationships
- planning for the future.

There are expectations to provide internationally comparable statistics to bodies such as the United Nations and OECD. For example, statistics are provided on population size and structure, fertility and life expectancy, unemployment rates, and income distribution.

Stakeholders include a correspondingly wide range of groups: central and local government agencies and service providers, non-government organisations, international bodies, researchers, and the New Zealand public.

No unifying conceptual framework has been developed for social statistics, as has been achieved with the System of National Accounts for economic statistics. Instead, social statistics are grouped across **domains** of interest. The social domains used in New Zealand are comparable to similar categories used internationally. Some domains, such as population, paid work, and economic standard of living, are long established – there is a good understanding of the conceptual issues and a history of producing official statistics. Other domains, for example culture and identity or social connectedness, are much less developed and our understanding around these is still evolving.

Domain plans will eventually provide a much fuller picture of enduring information needs for all aspects of social and population statistics. Some domain plans are completed, while others are in progress. In the meantime, the architecture must make some assumptions about the requirements for official statistics and their relative importance, and retain the flexibility to adapt to an evolving environment.

Social domains

- Population
- Housing
- Economic standard of living
- Knowledge and skills
- Paid work
- Health
- Safety and security
- Culture and identity
- Social connectedness
- Human rights
- Physical environment
- Leisure and recreation

Essential information needs

User needs cover four broad areas:[2]

- **Population statistics** describe the reference population and are fundamental to all other social statistics, and for per capita calculations for many economic applications.
- **Monitoring social outcomes and overall well-being** statistics measure **what** is happening, across each of the domains and in an overall sense.
- **Distributions and inequalities** describe **who** is affected; they show how an outcome is spread across the population, and measure the gaps between different groups.
- **Understanding causal relationships** is analysis that goes beyond the descriptive and attempts to answer the question **why** do we see this outcome?

The frequency of outputs needed to monitor progress varies considerably across domains, depending on factors such as the impact of change and the rate of change. What is common across all areas is the need for regular and consistent measures of change over time.

Population statistics: Independent, robust, and trusted information about the size and structure of the population is an essential underpinning of a modern democracy and economy. Population counts determine the number of electoral seats and are used to decide on electoral boundaries. Māori descent populations are required by the Electoral Act 1993. Population estimates are widely used by local government and inform population-based funding allocation in the very large health and education budgets. Population statistics provide context for most other measures by serving as denominators to create rates. Population projections are vital for future planning. Measures of population change – births, deaths, and migration – are important indicators and contribute to population estimates and projections. High-level national statistics are reported monthly or quarterly, more detailed measures annually.

Monitoring social outcomes and overall well-being: Each of the domains (apart from population) monitors an aspect of social outcomes related to well-being.

Measures of employment and unemployment are a key indicator of economic performance, are produced quarterly, and must meet internationally comparable definitions.

Measures of household income, expenditure, consumption, and wealth are needed to monitor material standard of living (Stiglitz et al, 2009). Income measures, and those housing costs required to calculate disposable income, are produced annually. Expenditure and wealth measures are produced less often, while at present expenditure is used as a proxy for consumption. Household expenditure data is required for the consumers price index and the national accounts.

Indicators are needed to measure outcomes in the main policy-related sectors: health, education, crime and justice, and housing. Indicators are generally updated every one to five years.

Well-being is multi-dimensional. Single sector statistics are limited because they miss the essential interrelatedness of phenomena in people's lives. For example, interactions between education, health, employment opportunities, and family circumstances may determine whether a particular level of income is sufficient for a person's well-being. Both objective and subjective dimensions of well-being are important. Thus information is needed from the same person about a range of areas, at a single point in time.

As well as single indicators of well-being, measures of overall well-being are needed. These combine outcomes across a number of areas of life, and provide a mixture of subjective, self-assessed measures of well-being (such as life satisfaction) and objective measures (such as the Human Development Index).

---

[2] See Brown (2003) and Brown (2010) for further discussion.

Distributions and inequalities: Leading indicators such as the national unemployment rate or median weekly income are important, but hide the variability underneath. Distributional statistics reveal the population subgroups that are worst affected (or most successful) and can be used to define measures such as poverty levels and measures of inequality.

Distributional statistics allow policy-makers to identify, target, and provide services or income transfers to specific groups.

Understanding causal relationships: The most difficult aspect is attempting to explain why certain outcomes occur. The association between factors needs to be understood in terms of cause and effect, while controlling for other characteristics such as demographic variables. However, in order to study causal relationships and linkages between policies and outcomes, cross-sectional data must be extended to longitudinal information. Longitudinal data follows the same person over time and is needed to study dynamics and transitions from one state to another. Information about attitudes, motivations, and values may also be needed to understand relationships between various factors.

Population subgroups and core variables

There is almost always a need for social statistics to be broken down by population subgroups. The main sub-groups are defined on the basis of:

- age
- sex
- ethnicity
- subnational geographies.

Coherence across social statistics is greatly enhanced when all data collections (including administrative data) are able to implement the standard classifications for these variables.

Of particular importance is the ability to carry out Māori / non-Māori comparisons, and also to provide information of specific importance to Māori. Sample surveys of the whole population are usually restricted (by sample size) in their ability to provide detailed breakdowns for Māori, Pacific, and other ethnic groups. This provides a key ongoing challenge for official statistics on these population groups.

Local government legislation enacted in 2002 has driven a growing demand for descriptive analysis of trends and patterns in outcomes at the regional and community level. Accountability requirements include the production of regular monitoring reports covering social, economic, and environmental dimensions. Again, sample surveys are restricted in the detail that can be provided at regional level, and are generally unable to produce information for local communities.

In addition to these population subgroups, a number of variables are commonly used when examining factors that influence outcomes. These include country of birth, household composition and family structure, labour force status, income, and educational qualifications. Ideally, all these core variables should be available at an individual level together with the main outcome (or outcomes) of interest.

Supporting analysis

Most of the information about population statistics, for monitoring outcomes, and some distributional statistics can be disseminated through a range of tables. However, data collections often hold a much richer set of information than can be seen in headline indicators or sets of standard tables. A range of statistical techniques can be used to more fully exploit the data. These include:

- multivariate analysis (to understand the interactions between multiple variables)
- longitudinal analysis (following the same person over time to understand dynamics and transitions)
- spatial analysis (to understand the role played by geography)
- microsimulation (providing the unit record and transition data to develop models of a process over time).

While traditional tables will continue to be core outputs, Statistics NZ products and services must also facilitate more complex analysis techniques. Mapping and other spatial analysis requires tabular data to be seamlessly connected to digital boundaries and other geo-spatial infrastructure. Administrative data should be coded and easily accessible at relevant subnational geographies.

Much of the analysis of the relationships behind the outcomes is carried out using data held at the individual level. Improving access to microdata is key to producing the evidence base for policy making and would significantly improve the return on investment in statistical information. This includes resolving issues related to accessing longitudinal microdata. Access to microdata for government, academic, and private researchers is crucial to meeting information needs.

