1. Introduction

In 2011 Statistics New Zealand initiated a project to replace the Business Frame which had been the statistical business register since 1995, but had legacy software components that were going to be unsupported in the next few years. This was one project within a 10 year programme of work called Statistics 2020 Te Kāpehu Whetū: Achieving the statistics system of the future.

Rather than build a standalone system as previous registers had been, it was decided to develop the new Business Register on the Micro-Economic Platform, also being developed as part of Statistics 2020.

A key objective of Statistics 2020 is to make maximum use of administrative data while continuing to produce economic statistics that are fit for use, with an aim for administrative data to be the first source of data for statistical outputs. This key principle of the new statistical architecture has been a driver for developing the new economic statistics platform.

The platform introduced two major innovations that are critical to the implementation of the new statistical architecture - it is designed for analysis, and has flexible data storage.

The decision to develop the Business Register on the platform meant that there had to be a different approach how some functions were implemented, but this was more than offset by the development of extra functionality to support the Business Register, which was then made available to economic collections using the platform.

The Business Register went live on the new platform on 5 May this year, realising the vision of seamless integration between economic data collected from administrative sources, direct surveying, and a statistical business register holding the classifications and size measures used to transform that data into coherent and relevant statistics.

In this paper we will describe the development of the Business Register, with particular attention to the aspects that relate to building it on the Micro-Economic Platform, and the benefits we expect to gain from having a single platform supporting our economic collections.
2. Statistical concepts used on the Business Register

2.1 Statistical units
The Business Register uses the Enterprise, Kind of Activity, Geographic unit statistical unit model. Enterprise groups and group top enterprises can also be derived from the ownership links held between Enterprise units.

Enterprise (ENT)
An Enterprise is an institutional unit (i.e. a legal entity or equivalent) or the smallest combination of institutional units that includes, and directly or indirectly controls, all necessary functions to carry out its production activities. In practice, it corresponds to legal entities such as companies, partnerships, trusts, estates, incorporated societies and sole proprietorships.

Geographic unit (GEO)
The Geographic statistical unit is an ENT or part of an ENT and is normally an unbroken physical area/site on or from which one or predominantly one kind of economic activity takes place on a permanent basis. Employment and location data must be able to be produced for a Geographic unit. Geographic classifications are applied at this level.

Kind of activity unit (KAU)
The kind of activity unit is an institutional unit, or part of an institutional unit, which engages in one, or predominantly one, kind of economic activity, without being restricted to a geographic area. Value added statistics must be able to be produced for a KAU, or be able to be readily or meaningfully imputed.

The Enterprise unit is the core of the Business Register. The Enterprise unit coverage of the Business Register is underpinned by the comprehensive administrative tax data used to maintain the Business Register. The Business Register has a robust coverage of enterprise units.

Enterprise births are automatically assigned a single KAU and single GEO. Approximately 99% of enterprise units have a single Geographic unit. There are less than 200 multi KAU structures on the Business Register.

2.2 Identifiers
All Enterprises are linked to a New Zealand tax number. This number is used when businesses and individuals are assessed for tax (both personal income, company tax and the Goods and Services Tax (GST)) and is the primary link to other administrative data used by Statistics New Zealand, e.g. employment data, Companies Office ownership and shareholding information.

The New Zealand government is in the process of implementing a new single business number, the New Zealand Business Number (NZBN), that will eventually be used by all government agencies and will apply to all business types (has currently only been rolled out to limited liability companies).

Statistics New Zealand applies additional numbers at ENT, KAU and GEO level for use within our systems and supplied to survey respondents.

2.3 Economic significance
Enterprises are continually monitored to determine whether they meet the economic significance requirements for inclusion. These enterprises are maintained on the Business Register and represent the target population from which Statistics NZ's economic surveys are selected. The criteria used are:

- Annual expenses or sales subject to GST of more than $NZ 30,000
- 12-month rolling mean employee count of greater than three
• Part of a group of enterprises
• Registered for GST and involved in agriculture or forestry
• Over $NZ 40,000 of income recorded in the IR10 annual tax return (this includes some units in residential property leasing and rental).

