

**Data standardisation project
for the development of
a national unit record
public housing data set**

WELFARE DIVISION
WORKING PAPER NO. 30

Data standardisation project for the development of a national unit record public housing data set

Hongyan Wang, Rosemary Karmel and David Wilson

August 1998

Australian Institute of Health and Welfare
Canberra

© Commonwealth of Australia 1998

This work is copyright. Apart from any use as permitted under the *Copyright Act 1968*, no part may be reproduced without written permission from AusInfo. Requests and enquiries concerning reproduction and rights should be directed to the Manager, Legislative Services, AusInfo, GPO Box 84, Canberra ACT 2601.

A complete list of the Institute's publications is available from the Publications Unit, Australian Institute of Health and Welfare, GPO Box 570, Canberra ACT 2601, or via the Institute's web site at <http://www.aihw.gov.au>.

Suggested citation

August 1998. Data standardisation project for the development of a national unit record public housing data set. Canberra. Australian Institute of Health and Welfare (Welfare Division Working Paper No. 30)

Australian Institute of Health and Welfare

Board Chair
Professor Janice Reid

Director
Dr Richard Madden

Any enquiries about or comments on this publication should be directed to:

David Wilson
Australian Institute of Health and Welfare
GPO Box 570
Canberra ACT 2601

Phone: (02) 6244 1202

This report was prepared by Hongyan Wang, Rosemary Karmel and David Wilson of the Housing Unit of the Australian Institute of Health & Welfare.

Contents

- Summary of findings 1
- Section 1: Summary of the project..... 3
 - 1.1 Background 3
 - 1.2 Project objective 3
 - 1.3 The data used in the project..... 4
 - 1.4 Project outline 4
 - 1.5 Future development of a national CSHA unit record data set..... 8
- Section 2: Specifications of standardised data..... 9
 - 2.1 Data set record content and formats..... 10
 - 2.2 Detailed specifications 13
 - 2.2.1 General data items..... 13
 - 2.2.2 Dwelling, or property, level items 14
 - 2.2.3 Household level items 25
 - 2.2.4 Income unit level items..... 37
 - 2.2.5 Person level data items 47
- Section 3: Documentation of merged data sets 53
 - 3.1 Introduction..... 53
 - 3.2 Contents of each data set..... 54
 - 3.3 Detailed classifications and explanations 57
- Glossary 92
- Appendix 1: The input data sets 94
 - Part 1: Description of public housing data files from four states..... 94
 - Part 2: Problems in reading in the data 98
 - Part 3: Recommendation 99
 - Attachment 1: List of data items on New South Wales public housing tenants and properties..... 101
- Appendix 2: The SAS code for reading raw data and producing the final merged nationally consistent data..... 104

Summary of findings

This report documents the work undertaken by the Australian Institute of Health and Welfare (AIHW) to develop data standards for use in building a national data set for Commonwealth-State Housing Agreement (CSHA) public housing. Currently, nationally consistent data is difficult to obtain as individual State and Territory data differs in terms of the counting rules, definitions and classifications used. The Department of Social Security (DSS) funded the AIHW to examine current data structures and prepare draft specification of a national data set to assist in future data development as well as developing software to allow common data to be derived from the data DSS currently holds.

This work has implications in the context of the long term development of data for the CSHA as it provides a starting point for discussion on issues such as data items and definitions and the structure of data required. This work compliments the CSHA performance indicator data development undertaken by the CSHA Performance Indicator Working Group. The compilation of national performance indicator data is currently based on aggregated State/Territory data while this report looks at data options for creating a national unit record data set.

This report summarises the main work of the project in developing meta-data and software to assist DSS in deriving nationally consistent unit record data for CSHA public housing.

This work is reported in three sections:

- | | |
|-----------|---|
| Section 1 | Summary of the project |
| Section 2 | Data manual (data dictionary) – defining the core set of variables identified in the project which could form the basis for the technical documentation for future development of a national public housing data set. |
| Section 3 | User’s guide – documenting the data that is produced from the use of the merging software on the 1997 data for four states. |

The two appendixes contain information on the input data sets (Appendix 1) and the SAS code for reading raw data and producing the final merged nationally consistent data based on the input data from the four states for which data was examined (Appendix 2).

The design of the data structure described in the report was based on two important areas of use identified by DSS:

- for analysis of the circumstances of public housing tenants in a uniform manner at both the jurisdiction and national level; and
- to examining the circumstances of public housing tenants in relation to their eligibility for DSS income support assistance.

This multiple use led to a data structure that accommodated both an emphasis on dwelling and household level data as well as a focus on persons and income units within households.

To capture these two types of data use a four level structure was proposed comprising person, income unit, household and dwelling levels. While this approach may appear complex several of the data items are similar and the levels build on each other to form consistent set of data at the person, income unit, household and dwelling levels. It presents an ideal data structure and would require further development and refinement to implement

across jurisdictions in terms of minimising data collection burden on providers. The relative importance of the four levels would change as the priority for use changed.

