Attachment 1

ABS LEADERSHIP PROJECT — STATISTICAL ANALYSIS IN THE ABS

Project Outcome

Ensure maximum business benefit to the ABS through the publication of statistical analyses.

Project Output

A report to ABS Division Heads Meeting (6 December 2002) which provides best-practice advice regarding the bounds and focus of analytical work appropriate for a National Statistical Service. The report should outline policies and practices which should be adopted when undertaking and disseminating statistical analyses.

Background

A key issue to be addressed is — How should the ABS engage with external analysts and other users, to maximise the benefit of its analyses and minimise the risks?

Potential questions arising from the research will include the following.

a. Bounds and focus of analyses undertaken by a national statistical office

i. What principles guide the sorts of analyses that it is appropriate (or inappropriate) for a national statistical office to undertake?
ii. What principles guide the sorts of analyses that it is most beneficial for a national statistical office to undertake?

b. External advisors and reference groups

i. What roles can external advisors and reference groups play during the stages of ABS analytical work — e.g. when designing the analysis program, when specifying individual analysis projects, during an analysis project, when analytical outputs are delivered?
ii. When can ABS standing committees (such as ESUG, LSAG and MAC) fulfil the external advisory role, and when would establishing a special committee be appropriate?
iii. How should external advisors and reference groups be chosen?

c. Quality management and peer review

i. What quality specifications should apply to a piece of analytical work — by whom and when should those specifications be agreed?
ii. What quality management / quality assurance processes should be applied? In particular, should external or other independent peer review be standard practice?

d. ABS clearance

i. By whom in the ABS should the different varieties of analytical work be cleared for release?
ii. What other policies and practices should apply to clearance?

e. Status and marque of analytical outputs

i. Should analytical outputs be published under the name of the Australian Statistician, or individual ABS employees with appropriate disclaimers, or should they be given some other mark of their status (vis a vis 'official' ABS statistical products)?
ii. Should analytical outputs be published as ABS catalogued publications?
iii. Should different practices be adopted for different varieties of analyses or at different stages of an analytical project's life?

f. Dissemination vehicles for analytical outputs

i. What vehicles are most appropriate — e.g. Web, professional and working papers (catalogued or not catalogued), publications (eg *Measuring Australia's Progress and Australian Social Trends),* or other vehicles?
ii. Should different vehicles be adopted for different varieties of analyses or at different stages of an analytical project's life?

g. Risks

i. What risks arise in ABS analytical work?
ii. What risk mitigation strategies are appropriate?

Attachment 2

LIST OF ABS ANALYSIS PROJECTS

PROTOTYPES OF ANALYTICAL PROJECTS — OUTPUTS DELIVERED SO FAR

Theme : **Product**
Administrative data : New monthly economic indicator based on ATO Business Activity Statement, EFTPOS and other by-product datasets
Administrative data : Analyses of FaCS/Centrelink databanks (Welfare dynamics for mature-age customers)
Progress : Measuring Australia's Progress
Income : Measures of neighbourhood income inequality for Australian cities
Wealth : Household wealth distribution
Crime : Synthetic estimates of crime victimisation for small areas
Crime : Model of propensity to report crime to the police
Human capital : Experimental estimates of human capital stock
Household economics : Dissecting changes in unpaid work into demographic and other influences
Macroeconomic indicators : TRYM macroeconometric model (quarterly release)
Productivity : Estimates of productivity and output for government services (health, education, police, courts and prisons)
Productivity : Estimates of labour adjusted for quality changes (educational attainment and experience)
Composite social indicators : Indexes of socioeconomic disadvantage for Indigenous areas
Prices : Price indexes for real estate agents' services
Prices : Hedonic price indexes for computer hardware

