

SOUTHERN REGION HEALTH & SOCIAL DATA SETS PHASE 2 FEASIBILITY REPORT

March 2008

Prepared by  **Insight** Social &
Health Research

Cathi Walker
Director, Insight Social and Health Research
Phone: (03) 9012 5291
E-mail: cathi.walker@insightsocial.com.au
Web site: www.insightsocial.com.au

Table of Contents

BACKGROUND AND PROJECT AIMS	1
GLOSSARY	2
DATA PRIORITIES	2
DATA PROCUREMENT	3
DATA MANIPULATION AND ORGANISATION	3
DATA PRESENTATION	4
PROJECT OUTCOMES	4
RECOMMENDATIONS	5
PHASE TWO, PART 1 PROJECT FLOW CHART	6
APPENDIX 1: LIST OF DATA SETS	8

Glossary

ABS	Australian Bureau of Statistics
ASGC	Australian Standard Geographical Classification
LGA	Local government area
SLA	Statistical local area
VLGA	Victorian Local Government Association
VPHS	Victorian Population Health Survey

Background and project aims

The Southern Region Community Planners Network consists of a group of local government social planners across the Southern Metropolitan Region (SMR), which collaborates on planning issues relating to community needs. In 2007, the network received Department of Victorian Communities funding for a project to provide key health and social data to the local government and community sectors.

The aims of the health and social data project are to:

- Phase One: Investigate the range of data (social, health, wellbeing and other selected data types) held by various State Government departments and other organisations.
- Phase Two: Organise the data of greatest relevance to councils and communities in an accessible format; and place that information onto a website for easy access by local councils and community groups.

The components of Phase Two are:

- Production of a feasibility study for a social and health data clearing house, and development of data sets for the clearing house.
- Development of a demonstration web site.
- Development of data usage guidelines and a data usage resource kit.

This report forms part of the first component – developing the data sets and producing a feasibility study. It makes references to the Phase One report and should therefore be read with that report as a reference if necessary.

Data priorities

Phase One of the project identified a range of data sets which were potentially relevant for inclusion in Phase 2 (see the Data Priorities Table in the Phase One report). There were prioritised from priority 1 (highest) to priority 3. Prior to the tendering out of Phase Two, the project steering group further identified a range of 'core' data priorities:

- Centrelink data.
- Drug and alcohol data.
- Accident data.
- Maternal and child health data.
- Housing data.
- Rental data.
- Education data.
- Cultural diversity data (including birthplace by age, new arrivals, language by age and suburb, and English proficiency).

All of these areas are – apart from education ¹ – covered to some extent by the data sets compiled for the project. There is probably the least information in the area of drug and alcohol data, with hospital data providing the most reliable and up-to-date data source. Customised cultural diversity data has been obtained for the project, with some areas (e.g. new arrivals) already covered by existing online data sources – primarily the VLGA social statistics website section.

Beyond these key areas, the Data Priorities Table from the Phase One report has been used as a guide in obtaining data, with all priority 1 items obtained where both feasible and within the project parameters (stating that the data be at local level, not be duplicating existing data, etc.). The priority 1 items not included as listed in [Appendix 1](#).

The sheer scope of the range of data identified in Phase 1 means that the desire of government researchers for a 'one stop shop' cannot be fully met. There is a wide range of standalone datasets available, and it has always been a stated goal of this project to not duplicate existing accessible data. Key examples of this would be the Burden of Disease data website and the VLGA social statistics library. However, the data compiled also includes a number of links which could be included on the website, meaning that the end product will serve as a comprehensive resource and portal for researchers.

¹ Education was dropped by the steering group during this phase of the project as part of a tightening of focus.

Data procurement

The project has been fairly straightforward in terms of obtaining data. Much of the data has been available free of charge, although a range of this data has been customised or unpublished data (e.g. hospital admissions, child dental health status). The main broad types of dataset consist of:

- Paid customised data sets (e.g. ABS data on cause of death and customised Census data; Centrelink data).
- Paid standard data sets (e.g. small area labour market data in spreadsheet format).
- Free customised data sets (e.g. hospital separations, emergency presentations, school dental health service data, immunisation rates).
- Online data already in available in a relatively usable format (e.g. ABS estimated resident population, service data, crime data).
- Data available online but not in a usable format; this data has needed to be converted to spreadsheet format from PDF and map format (e.g. data on diabetes, bulk billing rates, maternal and child health services, road accidents). This has been one of the most time consuming aspects of the project.

Mismatch of geographic types has been one issue in the data formatting process, with, for instance, very little data grouped by region. Some data is available by SLA and has needed to be grouped up to LGAs.

If funding is allocated at any future point to keeping the data sets up-to-date, then it would be worthwhile undertaking detailed negotiations with data providers about procuring more of their data in electronic table format (non-PDF) for the purposes of inclusion on the website.²

Development of data guidelines is being undertaken in part three of Phase Two. This component includes notes on the data sets, explanation of methodology (e.g. calculations of rates), notes on data limitations, and notes on how to use different data types and what the various datasets show.