2.4 Standard classifications
Statistical units on the Business Register have a number of standard classifications applied to them:
• Australian and New Zealand Standard Industrial Classification (ANZSIC) – developed from ISIC revision 4
• Business Type
• Institutional Sector and Control

3. Recent history of the statistical register

In 1996 Statistics New Zealand moved from a mainframe based system (the Business Directory) to the Business Frame. The Business Frame was designed with a Centura interface over a Sybase database, and is still in production today, although since May 2014 it has only been available as a sampling frame, with all updating and maintenance made on the new Business Register and replicated back to the Business Frame.

The development of the Business Frame coincided with the greater availability of administrative data from other government agencies, in particular the Client Registration File from the tax office (Inland Revenue). These sources were used to both create statistical units on the register and increasingly to maintain them. Through the 2000s there was a steady decrease in direct surveying to maintain register units – an annual survey of over 120,000 enterprises has decreased over time to be a survey of only 13,000. Maintenance of the register was targeted by the use of maintenance tiers based on business size (turnover and employment) and complexity (group ownership links, geographic locations in different regions), and the economic significance threshold referenced above was used to limit the number of enterprises that were recorded and maintained on the register (500k on our register out of 1.5 million entities with a tax number on the Inland Revenue client register).

The team that maintains the register has also come down in size as direct surveying has reduced. The current Business Register team is made up of 10 full time staff, half of whom are analysts who spend a large proportion of their time working on our annual business demography statistics (both the official release and servicing external clients who want more specific detail), and half of whom are expert updaters.

Part of the original vision of the Business Frame was that it would be open to all users to update, not just the core team, and a significant number of simple updates were made by other survey areas or the contact centre. By the mid to late 2000s the number of updates made outside of the core team had almost faded to insignificance. This was in large part due to decreasing survey volumes and changing priorities in other work areas (particularly the contact centre) where register maintenance was not seen as a core activity.

Since its deployment in 1996 a number of additional features were added to the Business Frame that supported related activity, but went beyond the role of the Business Frame as a purely statistical register. Foremost amongst these were the addition of tables related to survey management (to get a holistic perspective of respondent load across business surveys), the use of the Business Frame as a portal for access to tax data and survey images, and even survey specific functionality built directly into the user interface (sample selection button for Balance of Payment surveys). These functions have not been redeveloped on the Business Register.
The Business Frame user interface was developed in the days of 15” monitors and an optimal screen resolution of 800 x 600 – the standard across Statistics New Zealand in recent years has been for dual 24” monitors with 1920 x 1080 native resolutions. The relatively small interface of the Business Frame meant that limited information could be displayed at a time, but this was well arranged under different tabs so that users could usually see what they needed to on one screen (e.g. classifications, or size measures, or location address, etc.), although a significant limitation was the inability to compare two units side by side.

The new Business Register was created with a browser based interface (using Microsoft Silverlight) and allows for more information to be displayed on screen and greater customisation of personal views. This means that a casual user from a survey area can limit what they see to the relevant level of statistical unit (e.g. single enterprises), while an experienced updater can have multiple units from multiple businesses open and viewable at the same time.

A further development during the history of the Business Frame was the development of the Longitudinal Business Frame (LBF) – essentially a series of monthly snapshots of the Business Frame. The LBF was required because the way history records (e.g. changes of ownership or classifications) were stored on the Business Frame did not facilitate longitudinal analysis across the register as a whole. The new Business Register was designed with longitudinal analysis in mind, with a key requirement being able to quickly see what any unit looked like at a particular point in time – based on either the effective date of changes, or the actual date at which they were applied by the system or operator.

The other key development alongside the deployment of the Business Register has been the creation of a replication process to keep the old Business Frame up to date. This was required because survey sampling processes have not yet been reconfigured to read from the Business Register, and early on in the project it was decided that surveys would make this change at their own pace, possibly related to other redesign work planned in the next few years. We anticipate that the Business Frame will be kept up to date and the underlying tables made available until 2017.

4. How the Business Register is used

4.1 Maintenance sources
The primary source for maintaining the Business Register is the monthly client registration file provided by Inland Revenue. New units are birthed and existing ones ceased based on their registration status. Additional information is also updated including changes of legal name, postal addresses, and other indicators such as identification of Māori businesses. All these updates are made using the tax number as the primary linking key.