PROTOTYPES OF ANALYTICAL PROJECTS — WORK IN PROGRESS

Theme : **Product**	**Scheduled Completion Date**
Melded datasets : Socioeconomic analyses based on the 2001 Population Census and Agricultural Census	June 2003
Administrative datasets : Analyses of linked hospital - Medicare - Pharmaceutical Benefits Scheme data	March 2003
Composite social indicators : Review and update the Socioeconomic Indexes for Areas (SEIFA)	June 2003
Progress : Measuring Australia's Progress II	November 2003
Income : Synthesising unit record files for years when the Survey of Income and Housing Costs is not run	June 2003
Education and training : Modelling lifelong learning	March 2003
Productivity : Developing measures of output for the remaining government services (tax collection and social security administration)	September 2002
Productivity : Developing measures of industry multifactor productivity, based on annual input-output tables	March 2003

Productivity and performance : Information and communications technology and business performance	November 2002
Prices : "Spatial indexes" for Australian capital cities	February 2003
Prices : Price indexes for computer software	December 2002
Forward looking indicators : Review of the Composite Leading Indicator	October 2002
Crime : Analysis of repeat imprisonment	March 2003

CONSULTANCIES AND ANALYTICAL ADVICE

Theme : **Problem**
Administrative data : What is the quality of the metadata for the FaCS/Centrelink databanks (including quality of Indigenous coding)?
Administrative data : Can the ATO's Business Activity Statement data substitute for any part of the ABS monthly retail collection?
Income : How sensitive are findings about income distribution to the choice of equivalence scale?
Labour : What light can cohort analyses cast on changes in unemployment rates?
Disability : How do the full-blown disability survey and disability survey modules compare?
Crime : Can the National Crime and Safety Survey be used to analyse repeat crime victimisation?
Indigenous statistics : How can Indigenous demography be modelled to take best advantage of Census, survey and administrative data?
Prices : How can scanner data be used to inform CPI design and practices? (sample design and allocation)
Venture capital : What factors affect the take-up and performance of venture capital investments?
Forward looking indicators : What predictive power does the ABS job vacancies series have relative to private sector series?
Forward looking indicators : What predictive power do the ABS business expectations series have relative to private sector series?

Attachment 3

STYLES OF ABS ANALYTICAL WORK — SOME CASE STUDIES

(a) Exploiting By-product Datasets

More than half of our analytical work during the past couple of years has been based (or based in part) on such data. Broadly, we undertake six kinds of investigations :

- evaluating the potential for replacing or supplementing directly collected data
- confronting by-product data with directly collected data
- developing measures of socioeconomic concepts for which a direct collection would be too difficult or expensive or sensitive
- extending existing survey-based estimates to smaller geographic areas or subpopulations or to more frequent time periods than the samples will allow
- longitudinal and other dynamic analyses
- assessing the quality of the by-product data (and associated metadata) for analytical applications, and suggesting changes that would enhance their statistical value.

Where possible, these analyses are undertaken as collaborative projects with staff of the data custodian agencies; at the least, agency staff are on our project boards and peer review panels.

- *Analysing the FaCS-Centrelink databanks.* Since late 2000, we have been doing collaborative work with the Department of Family and Community Services. Under a Memorandum of Understanding, we have access to the Longitudinal Dataset that has been distilled from the social security customer databases. For our first tranche of analyses (completed a few months ago), we :

    - modelled welfare dynamics (ie, movement between payment types and movement in and out of the welfare system) for mature age customers

    - assessed the quality of the data and metadata, including Indigenous coding.

  The next project under the MoU umbrella is confronting information regarding social security payments in the FaCS-Centrelink databanks and in ABS surveys.

- *Analysing the Australian Taxation Office's Business Activity Statement data.* We have assessed the possibility of substituting BAS data for part of the monthly retail collection. Our interim conclusion is that substitution might reduce the provider load on medium-sized businesses but would probably not deliver large cost savings. We also explored the possibility of constructing a new monthly indicator of the pulse of economy activity using BAS and other by-product data (such as EFTPOS). This showed promise, but the time series are still too short to support definitive conclusions; the investigation has been suspended for a couple of years.

