Australian Bureau of Statistics

METHODOLOGICAL NEWS

A QUARTERLY INFORMATION BULLETIN FROM THE METHODOLOGY AND DATA MANAGEMENT DIVISION
September 2010

Research into Editing of Categorical Data in Business Surveys

There is increasing interest in collecting characteristics data along with quantitative data in business surveys in the Australian Bureau of Statistics (ABS).

For example, the ABS runs the Business Characteristics Survey which collects mostly characteristics data. Characteristics data is categorical in nature and is usually collected using tick-boxes. Whether a business accepts orders via the Internet is an example of a categorical or tick-box data item while the value of total earnings is an example of a quantitative data item. Values of 1 or 0 are used to indicate whether a box was ticked or not.

Efficient micro-editing of tick-box data cannot be achieved with significance editing. For quantitative data, the large range in values of the errors from responding businesses means a small proportion of the businesses tend to be responsible for most of the error generated in the statistics. It is desirable to concentrate most editing effort on these *significant* business responses. This situation does not occur for tick-box values because they can only be 0 or 1. Any error in a tick-box response is roughly as significant as an error in the next. It is not possible to obtain large gains in accuracy by editing only a small proportion of tick-box responses.

Also, due to the nature of tick-box questions, there tends to be too many tick-box edit failures. Those that cannot be corrected manually must be either corrected automatically (by using automatic editing techniques) or left as is. Automatic editing does not involve human intervention. The technique involves using algorithms to find the least number of data values (from those that failed the edits) which, when corrected, allow the complete set of failed data to pass the edits. The data items requiring correction are replaced by imputed values. Therefore, a tool for editing tick-box data must be able to select a minimum set of failed data requiring correction and create the imputed values required. The ABS currently does not have such a tool and current practice is to leave many failed tick-box responses uncorrected.

To address the problem of the cost of editing categorical data, the Statistical Services Branch is conducting research into the editing of categorical data (with a focus on business surveys). We recently completed a review of current methods and tools available for editing categorical data. The review concluded that the methods inspired by Fellegi and Holt's paper, "A systematic Approach to Automatic Edit and Imputation", published in the *Journal of the American Statistical Association* in 1976 are the best solutions currently available. The report provided a list of tools used by overseas agencies for editing categorical data which were considered suitable for further assessment to determine if any could be useful for the ABS. We are now commencing the assessment. In particular, we are interested in some systems built by Statistics Netherlands, Statistics Canada, and the U.S. Bureau of Census. We envisage that this preliminary assessment will provide guidance for further work in this field planned for the next year. For example, even if a suitable tick-box editing tool is found, there is still the need to find how to use it with significance editing

(since many business surveys will have a mixture of categorical and quantitative data requiring micro-editing).

For further information contact either Keith Farwell on (03) 6222 5889 or keith.farwell@abs.gov.au, or Kin Chung on (08) 9360 5286 or kin.chung@abs.gov.au.

ABS REEM Project: Update on progress

An article "Enhancing User Access to Microdata in Australia" appeared in the December 2009 Methodological News. It provided background information on future strategies for microdata access. Remote Execution Environment for Microdata (REEM) was identified as a possible replacement for Remote Access Data Laboratory (RADL).

Since then a number of milestones had been achieved by the REEM Project and one of them is the first stage of development of the Survey Table Builder (STB) which is a similar tool as the Census Table Builder (CTB).

The STB allows users to interact with the user interface or alternatively, the Statistical Data and Metadata Exchange (SDMX) Web service, to produce tabular outputs from categorical household survey data. The REEM Project team is finalising the database that can be loaded for the STB before REEM Stage 1 can go 'live'. That is, to make it available to a small number of external pilot users to gather feedback on its usability and functionality.

In addition to the first stage development of the STB, REEM Stage 1 also managed to achieve the following milestones:

- a functionality to describe ABS household categorical data using Data Documentation Initiative (DDI) in an automated way from existing ABS metadata stores;
- availability of the confidentiality routine for household survey data (that is, perturbation and additivity routines for Census data have been updated to confidentialise weighted count estimates);
- a fully functional SDMX Web service which is a machine to machine query service for generating tabular outputs; and
- Relative Standard Errors (RSE) calculation has been added to the STB (including the standard ABS annotation of the quality of estimates produced based on RSE values).