Following discussions with DSS and analysis of the data, a set of data items were identified as the core set of items that were to be developed for the project. These items covered:

- Market rent value
- Rent charged
- Rent subsidy
- Number of bedrooms
- Dwelling type
- Age of dwelling
- Location of dwelling – Postcode, Local Government Area, State or Territory
- Total weekly income (for each household, income unit and person in the dwelling)
- Principle source of income (for each household, income unit and person in the dwelling)
- Length of tenancy

Some of these data items were only required at one level such as dwelling while others were required for several levels such as total income and sources. In addition to the above data several characteristics that related the person, income unit, household and dwelling data to each other were also specified including:

Number of households in dwelling

Number of income units (in dwelling and household)

- Household type
- Household size
- Age and sex of each person and the household primary tenant
- Income unit size
- Number of Dependent children
- Rent charged to the income unit

The use of data standards for unit record data in the future will be determined by the role data plays in the next CSHA.

This project represents a useful first step toward identifying the formats and for key variables for public housing that could be used to produce national unit record data. Such a data set could be useful in several areas of the CSHA information reporting such as performance reporting, accountability and strategic reporting. The development of a single data set could also play an important role in reducing the duplication of reporting across these requirements thus reducing the response burden of data provision.

The Institute recommends that this work is further developed in the context of future CSHA data requirements and this unit record approach is linked to areas where aggregated data is used to ensure consistency. Standards should be developed for record structures, formatting conventions, data items and definitions that all stakeholders could usefully adopt to improve data quality and consistency.

Section 1: Summary of the project

" Exchange of any form of information, to be effective, must take place in an environment where it can be ensured that the receiver interprets the information in exactly the same way as intended by the sender. The information must also be easy to locate and retrieve. This is only possible where the meaning and method of representation of the information are known and agreed upon by the communication partners. " (from the description of meta-data contained in ISO 11179: Specification and Standardization of Data Elements)

1.1 Background

Recent focus on reforms to public housing and a greater emphasis on outcomes and accountability in the current CSHA have led to an increased need for improved national data on public housing. The more detailed examination of policy and program issues in this environment has illustrated the need for administrative data to provide information that is relevant, of good quality and timely.

In relation to unit record data, little work has been undertaken to date on a national scale to develop data standards for unit record administrative data for housing assistance. While standards have been developed by the ABS for Population censuses and surveys and also by the AIHW for performance reporting of CSHA programs they have not examined in detail how a consistent unit record data set for public housing assistance could be specified.

Currently unit record data are used by each State/Territory for national reporting for accountability, performance monitoring and strategic plans. The unit record data sources used may not be strictly comparable as no national standards for unit record data have been developed. Some unit record data is also provided to the Commonwealth under the CSHA Data Exchange Agreement which has as yet only been undertaken for the financial year 1996-97.

The issue of a lack of national standards for unit record data was highlighted when the Department of Social Security analysed the data obtained from several jurisdictions as part of their Data Exchange Agreement. The Department identified significant issues with the data that affected their ability to access and construct a coherent national picture. To facilitate the process of moving toward improved data quality and consistency and improving the availability of national data on housing assistance DSS contracted the AIHW to develop meta-data and software that would contribute to improved information.

Both DSS and the AIHW recognise there are a range of issues that still need to be addressed before moving toward improved national data and see this project as a first step in presenting options that could be further developed with other stakeholders to achieve improved data coherence.

1.2 Project objective

The purpose of the data standardisation project was to develop a system of meta-data and software to allow DSS to efficiently standardise CSHA public housing unit record data.

The objectives were:

- to report on the data issues to be resolved in receiving 'raw' input data based on the experience of DSS with data from the first CSHA Data Exchange Agreement;

- to propose a data structure, definitions and code values for the production of nationally consistent unit record data for CSHA public housing (this data structure, definitions and code values is often referred to as meta-data);
- to implement as far as practicable this data structure, definitions and code values on existing data through the development of SAS based software that DSS could use to derive common data form the current input data sets it held from the CSHA Data Exchange Agreement with jurisdictions;
- this system of meta-data and software should allow DSS to create the standardised data file on DSS mainframe/local system, for the four States covered by the project; and
- to report on these issues to promote discussion on longer term strategic directions for CSHA minimum data sets.

1.3 The data used in the project

The raw input data provided to the AIHW was based on the data files obtained by DSS as part of its data exchange agreement with the relevant State housing authority. This data was supplied to DSS as a unit record file of public housing tenants or dwellings at 30 June 1997. These data were only used for the testing and development of the meta-data and software and were returned to DSS following the completion of the project. The data provided was that for the four States of New South Wales, Victoria, Queensland and South Australia.