- *Analysing health by-product data.* Under an agreement with the Department of Health and Ageing, we have access to a dataset that links patient records from the hospital system, Medicare and the Pharmaceutical Benefits Scheme. Our first project concentrates on patterns of some key mental health conditions, and is examining co-morbidities and the contrasts between incidence and treatment patterns ("health pathways") in urban and rural areas. The project will deliver three outputs for Christmas 2002 — a report of the pilot study analyses; an assessment of the quality and analytical potential of the linked dataset; and a proposal for a *program* of collaborative research projects in 2003.

- *Assessing the potential for using business by-product data in price index compilation.* Scanner datasets record the prices and quantities of goods purchased through certain outlets, such as supermarkets. In the past couple of years, ABS analyses are concentrating on using the datasets to model and refine current compilation practices for the Consumer Price Index (CPI), such as the selection of outlets and commodities for which CPI price collectors gather data and the index formulae applied at various levels of aggregation. Recently, we have shifted our focus to supporting Prices Branch investigations of the direct use of by-product data (such as retail chains' price lists, Internet shopping prices, scanner data, and so on). Before any such direct use could be contemplated, the ABS would have to achieve a thorough understanding of such issues as: what proportions of outlets, commodities and transactions are covered by the scanner data and how the coverage may vary over time; whether the barcodes are assigned in such a way as to permit consistent tracking of commodities; and what costs may be entailed by acquiring and processing the very large scanner datasets.

- *Analysing prison statistics.* We have initiated a joint project with the National Centre for Crime and Justice Statistics to analyse patterns of re-offending prisoners. This use the National Prisoner Census, available from 1993 onward. The project, which will be guided by advisors from the Crime Research Centre (University of WA) and two State government departments, will be completed in May 2003.

(b) Exploiting Melded and Multiple Datasets

- *Estimating the distribution of household wealth.* The household balance sheet (part of the annual national accounts), shows net worth dissected by a couple of dozen assets and liabilities, such as dwellings, shares and loans. This project developed estimates of the distribution of household wealth across household types (including stages of household lifecycle), across States and across other dimensions for the years 1994 through 2000. The team drew on about a dozen surveys and administrative sources — the Survey of Income and Housing Costs, the Household Expenditure Survey, the Rental Investors Survey, and so on. In the future, the estimates will be used to refine and extend the household balance sheet, and will be integrated with estimates of household income distribution (as envisaged by the ABS some years ago in *A Provisional Framework for Household Income, Consumption, Saving and Wealth* (Cat. No. 6549.0)).

The household wealth project has drawn on the expertise of a wide variety of people in the ABS, in other departments (such as Treasury, FaCS and the Reserve Bank) and in universities (such as the ANU and NATSEM). In May 2002, we hosted a full-day evaluation workshop involving about a dozen experts. The work was favourably reviewed, and the team received valuable suggestions and cautions regarding the interpretation, validation and presentation of its estimates. The workshop was an expensive exercise, but so fruitful that we shall adopt the workshopping process for some of our other major analytical prototypes.

- *Estimating crime for small areas.* This project tried to develop estimates for Statistical Local Areas by modelling the relationship between crime victimisation (available from the National Crime and Safety Survey) and the characteristics of people and locations (from the Census). The project team did a thorough investigation, and we received valuable advice from other analysts inside and outside the ABS. But the problem defeated us — we have not been able to generate defensible small area estimates. So we have put the project to sleep, pending our finding a better approach. Multilevel modelling shows some promise; for a possible plan of attack, see the paper we presented to the recent congress of the International Sociological Association in Brisbane. It is important that we develop a competency for handling this style of problem, because of the growing demand for State and regional estimates.