Constraints

Information needs represent a client viewpoint. As well as considering those who use statistical products, other key stakeholders are: government as the funding source, the public as respondents and taxpayers, and the official privacy guardians that provide oversight to the broader private and public good information interests of society. The desire for more information can seem ever-expanding; however, the interests of these other groups can place constraints on the Official Statistics System in its ability to meet the demands of users. Hard constraints are imposed by the available financial resources needed to operate the Official Statistics System, as well as by the need to respect and manage the burden placed on respondents. A statistical agency must retain the utmost trust and goodwill of respondents, and so needs to ensure the security and confidentiality of data provided by individuals, and to work within publicly acceptable views of privacy and any limitations this might place on the uses of data.

Several strategies can be used to resolve tensions between the needs of users for information and other constraints:

- prioritisation, so that resources are directed to the most important uses
- rationalisation of content across census and surveys, to minimise respondent burden
- making the best use of administrative data for statistical purposes
- combining information from different sources using statistical models
- using data linkage to bring together information about the same person from different sources.

Other means of improving efficiency through statistical infrastructure (eg following a common business process model, using standard methods and common tools, more efficient development of IT systems, and new products) fall outside the scope of this paper.

Practical feasibility within realistic timeframes must also be taken into account. In developing the architecture, the focus is on a timeframe of about 10–15 years. The main architecture includes elements that are operationally achievable in this medium-term period. Options that clearly fall outside this range are treated as not viable. We discuss later our approach to options that may have merit, but are likely to be achievable only in the very long term.

3 The present: household surveys supported by administrative data

The architecture describes the structure of a set of data collections from an overall design point of view. The proposed future architecture will be a progression from where we are today. We therefore begin by describing the existing structure of official social and population statistics.

Models for statistical data collection systems lie somewhere on a continuum between a model fully based on household surveys, and on one based completely on administrative registers. The current structure of social statistics in New Zealand is based on a traditional model of household surveys supported by administrative data (Bycroft, 2008). The main features of the existing situation are outlined below in terms of the primary data sources, and the ways in which they are used in combination.

The census and sample surveys

The five-yearly Census of Population and Dwellings is the cornerstone of social and population statistics. The census is unique among household surveys in that it reaches every person in New Zealand. The census provides the basis for population estimates, and a range of socio-economic information for the whole population. The richness of the census derives from this ability to collect a wide range of variables at the same time, and to produce this data for small population subgroups and local communities. The Post-Enumeration Survey (PES) has been carried out immediately following the census since 1996, to measure census undercount, and contributes to the population estimates.

Sample surveys play a key role in producing statistics more frequently than is possible with a census, and can explore subjects in far more depth. Sample surveys are generally limited in their ability to produce detailed tabular breakdowns for subgroups because of sample size.

Sample surveys included under the POSS umbrella meet high quality standards. They are designed and conducted to produce unbiased estimates for the New Zealand resident population, and include measures of sample error.

One of the defining characteristics of sample surveys is the survey **frame**. The survey frame is some kind of list of the population, from which those to be included in the sample are selected. Most sample surveys conducted by Statistics NZ use an area-based frame to select household samples. Geographic areas (meshblocks) are formed into primary sampling units (PSUs). Households are selected from lists of dwellings enumerated in selected PSUs. One role of the household survey frame is to control overlap in household surveys so that respondent load is shared equitably.

Administrative sources may also be used as sample frames for specific populations; for example, the Longitudinal Immigrant Survey NZ (LISNZ) is selected from people approved for permanent residence in New Zealand, using data maintained by the New Zealand Immigration Service. Sample surveys conducted outside Statistics NZ have used a variety of ad hoc survey frames.

Because in New Zealand there is no complete and identified list of the resident population to draw samples from, sample surveys of people do not often select respondents directly from a list of the population units. The census does however produce a population listing once every five years, and post-censal surveys use the census as the sample frame to target particular sub-populations.

Other strong links exist between the census and sample surveys. Survey estimates are benchmarked to population estimates, which are based on census counts. The census provides the meshblock population sizes used in creating PSUs, and area characteristics for PSU stratification. The PSUs are updated and a new household survey frame is established after each census.

A degree of coordination across sample surveys is achieved through common use of the area-based survey frame. Greater coherence is being introduced through the development of a standardised set of questions. The main population subgroups (age, sex, ethnicity), plus other widely used socio-demographic variables, have been developed as 'core questions'. While most of these variables are included in some form in census and existing household surveys, new surveys are using a formally designated Core Questions Module.

POSS sample surveys

To meet greater demand for official statistics the number of sample surveys has increased over time and Statistics NZ has played a leading role in conducting sample surveys for official statistics. The Household Labour Force Survey (HLFS) has been producing official measures of employment and unemployment quarterly since 1985. The New Zealand Income Survey has been run annually as a supplement to HLFS since 1997. The Household Economic Survey (HES), running since 1973, produces detailed household income and housing costs data annually (HES-Income), and household expenditure data every three years (HES). The General Social Survey (GSS), introduced by POSS, covers topics across domains and includes subjective as well as objective measures of well-being. GSS is run every two years, with the first data being published in 2009.

Two longitudinal surveys have been conducted – the Survey of Family, Income, and Employment Dynamics (SoFIE) over eight 'waves', and LiSNZ over three waves. They completed their data collection cycles in 2010 and 2009, respectively.

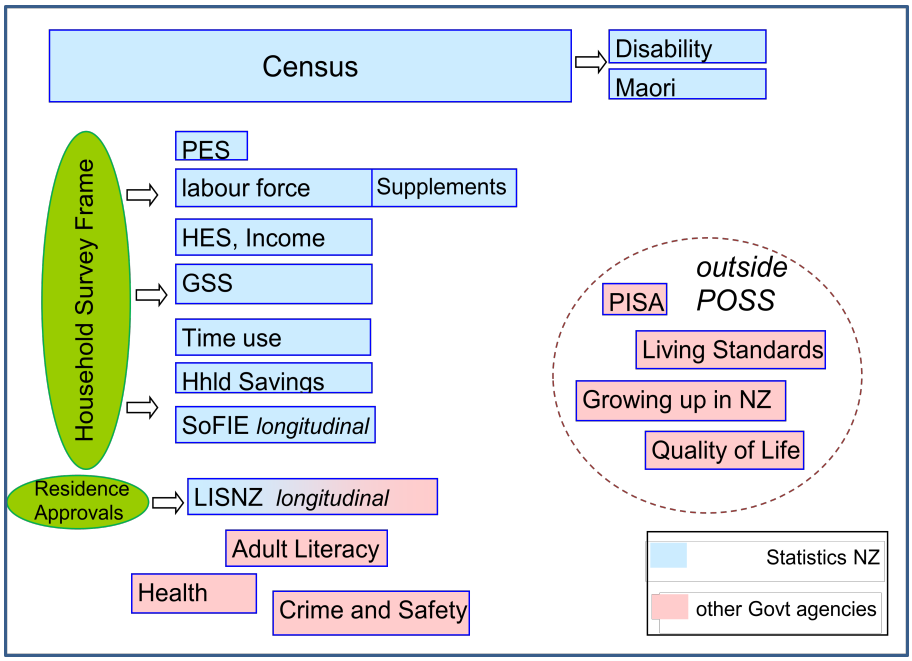
Other sample surveys have been carried out at less frequent intervals. The Time Use Survey was first run in 1996, and has been in the field again in 2009/10. A Household Savings Survey was run for the first time in 2001 to provide information on assets and liabilities (net worth). Surveys of two sub-populations, Māori and the disabled, have been conducted as post-censal surveys.

