

Innovative Production Systems at Statistics New Zealand: Overcoming the Design and Build Bottleneck

Allyson Seyb¹, Ron McKenzie¹, and Andrew Skerrett¹

In 2011, Statistics NZ embarked on an ambitious ten-year programme of change called *Statistics 2020 Te Kāpehu Whetū: Achieving the statistics system of the future*. This article outlines a key component of this approach to transforming how Statistics NZ delivers its statistics. This involves modernising all aspects of statistical production – from identifying need, right through to dissemination – to ensure the organisation has the functionality and capacity to deliver now and into the future. Standardisation of processes, methods, tools, and systems to increase flexibility and efficiency is a key component of this transformation.

Key words: Statistical production process; business architecture; administrative records.

1. Introduction

Statistics New Zealand (Statistics NZ) is the major producer of official statistics in New Zealand and leader of the New Zealand Official Statistics System (OSS). Like other national statistics offices, Statistics NZ faces many challenges in supplying the broad range of information that is required by modern users of official statistics. The organisation strives to maintain the quality of its statistics in terms of relevance, accuracy, timeliness, accessibility, consistency, and interpretability. At the same time, the organisation is in a tight fiscal environment that requires us to minimise costs and maximise effectiveness. Statistics NZ's aim is to produce statistics that are fit for purpose in a cost-effective and sustainable way.

In 2011 Statistics NZ embarked on an ambitious ten-year programme of change called *Statistics 2020 Te Kāpehu Whetū: Achieving the statistics system of the future* (Stats 2020 in short). The Stats 2020 programme of change is extensive. It involves changes to the way the organisation works; who it works with and how; to the systems, tools, and processes used; and to the skills, attitudes, and behaviours needed for success. These elements of change are viewed from four perspectives: transforming how statistics are delivered; leading the New Zealand OSS; obtaining more value from official statistics; and creating a responsive and sustainable organisation ([Statistics NZ 2010](#)).

This article outlines a key component of the approach adopted for the first of these: transforming how Statistics NZ delivers its statistics. This goal focuses on Statistics NZ's

¹ Statistics New Zealand, The Boulevard, Harbour Quays, PO Box 2922, Wellington 6140, New Zealand. Emails: allyson.seyb@stats.govt.nz, ron.mckenzie@stats.govt.nz, and andrew.skerrett@stats.govt.nz

Disclaimer: The opinions, findings, recommendations, and conclusions expressed in this article are those of the authors. They do not necessarily represent those of Statistics NZ, which takes no responsibility for any omissions or errors in the information contained here.

role as a producer of official statistics and involves modernising all aspects of delivery from identifying need, right through to dissemination, by ensuring that it has the functionality and capacity to deliver now and into the future. Standardisation of processes, methods, tools, and systems to increase flexibility and efficiency is a key component of this transformation.

The following objectives shaped this transformation project:

- Modernising information technology (IT) systems to leverage the opportunities that new technologies offer for working more effectively.
- Reducing risk by replacing legacy software.
- Standardising processes so that staff can move between outputs without having to learn new tools and systems. This will increase efficiency and remove a barrier to innovation.
- Updating and standardising methodologies to eliminate unnecessary diversity and duplication conceptually, in terms of business and information concepts, and practically, in terms of methods and technology.
- Creating statistical methods and systems that support a growing need for more detailed data analysis.
- Maximising the use of administrative data to reduce respondent load, minimise costs of data collection, and maximise data re-use.
- Developing and applying a quality framework and quality indicators to inform users about the quality of the data before and after statistical processing and to monitor and optimise the performance of the IT system.
- Maintaining the quality of the statistics produced by ensuring they continue to meet user needs.

While Statistics NZ is only one year into the official change programme, many of the initiatives had already been established in the organisation. These included creating common processing platforms for clusters of social and business collections, using standard statistical tools, and adopting a generic Business Process Model and Quality Management programme. Stats 2020 incorporates these existing initiatives into a cohesive, recognised, and measured programme, raising both the organisation's productivity and its flexibility. A key element of this is the replacement of legacy IT systems with fewer systems that are more flexible and more standardised, so that risks are eliminated and systems are cheaper and easier to maintain and manage.

The programme of change also builds on previous work to confirm the approach to data collection that will support future information needs. The existing structure of collections at Statistics NZ is based on a traditional model of sample surveys supported by a range of administrative data sources. By 2020, the organisation's aim is that administrative data will be the primary source of information, supplemented where necessary by direct collection. Since 2008, the organisation has focused on designing statistical architectures for both business and social statistics (see e.g., [Bycroft 2011](#); [McKenzie 2008](#)). A statistical architecture describes at a high level how future collections of survey and administrative data will be managed and combined to produce statistical outputs. These architectures have helped structure thinking and guide the development of an integrated statistical system to meet the goals set out in Stats 2020. The new IT platforms for

processing household and economic collections are designed to support the new statistical architectures and Stats 2020 goals.

The environment that Statistics NZ operates in has changed over the last ten years. Societal and economic changes as well as greater diversity have heightened demand for a broader range of statistics that will support more complex analysis at greater levels of detail (Glaude 2008). Users of official statistics want statistics to be more relevant to their needs. The role and contribution of Māori (the indigenous people of New Zealand) – economically, socially, and demographically – to the growth of the country is increasing significantly. Given this, there is an increased emphasis on information for and about Māori to support decision making and understanding.

Statistics NZ's role as the leader of the decentralised New Zealand OSS, as well as a producer of official statistics, has been reconfirmed and strengthened by the New Zealand government (Statistics NZ 2010). The organisation is mandated to coordinate statistical activity across government by driving the overall performance of the OSS and ensuring New Zealand both gets the information it needs and that this is value for money, at the lowest possible cost to government, the community, and suppliers of the data (State Services Commission 2011). This places greater expectations on the organisation to deliver a more cost-effective and productive statistical system. At the same time, the fiscal environment is becoming more and more challenging. Despite the growing costs of data collection and processing, perceptions of increasing response burden, and smaller budgets, the public sector is trying to increase and improve the services they offer.

Users of official statistics expect the national statistics office to improve the quality, quantity, and “time to market” of their statistics. The availability of vast stores of administrative data across government sectors, and the ability to easily store and manipulate the data, presents a wealth of opportunities to identify and reduce duplication, increase utilisation, and improve collection effectiveness.

Addressing these drivers has led to the developments described in this article. Section 2 describes the business statistical architecture that is helping Statistics NZ to achieve the desired changes to its systems and direction. The focus of the article from this point onwards is on economic statistics and the new economic platform. Section 3 focuses on statistical development, particularly the change in IT strategy that has seen statistical development move from being an expensive drawn-out process to a process that often does not involve IT staff at all. Standardisation and automation of statistical processing are discussed in Section 4. Section 5 focuses on governance. Section 6 discusses lessons learnt and the transformation in the culture of the organisation as a result of the changes in the way statistics are produced. Section 7 has some concluding remarks.