- *Analysing melded data from the 2001 Population Census and Agricultural Census.* We have initiated a joint project with Production Statistics Branch, Census, Demography and Geography Branch and the Rural and Regional Statistics National Centre. We are now developing a list of policy and research questions that might be investigated during the next year or two (such as the interaction between on-farm and off-farm income-generating activity, the relationship between demographic characteristics and land management practices, or locational choice of farm family members). From the list, we shall choose an entry-point project — addressing a tractable analytical question that will give us experience with constructing and manipulating the melded dataset. A key task for the project team and board will be finding analytical questions that are not just amusing to tackle but address issues that are relevant to policy and statistical priorities.

- *Synthesising income microdata for years when income surveys are not run.* We have initiated a joint project with Social Conditions Branch. Our aims are to develop methods for interpolating microdata between biennial surveys and extrapolating microdata beyond the most recent survey. The data synthesis would take account of such influences as changes to labour market conditions, tax and social welfare payments. If this project —due for completion mid-2003— is successful, the ABS will be able to provide a service to the research community much more cheaply than by running annual surveys. A key task for the project team and board will be understanding what "success" means for such a project, and how to bring the best technical and subject matter intelligence to bear on the work and its validation.

(c) Model-based Data Construction

- *Valuing the stock of human capital.* Human capital is the stock of knowledge and skills embodied in a nation's people. We have used a modelling framework (developed by Dale Jorgenson and Barbara Fraumeni in the USA) to estimate the present-day value of the lifetime income streams for Australians with given educational qualifications and other characteristics. The first round of estimates used Census data, but they can be extended to interCensal years. A presentation of the project at the Conference of Economists received enthusiastic critiques (from Professors Bob Gregory and Bruce Chapman), and we are now putting the work through more formal peer review (by Professor Jeff Borland, Barbara Fraumeni and others). Recently, policy departments have expressed interest in a satellite account for human capital / education and training that integrates stock and flow information.

- *Deriving industry productivity measures.* The ABS publishes estimates of multifactor productivity for the whole of the market sector. There is great interest among policy makers and researchers in obtaining MFP estimates at the industry level — and, now that the ABS is building up an annual time series of input-output/supply-and-use tables, there is some prospect of compiling industry productivity estimates in a defensible way. This project is one of great technical difficulty, and there are few external experts (either in Australia or abroad) who might guide us and vet our methods and findings. So we have submitted a paper on our plan of attack to the forthcoming International Conference on Input-Output Techniques in Montreal. We shall have experimental estimates around Easter 2003; we shall then assess whether the methods provide a viable basis for the regular compilation of industry MFP series.

(d) Analytical Products that Cut Across the Economic/Social/Environmental Domains

- *Measuring Australia's Progress.* In April 2002, the ABS released MAP, a new publication that presents and analyses indicators of progress in the economy, society and the environment. This was quintessentially a cross-cutting and collaborative project. The consultation and review processes took eighteen of the twenty-four months of development time. The reception of the publication has been overwhelmingly favourable, including very positive comments by the new ASAC chair at a recent international conference. And we have had one prominent critic.

- *Satellite accounts* are a more systematic way of drawing statistical threads together. They are an adjunct to the core economic accounts (such as the Australian national accounts) that can be used to :

    - highlight a particular aspect of economic life (such as tourism, which is not recognised as an industry in standard classifications)

    - display the results of different statistical treatments (such as household accounts, which treat the household as a producing not just a consuming unit; or income accounts adjusted for the depletion of natural resources)

    - analyse macro-micro links (such as dissected wealth matrices that link the aggregate balance sheet for the household sector with the distribution of wealth across types of household)

- analyse links between the economy and society and the environment (such as the link between the volume and composition of production and the emission of atmospheric pollutants).