1.4 Project outline

The project involved the AIHW developing a system that would enable DSS to take the raw data as received from a jurisdiction and convert it to a standard format. The development of a system of meta-data and software to enable DSS to create the standardised data file involved the following stages:

- examination of input data;
- development of standard data definitions and data structure; and
- production of software to enable uniform data to be created from the current data supplied to DSS.

1. Examination of input data

This stage of the project involved examining the data that DSS was currently using from jurisdictions and evaluating how this could be used to develop a common data set. Analysis of the data based on the structure of the input data sets supplied by jurisdictions was undertaken. This involved examining any technical documentation supplied with the data, accessing the data supplied to evaluate how it may have differed from the documentation and using check tables and record counts to verify the data.

This work identified a series of issues that were relevant to improving the quality and consistency of the input data that jurisdictions supplied. As expected the data structure of public housing data files varied across States and Territories in terms of the way data was held, the quality of the data and the way data was documented and defined. Developing a national public housing unit record database would obviously be much easier if a standard or common data form was sent by each state. The major issues identified at this stage of the project were:

- the need to standardise the data items in the State databases sent to DSS for a core or minimum set of variables for which national uniformity was a high priority;
- the need for standard documentation practices, where States were not able to send data in the same form. This documentation should describe the files provided. The development of a pro forma for documentation for each state to follow is highly recommended.

The detailed discussion of these issues is contained in Attachment 1. The need to standardise the data items held in state databases that form a core or minimum data set for national uniformity led to the major work of the project namely the development of a data manual or data dictionary.

2. Development of standard data definitions and data structure

In standardising the data items both for the current data and for any future national data collection activity the identification of key variables and their structure is a major task. For the project the development of uniform data was undertaken in consultation with DSS to cover three aspects:

- the identification of a core set of data items and a data structure that would meet DSS needs; and
- development of logical or ideal meta-data (data structure, variables and classifications) that would produce these data.

Following discussions with DSS regarding the quality of the current data and the short and long term use of such data several decisions were made on:

- the structure of a uniform data set;
- the level of detail required and core data items; and
- the appropriate classifications.

It was recognised by DSS and the Institute that this data development work was entering new ground and any proposals should be considered provisional as far wider discussion was required.

The identification of a core set of data items and a data structure

Currently there is a lack of reliable data at the national level on public housing. ABS Census and surveys often do not provide the level of detail required for policy and program analysis while the data provided for the Housing Assistance Act and for the Report on Government Service Provision is aggregate jurisdiction level information.

The design of the data structure was based on two important areas of use of this data identified by DSS:

- the analysis and modelling of the circumstances of public housing tenants in a uniform manner at both the jurisdiction and national level – a CSHA based use with emphasis on dwelling and tenant/household level data; and
- examining the circumstances of public housing tenants in relation to their eligibility for DSS income support assistance and in relation to assistance provided to private tenants through Rent Assistance – a DSS client approach with a broader focus on persons, income units and households.

These two quite different purposes – a dwelling/household approach and an income unit/person approach - led to the structure proposed in this report. To capture these two

types of data uses a four level structure was proposed comprising person, income unit, household and dwelling levels. This approach may appear complex but several of the data items are similar data items that build on each other to form consistent data at the person, income unit, household and dwelling levels. In addition to produce reliable data at the household level for variables such as household income or composition lower level data about persons or income units is required to be collected.

This design was seen as presenting an ideal data structure to stimulate discussion and indicate further development and refinement. This is particularly important in terms of minimising data collection burden on providers.

Developing the logical or ideal meta-data that would produce these data

As noted above, four levels were identified as a useful structure for standardised data sets. These were:

<i>Dwelling</i>	Containing information relevant to dwellings.
<i>Household</i>	Containing information relevant to households within dwellings.
<i>Income unit</i>	Containing information relevant to income units within households.
<i>Person</i>	Containing information relevant to persons within income units.

These four levels contain a core set of data that is consistent - that is household income is the sum of income unit income which in turn is the sum of person income. The structure and data items for each of the four data sets are defined in Section 2. In this project the development of classifications was wherever possible based on existing national standards. However for some data items these classifications are still under development or require some adaptation to fit in the context of public housing.

The data items identified by DSS as the starting point in specifying a national core data set are shown in Table 1. Using this list of draft data item specifications describing the derived data items were developed and discussed with DSS. Following clarification of definitions involved the final specification was developed for these data items. These are presented in Section 2. Based on this specification the currently available data was investigated to determine levels at which data sets can be produced and how different level data could be generated to give household, income unit and tenant level information.

