

Innovations in Data Collection– Three initiatives

Paper presented at United Nations 2009 Innovation in Official Statistics seminar,
New York, United States of America
20 February 2009

John Créquer, Michael Holland, Sean Keefe, Rebecca Merrington

Statistics New Zealand
P O Box 2922
Wellington, New Zealand
info@stats.govt.nz
www.stats.govt.nz

Liability statement: Statistics New Zealand gives no warranty that the information or data supplied in this paper is error free. All care and diligence has been used, however, in processing, analysing and extracting information. Statistics New Zealand will not be liable for any loss or damage suffered by customers consequent upon the use directly, or indirectly, of the information in this paper.

Reproduction of material: Any table or other material published in this paper may be reproduced and published without further licence, provided that it does not purport to be published under government authority and that acknowledgement is made of this source.

Introduction

Changes in both the internal and external environment have presented both opportunities and challenges for data collection at Statistics New Zealand. This paper describes Statistics NZ's strategic response to the environment, in particular its data collection strategies. The paper then outlines three innovative initiatives Statistics NZ is implementing to improve the productivity and quality of data collection.

These initiatives are:

1. Best practice field collection – Improving the performance of Statistics NZ's field workforce through delivery of the latest technologies, systems and processes.
2. Project Contact – One system for the management of all respondents to Statistics NZ's surveys.
3. Census optimisation – Statistics NZ's innovative approach to designing enumeration areas for the New Zealand Census of Population and Dwellings.

1. Environment

External environment

In recent years Statistics NZ has really begun to think about the business of producing statistics in terms of its component business processes. Until then, the approach tended to be survey specific and most aspects of producing statistics were managed by the owners of the statistical outputs the surveys fed into. Surveys were run successfully by the subject matter experts, and Statistics NZ enjoyed consistently high response rates. This willingness to respond, and respondents' accessibility, resulted in the organisation focusing less on the business of data collection and more on the analysis and use of the data. The result was a multitude of similar ways and infrastructures being developed to tackle the same problem.

In recent years, there has been both anecdotal and empirical evidence¹ to suggest that even with the statutory requirement to comply with our requests, this willingness is decreasing. There is also evidence that it is increasingly difficult to access individuals in hard-to-reach locations, and on the days and times of the day that Statistics NZ has historically operated.²

At the same time, there are increasing demands from users for Statistics NZ to provide more information. These demands have increased the number of statistical outputs Statistics NZ produces, and have slightly increased the number of surveys we conduct. (The increase in outputs is offset to some extent by producing statistics using administrative data, or a combination of administrative data and direct survey activity.) An example of the increase in outputs is the programme of official social statistics (POSS), which is introducing a suite of new surveys.

1 This is based on internal operations research on response to surveys before any reminder activity.

2 This is evidenced by the increased number of contacts required by the field force to solicit a response.

Busier lifestyles, and the more competitive business practices of our respondents, mean that the value exchange associated with giving up time or money to provide information is no longer accepted as a matter of course.³ For an increasing number of respondents, the benefits of complying (social good, participation in democratic process, interest in the results etc) are no longer greater than the cost of their time. This is particularly the case when requested data has previously been collected by either Statistics NZ itself or another government agency.

Internal environment

For Statistics NZ to maintain its historically high response rates, and the associated quality level, in an environment of reduced willingness and accessibility, an upward pressure has been placed on the costs of data collection. Evidence suggests that the cost of collecting data from the last few respondents is higher than the cost of response for those who are either willing or more readily available.⁴ These increased costs are in addition to ongoing increases in the costs of the inputs to collection (wages, postage, printing etc.). In summary, there is a group of respondents from whom it is easier to get a response, and a group from whom it is harder (but who in the end do still respond).

We do not expect the situation to change, and it is possible that with an even keener awareness of 'compliance costs' and changing lifestyles, this situation could worsen.

Within Statistics NZ's relatively fixed budget, there is a real need to reduce duplication and increase productivity in data collection, to maximise opportunities for efficiency and allow the organisation to reinvest in value-generating activity. There is also a need to be able to ensure the correct metrics are in place to allow us to measure any productivity increases.

Current situation

As mentioned earlier, data collection at Statistics NZ has been shaped by individual subject areas responding in different ways to collection issues. This approach has created three main issues.

The first is in the resourcing of data collection. What has occurred is that data collection staff have become specialised in one or two main surveys in order to be able to manage response, communicate effectively with respondents and master the technologies. This creates difficulty when we want to introduce new or casual staff to manage the commonly fluctuating work demands, or to introduce staff who work more flexible hours and days.