Throughout the article, the term “subject-matter specialist” refers to analysts who process and analyse particular surveys, such as the survey of Retail Trade, and the term “methodologist” refers to analysts who work in the methodology area of the organisation. Traditionally the roles of subject-matter specialists and methodologists were quite different from each other – methodologists primarily focussed on designing surveys and subject-matter specialists focussed on processing, analysing, and disseminating the data collected. More recently within Statistics NZ, the roles have begun to overlap – subject-matter specialists are developing skills in survey design and methodologists are taking a more active role in understanding the entire statistical production process (including the

technology solutions) and leading the cost-benefit analyses and decision making. When the term “analyst” is used in this article, we are referring to either a subject-matter specialist or methodologist, or both.

2. Statistical Architecture

A key objective of Stats 2020 is to make maximum use of administrative data while continuing to produce economic statistics that are fit for use. Further, the aim is for administrative data to be the first source of data for statistical outputs ([Statistics NZ 2010](#)). This key principle of the new statistical architecture has been a driver for developing the new economic statistics platform.

2.1. Background

Statistics NZ has had access to a variety of administrative datasets for many years. A tax-based business register was developed at the end of the 1980s after a value added tax (called “goods and services tax”, or GST) was introduced in New Zealand. Tax data was introduced into the quarterly and annual industry estimates of financial statistics by using tax data for strata containing the smallest units. The risks to quality from using tax data were mitigated by ensuring that the tax strata contributed less than 15% of the estimate.

A growing range of user needs created pressure for the development of a new approach. The existing survey-based method provided the information needed for the compilation of estimates of GDP, but could not meet the growing demand for detailed statistical information to support policy development and monitoring. For example:

- studies of business development need to focus on small businesses, where sample coverage is limited.
- longitudinal analysis needs to be able to track the performance of businesses over time.
- the non-profit satellite account of the national accounts needs information about units with relatively small financial flows and numbers of employees, which were not well represented in sample surveys.
- existing sample surveys could not provide the detailed information needed to estimate regional GDP.
- increased use of microdata analysis increased the demand for full-coverage unit record data.

To respond to these emerging needs without increasing respondent load, Statistics NZ embarked on a strategy that places a greater dependence on administrative data. A new high-level statistical architecture for economic statistics was formally approved during 2006 and a programme of methodological development, assessment, and implementation began. The architecture for economic statistics has now been subsumed within the broader Stats 2020 programme.

2.2. Principles

The new statistical architecture for economic statistics is shaped by the following guiding principles ([McKenzie 2008](#)):

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

These principles represent a significant change in the collection and production of business statistics (Kent 2011). Using administrative data wherever possible, with surveys filling gaps in information needs, is a reversal of the previous survey-based strategy. The developments currently underway are applying an approach that combines the use of survey data for large complex businesses with tax data from the rest of the population. Tax data is used directly for most of the population and surveys are targeted at businesses or data items that make a significant contribution to statistical outputs.

Statistics NZ has a long history of using administrative data, for example using customs data to create estimates of changes in overseas trade volumes and building consents data from local councils to create a sampling frame for the quarterly survey of building activity. However, most of the administrative data currently incorporated into business statistics has been tax data sourced from the New Zealand Inland Revenue Department. Tax data is relatively straightforward to use, because the reference numbers used by the administrative agencies are already mapped on the business register, so matching and coverage issues are easy to resolve. Furthermore, the data items are financial variables that are well defined. Limited use is being made of electronic card transaction data. In the future, this will be developed further and other data sources such as livestock records and store scanner data will be explored for use in our agriculture production survey and price indices respectively, and methods to make use of these new data sources will be introduced.

Developing a platform that can process administrative and survey data meant that a number of issues specific to administrative data had to be addressed:

- Administrative data comes in huge volumes that need to be processed and stored. An extract, transform, and load (ETL) workflow tool used can load large volumes of data quickly and efficiently.
- The processes for receiving administrative data are now well controlled. Duplicate records are often received. The administrative agency updates or corrects records without any warning. Methods for versioning data, eliminating duplicates, and managing revisions to data have had to be developed.

- The large volumes of data make traditional editing methods impractical. It is not possible for a statistical processor to examine every record. The development of automated editing and correction processes has been essential for the use of administrative data. An associated benefit is that the platform can apply the same automated processes to survey data.

2.3. *Range of Methods*

The Stats 2020 objective of maximising the use of administrative data means that statistics will be produced in different ways. The range of methods for using administrative data will change and expand as new data sources are introduced and assessed and new methods are developed and tested. The following are some of the ways administrative data is used currently or will be used in the future:

- A tax-based business register will continue to be used as the sampling frame from which survey samples are derived. For example, some information about business operations cannot be obtained from administrative sources.
- Financial collections will continue to use a mix of different methods for different parts of the population.
- When the administrative data definitions are closely aligned with the concepts required for the relevant statistical output and reporting is timely, administrative records will be used directly. This approach supports the production of aggregate estimates and provides additional unit-level data for microdata analysis.
- If the concept collected in a tax form does not exactly match the concept for a particular statistical output, some data modelling might be needed to produce robust estimates. In some cases, the statistical model can be estimated using data from other units in the administrative dataset. An example of this is combining data from two-monthly and monthly GST returns to produce quarterly statistics. In other situations, sample survey data might be needed for calibration of the model.
- Data records from several tax forms might be combined with survey data to produce the required statistical outputs. For example, Statistics NZ is planning to develop a measure of quarterly operating surplus as input to an income measure of quarterly gross domestic product. To minimise respondent load, estimates for small and medium-sized enterprises will need to be derived from two alternative sources: income and expenditure from their GST return and wages and salaries from a monthly return used to collect taxes on wages and salaries at source.
- Administrative data records will continue to be used in imputation for nonresponse in surveys.
- Administrative data will continue to be used as auxiliary data to optimise estimates from surveys (e.g., generalised regression estimates).

To support future uses of administrative and survey data, Statistics NZ needs systems that can store data in such a way that allows it to be re-used in a wide range of formats and for different purposes.

2.4. *New IT Platform to Support the New Statistical Architecture*

Our existing economic survey infrastructure had the following characteristics that made the new statistical architecture harder to support:

- Each survey had a separate stand-alone custom-made system.
- Survey-based systems were designed around small data volumes and could not easily be extended to handle the vast volumes of administrative data.
- Linking data from different sources was arduous.
- Data structures were inflexible, making it difficult to introduce new data sources. Data was stored in highly normalised, fixed tables optimised for transaction processing. The column names in these tables were often subject-matter specific and thus rigidly defined for a specific topic. The links between tables were also fixed and again were rigidly subject-matter specific. Reporting and analysis were customised to each production system. This allows fast processing, but changes to processes are complicated.

The microeconomic platform (previously known as BEST) developed a new storage system with entirely different data architecture. This platform has introduced two major innovations that are critical to the implementation of the new statistical architecture: firstly, it is designed for analysis, and secondly, it has flexible data storage.