3. Producing software to allow DSS to create uniform data from the currently available input data sets

Based on the data currently available to DSS it is not possible to produce a single national level unit record data set for CSHA public housing dwellings as:

- data is only present for four jurisdictions;
- documentation of the data was insufficient to identify the structure, contents and quality of the input data;
- the quality of the data supplied was often poor with logical editing identifying several inconsistencies in basic data items relating to household size, income and family type;
- the structure of the data varied with some data sets containing income unit data while others contained only dwelling level data;
- the coverage of dwellings is not consistent between jurisdictions in areas such as the treatment of vacant dwellings, head-leased dwellings, stock used for non-

accommodation purposes within public housing and stock allocated to other CSHA programs such as CAP or ARHP; and

- definitions used to describe the basic data items such as dwellings, households, tenants, family type, market rent and income are not consistent.

All these issues present challenges to produce consistent data and this stage of the project involved developing and testing the software to enable the merging of data sets and the derivation and recoding of data to generate a standardised data file structure. It also involved documenting major aspects of this work including the quality of the data produced and how it may be improved. The specifications describing the derived data items, the classification used and notes on the likely quality of the data are contained in Section 3 of the Report.

Table 1: Data set record content

Dwelling data	Household data	Income unit data	Person data
Housing program to which dwelling is assigned	Housing program under which household is assisted.	Income unit type	Relationship within household
Number of households	Number of income units	Income unit size	Relationship within income unit
Market rent value for dwelling	Household type	Dependent children in income unit	Age of person
Rent charged to dwelling	Household size	Other income unit members	Sex of person
Rent subsidy for dwelling	Dependent children in household	Total income unit weekly income	Total person weekly income
Vacancy flag	Other household members	Income unit principle source of income	Person principle source of income
Number of bedrooms	Total household weekly income	Market rent value for the income unit	
Dwelling type	Principle source of income for household	Rent charged to the income unit	
Age of dwelling	Market rent value for household	Rent subsidy	
Postcode	Rent charged to household	Age of income unit reference person	
Local Government Area	Rent subsidy	Sex of income unit reference person	
State or Territory	Rent subsidy/rebate flag	Income unit length of tenancy	
	Age of household primary tenant		
	Sex of household primary tenant		
	Household length of tenancy		

1.5 Future development of a national CSHA unit record data set

This project represents a useful first step toward identifying the formats and key variables for public housing that could be used to produce national unit record data. Such a data set could be useful in several areas of the CSHA information reporting such as performance reporting, accountability and strategic reporting. The development of a single data set could also play an important role in reducing the duplication of reporting across these requirements thus reducing the response burden of data provision.

The Institute recommends that this work is further developed in the context of future CSHA data requirements and this unit record approach is linked to areas where aggregated data is used to ensure consistency. Standards should be developed for record structures, formatting conventions, data items and definitions that all stakeholders could usefully adopt to improve data quality and consistency.

Section 2: Specifications of standardised data

Specifications of the standardised data sets are given below. Each jurisdiction should be requested to provide the data as described if possible. The data should be provided as four data sets with jurisdictions providing the required linking identifiers.

Data Sets

The data items to be included in the four required data sets are described below. It should be noted that some variables, in particular identifiers, need to be included at a number of levels so that records from different levels can be linked.

The four standardised data sets are:

<i>Dwelling</i>	Containing information relevant to dwellings.
<i>Household</i>	Containing information relevant to households within dwellings. Each occupied dwelling should have at least one household record.
<i>Income unit</i>	Containing information relevant to income units within households. Each occupied dwelling should have at least one income unit record.
<i>Person</i>	Containing information relevant to persons within income units. Each occupied dwelling should have at least one person record.

Three identifiers, D_ID, H_ID, and I_ID, are used to link the four data sets.

Jurisdictions should notify DSS if dwelling and household data are not available separately.

Data items for each of the four data sets are defined below. In a number of cases item definitions refer to concepts, such as *household* or *dependent child*. These concepts are written in italics within the item descriptions. Definitions for these are found in the Glossary.

Missing data

For each level a range of data items are available. However, for many variables complete information may not be available for all records. In general, there are two types of missing data:

- Cases where, in general, a jurisdiction's data does not contain the relevant information for derivation of the required variable. This may be true for all or a subset of tenants/dwellings.
- Individual cases where the data is not stated or unknown, but where, in general, a jurisdiction's data contains the relevant information for derivation of the required variable.

These two types of missing data are differentiated and coded correspondingly as thus:

99998	Relevant data not available from the jurisdiction
99999	Unknown.

2.1 Data set record content and formats

Dwelling data set:

LEVEL	Level identifier	A6
D_ID	Dwelling identifier	F12
D_PGM	Program to which dwelling is assigned	A5
D_NHHLD:	Number of households	F5
D_MRENT	Market rent value for dwelling	F5
D_RCHARG	Rent charged to dwelling	F8.2
D_RSUB	Rent subsidy for dwelling	F8.2
D_VAC:	Vacancy flag	F5
D_BED:	Number of bedrooms	F5
D_TYPE:	Dwelling type	F5
D_AGE:	Age of dwelling	F5
D_PCODE:	Postcode	F5
D_LGA:	Local Government Area	F5
D_STATE:	State or Territory	F5