While the household survey frame provides a common basis for sample selection, each social survey is designed and run separately, with typically long lead times for development. The HLFS provides some flexibility to be used as a vehicle for supplementary surveys. Apart from the annual income supplement, other topics have been appended to the HLFS in a given quarter on an ad hoc basis.

Three sample surveys conducted by other government agencies are included under the POSS umbrella – the Health Survey (Ministry of Health), New Zealand Crime and Safety Survey (Ministry of Justice), and Adult Literacy and Life Skills Survey (Ministry of Education). The sample surveys complement information available from administrative data within each of these sectors.

Figure 1

Major household surveys across government



Government agencies also undertake some national-level surveys outside the present POSS and Tier 1 statistics frameworks. These include the Living Standards Survey,[3] Quality of Life Survey,[4] and Growing up in New Zealand[5] (all supported by the Ministry of Social Development (MSD)). The Ministry of Education takes part in the international PISA study[6] that assesses the knowledge and skills of 15-year-old students.

Administrative data

A wealth of administrative data associated with government activities provides further opportunities for use for statistical purposes. Administrative data collected by government reflect the interaction of government with its citizens and are specific to the population of interest and the activities of each government sector. Examples include:

- registers of key life events that include births, deaths, abortions, marriages, civil unions, and divorces
- border control data is used to produce statistics on international travel and migration
- taxation of wages, salaries, and benefits provides some information on personal income and income sources. Local government taxation of ratepayers produces information on property
- providing services, for example by Health, ACC, Education, MSD, Labour, Housing, Police, Corrections, and Justice, produces statistics relevant to the services they provide. Land Information New Zealand maintains the basic geographical infrastructure.

A number of official social statistics sourced from administrative data are currently produced by Statistics NZ. Other agencies produce statistics from their own administrative sources, some of which are recognised as Tier 1 statistics. Government agencies are improving their

---

[3] Living Standards Surveys, 2000, 2004, 2008 aimed to develop a comprehensive description of the living standards of New Zealanders, using non-income measures of material well-being and hardship.
[4] Quality of Life project runs the only survey designed to provide data for New Zealand's largest urban areas.
[5] Growing up in New Zealand is a new longitudinal study of children.
[6] PISA OECD homepage.

own information systems, and some have introduced central indexes (National Health Index,[7] National Student Index[8]) or combined datasets for better analysis (the MSD Benefit Dynamics Database).

However, while the census and household surveys are referenced to the New Zealand resident population, many administrative data sources do not relate easily to this population. For example, excellent administrative information is available on young adult students in New Zealand tertiary education institutions but not for young adults not in tertiary education. Nor do they hold information for those who have gained tertiary qualifications overseas. Administrative data can be very useful when the focus is on providing information relevant to the administrative purpose, but producing statistics comparable to the statistically defined resident population can be difficult.

Administrative data is typically limited in its usefulness for statistical purposes when used in isolation, and sample surveys can help to build up a more accurate representative picture. At the same time, administrative data contains valuable detailed information that can enhance survey information. Maximum value is gained by appropriately combining the different primary data sources.

Combining data sources: data linkage

We distinguish two main ways of combining data. Firstly by data linkage at the unit record level, combining information about the same individual or unit. Other examples of combining data sources use data aggregated above the unit level.

Linked datasets created by Statistics NZ centre on either the census or the Linked Employer Employee Data (LEED) dataset.[9] The primary link in LEED, between Inland Revenue employers and businesses on the Business Frame, has been extended to include two additional links at the person level. Benefit data (from MSD) and tertiary student data (from Ministry of Education) are linked to employees in LEED. LEED is a longitudinal dataset, following individuals over time via the unique (anonymised) IRD number identifier.

The integrated Student Loans dataset links data from Inland Revenue, MSD, and the Ministry of Education to monitor student borrowing and is required by a Cabinet directive.

The New Zealand Census-Mortality Study links the census to the Deaths Register. Census adds socio-demographic variables to the deaths information, which then allows measurement of mortality differences by variables such as ethnicity, occupation, and cigarette smoking status. Similarly, the Cancer Trends project creates links between the census and the Cancer Register. In contrast to LEED, the two integrated datasets (Census-Mortality and Cancer Trends) are held separately, and there is no longitudinal link between the same individuals in successive censuses.

Linked administrative datasets for research and evaluation purposes have been created by other government agencies. One example is a research dataset of linked MSD and Inland Revenue administrative information constructed as part of the evaluation of Working for Families.[10]

In a somewhat different perspective on data linkage, sample surveys can be linked at unit record level to administrative data, replacing a survey question with information held about the respondent in administrative data. This is known as **data substitution**. The only example to

---

[7] NHI The National Health Index number is a unique identifier that is assigned to every person who uses health and disability support services in New Zealand.

[8] NSI The NSI is a database maintained by the Ministry of Education. The NSI allocates a unique identifier, a national student number, to every student enrolled at an education provider in New Zealand.

[9] LEED links Inland Revenue tax data from employers with the Statistics NZ Business Frame. Employees are all individuals who receive income from which tax is deducted at source. LEED includes most benefit payments. Guide to interpreting the LEED data.

[10] Known as the "WFF Research Datasets" (an example: Employment incentives for sole parents).

date is SoFIE, where alternate waves link individuals to administrative health information, with the consent of respondents.

In post-censal surveys, where the census is used as the sample frame, census variables are added to the survey dataset, thus reducing respondent burden.

Combining data sources: aggregate comparisons and statistical models

There are many examples of combining data sources at aggregate levels, and of a high degree of interconnectedness across government agencies. The census, and population estimates and projections are central to many uses.

The most straightforward method of adding value is by comparing outputs from different sources on the same geographic boundaries or for the same population subgroups. This is enabled by using standard classifications to define geographic areas and other variables. While Statistics NZ outputs use standard geographies based on the meshblock system, some administrative data is available only for non-standard administrative geographies, such as police stations or public health areas.

Population estimates serve as a denominator for derived official statistics, including indices such as fertility and mortality rates. Both census totals and population estimates are widely used as denominators or reference values for many purposes, including population-based weighting in funding models, determining electoral boundaries, and per capita economic statistics. The base estimated resident population itself is produced using a model that combines different data sources – where census counts are adjusted for net census undercount (via the PES) and residents temporarily overseas on census night (from migration data). Population statistics between censuses are estimated using an aggregate model that adjusts for population change through births, deaths, and migration.