2.4.1. *Designed for Analysis*

Our existing information systems were optimised for fast transactional processing using what is known as the online transactional processing (OLTP) model for storing data. This system design maximises processing speed and minimises storage with a high degree of normalisation. A strong focus on the data for the latest survey period meant that recording and understanding the events and decisions affecting data for earlier periods tended to be secondary. An emphasis on the most up-to-date value of each datapoint meant that when correcting data, the values replaced are either overwritten or made inaccessible in a log or separate history table. Analytical capability is often achieved by exporting data to a separate dataset which is structured in a way that supports analysis by tools outside the processing system.

When developing information systems, there is always a trade-off between speed, flexibility, and analytical capability. Existing survey systems were designed to favour speed over flexibility and analysis. Speed was important in a traditional survey processing environment, when large volumes of survey data had to be entered and edited by survey operators, but loss of history is a constraint in an environment where the emphasis is on analysis.

In developing the new platform, it was recognised that Statistics NZ is shifting from its narrower focus on data collection to becoming an organisation that provides a fuller service by collecting, analysing, and disseminating information. Data processing is becoming less important, while analytical power and flexibility in dissemination are increasingly important. To support this change in direction, the new economic statistics platform has been optimised for analysis, with analytical capability being built into the basic database design.

An online analytical processing (OLAP) methodology (Thomsen 2002) has been adopted to support this change in direction. OLAP is very common in the corporate world and many national statistics offices use OLAP as a read-only output data system that is exported from OLTP survey systems. The new production systems avoid OLTP altogether and process and store data within an OLAP analytical schema.

The software change that makes this approach possible is the ability to momentarily “pivot” a “narrow” single-cell table into a “wide” table, which allows the system to imitate an OLTP system when processing data. The wide table has variables as columns and units as rows, which is the format required by tools such as Banff and many other data-processing tools. The data is pivoted from the wide table back into the narrow table required for OLAP when bulk data processing is complete.

This OLAP approach provides a time-intelligent, version-controlled, metadata-rich datastore. The increase in computer processing power means that losses in processing speed are manageable. Large volumes of data (with versioning) are accessible to analysts without significant loss of speed in bulk data processing.

2.4.2. Flexible Data Storage

The microeconomic platform datastores are also designed for flexibility. In a traditional OLTP system, the data tables have column names that match the input and output data. For example, if a survey has a question about income, there will be a column called “Income”. This hardwiring of column names requires systems to be rebuilt to handle new surveys. Systems with hardwired columns are difficult to change when new data arrives. With large volumes of administrative data to process and statistics combining various survey and administrative data sources, this approach is no longer practical.

Statistics NZ has adopted a cell-based approach to data storage that eliminates the need for fixed column names. This means that building a stand-alone system for each new survey or administrative series is no longer necessary. The cellular data warehouse schema adopted by Statistics NZ takes OLAP to the extreme by adopting a single-cell table structure. In this model, each row in the narrow table contains one data item, called a fact. The other columns contain standard metadata items, such as description, frequency, and date. This ensures that all datasets have standard metadata including a record of the rules used to change the data. When data is processed, it is pivoted on the fly so that each variable becomes a column in a wide format table. When all the transactions are complete, the data is unpivoted back into the cell-based format of a narrow single-cell table. This process is fully automated.

The cellular approach has significant advantages for the production and analysis of statistics:

- Reports, queries, and processes developed for one collection work easily for another.
- Reports, queries, and processes developed for survey data can be applied to administrative data, and vice versa.
- A creation date on each row makes the data suitable for both point-in-time and longitudinal updating and analysis.
- Metadata is stored alongside data.
- The system is resilient to any change in the structure of the data.

- Administrative data and survey data can be processed and stored on the same platform.
- Metadata for survey and administrative data is stored in a way that facilitates data integration.
- Sharing data between teams is a matter of security settings allowing access to data, rather than an exercise in IT support and redesign.
- Subject-matter specialists can design and build production systems for new outputs, and maintain existing production systems, without being constrained by data storage or system barriers.
- Administrative and survey data are stored in exactly the same way, so solutions that combine survey and administrative data are easy to implement.
- The ETL Workflow can extract and load large volumes of data very quickly.

The disadvantages of this cellular format are:

- Processing performance will always be slower than a wide-column solution on the same hardware, but this becomes less of an issue as hardware gets cheaper and database technologies become more advanced.
- The cellular data structure presents new challenges and requires new skills from in-house IT developers as well as support in the form of expert advice and training from external expert data warehouse consultants.

A number of economy-wide administrative data series and surveys with simpler data structures have been migrated to the new platform. It now contains GST data and detailed financial data covering every business in New Zealand for several years, plus import and export transactions in detail for the whole economy dating back ten years, as well as quarterly surveys of manufacturers and wholesalers. The current holding is around four billion data points. Under development are the more complex collections, including the economy-wide annual enterprise survey, the annual agricultural production survey, and the business register. These latter systems demand more advanced statistical methods and require more complex integration of survey and administrative data.

Another key development currently underway is the creation of a prototype Integrated Data Infrastructure (IDI), an integrated dataset with person-level data from multiple sources. The IDI contains longitudinal microdata about individuals, households and firms.

The IDI is not a population register, but is able to be used to answer research questions about particular populations that are well represented in the linked dataset, for example students. By bringing together many datasets into one infrastructure, the IDI allows for longitudinal analysis across education, employment, migration, welfare, and business, with potential for further datasets to be added in future. This will help Statistics NZ to meet demand for more or new information, and allow more efficient responses to changes in existing data sources.

The prototype IDI is used as a research database by both government and academic researchers. Statistics NZ's standard security measures and protocols govern the management of the data used in the IDI. Statistics NZ is required to comply with the confidentiality provisions of the Statistics Act 1975 and also protocols for security

in the government sector ([Statistics NZ 2012](#)). Other security measures for the IDI include the following arrangements:

- All data collections and associated electronic workspaces will be secured (access will only be authorised for project personnel who need to access data for specific tasks, and to selected IT administrators who are required to maintain the IT system).
- Datasets will not be available to third parties.
- Regular audits of individuals able to access the dataset will be made.

Statistics NZ has developed microdata access policies and confidentiality standards for the IDI that apply to potential research outputs to protect the confidentiality of information providers. These rules also take into consideration the existing confidentiality rules for each contributing dataset ([Statistics NZ 2012](#)).

3. Statistical Development

Creating statistics is an intellectual, service-based activity, akin to practising law or engineering. Just as an engineer's bridge design is not the product of a computer-aided drawing system, a statistician's time series is not a product of the computer that calculated it. Productivity in our industry is a function of the efficiency and ability of the technical statistical staff involved. Systems can complement, but are not a substitute for, this intellectual and human productivity.