Household data set:

LEVEL	Level identifier	A6
D_ID	Dwelling identifier	F12
H_ID:	Household identifier	F10
H_PGM	Program under which household is assisted.	A5
H_QUAL:	Household data quality identifier	F5
H_NUNIT	Number of income units	F5
H_TYPE:	Household type	F5
H_SIZE:	Household size	F5
H_DEPT:	Dependent children in household	F5
H_OTH:	Other household members	F5
H_TOTINC:	Total household weekly income	F8.2
H_SOURCE:	Principle source of income for household	A5
H_MRENT	Market rent value for household	F5
H_RCHARG	Rent charged to household	F8.2
H_RSUB	Rent subsidy	F8.2
H_REB	Rent subsidy/rebate flag	F5
H_PAGE:	Age of household primary tenant	F5
H_PSEX:	Sex of household primary tenant	F5
H_LENGTH:	Household length of tenancy	F5

Income unit data set:

LEVEL	Level identifier	A6
D_ID	Dwelling identifier	F12
H_ID:	Household identifier	F10
I_ID:	Income unit identifier	F10
I_TYPE:	Income unit type	F5
I_SIZE:	Income unit size	F5
I_DEPT:	Dependent children in income unit	F5
I_OTH:	Other income unit members	F5
I_TOTINC:	Total income unit weekly income	F8.2
I_SOURCE:	Income unit principle source of income	A5
I_MRENT	Market rent value for the income unit	F5
I_RCHARG	Rent charged to the income unit	F8.2
I_RSUB	Rent subsidy	F8.2
I_PAGE:	Age of income unit reference person	F5
I_PSEX:	Sex of income unit reference person	F5
I_LENGTH:	Income unit length of tenancy	F5

Person data set:

LEVEL	Level identifier	A6
D_ID	Dwelling identifier	F12
H_ID:	Household identifier	F10
I_ID	Income unit identifier	F10
P_ID:	Person identifier	F10
P_RELH:	Relationship within household	F5
P_RELI:	Relationship within income unit	A5
P_PAGE:	Age of person	F5
P_PSEX:	Sex of person	F5
P_TOTINC:	Total person weekly income	F8.2
P_SOURCE:	Person principle source of income	A5

2.2.2 Dwelling, or property, level items

There are 13 data items which relate to dwellings.

Dwelling data items list

D_ID	Dwelling identifier
D_PGM	Program to which dwelling is assigned
D_NHHLD:	Number of households
D_MRENT	Market rent value for dwelling
D_RCHARG	Rent charged to dwelling
D_RSUB	Rent subsidy for dwelling
D_VAC:	Vacancy flag
D_BED:	Number of bedrooms
D_TYPE:	Dwelling type
D_AGE:	Age of dwelling
D_PCODE:	Postcode
D_LGA:	Local Government Area
D_STATE:	State or Territory

Dwelling data items details

D_ID	Dwelling identifier
<i>Name on merged file</i>	D_ID
<i>Data Type:</i>	Character <i>Print format</i> A12
<i>Field size:</i>	Min: 4 Max: 12 <i>Format on file</i> A12
<i>Purpose:</i>	This variable is used to identify individual <i>dwelling</i> s. The use of such identification is twofold: <ul style="list-style-type: none"> • It allows for follow-up if queries arise concerning data values. Such follow-up may either be through querying the data set as supplied or through returning to the states for clarification. • If there are multiple households in a dwelling this variable is used to relate households to dwellings.
<i>Definition:</i>	The dwelling identifier is a code which uniquely identifies each <i>dwelling</i> .
<i>Classification</i>	State/Territory identifier three digit alpha code – AAA – followed by the identifying code – NNNNNNNNNNNN – as provided by jurisdictions. This should never be ‘missing’ or ‘unknown’. State/Territory identifiers are: NSW New South Wales VIC Victoria QLD Queensland WA Western Australia SA South Australia TAS Tasmania ACT Australian Capital Territory NT Northern Territory
<i>Note:</i>	The State/Territory identifying prefixes avoid the problem of households and income units being associated with the wrong dwelling, and allows jurisdictions to use their own numbering systems.
<i>Validity checks:</i>	Must be unique within a State/Territory.
<i>Level:</i>	Dwelling.

D_PGM	Program to which dwelling is assigned			
<i>Name on merged file</i>	D_PGM			
<i>Data Type:</i>	Character		<i>Print format</i>	A5
<i>Field size:</i>	<i>Min:</i> 3	<i>Max:</i> 5	<i>Format on file</i>	A5
<i>Purpose:</i>	This variable is used to identify dwellings in particular programs to ensure that only dwellings in the program(s) of interest are included in an analysis.			
<i>Definition:</i>	This variable identifies the assistance program with which a <i>dwelling</i> is associated. The dwelling may be either owned by a state housing authority or head-leased from the private rental market.			
<i>Classification</i>	GEN	General public housing (including housing for pensioners)		
	ARHP	Aboriginal Rental Housing Program		
	CHP	Community Housing Program		
	CAP	Crisis Accommodation Program		
	OTH	Other, including state-specific housing programs		
	99998	Program cannot be determined (used when, in general, certain programs cannot be differentiated)		
	99999	Unknown (used when the program for a particular dwelling is not known)		
<i>Validity checks:</i>	Values must be restricted to those listed above. Check rent and bedroom ranges for those identified as GEN.			
<i>Level:</i>	Dwelling.			