In addition, REEM Stage 1 has acted as a catalyst for the broader ABS program of the implementation of DDI and SDMX. It is serving as an early demonstrator of the practical feasibility and benefit of applying these standards both internally and externally. It is also serving as an early pathfinder and exemplar for other ABS (and National Statistical Service) developments that will make use of DDI and/or SDMX in the future.

REEM Stage 2 is currently on-going and focuses on enhancing STB to produce key outputs from continuous household variables (such as income) by the end of this financial year. Stage 2 also continues to research automated confidentiality methods for analytical outputs. A user engagement strategy will be undertaken in the next two months to assess detailed user requirements in terms of analytical outputs from microdata.

If you would like further information, please contact Melissa Gare on (02) 6252 7147 or m.gare@abs.gov.au.

Paris Microdata Workshop – 23-25 June 2010

A second annual workshop was held in Paris, in June 2010 to discuss international collaboration in providing access to microdata. The workshop was attended on this occasion by representatives from the statistical organisations in Great Britain, the Netherlands, Germany, Canada, the United States of America, Italy, Eurostat, and was jointly hosted by ABS and OECD. Further workshops are planned for 2011 and into the future.

This year's workshop achieved some useful steps towards this long term goal, and also established an official purpose and list of objectives for the group.

The official purpose is for "statistical Institutes working together on practical steps to advance cross-border access to, and analysis of, microdata by leading the way and taking into account the needs of researchers and policy makers".

The objectives of the group have been established, as follows:

- Increase coordination and communication between institutes and other expert groups to adopt best practice, promote a common understanding and to minimise duplication of work in furthering international access to microdata;
- Advise and make recommendations to Chief Statisticians based on our own work and advice from other expert groups and practitioners;
- Undertake work commissioned by Chief Statisticians; and
- Serve as a forum for other countries or other expert groups to raise issues.

Other topics discussed at the workshop resulted in some very useful information sharing and agreement regarding the way forward. These include:

- The potential use of DDI and SDMX, with the group agreeing to trial the use of these standards by each of the attending National Statistical Organisations (NSOs). Labour Force Survey microdata is currently being considered as the best option for trialling DDI within each organisation as there are many similar concepts within this topic across NSOs;
- Working with the Labour Force Survey Harmonisation project, a project being coordinated by OECD. This work ties in with the topic of DDI implementation;
- Information sharing about new microdata developments within each of the NSOs as well as presentations from new participants about how microdata access is currently undertaken and any current issues; and
- Agreement within the group to co-write a 'directions' paper exploring the options and recommending the way forward for international microdata access. This paper is to be presented to the 2011 CSTAT meeting.

NSOs play a very significant role in the provision of access to microdata. Use of microdata by governments, academics and commercial organisations is an expanding business, and it is vitally important that NSOs stay relevant by actively facilitating access to high quality, timely microdata, both nationally and internationally.

For more information, please contact Michelle Gifford on (02) 6252 7499 or michelle.gifford@abs.gov.au.

Making Quality Visible Update- Data Quality online

In the June 2009 edition of Methodological News, an update was provided on one aspect of the *Making Quality Visible* project called the Australian Bureau of Statistics Data Quality Framework (ABS DQF), which was released on the ABS website in May 2009 (cat. no. 1520.0).

Since its release on the ABS website in 2009, information on the ABS Data Quality Framework has been expanded and now includes more detailed information on the various uses of the framework for making quality-informed decisions through an assistant. This assistant, called the Data Quality Online, helps people use the seven dimensions of the ABS Data Quality Framework to:

- define the quality of a data item or collection of data items (prepare a quality statement);
- assess the fitness for purpose of data in the context of a data need; and

- Identify data gaps and areas for future improvement.

Data Quality Online is located on the National Statistical Services website at http://www.nss.gov.au/DataQuality/

The assistant provides contextual information, in the form of questions, within each dimension of the ABS Data Quality Framework (Institutional Environment, Relevance, Timeliness, Accuracy, Coherence, Interpretability, and Accessibility) to clarify the dimension. Some of these questions vary depending on the context in which the ABS Data Quality Framework is being utilised. This includes the use of the framework with survey, administrative or a combination of the two types of data.