With the goal of making the subject-matter specialist role easier and more efficient, the following principles guided the design of the new economic platform:

- Subject-matter specialists must have direct control of all business processes and be able to make changes easily. This is explored in more depth in Subsection 3.1 below.
- Systems must support the experimentation and creativity of subject-matter specialists.
- Repetitive tasks should be automated.
- All data must be stored so that it can be accessed and integrated easily.
- Whenever business processes and systems interfere with analyst productivity, they should be reviewed and considered for change. For example, most existing processing systems could not produce graphs that meet Statistics NZ publication standards, so data was sent to a specialised publication team to create the required graphs. This process typically took about three working days. The solution is a processing platform that can produce graphs that meet publication standards as part of the usual quality monitoring.
- The platform must support continuous improvement and allow for timely monitoring of the quality of processes and statistical outputs.

The microeconomic platform already incorporates a module for calculating sample errors. If a new measure is developed for any other component of total survey error, a generic module to calculate and monitor it can be easily added to the platform. These measurable components of total survey error can be displayed along with other quality indicators in a dashboard, which is available for methodologists when optimising a collection design and

for subject-matter specialists monitoring the ongoing performance of the statistical production processes.

Data processing error is a significant component of total survey error in the quality models described by Groves and Lyberg and by Biemer (Groves and Lyberg 2010; Biemer 2010). It includes errors introduced at all stages of processing, for example during data editing and coding or during computation of weights and estimates. The versioning of data maintained by the platform allows the pre- and post-processing versions of a dataset to be compared and analysed to assess the significance of processing error.

3.1. *The Design and Build Bottleneck*

The generic Business Process Model (gBPM) was originally developed in New Zealand (Savage 2008) and is reflected in the Generic Statistical Business Process Model (Metis Steering Group 2012). It has, with subsequent modifications, been widely adopted by other statistics organisations.

The gBPM includes designing and building statistical outputs as a routine part of business activity. However, in practice, under the old approach a complex one-off development process and separate project was required to make substantive process changes in a given system.

“Design” and “build” have proved to be the primary performance weaknesses in the statistical supply chain. It can take years in a national statistics office to move from identification of need to the production of a new output. Large multidisciplinary development teams lead to high costs, a lengthy justification and approval process, and long development times. The cost and long lead time makes the organisation unresponsive to changing needs. Some needs are never met as they fall below cost/benefit thresholds. We call this problem the “Design and Build Bottleneck.”

To resolve this problem, the new platform has adopted an “analyst-driven” approach. In this analyst-driven model, creating new or improved statistical output becomes a routine activity for subject-matter specialists and methodologists. The role of the IT department is limited to building a generic platform that can be used by analysts for designing and building statistics as well as collecting, processing, analysing, and disseminating statistics. Incorporating analyst-driven design, build, and test functionality into the microeconomic platform allows new statistical products to be developed more quickly and at significantly lower cost.

In this model, subject-matter specialists and methodologists control all the business processes and methods. Examples include microdata editing, outlier detection, imputation, adjustment of weights, estimation of model parameters, and computation of standard errors. Common methods such as historic imputation are encapsulated in standard tools (in this case Banff, a data editing and imputation system developed by Statistics Canada (Statistics Canada 2005)), which are then combined into an end-to-end production process called a configuration. A configuration, or a group of processes that will be applied to a



Fig. 1. *The generic Business Process Model*

particular set of data, can be changed by the subject-matter specialist or methodologist in two main ways: firstly, the parameter settings that control a process in the configuration can be altered to improve the performance of the process, for example changing the minimum cell size criteria in an imputation cell; secondly, a process can be added or removed from the configuration, for example changing an imputation method from mean imputation to historic imputation.

This model offers subject-matter specialists and methodologists a great deal of flexibility when it comes to the design of collections and processes. This flexibility significantly affects the design of products, as it is no longer necessary to get the design absolutely right prior to the development phase. This is because it is easy to optimise the design once a configuration has been created. This flexibility also affects the production process itself. The ease with which configurations can be altered means that analysts are able to alter or replace processes as needed, with the constraint that there is a limited range of standard methods available for each process (e.g., analysts can choose between the imputation methods available in Banff). It also means that strict controls are in place to differentiate between configurations used in production and those under development. Only a very few staff are given the authority to make changes to configurations that are used in producing published products, and clear guidelines and procedures are in place to control changes. Also, for a tool to become available on the platform, it must first be validated as a standard tool by methodologists and be available in the statistical toolbox.

As outlined in Subsection 2.4.2 above, flexible and consistent data storage are critical as they give subject-matter specialists direct control over both existing and new data. Below we describe four features of the microeconomic platform that support subject-matter specialists to maintain the relevance of existing outputs and to design and build new outputs quickly and easily.

3.1.1. Business Register History

Statistics NZ maintains a comprehensive tax-based business register that is used for selecting samples and integrating and classifying administrative data. The business register records industry and institutional sector classifications for every economically significant unit in the economy. Classifications on the business register are derived from multiple sources: from administrative registers for simple units and via register maintenance surveys for larger and/or more complex units. Inconsistencies between sources are resolved by applying business rules during the register update process.

Before administrative data is used in the production of statistics, each administrative unit on the dataset is linked to the business register, so the classifications from the business register can be assigned to it. This ensures that all economic data is classified consistently by eliminating the need to rely on the classifications that may have been received from the administrative system.

The microeconomic platform stores point-in-time snapshots of the business register. Samples are selected from these snapshots and coordinated so that, for example, quarterly surveys are selected at the same time on the same basis. Immediate access to complete business register information facilitates the design and development of new statistical collections and outputs.

3.1.2. Configuration Store

The configuration store is an important innovation, because it makes the ETL workflows described in Subsection 2.4.2 user configurable. The configuration for a statistical output lists all the workflows and the parameter settings that must be applied during the production of a statistical output. It records all the instructions and information needed for statistical processing of a survey from collection to dissemination.

An individual workflow points to where data is stored and specifies a method or statistical process that must be applied to it. A simple workflow may consist of a datastep that derives a new variable. It specifies the data to be extracted, the code that is to be executed on the variables, and the new variables to be created. A more complicated workflow might extract a large dataset and pass it through a complex statistical tool like Banff or QualityStage. The workflow specifies all the parameters and rulesets required by the statistical tool.

A subject-matter specialist or methodologist builds a new statistical output by creating a new configuration in the configuration store of the microeconomic platform. This configuration will specify all the transformations and statistical methods that must be applied to the data to produce the output and the order in which they must be applied. The version-controlled configuration store ensures permanent traceability between raw data, the data process rules, the methods are applied, and the final outcome.

The components of a configuration can be copied easily and reconfigured for a different need. This means that processes, methods, queries, and reports that have been developed for one collection can be easily applied to another. For example, a generic editing and imputation process is available for any survey or administrative dataset by changing the instructions and parameters recorded in the configuration store. Standardisation is enforced, because new configurations are required to draw on existing standard configurations composed of processes based on standard tools.