Monthly employee count information is also received from Inland Revenue and this is used to apply size measures to employing businesses – again linked by use of the tax number. Monthly Goods and Services Tax (GST) data is loaded and used for financial size measures. Annual business tax returns (IR10 forms) are also used in this capacity, with both GST and IR10s being linked using the tax number. This financial data is also used to identify businesses ceases and reactivations.

The other main source of update information is from the New Zealand Companies Office. Name and address updates are received monthly, and shareholding information is used to identify and load ownership links between registered companies. This information is linked using the company number which in turn is linked to the tax number in the client registration file provided by Inland Revenue.
Administrative data is loaded to the register using configurations compiled within the micro-economic platform. These configurations are essentially SAS code and are mostly maintained by analysts within the Business Register team.

As mentioned above, register maintenance through direct surveying has decreased significantly over the last 10-15 years and the current focus is on the largest and most complex units. An annual survey is sent to approximately 13,000 businesses with questions on life cycle changes, operating locations, main activity, and overseas ownership/transactions/investments (to help specify the population for balance of payments surveys).

Most economic surveys that use the register also have questions around activity or life cycle changes, and a significant number of manual updates are initiated by direct respondent contact.

All register maintenance has been done on the Business Register since May 2014, with changes replicated back to the old Business Frame on a daily basis.

4.2 Sampling
The main purpose of the register is to enable sampling for economic surveys. Most surveys are targeted at the enterprise, but geographic units are also surveyed where a regional dimension is required. The standard classifications recorded on the register are used in conjunction with size measures (GST and employee count), and to manage respondent load, and a random number range (when they are created on the register, enterprises and geographic units are assigned a random number between 0.0000000001 and 0.9999999999) is allocated to each survey.

With the introduction of the micro-economic platform, a standard date within a given month has been increasingly used to enable more consistency across surveys. Initially this was achieved by creating a snapshot of the Business Frame each month – the snapshot was needed because it was not possible to independently read the Business Frame at a previous point in time. With the introduction of the Business Register and its improved recording of fact history, it is now possible to specify a date and select a sample directly from the register and a standard sample section process for surveys run from the micro-economic platform is being developed to do this.

4.3 Business demography statistics
Statistics New Zealand’s annual business demography statistics are produced from a dataset which is ultimately created from the register. The release consists of a continually revised longitudinal dataset with additional methodology used to create data on business births and deaths that aren’t available from a simple snapshot view.

The 2014 release will use a dataset created from the Longitudinal Business Frame as has been done for all previous releases. This is a convoluted process as all updating is done on the Business Register, which is then replicated to the Business Frame, from which the monthly copy of the Longitudinal Business Frame is created. We plan to create the 2015 dataset directly from the Business Register, again leveraging off the improved history, and cutting out the additional steps associated with the Business Frame and Longitudinal Business Frame.

4.4 User interface
Day to day users making reference to the Business Register have a flexible user interface (UI) which can be configured to their own preferences.
The default layout makes use of tiles for each unit (ENTS, KAU, GEOs) which can be expanded to show more detail. Users who may be focused on a different level of detail may configure the UI to only display enterprises, or to display any type of unit but only one on the screen at a time.

Each tile has a number of expandable headings allowing users to drill down into more detail relating to classifications, size measures, indicators, etc. Tiles can also be flipped to display other information such as the history of changes on different variables. Because the UI is browser based, users have the ability to bookmark units and see a list of recent units they have opened.

Three different levels of access control what users are able to do within the UI. Most Statistics NZ staff who work on economic surveys are Readers, which means they can only view units and cannot make changes. Readers can request a change to a unit by clicking a button on it which opens a dialogue box with a free text field. Any change request made this way is visible on the unit until it has been resolved, and it is also automatically added to a workflow management system within the Business Register so that it can be tracked and actioned.

A much smaller number of staff are Writers, this includes the core Business Register team (10 people) and a number of expert users (10-20) in other areas who make regular changes. Writers can update almost any unit, and are required to make the change effective from the correct date (almost always earlier than the date the change is made on) and tag changes with a source code.

A limited number of enterprises have an additional level of access associated with them and are only allowed to be edited by Writers who are named custodians of those units. This usually applies at an enterprise group level and has been implemented to ensure that these key enterprises are not accidentally updated incorrectly by a less informed user.