D_NHHLD: Number of households

Name on merged file D_NHHLD

Data Type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: The purpose of this variable is to identify the number of households, or tenancy agreements, in the dwelling. It will allow analysis to be carried out by number of households in a dwelling. It also facilitates reading in data.

Definition: This variable indicates the number of *households* in a *dwelling*.

Classification

0	Vacant dwelling
Number, no decimal places.	
99997	More than one, but exact number is not known
99998	Whether there is only one, or more than one cannot be determined
99999	Unknown (arising if variables from which this is derived is missing)

Validity checks: Compare program code with number of households. GEN and ARHP dwellings are unlikely to have more than 4 households per dwelling unless group households are treated as a group of households.

Level: Dwelling.

D_MRENT Market rent value for dwelling

Name on merged file D_MRENT

Data Type: Numeric *Print format* \$F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: This variable is used to determine the existence and value of rent rebates.

Definition: The market rent value of the dwelling is the rent the dwelling would receive if it were in the private rental market.

Classification \$ per week, value to be shown to nearest cent

99997 Market rent not relevant, for example for untenable dwellings

99998 Relevant data not available from the jurisdiction

99999 Unknown.

Note: The method of calculating market rent varies with jurisdiction; for example independent valuations, as a percent of capital value, and from newspapers. The method used should be noted by the jurisdiction.

Validity checks: Rents for GEN and ARHP dwellings should be less than \$300.00 per week. For multiple household dwellings this item should be the sum of the market rents (H_MRENT) for contributing households.

Level: Dwelling.

D_RSUB **Rent subsidy for dwelling**

Name on merged file D_RSUB

Data Type: Numeric *Print format* \$F8.2

Field size: *Min:* 1 *Max:* 8 *Format on file* F8.2

Purpose: This variable is the total value of rent rebates/subsidies received by inhabitants of the dwelling. It is used to examine the distribution of subsidies and the effect of subsidies on housing affordability.

Definition: This variable is derived as the difference between dwelling market rent and rent charged to the dwelling, that is

$$D_MRENT - D_RCHARG.$$

Classification 0 No rent subsidy
\$.00 per week, value to be shown to nearest cent
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Validity checks: $0 \leq D_RSUB \leq D_MRENT$
For multiple household dwellings this item should be the sum of the rent rebates (H_RSUB) received by contributing households.

Level: Dwelling.

D_VAC: **Vacancy flag**

Name on merged file D_VAC

Data Type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: This variable allows vacant and untenable dwellings to be excluded from analyses as required.

Definition: This variable identifies whether or not a dwelling is occupied, vacant or not tenable.

Classification 1 Occupied
2 *Tenable and vacant*
3 *Untenable and vacant*
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: Jurisdictions should note if they cannot use the above classification, or if they have only included certain types of dwellings in the data set.

Validity checks: $D_RCHARG = 0$ if D_VAC is 2 or 3

Level: Dwelling.

D_BED: Number of bedrooms

Name on merged file D_BED

Data Type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: This variable is used to allow analysis by dwelling size and to examine crowding.

Definition: This variable is the count of the bedrooms in each dwelling.

Classification 0 None (includes bedsitters)

Number, no decimal places, giving the number of distinct bedrooms.

99998 Relevant data not available from the jurisdiction

99999 Unknown.

Note: The number of bedrooms is based on rooms whose original purpose were as bedrooms and not as they are currently used. Exceptions to this should be noted.

Validity checks: Dwellings in GEN and ARHP generally have less than 5 bedrooms.

Level: Dwelling.

D_TYPE: Dwelling type

Name on merged file D_TYPE

Data Type: Numeric *Print format* F5

Field size: Min: 1 Max: 5 *Format on file* F5

Purpose: This variable allows analysis by dwelling type.

Definition: This variable identifies the structure of private dwellings.

Classification

1	Detached house
2	Semi-detached house, townhouse, terrace house, duplex
3	Flat, unit, apartment
4	Boarding house, hostel
5	Other (that is, none of the above, and includes caravans and movable units)
99997	Not known (includes dwellings whose original dwelling type code is based on something other than structure, for example a program)
99998	Relevant data not available from the jurisdiction
99999	Unknown (missing/not stated).

(Source: DSS specified)

Note: Dwellings included in category 99997 should be described.