The configuration store allows subject-matter specialists and methodologists with a basic understanding of computer logic to set up all the statistical processes needed to produce new statistical outputs with minimal IT support. The configuration for an output breaks the end-to-end process into a large number of small steps, so complex data transformation and statistical processing is achieved by combining well-tested data processes and statistical tools. An advantage of this approach is that the production system is created as the new output is being developed, assessed, and tested.

For example, the configuration store contains a generic automated edit and imputation process that can apply a wide range of edit types and all standard imputation methods to any survey or administrative dataset. To apply this process to a new dataset, the subject-matter specialist or methodologist amends the configuration to record the methods, parameters, and other information that is needed to apply the required edit and imputation process to the new data.

New methods can be developed, tested, and implemented within the existing processing system. They can be introduced to other outputs by changing the relevant configurations.

The configurations make extensive use of SAS for three reasons:

- SAS is a high performance statistical tool for data processing and statistical analysis that manages large-scale data manipulations with ease.

- A library of standardised and tested SAS-based statistical methods is already available within the organisation.
- Many subject-matter specialists are trained in SAS.

The disadvantage of SAS is that subject-matter specialists and methodologists can rewrite complex code again and again in an unstructured way, defying standardisation and efficiency. They can create new SAS datasets easily, making security and version control extremely difficult.

The microeconomic platform controls and limits the use of SAS by using the configuration store as a facility for writing and managing SAS in a structured and standardised way. The benefits of SAS described above are retained, but the pitfalls are avoided because the platform controls the way SAS is executed and securely manages the resulting data.

Although the new microeconomic platform uses SAS, it is not tied to this particular software. When using SAS, the workflow sends the code stored in the configuration store to a SAS server for processing. Other calculation engines, such as R and Excel, can also be used in the same way.

3.1.3. Scenario Store

The scenario store is a key element in an analyst-driven statistical system. Whenever an automated process to change data is applied, the entire dataset is extracted to a temporary area, called the scenario store. If the analyst selects the current approved configuration ruleset for the scenario (i.e., the version used in producing outputs) then, after confirming that results are as expected, the data can be written back into the main permanent store and from there be used in outputs. However, the subject-matter specialist can copy an approved ruleset, make changes and create an experimental scenario at any time. This experimental scenario can be used for evaluation. The approved and experimental scenarios sit side by side in the store for easy comparison using standardised views and tools.

This means that as well as changing the process rules, analysts can easily test those changes with live production data without affecting production systems in any way. In this way, existing rulesets can be continuously improved or completely replaced. The scenario area can be used to prototype entirely new outputs or to make small changes in existing rules. The rights of subject-matter specialists to make changes are managed by security. There is version control of the configuration rulesets and the scenario store. A change management process ensures that all changes are properly tested and approved before they are implemented in published statistics.

3.1.4. Reduction in Staged Processing

Reviews of existing systems to inform the new design showed that data processing was often broken down into many stages. Each stage had its own data preparation followed by manual inspection. Sometimes data is repaired manually when, if left alone, it would have been corrected by automated repair processes at a later stage. Processing analysts often had to carry out the same manual repair multiple times if data was reprocessed. If the output result was several stages away, more serious quality issues could easily be missed. Later detection meant more rework. In summary:

- Each process operated independently of the others, often in different parts of the organisation. Problems and errors arose from the transfer of data between environments.
- Inspection and manual repair was inefficient and duplicated.
- The contribution that each process made to the whole was not always clear to those responsible for designing and monitoring it. It is often not clear whether they are adding value.
- The effect of actions cannot be seen until the end of the last process is complete and an output is produced. This may be days or weeks after the first process began.
- When errors are detected in later stages, the process is very slow to repeat, or in the worst cases cannot be repeated in the time available before output deadlines.

These problems have been resolved by building larger, more holistic processes that can be run repeatedly until quality targets are met.

In this iterative approach, each time the subject-matter specialist runs the configuration that controls the statistical production process, the complete set of workflows is run over the data and the results are available for inspection via a standardised reporting suite. Over time, as more data is received in sample surveys, or rules are adjusted in administrative collections, scenarios move closer and closer to the final outcome. Approximate output results are always available. This methodology is born of necessity as the organisation moves into processing more large-scale administrative data where manual intervention and manipulation is impractical. In summary:

- Data is managed in a single very large cohesive environment.
- Processing is, as much as is possible, done in a single pass and in a single scenario set.
- Manual editing is post-process and driven by significance and by output needs.
- Processing and revision of data is easily repeatable.
- The configuration is designed and built by subject-matter specialists as a cohesive single sequential process. The configuration specifications are improved over time as they are fine-tuned by real world data experience.

Strategically, the microeconomic platform makes product quality highly visible. It also allows changes to maintain quality, for example updating an industry classification to the latest version, to be carried out quickly and cheaply. Operationally, product quality is controlled through security settings that allow only a very few people with appropriate authority to make changes to configurations used to produce outputs. Change control policies and procedures ensure proper testing and review has accompanied proposed changes.

The microeconomic platform is making statistical development a more standardised, more cost- and time-effective process. These attributes are essential for achieving the goals of the Stats 2020 programme of change.

4. Statistical Production Process

The development of the new platform has had a significant impact on the statistical production process, particularly editing and imputation. Modern designs typically include

administrative data in some way, often in large quantities, and use standard concepts, methods, and tools.

A major part of the regular government audits of Stats 2020 has to do with measuring improvements. The most significant savings will be made through operating efficiencies in relation to the new statistical production platforms that have started to replace the outdated and dissimilar existing systems. The four perspectives referred to in Section 1 are the strategic goals of the organisation for the next seven years (transforming how statistics are delivered, leading the New Zealand OSS, obtaining more value from official statistics, and creating a responsive and sustainable organisation). The programme is planned to reduce labour input by just over 20% in this time frame. Each of the Stats 2020 projects has short-term performance targets, and a time frame in which expected benefits should be achieved. Early indications for the microeconomic platform are ahead of this target in both labour savings and the time taken to achieve them.

For example, a study by the methodology group found that the number of units needing manual edit intervention decreased by 66% in repeating business surveys. This was based on the use of the standard significance measure deployed in the new system. The business area has confirmed this saving in practice with operational changes. Preliminary results for another survey, using Banff to do automated repairs, combined with the impact assessment above should see realised labour savings of 20% or more.

Development of the microeconomic platform has not involved a substantial upfront investment. It has however been completed over quite a long period of time, taking four years to deliver a full suite of business processes. Delivery was deliberately planned in phases, with each phase of the programme being self-justifying in terms of costs and benefits.

The first stage focused on a large administrative data series that was not an output in its own right, but is used to replace small units in a major postal survey. This made the initial phase and goal much simpler. A target of reducing postal units in the survey by around 60% over a five-year period was achieved by using more of the administrative data, developing a more efficient data handling capacity and automated editing and imputation. At the close of the first phase, a cashflow analysis (project cost input, reduced postal costs output) showed an investment return on capital of 18%. In addition, the project had created generic data handling and edit and imputation processes that could be used for other collections.