Specific information on the ABS Data Quality Framework in regards to the reporting requirements of the Council of Australian Governments (COAG) is also available within Data Quality Online.

For more details on the dimensions and uses of the ABS DQF please see the ABS Data Quality Framework (cat. no. 1520.0) or the Data Quality Online assistant. More information on the ABS Data Quality Framework can be obtained from Narrisa Gilbert on (02) 6252 5283 or narissa.gilbert@abs.gov.au. For any queries relating to the Data Quality Online assistant please contact Kellie Browning on (02) 6252 5389 or kellie.browning@abs.gov.au.

Donald Rubin to Visit Australia

Professor Donald B. Rubin, John L. Loeb Professor of Statistics at Harvard University, has accepted an invitation to visit the Australian Bureau of Statistics and CSIRO Mathematics, Informatics and Statistics Division (CMIS) in January 2011.

Professor Rubin's research interests include:

- Causal inference in experiments and observational studies;
- Inference in sample surveys with nonresponse and in missing data problems;
- Application of Bayesian and empirical Bayesian techniques; and
- Developing and applying statistical models to data in a variety of scientific disciplines.

Current plans are for Professor Rubin to visit the ABS and CMIS Canberra during the week of 10-14 January 2011 and CMIS Sydney during the week of 17-21 January 2011.

In Canberra, Professor Rubin will give two one-day short courses at the ABS:

- A short course on Causal Inference on Wednesday 12 January, and
- A short course on Missing Data on Thursday 13 January.

A one-day workshop on Data Confidentiality is also proposed for Tuesday 18 January at CMIS (Macquarie University campus).

Further details of the visit will be circulated when confirmed. For more information, please contact Christine O'Keefe (CMIS Canberra) at <Christine.O'Keefe@csiro.au>, or Peter Rossiter (ABS) on (02) 6252 6024 or <peter.rossiter@abs.gov.au>.

How to Contact Us and Subscriber Emailing List

The Methodological Newsletter features articles and developments in relation to methodology work done within the ABS Methodology and Data Management Division. By its nature, the work of the Division brings it into contact with virtually every other area of the ABS. Because of this, the newsletter is a way of letting all areas of the ABS know of some of the issues we are working on and help information flow. We hope the Methodological Newsletter is useful and we welcome comments.