Validity checks: Codes should be in above range. Dwellings of program GEN (and ARHP) should not have code 4.

Level: Dwelling.

D_AGE: Age of dwelling

Name on merged file D_AGE

Data Type: Numeric *Print format* F5

Field size: Min: 1 Max: 5 *Format on file* F5

Purpose: This variable allows analysis by age of dwelling stock.

Definition: Age of dwelling is measured in terms of the number of completed months since the dwelling was built.

Classification

Number of months, no decimal places	
99998	Relevant data not available from the jurisdiction
99999	Unknown.

Validity checks: Non-missing values should be between 0 and 600.

Level: Dwelling.

D_PCODE: Postcode

Name on merged file **D_PCODE**

Data Type: Numeric *Print format* F5

Field size: *Min:* 4 *Max:* 5 *Format on file* F5

Purpose: This variable allows analysis by location in terms of postcode.

Definition: This variable indicates the postcode in which the dwelling is located.

Classification Four digit numeric code

99998 Relevant data not available from the jurisdiction

99999 Unknown.

Note: Where postcodes have changed over time, the existence of postcodes that have not been updated should be noted.

Validity checks: These values should correspond to official postcodes.

Level: Dwelling.

D_LGA: Local Government Area

Name on merged file **D_LGA**

Data Type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: This variable allows analysis by location in terms of local government areas.

Definition: This variable indicates the Local Government Area in which the dwelling is located.

Classification As provided by jurisdictions.

99998 Relevant data not available from the jurisdiction

99999 Unknown.

Note: Where LGAs have changed over time, the existence of LGA codes that have not been updated should be noted. States/Territories should provide a format list to associate codes with names.

Validity checks: Each LGA code should have a corresponding name. Missing names suggest problems with the data, for example that LGA codes have not been updated as LGAs have changed boundaries/names.

Level: Dwelling.

D_STATE: State or Territory

Name on merged file D_STATE

Data Type: Character *Format on file* A1

Field size: Min: 1 Max: 1 *Print format* A1

Purpose: This variable allows analysis by location in terms of states and territories. It also facilitates for follow-up if data queries arise that require clarification by a State/Territory.

Definition: This variable indicates the State/Territory in which the dwelling is located.

Classification

1	New South Wales
2	Victoria
3	Queensland
4	Western Australia
5	South Australia
6	Tasmania
7	Australian Capital Territory
8	Northern Territory

This variable should not be missing.

(Source: order as used in CSHA Performance Indicator manuals)

Note: This order is different from that used in the ABS Census 1996 Data Dictionary, in which SA and WA are reversed.

Validity checks: Counts of dwellings by State/Territory should be compared to previously reported numbers.

Level: Dwelling.

2.2.3 Household level items

There are 18 data items specified at the household level.

Household data items list

D_ID	Dwelling identifier
H_ID:	Household identifier
H_PGM	Program under which household is assisted.
H_QUAL:	Household data quality identifier
H_NUNIT:	Number of income units
H_TYPE:	Household type
H_SIZE:	Household size
H_DEPT:	Dependent children in household
H_OTH:	Other household members
H_TOTINC:	Total household weekly income
H_SOURCE:	Principle source of income for household
H_MRENT	Market rent value for household
H_RCHARG	Rent charged to household
H_RSUB	Rent subsidy
H_REB	Rent subsidy/rebate flag
H_PAGE:	Age of household primary tenant
H_PSEX:	Sex of household primary tenant
H_LENGTH:	Household length of tenancy

Household data items details

H_ID: Household identifier

Name on merged file H_ID

Data Type: Numeric *Format on file* F10

Field size: *Min:* 1 *Max:* 10 *Print format* F10

Purpose: This variable is used to identify individual *households*. The use of such identification is twofold:

- It allows for follow-up if queries arise concerning data values. Such follow-up may either be through querying the data set as supplied or through returning to the states for clarification.
- If there are multiple income units in a household this variable is used to relate income units to households.

Definition: The *household* identifier is a code which uniquely identifies each *household*.

Classification As provided by jurisdictions. This should not be missing.

Validity checks: This should not be missing.

Level: Household.

H_PGM Program under which household is assisted.

Name on merged file H_PGM

Data Type: Character *Print format* A5

Field size: *Min:* 1 *Max:* 5 *Format on file* A5

Purpose: This variable is used to identify households in particular programs to ensure that only households in the program(s) of interest are included in an analysis.

Definition: This variable identifies the assistance program through which a household is assisted.

Classification

GEN	General public housing (including housing for pensioners)
ARHP	Aboriginal Rental Housing Program
CHP	Community Housing Program
CAP	Crisis Accommodation Program
OTH	Other, including state-specific housing programs
99998	Relevant data not available from the jurisdiction
99999	Unknown.

Validity checks: Values must be restricted to those listed above. Check rent range for those identified as GEN.

Level: Household.