Data manipulation and organisation

The data has been organised so that consistent ASGC geographic labels and codes are used throughout for LGAs and other geographic area types, and the LGA data is also grouped by region where feasible, and also by metropolitan or rural area.³ This also allows the data to be geocoded if necessary; the feasibility of having mappable data will need to be explored during the website development. Options would include having thematic data maps on the website and/or providing people with geocoded data so that they can map it themselves if they have the appropriate software and expertise. As the data stands, LGA data codes as well as labels are included, meaning that the data can be geocoded if users wish to, for purposes such as thematic mapping of data spread. Consistent groupings and field labels have also been used where possible for common fields such as age groups.

² Note that this option was pursued in some areas, e.g. diabetes and road accident data, with data providers either unwilling or unable due to lack of time to provide the data in spreadsheet/customised format.

³ Due to missing data for some LGAs, it is not, for instance, feasible to calculate regional sub-totals for the hospital data.

Rates per 1,000 estimated resident population have also been added to selected data sets, to allow valid comparisons, in addition to inclusion of percentage data where useful (e.g. population by age).

Data presentation

The final data sets are in the format of Excel spreadsheets, with common identifiers at geographic level (primarily LGAs and regions). It is envisaged that the data will be distributed both via CD-ROM and via a website. The data may also be collated in database format at a later stage. Note that the customised Census data used is too detailed to be manipulated in its current form and would require transfer to a database to be used in a comparative way.

The development of a demonstration website is being undertaken separately.

Project outcomes

The project has identified various data sets from the original Phase 1 list which did not turn out to be appropriate for the project in terms of its stated parameters. These types of data sets included:

- Data sets where the geographic areas were not local (e.g. GP data by GP Division; absences from school data by region).
- Data sets where the amount of data provided varied by area (e.g. some HACC and aged care data).
- Data sets which were out-of-date and/or superseded by new data sets (e.g. school entrant children's health status; lifestyle data from the VPHS, small area estimates of disability, housing stress).
- Data sets which fell outside the priority list of data types identified by the steering group (e.g. environmental data).
- Data sets which turned out to have inadequate information or detail at local level (e.g. childcare funding, aged services data).
- Data sets with confidentiality issues (child protection notifications, assessment of reading).
- Data sets which were already available at an adequate level of detail (e.g. new arrivals).

The main data gaps relate to aged care (where the data has been affected by the change in federal government and shifts in website navigability), dental health (where little data is available at local level), and drug and alcohol usage (where the data available tends to be non-representative of the broad problem – e.g. service usage, ambulance data – or only at the extreme end, such as hospital data and deaths data).

Recommendations

One of the tasks included in the original Phase Two project brief was to “Develop methodology to ensure accessibility including technical support needed”. A key issue for the project is the apparent lack of resources available for the provision of such technical support (particularly in terms of answering queries, although also for website technical queries and glitches). The Phase One report recommended providing a feedback forum so that changing data needs and site gaps can be addressed. Whilst the means of providing such a forum is part of the web site design component, the issue of a lack of resources to respond to feedback is an issue for all elements of Phase Two. It is recommended that there be one contact point for addressing data issues/queries and a separate contact point for addressing technical issues.

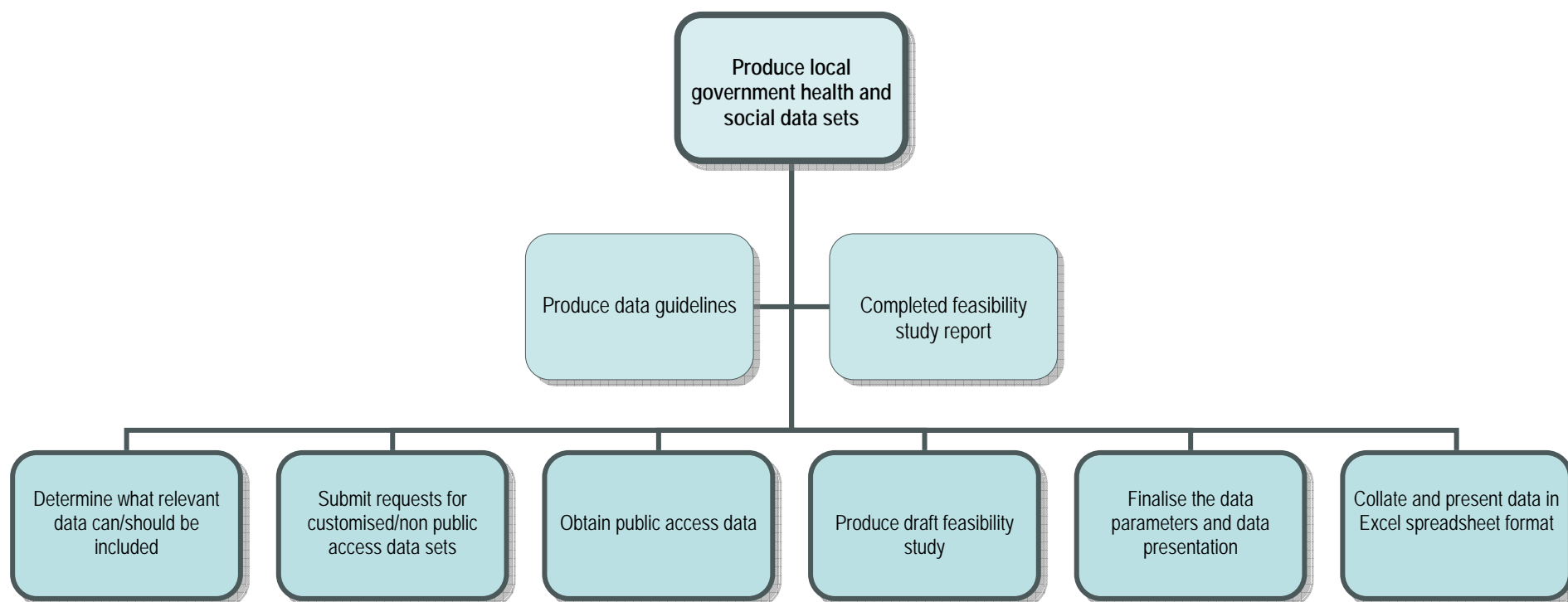
A key recommendation from Phase One was that “The data and web site to be kept up-to-date”. This was considered crucial across the board during the Phase One consultation. This recommendation relates mainly to the actual data sets being kept up to date, but it is also important that any dead/altered links be removed or updated as necessary; that new data sets be added as they become available; and that one-off data sets which become out-of-date be removed or archived. It was therefore recommended that “Staff or financial resources be set aside for updating the site regularly”. This recommendation stands as central to ensuring that the web site continues to be usable beyond its first twelve months. If the web site was to be maintained, funding would also need to be available for updating customised data sets, however, this would be a minor cost component (less than \$2,000 so far in the project).