The UI allows users to link directly through to other information stored on the micro-economic platform for a particular unit – this includes tax reports (graphs and tables), survey images and survey data. Unit reports on the platform can also be configured to include commentary added by users, e.g. the reason for a change of industry, or why retail sales in a particular month are so high.

5. The Micro-Economic Platform

(this section has been abridged from Innovative Production Systems at Statistics New Zealand: Overcoming the Design and Build Bottleneck, published in the Journal of Official Statistics, 2013, vol. 29, issue 1, pages 73-97)

Statistics New Zealand’s aim is for administrative data to be the first source of data for statistical outputs. This key principle of a new statistical architecture has been a driver for developing the new economic statistics platform.

Statistics New Zealand’s existing economic survey infrastructure had the following characteristics that made a new statistical architecture harder to support:

• Each survey had a separate standalone 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.
The micro-economic platform 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.

5.1 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.

In developing the new platform, it was recognised that Statistics NZ is shifting from its narrower focus on data collection to being an organisation that provides a fuller service by collecting, analysing, and disseminating information. Data processing is becoming a 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 was 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.

5.2 Flexible data storage
The micro-economic platform data stores are also designed for flexibility. In a traditional OLTP system, the data tables have column names that match the input and output data. 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. 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 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.

6. Reasons for developing the Business Register on the Micro-economic platform

The original register redevelopment project was based on developing the Business Register as a standalone system similar to the Business Frame. But as planning progressed and business requirements were documented, the project was reassessed in light of the way the micro-economic platform project began to successfully deliver infrastructure for a number of key administrative data collections and economic surveys.

A key benefit was the reduced development effort required as a large number (83%) of the original business requirements were considered to be delivered by existing platform functionality, or could be delivered with a small amount of additional work. Other benefits were related to the ongoing support and functionality of a register tightly integrated with an economic statistics platform.

6.1 Standardisation

The survey silo environment that Statistics New Zealand was trying to move away from was characterised by non-standard processes and software. A consequence of this were specialised operational and support teams that were not as easily able to reassign resources during peaks and troughs of activity through the survey cycle. While subject matter expertise is a valuable asset, system expertise was an increasingly distracting overhead and a barrier to achieving greater efficiency.

By developing the register on the platform, Statistics New Zealand has been able to ensure that this core piece of infrastructure is understood and supported by a team that also understands how to support other systems on the same platform.

This has also meant that the development of the register has not had to cover all business requirements – many were already met by the platform and others were developed as platform solutions. A good example of this has been the way in which Statistics New Zealand’s new geospatial infrastructure has been linked to the platform and the associated services (e.g. geocoding – an important activity on the register) have been able to be quickly adopted by all platform users, even though the original functionality was worked on by a team migrating the building consents system onto the platform. Likewise, survey image viewing functionality was developed as part of the register project, reducing the business requirement scope for other subject matter areas migrating to the platform.

In some cases standardisation has meant compromise, or at least a need to review a process or procedure to understand why it needs to work in a particular way. Small survey specific inefficiencies associated
with adopting standard processes are meant to be offset by the wider efficiency gains made by consolidating the support team and improving the ability of analysts to move between subject matter areas without learning new systems from the ground up.

6.2 Reuse of data
A key enabler for register development on the platform was that many of the administrative sources (tax and employment data) used for maintenance were already being managed on the platform. This meant that links between processes and data sources were much more easily managed using standard tools and configurations.

A further benefit of doing this was to make sure that administrative data was being used consistently. For example, employment data provided by Inland Revenue comes in a complex dataset with different columns representing different statuses of the employee count. As configurations were written to use this data to maintain the register, it was found that the labour market area was using the same dataset but different columns – and a clear opportunity for further standardisation was identified.

6.3 Closer links with economic statistics
As more of the core economic surveys were migrated to the platform, and plans put in place to migrate the rest, the role of the Business Register as the backbone for New Zealand’s economic statistics was embedded more within our core statistical processes by virtue of being developed on the platform.

In many cases the same data (employment, annual and monthly tax returns) is used in the same processing environment. While this could have happened with a standalone system, full integration with the platform has meant that users seamlessly move between looking at data from an economic collection perspective, to looking at the data in the context of the register units it is associated with across a wider range of collections.

7. Development challenges
Alongside the benefits, redeveloping the register on the platform presented a number of challenges.