The second issue is that each survey has been developed with its own unique end-to-end system that covers all activities, from sample selection to collection, processing, analysis and to a lesser extent dissemination. For data collection, Statistics NZ uses well over 10 distinct systems of various vintages. This is an issue because there are so many computer systems to maintain, as well as significant costs in developing systems from scratch.

The third problem is that Statistics NZ does not have one definitive view of our interactions with respondents. This is an issue when trying to developing strategies to

3 This is based on internal qualitative research designed to better understand respondent's attitudes to surveying.

4 This is based on internal analysis of the cost effectiveness of each call when conducting the Household Labour Force Survey.

promote response by respondent groups, providing across collection metrics, and maintaining strong relationships with our respondents.

Since this situation was not caused by one factor alone, Statistics NZ's response must also be multifaceted. In managing the cost of data collection, many responses involve statistical techniques, including increasing the use of administrative data to replace direct surveying, using statistical methodologies that reduce sample sizes, and removing burden in certain sub-populations of New Zealand's economy (e.g. manufacturing) and society (e.g. Māori and Pacific people).

The easiest way to reduce the cost of collection and respondent load is by limiting direct survey activity. To do this, for the last 15 years Statistics NZ has increased its use of data from administrative sources. The greatest source of administrative data has been from the tax department, which is able to supply detailed goods and services tax (GST) data, and financial and salaries and wages information. Even with strong use of administrative data, there is still more that could be achieved through its further use.

2. Vision

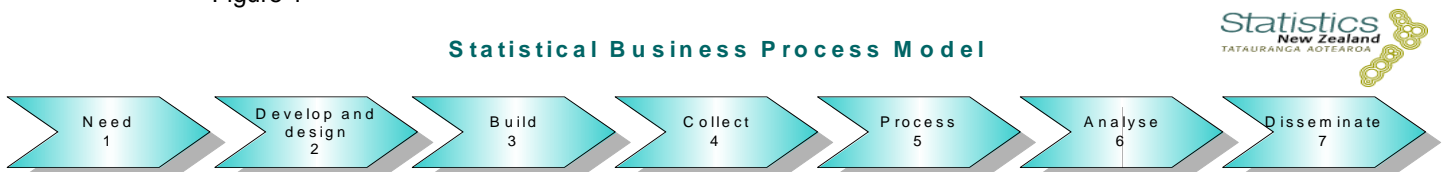
Overarching frameworks

Over the last few years, two driving strategies have shaped Statistics NZ's approach to the business of statistics. The first is the establishment of a generic business process model (gBPM) and the second is Statistics NZ's statistical architecture.

Generic business process model (gBPM)

In order to better understand the business of Statistics NZ, to inform strategy, and to identify opportunities to reduce costs, Statistics NZ used business process modeling techniques to identify all its separate business processes. These processes are described in figure 1.

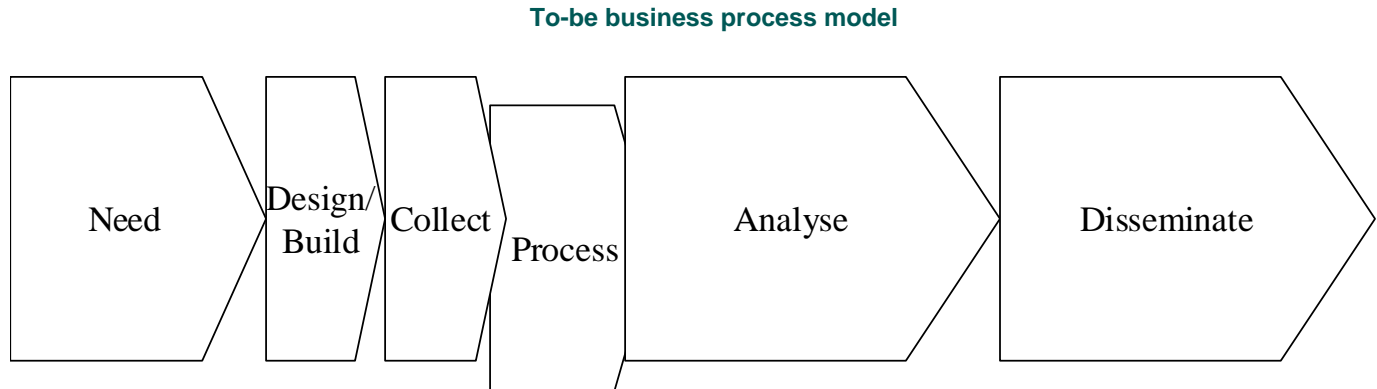
Figure 1



Once the core processes were established, the largest areas of activity and cost were determined. These were data collection (4), develop and design (2), build (3) and data processing (5).