The use of more complex statistical models in New Zealand is rare. At present there are no examples of small area estimation models being used in official statistics outputs. Microsimulation techniques are used by some agencies. For example, Taxwell[11] is a microsimulation model based on HES unit record data that simulates the effect of changes to income tax rules, and is an important policy analysis tool for the Treasury. The Ministry of Justice has developed a microsimulation tool called 'Pipeline' for simulating the effect of different types of intervention. MSD is developing Betsim, a new static microsimulation tax-benefit model based on SoFIE (Ota and Stott, 2007).

[11] Taxwell is the successor to Taxmod.

4 The future social statistics architecture

We now describe a proposed architecture for future social and population statistics that builds on the current system and is achievable in the medium-term timeframe of 10–15 years. The aims are to achieve greater efficiency in conducting census and sample surveys, more flexibility and responsiveness in meeting user needs, and to obtain better value from existing data sources.

The existing system has served us well in many ways. It is not broken, but it can be improved. Continuing challenges include the high development cost and length of time needed for new household surveys, meeting the needs for detailed and complex information without over-burdening respondents, providing sufficiently detailed information about small areas and population subgroups without major increases in sample sizes, and in continuing to provide longitudinal information in a sustainable way.

The key element of this proposal is that information needs for social and population statistics will continue to be met through household surveys supported by administrative data. The main reason for continuing this model is the absence of a New Zealand population register. Without a population register there is no independent population frame and so population estimates must continue to rely on a regular census. Regardless of the survey frame, sample surveys are still required for information that is not captured by administrative data.

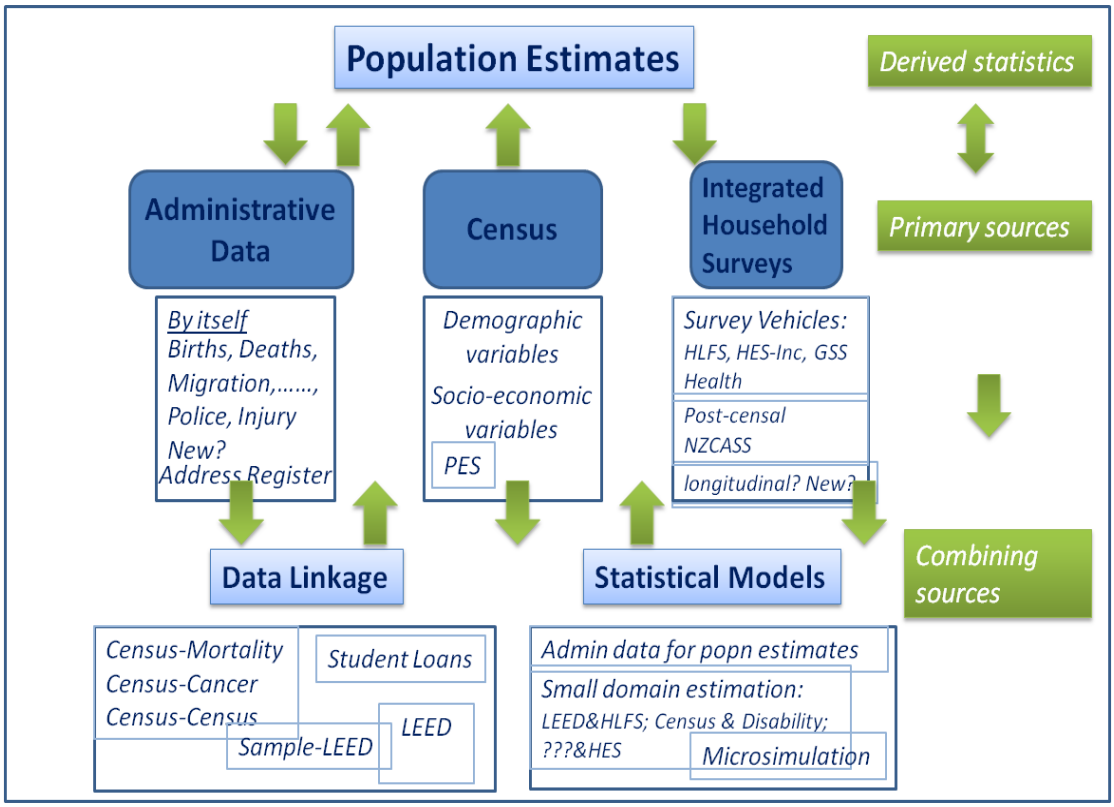
Administrative data is important, but is of limited value in replacing either the census or sample surveys. However administrative sources do play an increasing role. Combining administrative data and surveys, through data linkage, data substitution, or via statistical models, takes advantage of the strengths of each data source, and increases the value of the statistical information that can be produced from the collections.

Alternative models based on administrative registers have also been considered and are discussed in section 5. It is unlikely that statistical use considerations alone will be a sufficient driver to introduce the major changes to government infrastructure that would be required. The development of a population register for example would be more likely as part of a wider redevelopment within the public administration infrastructure in New Zealand, something that would need to be driven by the government sector more broadly and represent a much greater community of interest. However the proposals presented below for making more and better use of existing administrative data for statistical purposes are also foundation steps on a pathway towards an alternative register-based system.

The proposed future social statistics architecture is shown in figure 2.

Figure 2

Future official social statistics architecture: high level



The main elements are the primary data sources and how those data sources are combined.

Primary data sources:

- a census of population and dwellings
- sample surveys: integrated household surveys (cross-sectional); longitudinal survey(s)
- administrative data sources producing statistical information.

Combined data sources:

- population statistics derived from the census, the PES, and administrative sources
- data linkage: to the census; to LEED; Student Loans dataset
- statistical models combining administrative and survey data: small domain estimation; inter-censal population estimates; microsimulation.

The main new features are:

- an integrated approach to cross-sectional household surveys using ‘survey vehicles’
- statistical models combining survey and administrative sources
- further data linkage: sample-LEED database; longitudinal census database
- an address register.

These new features are highlighted because they have been identified as the most obviously beneficial and viable new developments within Statistics NZ. Each is discussed further below.

The architecture is designed to be adaptable and is not intended to fully prescribe the details of these elements. Other household surveys may be needed, new administrative data sources may be found to be useful for statistical purposes, other opportunities for linking administrative data or using statistical models may arise. For example, the review of crime and justice statistics (Statistics NZ, 2009) recommends data sharing across Justice, Police, and MSD to improve youth justice statistics.

Meeting challenges for ethnic, regional, and longitudinal information

Population statistics will continue to be provided by a population census and the use of administrative data, and, combined with the sample survey programme, a broad range of social outcomes and well-being will be **monitored** at regular intervals. But, given this, how does the architecture address the information needs that are the most challenging to meet within constraints?