The second phase of work added the processing of GST (or value-added tax) unit records to the platform. This phase enhanced the ability of the system to handle longitudinal data. It also increased the confidence of the organisation in the development and led to the organisation committing to further investment. These very large administrative datasets are routinely transferred electronically from collecting agencies to Statistics NZ through a standardised process. The wide variety of types of administrative data and of statistical issues encountered, such as common mistakes that occur when filling out a form, encouraged a pragmatic approach. Editing strategies aimed to resolve the most common errors in form filling, such as ignoring a request to use thousands of dollars when answering financial questions. Errors that are less common, but are material, are also resolved. For the sake of platform performance, errors that are less common and immaterial are ignored. The way statistical issues such as definitional differences and temporal or cross-sectional aggregation effects are resolved depends on their importance

with respect to the statistical product being produced using techniques such as statistical modelling.

The third phase of work brought the first postal business surveys onto the platform and the very large Overseas Trade series. During this phase, the investment costs for development and implementation rose dramatically, as sample selection and survey management functionality was added to the platform. However, Statistics NZ was willing to commit to this investment, because the core elements of the platform were already in place and proven. Efficiencies were able to be achieved immediately for successive products, because the generic editing and imputation tools were already available. The platform is able to cope with a wide variety of data formats, data quality, and processing methods. In tandem with the platform development, the organisation is focussed on developing a proven set of methods and tools for evaluating, collecting, processing, analysing, and disseminating products based either entirely on administrative data or on a combination of data from multiple sources. Techniques such as statistical modelling and data integration are especially important for making extensive use of administrative records for statistical work.

The fourth phase of the project will see 19 statistical outputs moved to the platform in the next two years. For some of these outputs, new functionality will have to be developed. For example, the Business Register has been incorporated onto the platform, which has created a need for new functionality to enable handling of units with data that does not change every month, but does have complex ownership relationships that need to be maintained in real time without creating performance issues. The investment for development and implementation in this phase is greater as the platform is being scaled up from a few users to hundreds of users. This requires intensive investment in reliability and robustness of performance, with separate teams being established to support survey migration activity and provide support for the new standardised platform. One result of this is a change in focus from teams supporting individual surveys to teams supporting clusters of surveys using common processing systems and common data storage.

The project to develop a microeconomic platform was not carried out in a way that required a large investment upfront; rather, it was established and proven in stages over a long period. A clear overarching vision guided the development but clear targets were set for results in each phase. In the past, development would have been preceded by carefully focused and relevant analyses by methodology teams before a start was made on the IT requirements; these recent developments were approached in a different way that allowed for a much more pragmatic decision-making process. Standard tools were predominantly off-the-shelf rather than built in-house, so development work was more in the nature of adding tools to the platform and optimising tools for particular collections rather than starting from scratch and redesigning entire collections. All collections migrating to the new platforms are expected to adopt the available standard tools and there are very few exceptions. Teams are multidisciplinary and include methodologists, subject-matter specialists, and IT developers. IT business analysts provide a link between methodology and technology. They play a key role in ensuring that all business requirements are articulated and that the implications of requirements in terms of statistical and business architecture are understood.

Changes to collections on the new platforms are carried out incrementally in small steps rather than via a major effort in system development and implementation. Statistical redevelopment is planned and implemented in a programme of work called the Statistical Infrastructure programme, a ten-year rolling programme of statistical redevelopment to maintain outputs at the desired quality where quality is considered in all its dimensions. Examples include updating the industry codes in business surveys when the business activity classification has been updated, changes in methods, and responding to real world events such as the recent change in New Zealand's rate of GST. A key goal for this project is that all investment decisions around statistical development priorities take an organisation-wide view. Until recently, redevelopment decisions were made independently of quality and priority frameworks. A key difference between development in the past and development on the new platforms is that in the past a change would need to be applied in each affected collection separately, typically sequentially because of dependencies between collections. Now many changes to collections can be applied to all collections on the platform simultaneously.

A key to minimising production costs is the statistical toolbox, which includes standard tools for common business processes, many sourced from other national statistics offices. For example, editing and imputation uses the Banff and Canceis programs developed at Statistics Canada (Statistics Canada 2005; Bankier et al. 2001). X-12-Arima is used for seasonal adjustment, and QualityStage™ is used for data integration. In this way, the organisation promotes the use of standard tools for common processes including detecting outliers, deterministic imputation, historic imputation, donor imputation (and many other standard imputation models), seasonal adjustment, and data integration.

Tools in the toolbox are categorised according to whether they are current, emerging, legacy, or graveyard tools. Emerging tools are under evaluation for inclusion in the toolbox; current tools are those actively promoted for use within the organisation and supported by the organisation; legacy tools have been identified as needing to be replaced with more up-to-date versions; and graveyard tools are being phased out. The corporately supported tools are not always the most efficient, but they usually represent the standard methodological approach for common problems. Exceptions to the use of current tools are only permitted by agreement at the highest levels of the organisation. The tools are accessible from the platforms as a tools library that allows analysts to choose a tool and use it in the production process the analysts are creating.

Standard tools have been made available on the production platforms as the need arises – not all tools are currently available for use in the new production environments. Methodologists are still exploring some tools; for example, Statistics NZ has been evaluating the selective editing tool SELEKT developed by Statistics Sweden. However, it isn't clear yet how SELEKT could be used in a fully automated processing system. If the goal is to score records based on the significance of any change, and then review the records with the highest scores, then a simple scoring mechanism may suffice. In a number of cases, tools have been modified to be of more use. For example, a version of Banff's Proc Outlier that includes significance criteria for identifying outliers has been developed, because of the prevalence in business tax data of units that are extreme but also extremely small.

To achieve the organisation's efficiency goals, production processes for designs using administrative data are predominantly automated, as the volume of data precludes the

possibility of manual data editing. Unlike data the organisation collects itself, contacting the data supplier to clarify data anomalies is not possible with administrative data. Full end-to-end automation of business processes reduces manual editing of records, improves data quality, and substantially reduces the time and staff required to produce statistics.

Adopting an analyst-driven approach has changed the underlying architecture of the microeconomic platform. In the role-based approach, handovers of data and manual intervention are minimised. Reliable large-scale processes automatically process data that is stored in a central datastore. The goal is to make it easier for analysts to build and test new configurations or make changes to existing production systems. The emphasis is on building a highly flexible, rapidly repeatable process. Rule changes and manual repairs are implemented at the end of each cycle and the whole process is repeated until the desired outcome is achieved.