The next step was to design a 'to-be' business process model. This model (or vision) describes a reduction in the dominance and cost of collect (4), develop and design (2), build (3) and process (5), and an expansion in the activities of need (1), analyse (6) and disseminate (7). This to-be business model is described in figure 2.

Figure 2



To enable this vision, it was clear that the existing organisational structures and statistical infrastructure would need to be rationalised to provide generic pieces of infrastructure that cut across traditional survey processes. It was also implied that a strong understanding of the cost drivers of the organisation was needed.

Statistical architecture

In addition to the proposed to-be business process model, the statistical architecture outlines the following design rules for the development and review of statistical outputs.

- Information can only be collected if a clear user need has been established.
- Information should only be collected once.
- Administrative data will eventually be used as the primary source of data.
- Surveys will only be used to fill the gaps that cannot be met from administrative sources.
- Survey and administrative data will be integrated using a comprehensive business register.
- Information should only be collected from units that can readily provide reliable and meaningful estimates.
- Large complex business units will be closely managed, to facilitate the collection of all the data that is needed from them.
- Information quality will continue to be fit-for-purpose.
- Reliance on administrative data will increase in incremental steps, beginning with the parts of the population for which tax data is robust and then expanding into more difficult areas as data issues are resolved.

Using administrative data first, with surveys filling the gaps, is a reversal of the current practice of using surveys for important information with administrative data being used where the contribution to the output is less significant. The premise of the statistical architecture is that wherever possible, administrative data should be used to replace direct surveying. Statistics NZ has been successful with this approach so far, and there are still many opportunities remaining to make better use of these data sources.

While the use of administrative data can ease the cost of data collection and improve the range of statistical outputs, direct surveying is likely to continue for some time. This paper focuses on strategies and initiatives that can improve productivity in data collection, while providing opportunities to improve data quality and reduce respondent load.

Statistics New Zealand's vision

Statistics New Zealand's vision is “turning data into relevant knowledge, efficiently”.

Within this broader vision, and from the to-be business process model described above, there is a clear direction for data collection. This is that a statistical agency adds value through processes of data analysis and not through data collection.

3. Data collection

Vision for data collection

The vision Statistics NZ has for data collection describes an integrated collection operation and integrated infrastructure. The vision has three key themes:

- one collection division
- one view of respondents
- one collection system.

Each of these themes is described in more detail below.

One collection division

Business process modeling and the gBPM have confirmed that all surveys follow very similar business processes. Therefore a strong vision is to combine the operation of all surveys (other than census) under one management structure. More importantly, is the aim to combine activities that define the data collection business process, to create a division with sole responsibility for data collection in order to:

- clarify the boundaries of data collection and Statistics NZ's other business processes
- understand the cost of, and the drivers of the cost of, data collection
- create efficiencies through standardising activity between social and economic data collection areas
- provide advocacy within the organisation for respondent issues
- provide one point of contact between the data collection division and the subject areas
- have a group to own and develop best practice in data collection, to include ideas from both the statistical community and relevant public sector disciplines, for example contact centre management, market research, print and distribution.

One view of respondents

The vision of ‘one view of respondents’ is that through more tailored engagement, there will be increased willingness to participate in our collections. This can only come about if we have information on all our interactions with our respondents. This perspective

allows collection managers to target specific subgroups, for example non-respondents, and tailor a specific response to their needs. It also allows data collection staff to communicate with respondents with full information on previous communications. Having this information will also allow us to better manage overlap across surveys for respondents. For surveys of individuals and households this will mean engaging more closely with representative groups, in particular groups that are over-sampled in aggregate (e.g. Māori and Pacific peoples). Due to sampling techniques and operational procedures, it would be unusual for individuals to be in more than one survey.

The strategy is to encourage willing cooperation from respondents, through a respondent contact service that is both universal and personal. 'Universal' in that we are aware of the broad range of interactions of the respondent, 'personal' in that we know the people involved in the process and recognise their value – we aim to build a trusted ongoing relationship with our respondents.

The vision is that targeted strategies will improve respondents' willingness to participate and reduce expensive follow-up activity.

One collection system

As described above, data collection has more than 10 data gathering systems. A clear vision is to have one system supporting all data collection.

Given business processes are so similar across collections, having just one system makes running data collection and the development of collection systems a lot easier, and can contribute to:

- increased productivity through using professional collection technologies
- reduced development costs with only one system to support and bring new collections into
- reduced training activity with only one technology and set of procedures
- increased quality in collection through associated quality monitoring.