Providing information from sample surveys on **distributions and inequalities** for important ethnic subgroups and regional and local communities is limited by sample size. Longitudinal surveys can provide excellent, in-depth information to inform **causal relationships**, but they are expensive and impose a high respondent burden. Small, high needs population groups also pose difficulties. The proposed architecture may not meet all information needs for all groups, but through the census and use of linked data and statistical modelling, cost-efficient solutions can be found for most of these. Opportunities for providing information for Māori, Pacific, or other ethnic groups, for subnational geographies, and for longitudinal data, are outlined below.

Statistics about Māori (and other ethnic groups):

- population statistics by ethnicity
- the census
- some weighting in sample designs towards high Māori, Pacific, or Asian areas
- small domain estimation for samples
- pooling sample survey information over time
- post-censal surveys targeted directly at Māori
- linked data: sample-LEED; Census-Mortality; Cancer Trends

Regional and local community statistics:

- population statistics by subnational geographies (to area unit level)
- the census (to the smallest output area)
- small area estimation for samples (regional council or territorial authority level)
- pooling sample survey information over time
- administrative data (where possible depending on the source).

Longitudinal data:

- linked databases: Student Loans, LEED, sample-LEED; longitudinal census
- longitudinal sample survey(s) – to be confirmed.

A further challenge is posed by the need for information about very small groups that are of high policy interest, dealing with issues such as serious physical and mental health problems, end of life care, extreme poverty, crime, social and physical isolation, truancy, and illegal migration. These high needs or marginalised groups are more likely to be in hospitals, rest home care, prisons, or drifting, or otherwise homeless and thus not captured in the household frame. Also, because the groups have small numbers, very few would be selected in a general household survey. For some, options are available, while other groups, such as illegal immigrants, it is recognised that they may never be covered well by official statistics.

Census is the only comprehensive source of information about those living in non-private dwellings, but it is limited to a set of fairly simple questions. Including small groups in a sample survey would require specific targeting and a suitable sample frame. The Residential Disability Survey does just that as it surveys residents of rest homes etc, and is run at the same time as the Household Disability Survey. More use could be made of some administrative data. Systematic linkage across administrative datasets (see below) would identify small groups interacting with particular government services, and their outcomes could be followed over time. The linkage between MSD benefit data and LEED is an existing example.

Outstanding issues

While further investigations are needed to test the feasibility of these approaches, they appear to be largely achievable given the right expertise and appropriate resourcing. However, key questions remain to be addressed.

Longitudinal survey: What are the priorities for longitudinal data? What are the nature and objectives of any longitudinal survey(s)? Can these be managed within reasonable costs? Who should be responsible for running the survey? The answers depend partly on how far the new sample-LEED and longitudinal census datasets go towards meeting needs for longitudinal information, as well as on the potential of more systematic linkage of administrative data.

Systematic data integration for research: The two new data linkage datasets are important, but fairly modest additions to data linkage work already carried out. They fit within the existing Data Integration Policy and continue the approach of assessing new data linkages on a case by case basis. There are advantages in a more systematic approach to data linkage. In order to provide flexibility in developing new statistics, Van Tuinen (2009) recommends that statistical agencies collect as much microdata as possible (preferably register data), and make it standardised and harmonised so it can be linked at micro-level.

A dataset that links data across multiple administrative sources (such as tax, education, benefit, health, justice, births, deaths, and migration) should be technically feasible in New Zealand if the aim is to support valid research purposes. Such a resource is likely to be more efficient to create and maintain, and opens up the potential for exploring a wider range of research questions. This approach assumes *a priori* that administrative data linked across multiple sources is a valuable resource. The linked data provides the flexibility to respond rapidly to new information needs, is particularly advantageous for longitudinal information, and stimulates new lines of research to help answer complex research questions. For example, data linkage across multiple sectors is seen as critical to developing effective intervention measures in the justice sector (Statistics NZ, 2009, 2.4.2). The benefits are demonstrated within the business sphere by the use being made of the Longitudinal Business Database (LBD).[12]

Systematic linkage across many sources is also a feature of fully register-based statistical systems. The experience gained in developing and managing more extensive linked data will inform decisions around proceeding further towards greater reliance on administrative data.

The question is whether whether systematic data linkage of personal data fits within the public's tolerance over privacy. There is a choice of models for managing privacy. Rules and protocols can address specific concerns over the privacy of individual data records where they are used in the context of compiling and analysing official statistics. In the register-based statistical systems of Nordic countries, for example, legislation gives the statistical agency the right to hold and link administrative data. In Austria, data is matched across government branches by their Data Protection Agency (rather than by Statistics Austria), using a complex system of encrypted identifiers (Statistics Austria, 2008). Another option is the West Australia Data Linkage System,[13] where the linking is done by a small central team that holds concordances between different data sources, but the data itself remains with the source agency.

---

[12] The Statistics NZ LBD integrates longitudinal administrative and survey data, at the enterprise level, that will meet users' needs to better understand the dynamics of enterprise performance without increasing respondent load. "From a research perspective, the breadth of data included within the prototype LBD enables advances to be made in many of the microeconomic studies previously investigated in New Zealand, as well as opening up many new avenues for investigation." (Fabling et al, 2008)

[13] West Australia Data Linkage home page www.datalinkage-wa.org.au/ (Accessed 9 Feb 2010)

Census of Population and Dwellings

The Census of Population and Dwellings retains its central place in this architecture, providing the basis for population estimates, and socio-economic data for small sub-populations and for geographical areas down to small, local communities. The census would still be one in which the whole population is enumerated at a single point in time. However, the architecture leaves open questions such as the frequency of census, content beyond the main demographic variables, and operational issues such as the modes of collection and delivery.

Consideration of changes to census frequency or reduced content would have to take into account the impact on the sample survey programme. Any information loss from census is likely to see greater demands placed on sample surveys.

Section 5 considers potential replacements for the full enumeration census. If any of these were to be supported, it is likely to take at least 15–20 years development, and major investment, before they could be tested against census results. Thus full enumeration censuses will be needed until 2021 and more likely to 2026 or beyond.