Examples of cost savings and improvements in quality include:

- The three administrative data collections and two business surveys currently using the microeconomic platform are all using standard tools for each of their business processes.
- Moving to the new platform has seen either a reduction in the size of each survey or administrative data collection's processing team; or a reduction in processing time; or both.
- Analysts using the platform are able to move easily between the outputs because of the common look and feel of production systems and the use of common tools.
- Complete audit trails are available of the changes to each record and this information has allowed managers to easily review each output's business processes and make improvements in time and cost of production. For one of the first four outputs introduced to the new platform, the publication has been brought forward.
- Statistical development for those outputs already on the platform is able to be carried out without IT intervention in most cases, thereby reducing the cost of the work significantly. The New Zealand Harmonised System Classification 2007 (NZHSC2007), used in overseas merchandise trade statistics, was able to be updated to NZHS2012 by a single subject-matter specialist. Data processing and checking is a single continuous process, and all analysts access the same dataset. This has eliminated the problem of different analysts holding and using different versions of the same dataset.

The platform has a high degree of flexibility in accommodating new methodologies. The platform is able to load and process most common types of survey and administrative data. Changes to editing and imputation strategies, such as adding new edits in response to changes in the underlying data and using alternative imputation methods, are simple to test and deploy. Standard statistical tools are available for most common business processes.

5. Governance

Significant governance is key to ensuring that generic solutions are delivered and that Statistics NZ meets its organisational efficiency goals. Stats 2020 is a large and costly programme of work, depending upon the success of over 100 multi-year projects. The New Zealand government has made a substantial investment in the first four years of the

programme, with additional funding of approximately NZ\$60 million for operating costs over that time, primarily funding common statistical infrastructure and common platforms, methods, processes, and tools (Statistics NZ 2010). For that reason, there is more oversight of the organisation by the government and a higher level of reporting requirements than ever before.

The governance framework for the oversight and assurance of the Stats 2020 programme consists of a set of functional responsibilities that have been developed for project management, governance, and project assurance. The Stats 2020 governance framework is additional to the existing organisational governance model and is intended specifically to support the delivery of the Stats 2020 programme. Such a formal approach to monitoring the programme's progress is necessary to ensure that the organisation is well placed to report on progress to the government.

Governance bodies called portfolio committees have been established for clusters of dependent and/or related projects. The portfolio committees monitor benefit realisation, for instance that all projects in the portfolio are making their expected contributions to the strategic priorities.

Each project has a project owner and a project manager who are focused on project outcomes and project outputs respectively. Steering committees focus on the project scope, financial risks, and roadblocks; they provide operational direction for the projects and have a quality assurance role. Steering committees have the authority to make decisions on specific project components, and enforce the organisation's standards for quality and risk management.

Decisions that result in changes to a project's scope, timescales, or budget that are outside agreed project tolerances are raised to the appropriate portfolio committee for approval. The portfolio committee structure is different from the usual line management; decision making is therefore independent of the usual line management considerations and takes an OSS perspective.

Personnel are managed using a matrix-management model: line managers are responsible for the day-to-day performance and development of their staff, and project managers are responsible for the project management aspects of staff assigned to projects. All projects are prioritised annually and decisions about resource allocation take into account the priority of each project, always aiming to keep projects that have a higher priority fully resourced.

A Transformation Office provides programme assurance. The Transformation team monitors the programme from a benefits perspective and provides independent strategic oversight and advice to the Government Statistician. Key members of this team are portfolio programme managers whose role is to provide strategic oversight across portfolios, advise on progress, and identify opportunities to leverage change. In addition, a project management team has been established to provide internal project management assurance. The team provides training and support on project management and the organisation has invested significantly in project management training for all project leaders.

This governance model is supported by:

- a Standards Review Board, which oversees the corporate standards and prescribes their use.

- a statistical toolbox, which contains the current set of standard statistical tools that are to be used in systems developments.
- steering committees that provide operational direction and quality assurance to the clusters of projects.
- training for staff in how to assess the quality of administrative data, process administrative data, and develop statistical products based either entirely on administrative data or on a combination of data from different sources, and
- training for staff in how to use the standard tools and design statistical production systems (e.g., common editing and imputation strategies for survey data and administrative data, typical order of business processes, and monitoring process quality).

Performance indicators associated with each of the Stats 2020 projects show baseline and target information and an indication of the time frame in which benefits are expected to be achieved. An example of a performance target might be lower end-to-end production costs. When performance indicators measure end-to-end production costs of products, the baseline is the direct operating cost of producing each product. Direct operating costs are available from the finance system, which is detailed enough to allow direct operating costs, such as time spent preparing for a release, to be separated from costs associated with development work, since these activities are “coded” differently in the time-keeping system.

6. Lessons Learnt

The move from stand-alone production systems to a model where generic components are developed and delivered has been difficult. In the mid- to late part of the last decade, the organisation tried to implement a large-scale project known as the Business model Transformation Strategy (BmTS). It aimed to create an organisation-wide framework of standard processes and systems for statistical production. The BmTS was designed to achieve this through a programme of projects – each project delivering different components of a complete end-to-end processing system (Savage 2008). For example, one project would deliver an input data environment for physical data storage; another project would deliver an output data environment for data dissemination.

The Legolution system is one example of a processing system developed using BmTS principles (Savage 2008). It was developed in 2005 to process a suite of surveys measuring business performance (e.g., the Research and Development Survey). Legolution provided generic data editing and imputation, using Banff, a data editing and imputation system developed by Statistics Canada (Statistics Canada 2005).

Many of the principles that have been adopted for the microeconomic statistics platform had been tried out in the BmTS-based systems. For example, every output processed in the Legolution system had a documented data editing and imputation plan, which included measures of processing quality to enable monitoring of the system and allow continuous improvement. Unfortunately, the project had to be de-scoped during the initial development because the work involved in creating a generic solution applicable to more than one survey had been underestimated and, as a result, the solution was not as generic as first specified.

The BmTS project struggled to deliver some key elements within the time, scope, and budget constraints imposed by user needs. Ultimately the programme was cancelled (Savage 2008). The short time frame for delivery was at odds with the large scale of the components being developed. Also, building generic components in separate projects resulted in incompatible solutions. The generic lost out to the specific, causing affected projects to reduce their scope and diverge in order to meet the immediate business goals of each project. At the time, it seemed that the drive for standardisation was lost. However, the ideas and lessons learnt had been carried forward:

- The BmTS goals of standardisation and an integrated data environment are kept in mind but reprioritised to make business delivery the focus and generic features more staged.
- Development teams adopt agile software development techniques (e.g., Highsmith 2004). The system is developed and delivered incrementally so that analysts are able to conduct tests very early on and their feedback is used to improve the product. Prototypes are developed using real data to ensure the final product is fit for purpose.
- Subject-matter specialists and methodologists adopt a more hands-on and involved approach, becoming part of the development team and involved in all aspects of testing and delivery. The team responds to change rather than working to plan. The incremental approach allows for changing user needs while minimising costs.

Now, seven years on, the goals of the BmTS are beginning to be realised through the microeconomic statistics platform.

Providing subject-matter specialists with generic highly automated production systems is not enough to change an organisation's culture. Traditionally, Statistics NZ has had what is known as a *process* culture: there was little or no feedback and people focused on doing things right rather than on doing the right things. The organisation was a typical hierarchical government department where predictable and consistent systems and procedures were highly valued and people had delegated authorities within a highly defined structure.