Implementing the vision

Clearly, moving the organisation from survey specific management and infrastructure to having an integrated collection operation, and integrated infrastructure, takes time. The following section describes the initiatives Statistics NZ is undertaking to achieve this vision. Some initiatives are in place, some are well-advanced and others are just starting. Measuring the expected and actual value of these initiatives is also important.

One collection division

The first step in achieving the overall vision was to establish an integrated data collection division (IDC) within Statistics NZ. This division was established in 2002 and has operated in essentially the same form since then.

The IDC vision statement is:

"IDC aims to provide a cost effective collection service to internal Statistics NZ customers, while maintaining New Zealanders willingness to provide information."

IDC brings together all Statistics NZ's data collection activity. The exception is the Census of Population and Dwellings which, by virtue of its size and cyclical nature, is run independently. Bringing the collection activity together has allowed IDC to:

Innovations in Data Collection – Three initiatives

- own the business processes and best practices associated with data collection
- provide an unified voice on data collection issues
- manage relationships with internal subject areas in a holistic way.

The main areas of the IDC division are:

- the contact centre
- IDC operations
- the provider relations unit
- collection strategies and process improvement
- field collections.

The contact centre

The contact centre is responsible for any one-to-one relationships with respondents. This includes query resolution, following up those who have not yet responded, and running computer assisted telephone interviewing (CATI), for Statistics NZ's Household Labour Force Survey in particular. The contact centre has 52 full-time equivalent staff and works six days a week (Sunday to Friday) running two shifts between 7am and 9pm most days. These hours allow Statistics NZ to make better use of times when individuals are available for interview. The contact centre employs both permanent and casual staff, allowing us to appropriately size the centre as needed.

Basic contact centre statistics for the 2007/08 financial year:

- 51,818 inbound calls
- 447,063 outbound calls
- monthly averages:
 - 4,318 inbound calls
 - 37,255 outbound calls

IDC operations

IDC operations manages the dispatch and receipt of paper questionnaires, manages large administrative data feeds, manages our imaging and recognition centre, and deals with non-standard data. The section has 32 staff, both permanent and casual.

IDC operations statistics for the 2007/08 financial year:

- 8,841,111 migration cards scanned
- 500,000 items mailed (including reminder runs)
- 330,000 surveys captured through imaging
- hundreds of thousands of administrative data unit records loaded annually.

Provider relations unit

The provider relations team of five manages Statistics NZ's relationship with the department's top 50 corporations. Their core function is to ensure the willing supply of information from New Zealand's most economically significant corporations.

Collection strategies and process improvement

The collection strategies and process improvement (CSI) section is IDC's research and development (R&D) team. The section has 11 staff of project leaders and researchers. The R&D team allows IDC to look beyond the day-to-day issues of data collection. While the CSI work programmes changes, based on the priorities of the IDC division, in general the team works on:

- collection development projects (including Project Contact described below)
- respondent-load management strategies
- measurement of respondent load and overall data collection performance
- development of electronic scripts for the contact centre (e.g. CATI)
- specific process-improvement initiatives
- respondent and operations research.

CSI are seen as 'agents of change' and are the promoters of data collection best practice.

Field collections

Field collections manages the field workforce that interviews individuals, households and some businesses (e.g. for the consumers price index). There are 22 administrative staff and 160 field interviewers. In the 2007/08 financial year, the team managed 88,065 cases. Improving the professionalism and productivity of the field workforce is described below in initiative 1

Other

IDC also manages the work of Statistics NZ's respondent advocate. The respondent advocate's role is to promote respondent interest in the work of Statistics NZ, assess new statistical collections, and assess the validity of respondent-load-related exemption requests.

One view of respondents and one collection system

To manage respondents personally, we must have one version of our interactions with our respondents. The contact centre aims to have information about a respondent displayed across collections, to help staff engage knowledgeably with the respondent in any interaction. 'One view of respondents' also allows us to target collection activity to many more subgroups, thereby improving relationships and increasing response. Having one collection system will improve productivity and reduce IT development and maintenance costs.

Initiative 2 below describes the initiative to achieve this vision.

4. Initiatives

Initiative 1 – Best Practice Field Collections Project

Background to the project

The Best Practice Field Collections (BPFC) Project was initiated to ensure that the field collection workforce is professional, productive, skilled in the use of the latest methods and technologies, and supported by appropriate systems and processes. The BPFC project consists of initiatives that are designed to deliver this vision, and cover key aspects of the field operation. The initiatives include:

Innovations in Data Collection – Three initiatives

- Best practice standards for interviewers
- A new performance management system for interviewers
- A new recruitment process for interviewers
- Management information reports on the field operation
- Updated employment terms and conditions for interviewers
- Development of an e-learning training module for the laptops
- Review of the field collections structure
- Review of internal field business processes
- Technology initiatives (mobile phones, email on laptops)
- Leased vehicles for interviewers.