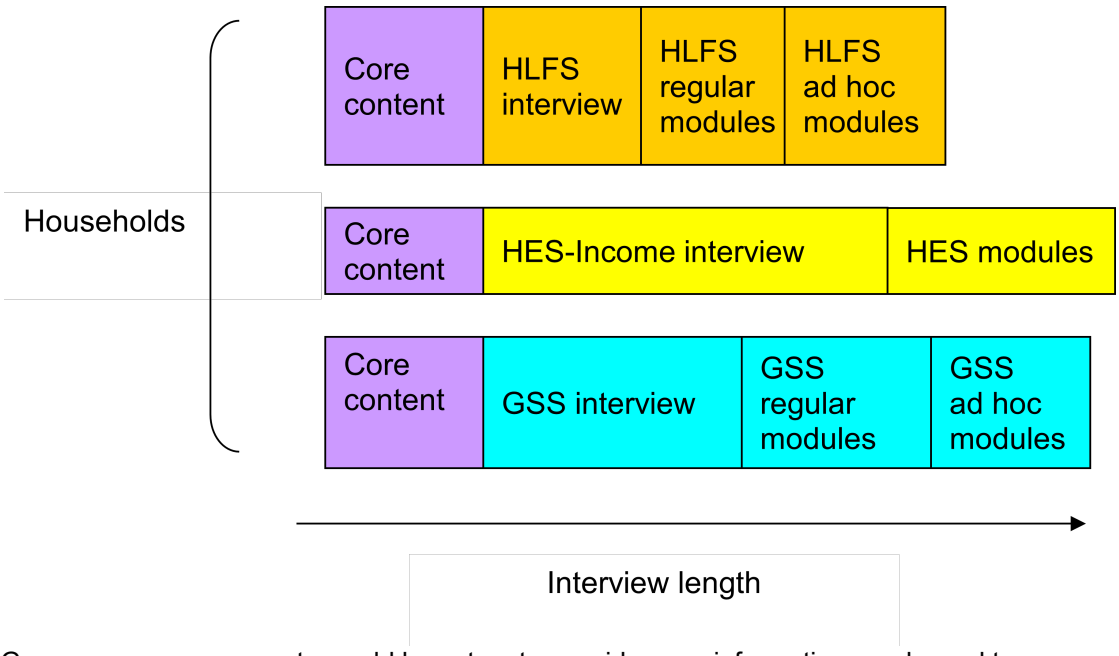
Integrated household surveys

The integrated household surveys approach consolidates the household survey programme around the three main surveys: HLFS, HES-Income, and GSS (Bycroft, 2010a). Each of these surveys is not simply an individual stand-alone survey, but acts as a survey 'vehicle' for a flexible, coordinated programme of topic modules. Each survey vehicle would consist of the core socio-demographic questions and primary interview questions repeated in each period, with additional rotating topic modules and some ad hoc supplementary topics (figure 2). The proposed integrated Health Survey (Ministry of Health) effectively acts as a vehicle for health-related topics.

Using core questions ensures that all key variables are included in surveys and that comparisons between surveys are valid.

Figure 3

Statistics NZ's three survey vehicles



Governance arrangements would be set up to consider new information needs, and to determine the most suitable survey vehicle. This should see better prioritising of investment in new information. An integrated approach to conducting household surveys should improve responsiveness to changing information needs and reduce costs in the overall end-to-end

business process, while managing respondent burden. Increased and/or better analysis should result from having fewer datasets with a common infrastructure.

Within this model, there is potential for changes to current sample sizes and frequency. For example, HES-Income could increase in size periodically to accommodate a wealth module; GSS could be run continuously rather than every two years; a longitudinal component could be included within the vehicles.

The survey vehicles strategy does not necessarily improve our ability to target small sub-populations, nor does it make full use of the common core questions across surveys. Full integration of the survey vehicles into one, large, master sample (Ramsay, 2006) might address these issues, but this requires a further investigation of its feasibility and practical benefits.

Pooling results from the same survey over time will increase the number of respondents from smaller sub-populations that are available for analysis. However it is uncertain how far this will improve the reliability of estimates for smaller sub-populations (eg ethnic groups) or regions.

Data substitution, where linked administrative data replace survey questions, has the potential to reduce respondent burden and improve data quality. The feasibility of replacing some survey income questions by linking to LEED data is the most obvious example to investigate.

Stand-alone surveys may still be needed in some situations. Existing survey vehicles have limited longitudinal capacity. Other examples include when it is more efficient to target particular sub-populations via another sample frame (eg the census), or where particularly long, complex, or sensitive topics would put the main purpose of the survey at risk.

The programme of rotating modules in the main surveys will have some spare capacity, providing an opportunity for new topics to be added. We have also considered whether a more rapid response to emerging issues should be included in the overall survey programme. The UK Office for National Statistics runs the Omnibus Survey, a monthly telephone survey that offers very quick turn-around for simple topics. This type of survey is useful in responding to issues that are important, but “of the moment” and probably not enduring. Reaction from a range of government departments that had need to undertake this sort of activity suggested that most were happy to continue obtaining this type of information through existing market research companies.

Address register

An address register is included as part of the proposed architecture because of the tangible benefits of having a frame of dwellings (if not people). An address register that identifies residential dwellings would give census the option to post out forms, reducing delivery costs. For sample surveys, it would provide a list of dwellings in selected PSUs, reducing enumeration costs, and would also have the potential to be used directly as a list frame.

Longer term, an address register is one of the essential base registers that make up a fully register-based system (if we were to eventually adopt such a system).Ideally there would be one definitive national address register forming part of basic government infrastructure. Statistics NZ would be a major client rather than responsible owner of a national address register.

Statistical models

Results from household sample surveys are estimated directly from the weighted survey responses. However, because the size of sample surveys is relatively small, reliable estimates for subnational geographies are produced only for broad regions or not at all. Given the need for local government to monitor outcomes for their areas, and the prohibitive cost of increasing sample size, another solution is needed.

Sometimes other data is available at lower geographies and measuring a concept that is related to, but not appropriate for, the purpose of the survey. For example, those receiving the

unemployment benefit are often unemployed, but benefit data is not equivalent to the official unemployed definition.

Small area estimation is a statistical technique that combines direct survey estimates and related auxiliary data in a model. The small area model makes best use of the survey information, and the geographic patterns present in the auxiliary data, to produce survey estimates at regional council level or possibly territorial authority (or similar) level. The auxiliary data from the census or administrative sources must be available at the required geographic level and should show a strong relationship with the variable of interest.

Fairly standard methods have been developed and software is available so that it is now practical to implement small area estimation and produce modelled outputs, perhaps as experimental statistics at first. Using small area estimation to improve geographic detail for HLFS, using unemployment benefit data, appears feasible. The same technique (more widely named 'small domain estimation') may be applied to produce better estimates of other sub-populations, such as ethnic groups and disabled people. Further work is needed to identify suitable auxiliary data sources that could potentially be used with other sample surveys.

In a similar way, inter-censal population estimates have used births, deaths, and external migration data as direct measures of population change. There are no direct measures of internal migration, which is needed for subnational population estimates, and the process now relies on subjective judgment. However, a number of administrative datasets (electoral rolls, school enrolments, LEED) do capture some movement between areas. A modelling approach would combine the partial information from different sources to produce subnational population estimates.

Data linkage: connecting the social and business architectures

The architectures for social statistics and economic statistics have been developed separately, partly because the basic units are different – units for social statistics are people and households, while businesses form the main units of the economic statistics architecture. However the two architectures are closely connected. At an aggregate level, HES household spending patterns feed into the consumers price index (CPI), and household-based measures of expenditure and income are used in the national accounts. Time use data is used in the not-for-profit satellite accounts. Census dwelling and household tenure and rent variables feed into the national accounts and CPI for weighting purposes. All these inputs occur at an aggregate level.

The LEED dataset provides the interface for linking people and businesses at the unit record level.[14] The usefulness of the employee data is limited by the small number of person characteristics available,[15] and because there is no means of directly relating the LEED population to the reference usual resident population.