In late 2000, the ABS published a tourism satellite account for Australia, which showed aggregate activity, the supply and use of commodities and employment associated with tourism activities (ABS Cat. No. 5249.0). Australian water accounts and energy accounts have been released (ABS Cat. Nos. 4610.0 and 4604.0). The ABS has also developed experimental estimates of national income adjusted for the depletion and some discoveries of subsoil assets. Satellite accounts that are being developed or contemplated include: household wealth distribution; non-profit institutions; household production and consumption; information technology and telecommunications.

Satellite accounts are a very powerful tool for analysis, but constructing them is expensive and time consuming. When deciding whether to develop a new satellite account, the ABS takes into consideration such criteria as: the strength of user demand; the availability of international standard concepts, frameworks and procedures; and the quantity and quality of data.

(e) Reviews of Analytical Methods and Products

In the interests of brevity, I just list some of this work :

- Completed — reviews of the predictive power of the ABS job vacancies and business expectations surveys.
- In train — a review of the composite leading indicator (which is intended to give early signals of turning points in aggregate economic activity).
- For the coming year — a review of the Socioeconomic Indexes for Areas (a suite of indexes of advantage and disadvantage, derived by applying principal components analysis to Census data).

Attachment 4

HOW THE ABS ANALYSIS PROGRAM IS DECIDED

The program is overseen by the ABS Analysis Board (chaired by Geoff Lee, the head of Methodology Division, with the heads of the economic and social subject matter divisions as members). Four times a year, the Board considers a costed menu of projects and decides which projects it will commission.

Below we outline how projects get onto the menu and how the board chooses between them.

Sources of Ideas for Analytical Projects

A paper written by senior ABS managers in late 1999 listed twenty themes for analysis work. These include the distribution of household wealth, national progress, human capital, the productivity of government and the market sector, and disability. During the past couple of years, that paper has been a major source of project ideas.

Other suggestions have come from :

- external clients — for example, modelling company income and tax (ATO), modelling the dynamics of participation in the social security system (FaCS), and constructing an index of Indigenous socioeconomic disadvantage (Commonwealth Grants Commission)
- Analysis Board members and periodic meetings with senior staff of the subject matter divisions — for example, spatial price indexes, productivity-adjusting labour price indexes, modelling Indigenous births and deaths, and melding the 2001 Census and agricultural data, synthetic estimates of crime victimisation, a price index for real estate agents' services.

Recently, we have adopted a more systematic approach to developing project proposals — these flow from the suite of "Information Development Plans" that the ABS is generating in co-operation with government departments, researchers, community organisations and other users.

Information Development Plans as a Source of Project Ideas

Schematically, an IDP embodies three kinds of knowledge and shared commitment to statistical development activity.

First, an IDP encapsulates the demand for statistics in a given field. Through close consultation with users —government agencies, business, academics and community organisations— a picture is developed of the statistics that would, ideally, support informed design and evaluation of policy, other decision-making, research and community discussion.

> In the early stages of IDP work, the picture of demand might be as simple as a list of the major uses, such as the key policy issues, research questions or matters of community debate. As the work on an IDP matures, it would be possible to develop a more sophisticated 'information demand model' that captures the key entities (such as

persons, households, business and government agencies) about which data are needed, the relationships between them (such as service provision or transactions) that should be given statistical expression, the classifications and dissections (such as region or industry or subpopulation) and the data items and metadata that, ideally, would be available to decision-makers, researchers and the wider Australian community.

Second, an IDP encapsulates the supply of statistics (and of raw data that might be used to create statistics) in the given field. Through close consultation with data custodians —chiefly government service agencies, but also some business and other organisations— a picture is developed of the databanks that might become part of the national statistical service.

In the early stages of IDP work, the picture of supply might be just a list of the major data custodians, the databanks that they own and the broad kinds of data contained therein. As the IDP matures, it would be possible to develop a more sophisticated 'information supply map' — a quality-annotated data census for the field that covers ABS and non-ABS direct collections, transactional and other by-product databanks assembled by government agencies and business organisations, as well as other potential raw material for statistics.