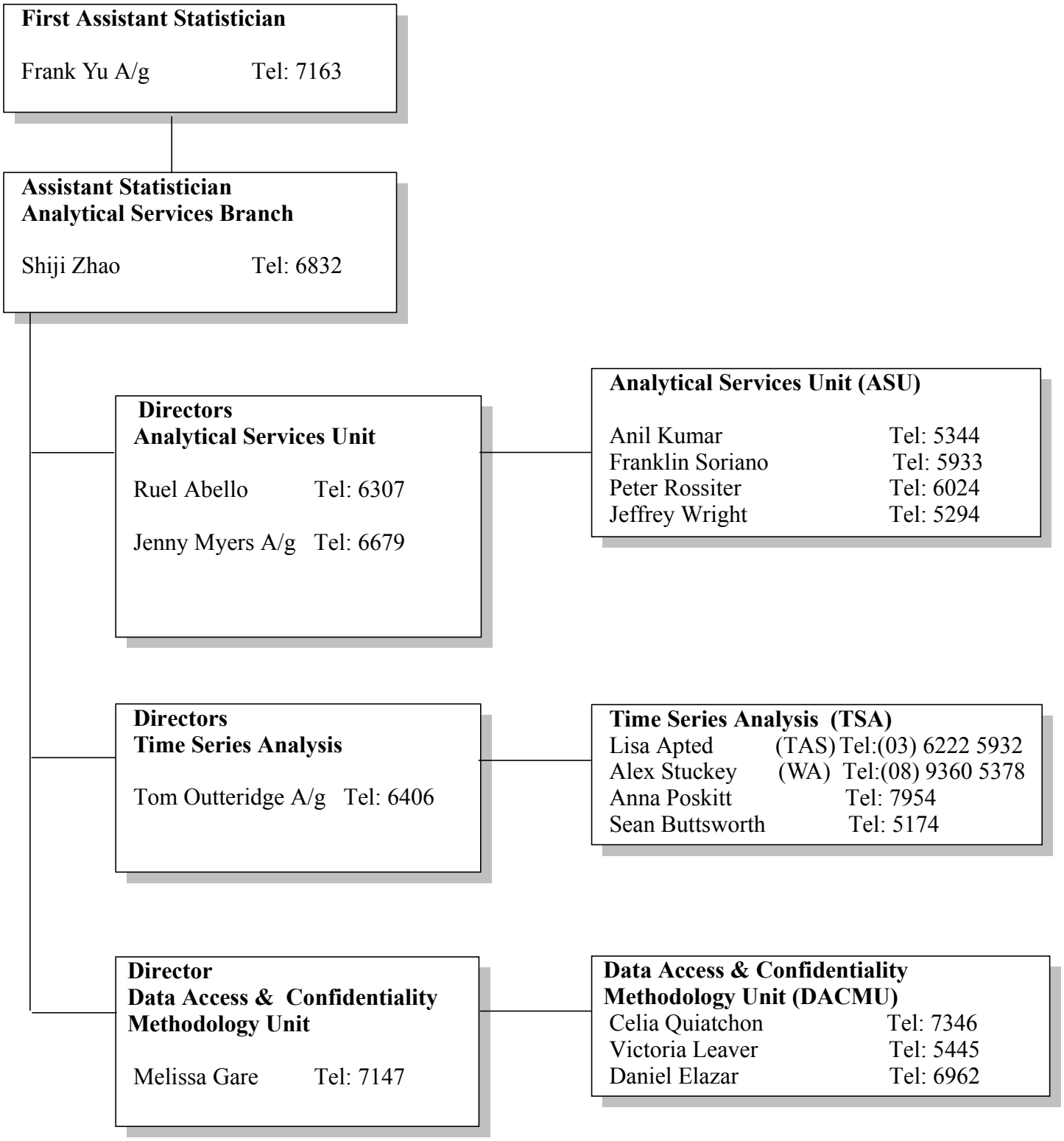
If you would like to be placed on our electronic mailing list, please contact:

Valentin M. Valdez
Methodology & Data Management Division
Australian Bureau of Statistics
Locked Bag No. 10
BELCONNEN ACT 2617