The architecture proposes a new unit-record link between a sample survey and LEED that will add a range of demographic variables to LEED, such as ethnicity, qualifications, occupation, and hours worked. Just as linking census demographic variables to deaths opened up many new avenues of research in the health related sector, the sample-LEED linkage will broaden the opportunities for labour market analysis and productivity measures, and improve understanding of the consistency between survey measures and administrative reporting. HLFS, SoFIE, and a sample taken from the census are the main candidates, though further work is required to determine which is most suitable.

Further extension of LEED is possible by bringing through more variables from the existing linkages. For example, it may be possible to include other investment income or Working for Families information from Inland Revenue, or to add non-taxable benefits from MSD data.

---

[14] The basic reporting unit in LEED is the job, which is a unique employer-employee combination.

[15] Only date of birth, sex, and TA-level geography.

Data linkage: a longitudinal census database

The census reaches the whole population every five years. A longitudinal census database links responses from the same person over successive censuses, providing a view of the main socio-demographic changes that occur over a long timeframe. The UK Longitudinal Study has linked a 1 percent sample from censuses since 1971, and included links to health datasets and birth, death, and migration data.[16] In Australia a longitudinal census is planned, firstly by linking a 5 percent sample from the 2006 Census to the 2011 Census (Australian Bureau of Statistics, 2010).

In New Zealand, while we have valuable cross-sectional datasets linking the census to administrative data, longitudinal linking between censuses has been carried out only for studies limited to assessing census quality (in particular the consistency of ethnic responses). It seems likely that a longitudinal census database would also be useful in New Zealand, for example to study long-term social mobility. Further work is needed to determine objectives, risks, and benefits, and what other data (if any) should be included. It is also likely that any longitudinal linking would be restricted to a census sample, to allay privacy concerns.

[16] The ONS Longitudinal Study, available from www.ons.gov.uk.

5 Other options

The main alternative to the household survey-based model is a register-based statistical system. Nordic countries have led development of this new system that makes use of administrative registers and avoids the need for a traditional census. A population register provides the reference population base and is linked to an address register, to birth and death registers and also to other administrative sources such as tax, health, and education data. Geographic information is obtained by linking individuals to their place of residence through the address register and updating change of address needs to be comprehensive and done on a timely basis. The availability of a common personal identifier makes linkage between sources relatively easy. The central population register holds basic demographic information about every person, and serves as a frame for sample surveys. Information about people can be linked to business data through employee identifiers.

When well developed, as in Finland for example, this register-based system has replaced the traditional census. Census information is produced from the linked registers without the need to enumerate the whole population (figure 4). The primary advantages of a register approach are reduced census costs for the statistical agency and greater frequency of data. The use of administrative sources also involves some drawbacks including the fact that the data that can be compiled is limited to what already exists in the registers (United Nations 2011).

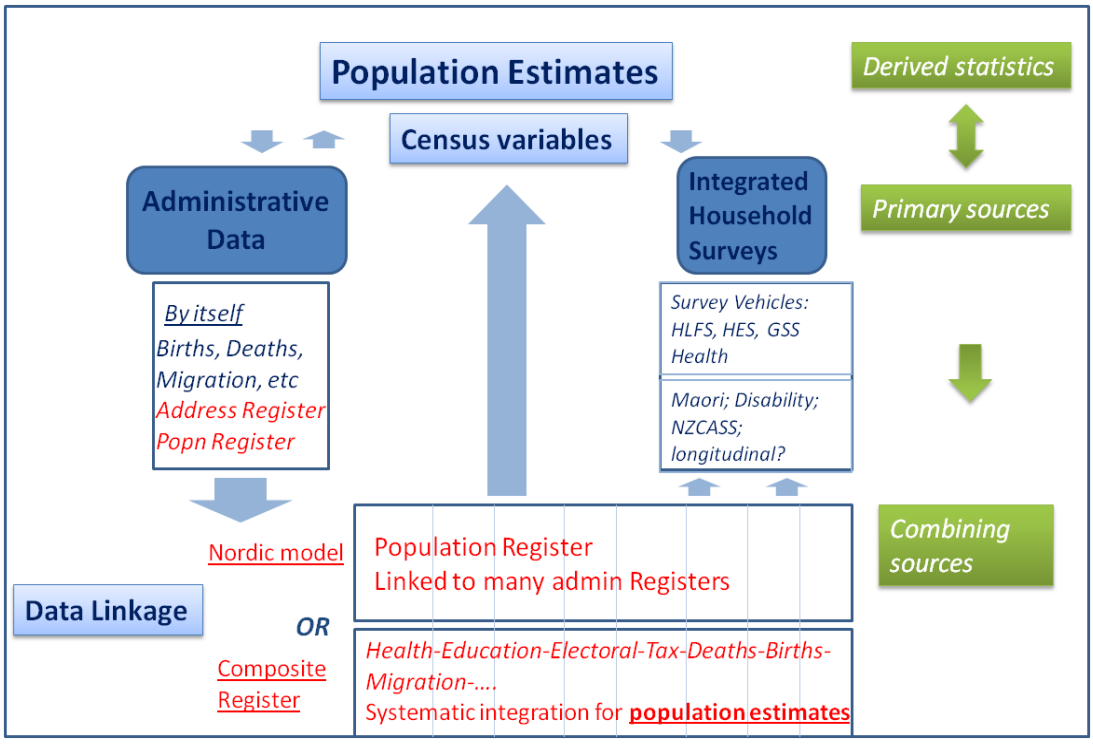
Bycroft (2010b) considers how well-placed New Zealand is to meeting the necessary pre-conditions for implementing this register-based approach. The pre-conditions are: a strong legal basis, public approval, unified identification systems, and comprehensive and reliable register systems developed for administrative needs (Tonder 2009). At present none of these pre-conditions are met in New Zealand. One of the essential pre-conditions of a register-based census is that the country should have an established central population register of high quality and good coverage linked with a system of continuous updating (United Nations, 2008. 1.65).

The main barriers to a register-based statistical system in New Zealand are the absence of such a population register, and uncertainty over public acceptability. Countries with population registers see major benefits for government administration and efficiency, and in having a rigorous system for personal identification. A key task would be to address the privacy concerns that people might have which may be grounded in historical and cultural attitudes towards a compulsory national register, unique personal identifier and a requirement to notify change of address with authorities.

Establishing and managing a population register for administrative purposes would have to be championed by government, and a vigorous public debate over privacy concerns versus potential benefits could be expected. New legislation would be needed. The major drivers are unlikely to be statistical, but would come from a broader consideration of benefits across government. This would be a long-term process. Countries that have replaced their traditional census have done so over a 20-30 year period. It appears most unlikely that a population register linked to other administrative sources could provide a viable alternative to the census in New Zealand in the short to medium term.