Third, an IDP encapsulates the agreed statistical development activity in the given field. The comparison between the demand and supply of statistics will reveal :

- Information gaps — such as key variables arising in policy, decision-making, research or debate that have not yet been given statistical expression. In principle, such gaps might be addressed by new or expanded statistical collection activity. But they might also be addressed by distilling statistics from by-product databanks or by data modelling. Not all information gaps can be addressed; resources would be directed to those gaps that have the most serious consequences for policy, other decisions, research and community discussion. And not all gaps need be addressed by the ABS; some of the gap-filling activity will might be undertaken by colleagues in other agencies or in universities.

- Information overlaps — such as variables for which competing or inconsistent measures are available. Such overlaps might be addressed by undertaking a data confrontation study to understand the differences between the statistical pictures painted by the multiple data sources and, perhaps, to construct statistics that make best use of the information embodied in all the sources. Again, not all information overlaps will be addressed, and not all need be addressed by the ABS.

- Other information deficiencies — such as poor quality (relative to the analytical purpose to which the data are being applied) missing disaggregations by region or industry or subpopulation, differing definitions or counting rules, or only rough approximations to the desired socioeconomic concept. Direct collection, distillation from by-product databanks, data modelling, and development of national standards and minimum datasets might be used to address such deficiencies.

Criteria for Choosing Analysis Projects

The selection of projects from the costed menu is done by judgment, not by formula. But the Analysis Board refers to the following broad criteria, among others :

- How important is the client and how important is the issue to the client?
- Is there a niche for us — do ABS analysts have an advantage over nonABS analysts; is the analysis being undertaken or could it be better undertaken by someone else?
- Are suitable data and adequate metadata available?
- What is the likelihood that the technical problems can be solved?
- Do we have staff with the necessary analytical skills? Do we have a project leader?
- Are staff with the necessary subject matter expertise available to join or guide the project team? Who is available to peer review the technical and subject matter content of the project?
- What is the likelihood that the prototype product would be incorporated in the ABS's (or our external clients') ongoing work program?

Attachment 5

PEER REVIEW AND ADVISORY PANELS FOR ABS ANALYTICAL PROJECTS

Project	Universities	Government agencies	Other
Administrative data	National Centre for Social and Economic Modelling; Social Policy Research Centre (UNSW); Australian National University	Australian Taxation Office; Department of Family and Community Services; Department of Health and Ageing; Centrelink	Statistics New Zealand
Progress	Australian National University; Swinburne University; National Centre for Social and Economic Modelling; University of Canberra; University of Sydney; University of New South Wales	CSIRO; Treasury, Environment Australia; Productivity Commission; Department of Family and Community Services; Department of Health and Ageing; + several dozen other Commonwealth and State agencies	Several hundred other community agencies and individuals
Income	National Centre for Social and Economic Modelling; Social Policy Research Centre (UNSW); Australian National University	Department of Family and Community Services	Statistics Canada
Wealth	Institute of Applied Economic and Social Research (Melbourne U); National Centre for Social and Economic Modelling; Social Policy Research Centre (UNSW); Australian National University	Reserve Bank; Treasury; Department of Family and Community Services; Bureau of Rural Sciences	
Crime		Australian Institute of Criminology; NSW Bureau of Crime Statistics and Research; WA Crime Research Centre; SA Office of Crime Statistics; Police statisticians in the other States	
Human capital	Australian National University, University of Melbourne		US Bureau of Labor Statistics; US Bureau of Economic Analysis
Lifelong learning	University of Canberra	Department of Education, Training and Youth Affairs	
Forward-looking indicators		Treasury; Reserve Bank	
Productivity	Australian National	Productivity Commission;	Statistics Canada;

and business performance	University	Department of Industry, Science and Resources; Australian Taxation Office, Centrelink, Department of Family and Community Services	US Bureau of Labour Statistics
Venture capital	University of Alberta	Department of Industry, Science and Resources	