A range of recommendations were made in the Phase One report in regard to potential for improving data quality and accessibility in a range of areas, and will not be replicated here.

Phase Two, Part 1 project flow chart

Phase One of the health and social data project involved investigate the range of data (social, health, wellbeing and other selected data types) held by various State Government departments and other organisations. The key tasks from this phase were investigating the available data, consulting about what data was both available and useful, and assessing the relevance and value of the data sets identified.

The main tasks involved in Phase Two, producing the final data sets, are shown in the chart below.



The project has not followed a consecutive order for these tasks, with the list of to be included being extended in some areas and reduced in others as the project progressed. Thus the process of requesting and obtaining data has also been ongoing. The starting point for the list of data to be included was the Data Priorities Table from Stage One, and a summary of key areas developed by the steering group. Some of the data from the Data Priorities Table did not get included in the final data set due to issues such as lack of LGA level data, out of date data, and lack of data usefulness, or lack of availability within the project time frames. Some data (such as service data) was excluded following discussions by the steering group.

Following this part of Phase Two, a data guidelines report was developed as an accompaniment to the data sets. The final stage is for the data to be loaded onto a web page.

Appendix 1: List of data sets

Data sets included:

- LGA and region list
- ERP
- ERP notes
- Population by age by postcode
- Small area labour market data
- CentreLink payments by LGA
- CentreLink metadata
- CentreLink postcodes assigned to LGAs
- VAED - LGA by age group
- VAED - LGA by MDC
- VAED - metropolitan LGA by MDC by age
- VEMD - LGA by age group
- VEMD - LGA by diagnosis
- VEMD injury data - LGA by injury cause by human intent
- Hospital data notes
- Road traffic casualty accidents
- Diabetes prevalence
- GP bulk billing
- Public housing stock
- Rental affordability data
- Family violence incidents
- Maternal and child health indicators
- Immunisation by LGA
- Immunisation by postcode
- Child dental health status
- Birthplace of person by age (detailed)
- Language spoken at home by age (detailed)
- Persons in non-private dwellings by dwelling type
- CentreLink data by postcode
- Cause of death by age and LGA.

Data excluded:

- Aged care assessment service data
- HACC service data
- Education data
- Kindergarten participation rates
- Mental health service clients by service type
- Alcohol and drug service clients
- Environmental data
- Best Start data
- Crimes – total offences reported by postcode
- Ambulance attendances at drug overdose.

Data not available within project time frame:

- Median house price by municipality and suburb
- SEIFA data at LGA level
- Detailed road accident statistics (data provided online currently being revised, custom data not available at present)
- Child health and wellbeing indicators from Local Level Child Health and Wellbeing Survey.

Data not relevant to project aims:

- Housing stress data from NATSEM (superseded by 2006 Census data)
- Lifestyle data from VPHS.

Data already available:

- New arrivals
- Births (including teenage births)
- Crimes and crime rates by LGA
- Disability data (Census data now available).

Data quality/coverage inadequate

- Aged care places
- Bankruptcies, debt agreements & personal insolvency agreements by statistical subdivision
- GP distribution
- Public housing waiting list
- Childcare funding
- VHPS regional data
- School entrant children's health status
- Primary health services data.