The BPFIC is a three-year project that began in 2007 and will be completed by the end of 2009. The majority of the initiatives have been completed or are underway.

Project objectives

The key objectives of the project involve identifying and implementing initiatives that support the vision described earlier. In particular to:

- review and recommend changes required to ensure field collection systems and processes support a 'best practice' field collection model
- identify a field collection business unit structure to support the field collection vision
- ensure the field collection workforce has the capability to deliver, in a cost-effective manner, high quality data across multiple types of surveys
- integrate field collections within the wider IDC division
- implement changes that result in interviewers who are committed to:
 - working a minimum number of interviewing hours per week (on average)
 - undertaking a variety of work covering Statistics NZ activities
 - being involved in programmes for continuous improvement
 - a performance management system
 - a linkage between performance and pay
 - being engaged in their work.
- review and recommend, in consultation with the Public Service Association (PSA) (union), any changes required to the employment terms and conditions for field interviewers as a result of the proposed field collection vision.

Successes

With many of the initiatives completed (or nearing completion), a number of significant project outcomes are being realised.

- Best practice standards for interviewers – A suite of 12 quality assurance processes have been developed for systematic implementation. A corresponding quality assurance role has been created in the field collections section.
- A new performance management (PM) system for interviewers – The new PM system is based on tangible measures such as wage cost per survey, or response rates; and behaviours that support the organisation's vision, mission and values.
- A new recruitment process for interviewers – Integration with Statistics NZ's Internet-based recruitment process, including updated documentation to reflect other project initiatives and outcomes.

Innovations in Data Collection – Three initiatives

- Management information reports on the field operation – A suite of real-time reports covering key field metrics (cost by survey, etc).
- Updated terms and conditions for interviewers – Work with the PSA gained settlement on key employment terms and conditions for interviewers, which are consistent with the project vision.
- Technology initiatives (mobile phones, email on laptops) – Implementation of technology initiatives to improve communication and efficiency in the field.
- Leased vehicles for interviewers – Interviewers who travel a significant number of kilometers annually have a leased vehicle as a cost-efficiency initiative.

Learning

Interviewers, by the nature of their work, are isolated from both the organisation and other interviewers. The effect of this should not be underestimated. Measures to compensate for the isolation need to be evaluated and applied where possible. The project has identified the following opportunities to help bridge the gap.

- Regular face-to-face meetings with interviewers to discuss issues and update them on project activity are important for the success of project initiatives.
- On-going, regular collective meetings of interviewers are important to help them feel part of both the team and the wider organisation. Ideally, regional meetings would be held every three or six months.
- Information on how field interviewers fit into the collections division, and regular updates on the organisation's activities, are provided. Technology should be used to disseminate this information (e.g. laptop email).
- Collective training sessions should focus on professional development, rather than survey specific content.

Where to from here

From a project perspective, all initiatives will be fully implemented by the end of 2009. Some, such as the performance management system, require bedding-in over time and require collective review and input from the field management team. Cost-efficiency initiatives, such as leased vehicles for interviewers, will be expanded over time (provided the expected mileage of interviewers remains high). For all initiatives, an on-going monitoring programme will be established to report on project outcomes.

From a business perspective, implementation of the BPFC project will place field collections in a strong position to manage the increased demands of the work programme. There are a significant number of new social surveys to be introduced to the field over the next five years.

Initiative 2 – Project Contact

Background to the project

The initiative to implement the vision of 'one collection system' is Project Contact. Project Contact was established in 2004 as part of a wider suite of initiatives to streamline the operation of Statistics NZ.

Project Contact aims to increase productivity in the following ways.

- Increase the contact centre staff's productive time by automatically presenting the next contact. Staff can be moved easily between calls, the system manages appointments, and the staff member does not need to record and monitor. The

result is that we can make more calls with the same number of staff. There are also reduced training costs for moving between collections.

- A move to electronic methods for reminders reduces costs associated with sending letters.
- A reduced development cost through new surveys becoming a new collection within the infrastructure. No new system needs to be built.

Project Contact will deliver a standard set of tools for the collect phase described in figure 1.

Project objectives

The vision of Project Contact is to create a solution that supports users to interact efficiently and professionally with respondents.

There are six key outputs to Project Contact.