Tel: (02) 6252 7037
Email: methodology@abs.gov.au

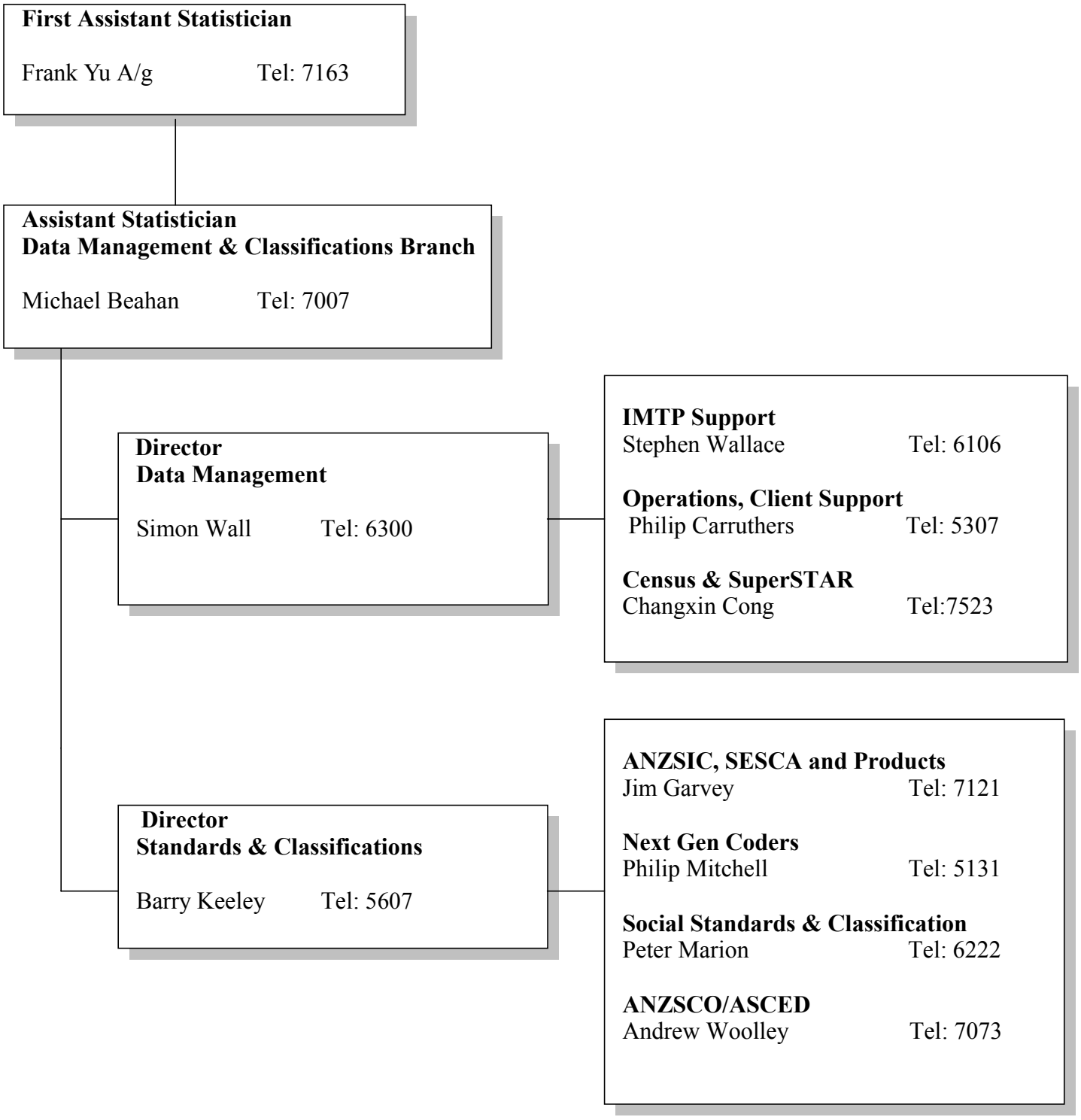
Methodology & Data Management Division
Management Structure