Figure 4

Alternative scenarios replacing the traditional census with linked administrative data



Another approach could be to create a pseudo or 'composite' population register by linking existing administrative datasets that already have high coverage of the population. Possible data sources include the National Health Index, National Student Index, IRD Client Register, and the Electoral Roll. While combining data sources such as these would provide a rich source of information for research across sectors, it is much less clear whether sufficiently accurate population estimates could be produced. The successful data linkages for research purposes mainly use just the linked records; that is, the intersection of datasets. The demands for linkage accuracy are much higher if the goal is to create a complete list of the New Zealand resident population (without duplicates and excluding non-residents), from the union of several datasets. Highly accurate linkages are difficult to achieve without a common unique identifier. We are not aware of any country that has successfully implemented such an approach.

As well as the methodological challenges and costs, Statistics NZ would need to consider privacy issues, the impact on the integrity of the whole Official Statistics System, the impact on agencies providing the data, and legislative changes. Accurately recording change of address to ensure on-going reliable local area populations would be a considerable challenge, especially given the high rates of mobility in New Zealand. Development and testing would require major resourcing over a long period, and this approach could not be expected to provide a replacement for census in the short to medium term.

The architecture described in section 4 does not rule out an eventual progression to a register-based system, and more positively, does support our understanding of whether this is likely to be a viable alternative in the long term. Use of an address register, more data linkage, and improved quantitative use of administrative data for subnational population estimates will all help us consider whether some combination of linked administrative data, a coverage survey, and statistical models could eventually replace the full enumeration census.

Conceptually, ‘counting the population’ may sound simple, but this impression is deceptive. Total population is relatively straightforward to measure in New Zealand since we have good administrative systems for births, deaths, and external migration. But in practice, it is the detailed distributions at subnational geographies (including small local areas), and by single year of age and sex, that are both critically important and difficult to measure accurately. Internationally successful systems are purpose-built, involving either a periodic full enumeration census, or an administrative population register.

6 Discussion

The test for the architecture is whether it will be capable of meeting information needs in the medium term, within constraints imposed by funding, privacy concerns, respondent burden, and capability.

The census, the main household surveys (HLFS, HES, GSS, Health), and births, deaths, and migration data are essential to meeting information needs for population statistics and monitoring key outcomes, and as input to national accounts. Administrative data is also used to monitor outcomes within sectors. Distributional information can be produced from census and sample surveys through appropriate design of questions related to the main topics and the use of core questions across all surveys. These surveys are designed to ensure that relationships between various factors can be explored. Longitudinal sample surveys provide information on dynamics and transitions. LEED is an example of linked administrative data that permits the study of outcomes across employment, education, and benefits, and is also longitudinal in nature.

All these sources have their limitations. The census reaches the whole population but is infrequent and must be simple enough for self-completion. Sample surveys can provide depth, but not detailed breakdowns for subgroups. Asking respondents directly produces high quality information, but imposes costs on both the survey agency and the respondent's time. Administrative data incurs no new respondent burden, but is limited to information about the sub-populations and variables related to the specific administrative purpose.

This paper has addressed these issues with proposed improvements that are two-fold: be as efficient and effective as possible in conducting the census and integrated household survey programme; and combine administrative and survey sources using statistical techniques that make up for the weaknesses of each source through the strengths of the other.

The architecture described above is a high-level one, and is not intended to be fully prescriptive of the detail underneath. This in itself ensures flexibility for coping with a future that is inherently uncertain. Household surveys will remain a strong component of the system, but other surveys may be needed (or some dropped). Census will continue, but the interdependency between the content of the census and sample surveys may change. Other administrative datasets or commercial data may emerge as useful for statistical purposes, data linkage could be more or less extensive, statistical models could prove to be more or less successful than hoped.

Finally, we look at where this approach will take us in future. In the medium term, we would hope to have established an ongoing, sustainable suite of integrated household surveys plus the census, which is both cost-effective and responsive to changing information needs. Data linkage will have been extended in a manner consistent with preserving privacy, and outputs based on good statistical models will be understood and accepted.

There is a strong emphasis on making better use of administrative data. In the process of making more use of administrative data, either on its own, in statistical models, or linked with other sources, we will be much better placed to understand whether New Zealand could indeed move towards a register-based system in the long term.

A work programme will set out the research and development required to further prioritise information needs and test the feasibility of the options outlined above for meeting those needs. Establishing the architecture is a key step in the overall context of mapping out the future approach for social and population statistics and delivering those statistics in a sustainable manner.

7 References

Australian Bureau of Statistics (2010). *Census data enhancement project: an update, Oct 2010* Catalogue number 2062.0. Available from www.abs.gov.au.

Brown, Denise (2003). *A social statistics programme for New Zealand. Discussion paper on information needs – issues and gaps*. Unpublished Statistics NZ report.

Brown, Denise (2010). Social and population statistics information needs: Towards 2020. Wellington: Statistics NZ. Available from www.stats.govt.nz.

Bycroft, Christine (2008). *The current structure of official social statistics data collections.* Unpublished Statistics NZ report.

Bycroft, Christine (2010a). Integrated household surveys: a survey vehicles approach. Wellington: Statistics NZ. Available from www.stats.govt.nz.

Bycroft, Christine (2010b). A register-based census: what is the potential for New Zealand? Wellington: Statistics NZ. Available from www.stats.govt.nz.

Cornish, John (2007). *Review progress with development of a statistical architecture of New Zealand social statistics*. Unpublished report prepared for Statistics NZ.

Fabling, R, Grimes, A, Sanderson, L, & Stevens, P (2008). *Some rise by sin, and some by virtue fall: Firm dynamics, market structure and performance*. Ministry of Economic Development occasional paper.

Statistics Austria (2008). Register-based census 2010 and census test 2006 in Austria. *Joint UNECE/Eurostat meeting on population and housing censuses,* Geneva. Available from: www.unece.org/stats/documents/2008.05.census.htm

McKenzie, Ron (2008). *Statistical architecture*. Unpublished Statistics NZ report.

Ramsay, Diane (2006). *Towards a new household survey strategy*. Unpublished Statistics NZ report.

Ota, Rissa, and Stott, Helen (2007). *A new New Zealand static microsimulation model – challenges with data*. Presented to the First General Conference of the International Microsimulation Association, Vienna. Available from: www.euro.centre.org/ima2007/programme/papers/Stott2_paper.doc

Stiglitz, Joseph, Sen, Amartya, & Fitoussi, Jean-Paul (2009). *Report by the commission on the measurement of economic performance and progress*.

Statistics NZ (2009). Review of crime and criminal justice statistics report 2009. Available from www.stats.govt.nz.

Tonder, J-K (2008). The register-based statistical system – preconditions and processes. IAOS Conference, Shanghai http://isi.cbs.nl/iaos/conferences/2008Shanghai.htm

United Nations (2008) Principles and Recommendations for Population and Housing Censuses Rev2 unstats.un.org/unsd/Demographic/sources/census/docs/P&R_Rev2.pdf

United Nations (2011) Report on the Results of a Survey on Census Methods used by Countries in the 2010 Census Round Working Paper: UNSD/DSSB/1

Van Tuinen, Henk K (2009). Innovative statistics to improve our notion of reality. *Journal of Official Statistics* Vol 25(4).