1. Contact centre – Maximising staff productivity by providing a collection management system to schedule data collection staff, and effectively manage both inbound and outbound calls. Through this system, work can be planned for and allocated to agents, helping them to easily move their work focus between collections as required.
2. CATI/Web/CAPI/CADI – Questionnaire creation tools provide the ability to easily create questionnaires and outputs in various formats, depending on the mode required – CATI (telephone-based interviewing), CAWI (web-based interviewing), CAPI (for face-to-face interviewing), CADI (for any incoming mail that cannot be scanned).
3. Collection management – A system that manages the survey collection process and respondents end-to-end. A survey is treated as a collection, including the request for information (egg distribution of mail-outs by a mail house, request for administrative data, or the deployment of field staff to visit households), mark in, and follow-up activities.
4. Customer relationship management (CRM) – A CRM solution that includes delivery of a data repository for contact information, as well as an instinctive user interface through which to view the information. Through this, providers will be managed on an individual rather than a survey basis, allowing for management of respondents across all surveys.
5. FAQ – To provide a tool for agents to easily find answers to customer queries.
6. Integration – To provide seamless integration between the different tools used for Project Contact, as well as between these tools and the rest of the Statistics NZ architecture.

Project Contact has four distinct pieces of infrastructure.

1. Microsoft Customer Relationship Management (MS CRM) – the heart of Project Contact. There was a deliberate decision made to use an off-the-shelf tool. MS CRM has a great deal of flexibility, which better allows it to be integrated.

Innovations in Data Collection – Three initiatives

2. QMaster telephone system – manages queues and staff assigned to queues.
3. Blaise – used to develop non-response follow-up script for CATI and field interviewer administered questionnaires CAPI.
4. Lotus Notes-based computer assisted interviewing system (CAI) – used to manage replication to the field workforce. Interviewers see this system.

All the different tools are loosely coupled, meaning that it should be a simple process to add or remove aspects of the system. Previously, to add new functionality into the 'collect' space would mean at least 10 separate service requests, to allow for all the different systems – now it's one.

Successes

One of the key successes was achieved early in the project. This was the delivery of a new telephony system to the contact centre. This system has professionalised the operation of the contact centre, and allowed for management and monitoring of inbound and outbound queues.

In terms of overall operation, the Project Contact system has been rolled out to the following surveys:

- General Social Survey (GSS) – field-based social survey
- Time Use Survey – field-based social survey
- Annual Frame Update Survey – paper-based business survey, with CATI non-response follow up
- Monthly Frame Update Survey – paper-based business survey, with CATI non-response follow up
- Business Performance suite of surveys – four paper-based business surveys.

Another success is the speed at which new collections can be added to the Project Contact system. A new collection, the Energy Use Survey, is being delivered to the infrastructure for far less cost than developing an end-to-end system for it.

Learnings

The project has been operating since 2004. Keeping to the vision has been difficult with many managerial changes over the course of the project.

Most of the delay in the project was due to the selection of the CRM. Statistics NZ is currently on its third CRM. MS CRM was proven to be far easier to integrate into our existing infrastructure, because it can work both as a CRM and a platform. This means we can be confident that we can evolve the technology to fit any changes to the business process or new technology opportunities.

The delay in bedding-in the CRM impacted on the expected return on investment of the project. In addition, over the years the project has been running, the data collection approach has changed to take up some of the principles of standardisation, while still operating within separate systems. This has meant that some savings from the project have been achieved through different means. An example is the intention to use electronic reminders to reduce the cost of data collection. In the absence of this system,

the IDC operations area has streamlined its postage activity and made savings, but in a different way.

The learning is that savings need to be very clearly articulated in order for them to be achieved. It is also important that savings are gathered or accounted for, before quality or cost increases dilute them.

However, the core vision has remained intact, and the project has continued to have the support of its stakeholders.

Another key learning is that with the tools we chose, there were often overlaps in functionality. For the architecture to work, important decisions needed to be made about which components were master and which were slaves.

Until all collections are included in Project Contact, the project is simply introducing another collection tool for users. However, so far, user response has been more positive than expected. There is strong understanding of the goal to replace the myriad of current systems with one system, and one standard process.

Project Contact is requiring considerable process re-engineering to support a much more cost-effective contact operation. The issue with standardisation is that someone has to give something up. Managing this remains difficult, but worth the effort.

Where to from here

The project will continue for the next two years, over which time we have a systematic process for adding the remaining collections into Project Contact, which will constantly increase the wealth of information available.

Initiative 3 –Census of Population and Dwellings

Introduction and scope of the New Zealand Census

Statistics NZ has developed an innovative approach to designing enumeration areas for the New Zealand Census of Population and Dwellings.

The next five yearly census will be undertaken in March 2011. The geographic coverage includes all dwellings and population (including overseas visitors) in the two main islands and a number of smaller offshore islands. It is estimated that there will be 1.7 million dwellings and 4.5 million people in New Zealand at the time of the 2011 Census. A temporary labour force of approximately 7,000 people is recruited to undertake the delivery and collection of census forms for each dwelling.