Current at Sept 2010



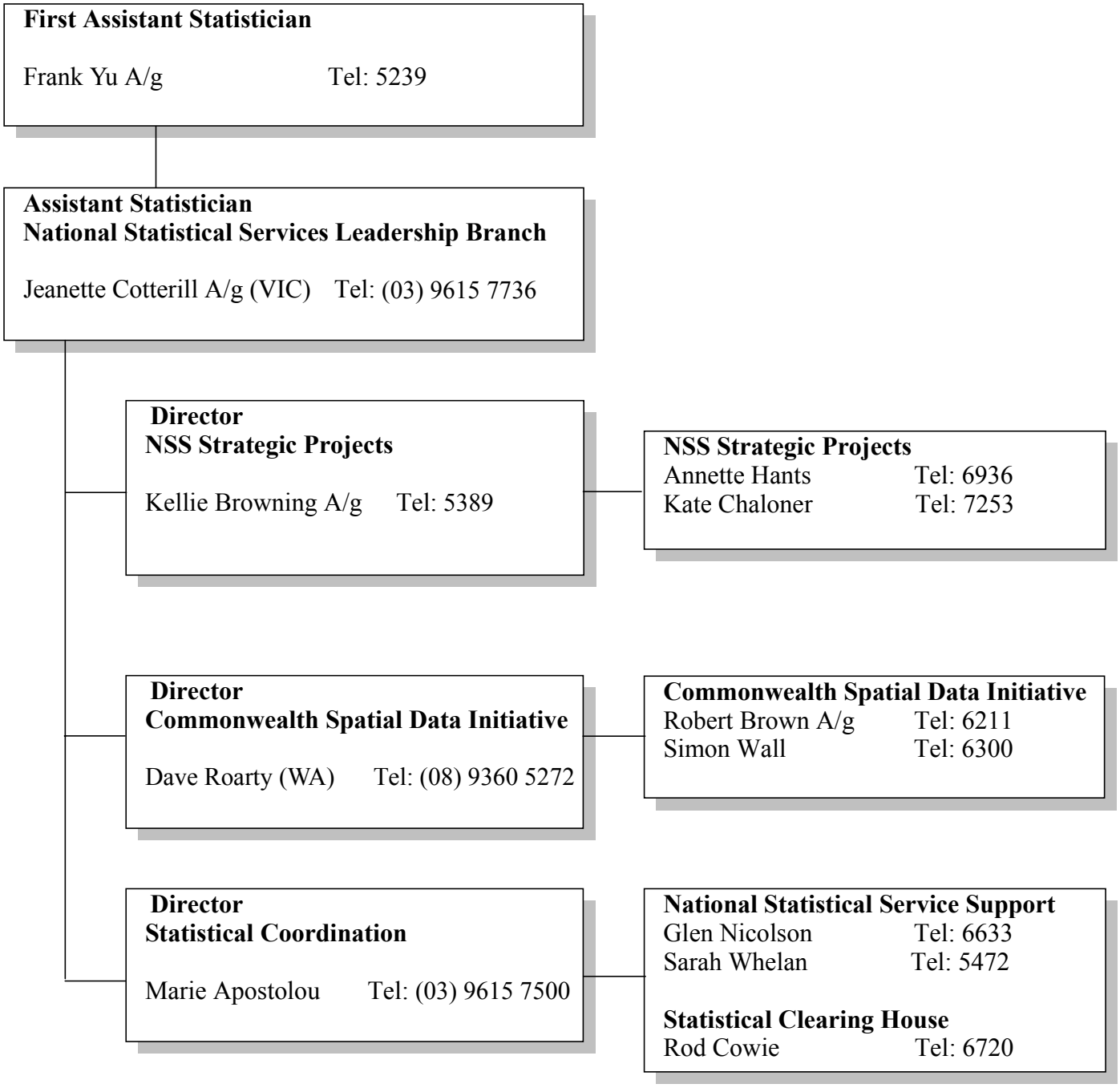
Methodology & Data Management Division
Management Structure

Current at Sept 2010



Methodology & Data Management Division
Management Structure

Current at Sep 2010



Methodology & Data Management Division
Management Structure

Current at Sep 2010

First Assistant Statistician
Frank Yu A/g Tel: 7163

Assistant Statistician
Statistical Services Branch
Paul Schubert A/g Tel: 6591

Director
Household Survey Methodology

Alistair Rogers Tel: 7334

Household Survey Methodology (HSM)
John Martin Tel: 7006
Ross Watmuff Tel: 7084
Justin Lokhorst Tel: (08) 8237 7476

Director
Methodology Development Unit

Bill Gross Tel: 6302

Methodology Development Unit (MDU)
Carmen Kong Tel: 5944
Philip Bell (SA) Tel: (08) 8237 7304
Carl Mackin (WA) Tel: (08) 9360 5250
John Preston (QLD) Tel: (07) 3222 6229

Directors
Business Survey Methodology

Justin Farrow A/g Tel: 5795

Business Survey Methodology (BSM)
Amanda Norton Tel: 5705
Keith Farwell (TAS) Tel: (03) 6222 5889
Brett Frazer (QLD) Tel: (07) 3222 6028
Irina Pribil (VIC) Tel: (03) 9615 7566

Director
Data Collection Methodology

Emma Farrell Tel: 7316

Data Collection Methodology (DCM)
Narrisa Gilbert Tel: 5283
Kettie Hewett (VIC) Tel: (03) 9615 7843

Director
Operations Research and Process Improvement Unit

Mark Zhang Tel: 5132

Operations Research and Process Improvement Unit (ORPI)
Melanie Black A/g Tel: 7241
Rosslyn Starick (VIC) Tel: (03) 9615 7055

Divisional Support Unit (DSU)
Jackie Bishop Tel: 6790