7.1 User interface vs platform
The Business Register is unique in that it is the only collection on the platform with a custom interface. Once the decision was made to develop the register on the platform, the original development team (a project manager, 3 developers, and a tester) was complemented by development resource assigned from the platform project, but there was a geographical and functional split between the teams. The core register team was based in Auckland, was managed from there using Agile methodology, and was focused on developing the user interface, while the platform development team was based in Christchurch (almost 800 km away) and was focused on the underlying platform. The Christchurch team was working on other platform functionality at the same time, whereas the Auckland team was fully allocated to the project.

The geographically split team actually worked very well, with daily standup meetings and regular travel for key sprints (arranged in 3 week blocks). Slightly harder to manage were issues where the UI and platform architecture were not originally designed to work well with each other.

An example of this was the fact lifecycle already implemented on the platform, which had been designed around fixed period collections where facts moved from an early ‘raw’ state to a final ‘published’ state. This was at odds with register requirements for facts to be continually updated or changed back to a
previous state. This was resolved by developing the concept of a continuous collection on the platform, but elements of the UI also had to be redesigned to work properly with the fact lifecycle design.

A significant UI element affected by this was the workflow interface which is used by operators to select units that require manual intervention (usually failed auto-classifications). Additional ‘dummy’ facts had to be created to store the triggers for workflow inclusion. This was because if the primary fact (e.g. the industry classification) was ever referenced in an automated process, the fact lifecycle would be updated even if there was no change, and the workflow was overloaded with unnecessary alerts.

7.2 Performance
While the performance of the platform is an ongoing operational concern, the first indications that performance might be an issue were highlighted as the register was developed and tested on the platform. As was stated in section 4, the platform was designed to facilitate bulk processing and analysis rather than individual transactions. The original performance requirements for the UI were based on those for the old Business Frame application (up to 5 seconds to save a record), and were unable to be consistently delivered in the early stages of testing.

A lot of effort was devoted to resolving this, bringing in wider infrastructure experts, implementing more monitoring and reporting, and running carefully managed stress tests. Eventually possible causes were narrowed down and hardware was upgraded and the end result is a much more responsive interface. One improvement made to the UI was to present facts and enable user editing before all facts are loaded, and to enable a user to ‘save’ a change and move on to another unit before that save was actually committed to the database. A slightly negative consequence of this is that during periods of peak load, a change made by a user can be queued for a few minutes, meaning that a user going back into a unit soon afterwards may not see the changes they have made.

By virtue of being on a platform and closely sharing infrastructure with other collections, the register is subject to more performance slow-downs than if it had been developed as a standalone system. But this has been offset by the improved performance of bulk maintenance processes, which is directly related to the fundamental design of the platform.

7.3 Replication to the Business Frame
Because there hasn’t been a requirement for survey areas to immediately reconfigure their processes to work from the Business Register, the old Business Frame has to be kept up to date with changes from bulk processes and manual updates made on the Business Register.

Until the Business Frame is finally decommissioned and all economic surveys are referencing the Business Register, Statistics New Zealand has to support two complex applications with additional points of failure.

Keeping the Business Frame up to date has proved to be a much more complex task than was originally anticipated, in large part due to the fundamentally different way in which facts are stored, and changes to those facts are recorded as history.

The replicator had been subjected to limited testing, but a number of serious issues were encountered when it was put into production after the Business Register went live. Initially the replication process was unable to cope with the volume of fact updates (approximately 6 million in the first replication run) in a reasonable time – it was taking days rather than hours. As a contingency measure, the facts that were the root cause of the problem, employee count and GST size measures, were removed from replication, and the old maintenance processes were run directly on the Business Frame. The number of fact updates was further reduced by better identification of changes as opposed to confirmation of current values.
A further issue arose as replication progressed and was related to the timing of fact changes made on the Business Register. When replication was working in ‘real time’, different facts associated with a unit (e.g. industry classification vs the ownership link between and enterprise and a geographic location) could be processed separately on the Business Register, and replication would only identify and action the first of the changes to the unit and discard the rest.

As well as issues replicating the Business Frame, there were flow on effects to the Business Frame snapshots used for sampling and the Longitudinal Business Frame. Both issues were associated with ordering problems in the replicated change history, which only manifested when those snapshots were created.