Historically New Zealand has experienced a high participation rate in the census. For the 2006 Census the coverage was 98 percent, and the response rate was almost 95 percent.

The results from the census provide an invaluable source of small-area data that informs decision making relating to New Zealand's economic and social well-being. Census data is used to ensure fair electoral representation, and to distribute government funding for health and education services.

Enumeration design

Approximately 6,500 enumeration areas, known as sub-districts, are created. Census collectors undertake the door-to-door delivery and collection of census forms for each sub-district. New Zealand is divided into approximately 45,000 meshblocks, which are the smallest geographic area classification and the lowest level for the collection and dissemination of final output. Meshblocks vary in size, from city blocks in urban areas to large tracts of land in rural areas. The number of dwellings in each meshblock varies from none to more than 700.

Sub-districts are made up of geographically contiguous meshblocks. A collector delivers and collects census forms for the dwellings and people in each sub-district. Sub-districts are organised into approximately 412 districts, with each district having a district supervisor who manages a team of collectors. The workload of a Collector involves delivering census forms to all dwellings in their respective sub-districts and collecting forms from households that have not returned forms by internet or mail.

In the past, sub-districts and districts have been developed through a resource intensive clerical process. The process required the physical checking of dwelling estimates and paper maps, to ensure that the required workload was achieved and that the enumeration areas were contiguous with no physical features restricting a collector's travel. The manual nature of the process resulted in many sub-districts having disparate workloads and weak contiguity relationships between neighbouring meshblocks. For example, two meshblocks were neighbours but were divided by an impassable motorway.

Early developments for the 2011 Census recognised the deficiencies with the historical sub-district design model. A small multi-disciplinary team worked with the University of Canterbury to determine if this type of problem could be solved using operational research.

An innovative solution

Previous experiences in literature

The literature related to the collection area problem is quite expansive, with papers going back decades, and spanning the fields of geography, graph theory, computational complexity, and combinatorial optimisation.

Terms given for the related problems include spatially constrained aggregation methods, zone design, multi-site land use allocation, constrained clustering, districting, sales territory alignment, voting tract creation, and supervised regionalisation methods. Some papers give evaluation techniques only, which can rank solutions but not generate them. The problem at hand is one of design, for which solutions must be generated.

It was concluded that there was no obvious solution to the problem in the literature. It is extremely difficult to solve a problem in which a high degree of contiguity is desired, with strong constraints on the aggregated regions (such as having all workloads be nearly identical).

However, the collection-area problem Statistics NZ faces is somewhat simpler than the problems attempted in the literature, since the workload need not be tightly constrained. The problem is one of ensuring contiguity while maintaining the workload within a range.

Researchers have used true optimisation only for small cases, and used approximating heuristics for realistic cases. The approach here falls in the category of heuristic optimisation – optimisation tools are used to find the solutions, but solutions are not guaranteed to be optimal. However, optimisation provides a powerful tool, and the solution method has some theoretical justification.

Method

A two-stage solution process was adopted by Statistics NZ. In the first stage, a variant on Kruskal's graph algorithm (1956) was employed. This is used to generate proposed solutions, in the manner of column generation. The second stage is the selection of the best set of solutions using integer programming methods. The algorithm for the first part is in custom code (C++) while the second part is written in AMPL and uses CPLEX.

The algorithms require input data from three sources. These comprise information on the contiguity of meshblocks, the workload for each meshblock, and the solution parameters for the operational research methods.

Contiguity

One of the key aims in respect of the design of sub-districts is to make enumeration as efficient, easy and safe as possible for collectors. Key factors are to minimise travel time by providing collectors with compact and logical areas to enumerate, to avoid the need to cross difficult or hazardous barriers, and to enable the enumerating of their sub-district without having to leave its confines. The clerical processes previously used did not allow for analysis of the rich geographic data available by using geographic information system (GIS) technology.

A matrix showing the contiguity relationship between each meshblock was developed. A score was assigned to each relationship, and this score was used by the first stage of the process to determine the best possible combination of meshblocks.

The variables used to determine the contiguity between meshblocks were: the proportion of the border perimeter shared between meshblocks, centroid-to-centroid distance (geometric centroid), and road connections. The contiguous relationship between meshblocks was broken where neighbouring meshblocks were divided by a motorway, busy road, river, or regional administrative boundary.

Workload

The workload for each meshblock was formulated using the number of dwellings and an approximation of the time required to enumerate each dwelling. Meshblocks were combined to form sub-districts where the workload for each was within an acceptable predetermined range. Further research is underway to establish a more complete workload model, incorporating factors such as difficulty of enumeration and topography.