In this process culture, decisions about statistical maintenance of outputs were made independently of prioritisation and quality frameworks. The decisions were driven by the need to replace aging production systems. Because they were reactive rather than planned for, IT systems developments were exceedingly expensive and, because they happened at widely spaced intervals, resulted in production silos, unique production systems for each survey run by Statistics NZ where any further changes to systems (e.g., updating classifications) could only be made by IT teams. Different production systems for each survey meant it was difficult for people to move between surveys without a significant amount of training and analysts became experts in processing data and troubleshooting IT systems rather than in data analysis.

The new platforms allow statistical maintenance to be carried out in small regular increments and use standard tools for common processes. The advantage of the change to a cluster approach (i.e., groups of surveys and administrative data collections processed on the same IT platform using the same tools for editing, imputation, confidentiality, data dissemination, and analysis) is that analysts have more time to concentrate on data analysis. When statistical maintenance is required, the new platform is highly configurable

so changes are easy to implement and only minimal IT intervention is needed. For example, a new classification can be introduced to the platform by a subject-matter specialist. All they need to do is to change the metadata and business rules in the configuration store. The impact of the classification changes can be tested in the scenario area without any risk to existing statistical outputs.

It has not been easy to move from a process culture to a more constructive culture – where quality is valued over quantity, creativity is valued over conformity, cooperation is believed to lead to better results than competition, and effectiveness is judged at the system level rather than the component level (Cooke 1987). Statistics NZ is still in the early stages of achieving this transformation. The organisation has developed a strategy to help teams understand and explore what the new environment means for them. The people strategy has been broken down into two areas of work:

- a workforce programme, which focuses on building on people's skills and capabilities, and makes sure systems are in place to attract, develop, and appropriately reward people, and
- a culture and change leadership programme, which focuses on supporting people through change, developing strong leaders, and building a performance culture.

A key enabler of the culture change to date has been the adoption of an agile project management approach for IT development projects. The agile approach involves setting up multidisciplinary teams, planning the team's work in manageable pieces that take a short period of time to complete (e.g., two weeks), and then checking in with the team regularly to monitor progress and remove impediments. Progress is very fast under this approach and impediments to progress are identified and resolved quickly. While developed primarily as an iterative method of determining requirements in engineering development projects, it can be used in any team situation and has had the added benefit at Statistics NZ of blurring role boundaries, encouraging all team members to help solve any problem. Team capability develops very quickly and the team environment becomes more conducive to the creation of innovative solutions.

7. Conclusion

Statistics NZ aims to produce statistics that are fit for purpose in a cost-efficient and sustainable way. With the Stats 2020 programme of work, the organisation is developing and implementing business architecture that will support the organisation's goals. Statistics NZ is sharing the knowledge and experience gained with other national statistics offices both informally and formally, through the international collaboration network instigated by the Australian Bureau of Statistics in 2010.

The move from stand-alone systems to large-scale IT platforms is underway. The large-scale platforms will support the processing and analysis of multiple outputs and are based around standard business and information concepts encapsulated in standard methods and systems. Opportunities identified during the move include providing systems built in such a way that subject-matter specialists can easily create new or modified outputs by applying standard statistical tools and processes, or combining data from different sources, without the help of IT teams. The use of standard tools and processes ensures that production

systems are automated as fully as possible. Governance structures, a people strategy, and a change leadership programme focused on moving towards a more constructive culture and supporting people through change are in place to support it.

8. References

- Bankier, M., Poirier, P., and Lachance, N. (2001). Efficient Methodology Within the Canadian Census Edit and Imputation System (CANCEIS). Proceedings of the Annual Meeting of the American Statistical Association, August 2001. Available at: <http://www.amstat.org/sections/srms/proceedings/y2001/Proceed/00202.pdf> (accessed December 2012).
- Biemer, P.P. (2010). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly*, 74, 817–848.
- Bycroft, C. (2011). Social and Population Statistics Architecture for New Zealand. Available at: <http://www.stats.govt.nz/~media/Statistics/surveys-and-methods/methods/research-papers/topss/Social%20and%20population%20statistics%20architecture-jul11.pdf> (accessed December 2012).
- Cooke, R.A. (1987). *The Organizational Culture Inventory*. Plymouth, MI: Human Synergistics, Inc.
- Glaude, M. (2008). A Strategy for the Development of Social Statistics in the European Statistical System. Eurostat. Available at: http://unstats.un.org/unsd/statcom/statcom_08_events/special%20events/New_directions_social/Michel_Glaude_Paper.pdf (accessed December 2012).
- Groves, R.M. and Lyberg, L. (2010). Total Survey Error: Past, Present and Future. *Public Opinion Quarterly*, 74, 849–879.
- Highsmith, J.A. (2004). *Agile Project Management: Creating Innovative Products*. Boston, MA: Pearson Education.
- Kent, J. (2011). Availability of Data from Scarcity to Profusion. Paper presented at UNECE Conference of European Statisticians, Geneva. Available at: <http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2011/wp.33.e.pdf> (accessed December 2012).
- McKenzie, R. (2008). Statistical Architecture for a New Century. Paper Presented at International Seminar on the Use of Administrative Data in Daejeon, Korea. Available at: <https://www1.oecd.org/std/41142889.pdf> (accessed December 2012).
- Metis Steering Group (2012). The Generic Statistical Business Process Model. Available at: <http://www1.unece.org/stat/platform/download/attachments/8683538/GSBPM+Final.pdf> (accessed December 2012).
- Savage T. (2008). Statistics New Zealand's Move to Process-oriented Statistics Production: Progress, Lessons and the Way Forward. Paper presented at the IAOS Conference 2008: Smart Data, Innovative Uses – Reshaping Official Statistics in Shanghai, China, October 2008. Available at: http://isi.cbs.nl/iaos/conferences/Papers/CS_19_2_Savage_Gretton.doc (accessed December 2012).
- State Services Commission (2011). Performance Improvement Framework: Formal Review of Statistics NZ. Available at: <http://www.ssc.govt.nz/sites/all/files/pif-statisticsnz-review-oct11.pdf> (accessed December 2012).

Statistics Canada (2005). Functional Description of the Banff System for Edit and Imputation V1.04. Ontario: Statistics Canada.

Statistics NZ (2010). Statistics New Zealand Strategic Plan 2010–20. Available at: <http://www.stats.govt.nz/~ /media/Statistics/about-us/corporate-publications/strategic-plan/snz-strategic-plan-2010-20-new.pdf> (accessed December 2012).

Statistics NZ (2012). Integrated Data Infrastructure and Prototype. Available at: http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/integrated-data-infrastructure-and-prototype.aspx (accessed December 2012).

Thomsen, E. (2002). OLAP Solutions. Building Multidimensional Information Systems. New York: Wiley.

Received December 2011

Revised July 2012

Accepted November 2012