Solution parameters

The third aspect of the inputs comprises the solution parameters for the operational research algorithms. These include acceptable workload range, target workload, and the penalties for being outside the specified range. Statistics NZ has preliminary estimates of these parameters from the development stage. As we progress through sensitivity analysis we expect to further refine our parameter estimates.

Current status

Statistics NZ is at the pre-implementation stage of the project. Developments have been encouraging, both in the design of sub-districts and the operational aspects. We expect it will be possible to produce a solution for evaluation within a few hours (by region) compared with months using the previous methodology. This reduction in solution time means it is possible to try different options for the design of sub-districts. Consequently, we can assess the sensitivity of the solution to changes in the input data within a useful timeframe. This enables us to assess the risks and effects of changes in those inputs. For example, the sensitivity of the solution to the way workload is calculated, the effect of difficulty on workload, and the effect of different rates of take-up of the Internet and mail-back response modes.

Savings for the generation of sub-districts are approximately 50 percent or \$150,000. The solution methods will also be applicable for other layers of census field areas, future censuses, and other similar projects. The design of the sub-districts appears to facilitate more efficient enumeration, which leads to a reduction in the number of sub-districts and consequently the number of collectors required. Early indications are that this reduction in sub-districts is around 10 percent.

The benefits described above will enable better informed management decisions for field operations. Interestingly, all these benefits are a consequence of addressing the sub-district creation problem.

The operational benefits will include narrowing the range of workloads, and creating more sub-districts of an ideal size. This results in more sub-districts with the target workload and fewer sub-districts that are too small. This should lead to better value from collectors and better value for collectors. There are also benefits in improving the ease of enumeration for collectors, and their safety. Both these elements make it easier for collectors to complete the enumeration within the required timeframe.

Conclusion

Statistics NZ has used innovative methods and made better use of existing data, technology and skills to provide an improved solution to an old problem. The work described above has the potential to be applied to similar problems.

5. Conclusion

It is clear that data collection has to face issues of reduced respondent willingness, reduced access to respondents, and increasing costs of inputs to data collection. This occurs within Statistics NZ's relatively static budget, an increasing demand for information, and a desire to maintain quality levels.

To continue to operate in this environment, many initiatives have been established to improve data collection quality, but more importantly to reduce the cost of data collection. The case studies presented highlight three of the largest initiatives.

However, as described in the initiatives, benefits are difficult to quantify (census example), or will take some time to be delivered (BPFC and Project Contact). This is not because there has not been work to quantify these benefits, but because the factors

Innovations in Data Collection – Three initiatives

influencing costs and quality of data collection are complex or beyond the control of the statistical agency. This is particularly the case with respondents' willingness to supply information, and their availability. Initiatives can be put in place to better manage target groups (Project Contact), but issues of compliance are wider than this and require an organisation-wide response.

None the less, it is important to be able to clearly state the benefits, and measure them to confirm they have occurred. This is something Statistics NZ needs to do better. Making more use of the business process model, along with having good information on cost drivers at each stage of the model, can help.

With initiatives in train to improve productivity within data collection, now is the time to focus back on the statistical techniques we can use to reduce costs, as well as fulfilling the aims of the statistical architecture. This should include re-examining quality and its relationship to the need of the statistical output, further increasing the use of administrative data, and using better sampling techniques to ensure respondent load is fairly distributed. From this base, Statistics NZ can be seen to be credible by our respondents and their willing supply of information can be ensured.

References

Ang L (2008). "Analysis of call backs made by the call centre", unpublished manuscript Statistics New Zealand.

Dunnet G (2004). "End-to-End Business Model v3", unpublished manuscript Statistics New Zealand.

Kruskal, Joseph. B., "On the Shortest Spanning Subtree of a Graph and the Travelling Salesman Problem," *Proceedings of the American Mathematical Society*, Vol 7, No. 1 (Feb, 1956), pp. 48-50,
[http://www.jstor.org/sici?sici=0002-9939\(195602\)7%3A1%3C48%3AOTSSSO%3E2.0.CO%3B2-M&cookieSet=1](http://www.jstor.org/sici?sici=0002-9939(195602)7%3A1%3C48%3AOTSSSO%3E2.0.CO%3B2-M&cookieSet=1), accessed 3 June 2008.

McKenzie R (2008). *Statistical Architecture*, unpublished manuscript, Statistics New Zealand.

Statistics New Zealand (2008). *Integrated Data Collections Report: September 2008 quarter*, unpublished manuscript, Statistics New Zealand.

Thorne H (2008). *Understanding Respondents' Attitudes to Surveying*, unpublished manuscript, Statistics New Zealand.