H_QUAL: Household data quality identifier

Name on merged file H_QUAL

Data Type: Numeric Print format F5

Field size: Min: 1 Max: 5 Format on file F5

Purpose: This variable is used to identify those households which have (reasonably) up-to-date data.

Definition: Households for which data have been updated or collected within the last twelve months are said to have up-to-date information.

Classification

0	Not updated within the last 12 months
1	Updated within the last 12 months
99998	Relevant data not available from the jurisdiction
99999	Unknown.

Note: Most jurisdictions do not obtain up-to-date information on households not receiving a rent rebate/subsidy.

It is assumed that the data for ALL household members/income units are updated at the same time. Jurisdictions should indicate if this is not so.

Validity checks: If H_REB=1 this variable should have a value of 1.

Level: Household.

H_NUNIT: Number of income units

Name on merged file H_NUNIT

Data Type: Numeric Print format F5

Field size: Min: 1 Max: 5 Format on file F5

Purpose: The purpose of this variable is to identify the number of income units in the household. Until complete income unit data is available this variable will allow analysis to be carried out on single income unit households only, if desired. It also facilitates reading in data.

Definition: This variable indicates the number of *income units* in a household.

Classification

Number, no decimal places.

If the number of income units is unknown or cannot be determined the following codes should be used:

99997	More than one
99998	Relevant data not available from the jurisdiction
99999	Unknown.

Validity checks: This variable should be of value 1 or more.

Level: Household.

H_TYPE: Household type

Name on merged file H_TYPE

Data Type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: This variable allows analysis to be undertaken by household type

Definition: This variable describes the type of household. Households can contain dependent and non-dependent children as well as non-family members.

Classification

- 1 Person living alone
- 2 Couple only
- 3 Couple with dependent children only
- 4 Couple with non-dependent children (with or without dependent children)
- 5 Sole parent with dependent children only
- 6 Sole parent with non-dependent children (with or without dependent children)
- 7 Group (unrelated adults)
- 8 Other, including multiple family
- 99998 Relevant data not available from the jurisdiction
- 99999 Unknown.

Note: If jurisdictions cannot, for example, differentiate between households with dependent children only and those with non-dependent children, a two tier classification is suggested. Otherwise there may a large number of households classified to 'Other'.

Level: Household.

H_SIZE: Household size

Name on merged file H_SIZE

Data Type: Numeric Print format F5

Field size: Min: 1 Max: 5 Format on file F5

Purpose: This variable allows for analysis by household size. It is also useful for examining crowding or for making adjustments for household composition when trying to compare the circumstances of different households.

Definition: This variable gives the total count of people in the household.

Classification 0 An 'empty' household
Number, no decimal places.
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Validity checks: This variable is the sum of all people in contributing income units. If the household is in GEN, household size is generally less than 10.

Level: Household.

H_DEPT: Dependent children in household

Name on merged file H_DEPT

Data Type: Numeric Print format F5

Field size: Min: 1 Max: 5 Format on file F5

Purpose: This variable allows for analysis according to the presence or absence of dependent children. It is also useful for examining crowding and for making adjustments for household composition when trying to compare the circumstances of different households.

Definition: This variable gives the total count of *dependent children* in the household.

Classification 0 No dependent children in the household
Number, no decimal places.
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: Not all jurisdictions identify dependent children the same way. Jurisdictions should note where a definition different to that given in the Glossary is used.

Validity checks: If the household is in GEN, household size is generally less than 10.

Level: Household.

H_OTH: Other household members

Name on merged file H_OTH

Data Type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: This variable allows for analysis according to the presence or absence of 'other' household members. It is also useful for examining crowding and for making adjustments for household composition when trying to compare the circumstances of different households.

Definition: This variable gives the total count of people in the household who are neither the *household primary tenant* nor their partner/spouse nor a *dependent child*. Such people include non-dependent children, other relatives and unrelated household members. Dependent children of secondary income units, for example the children of a single mother living with her parents, are counted as *dependent children*.

Classification 0 No 'other' people in the household
Number, no decimal places.
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: Not all jurisdictions identify dependent children the same way. Jurisdictions should note where a definition different to that given in the Glossary is used.

Validity checks: This variable is less than (H_SIZE - H_DEPT). If the household is in GEN, household size is generally less than 10.

Level: Household.

H_TOTINC: Total household weekly income

Name on merged file H_TOTINC

Data Type: Numeric *Print format* \$F8.2

Field size: *Min:* 1 *Max:* 8 *Format on file* F8.2

Purpose: Household income is used to examine the affordability of housing costs.

Definition: This variable is the sum of the personal *gross weekly incomes* of each member of the household.

Classification

0	No income
\$.00	Dollar amount per week (to nearest cent)
99998	Relevant data not available from the jurisdiction
99999	Unknown.

Note: This variable should include all income, rather than what is considered 'assessable' by jurisdictions. Jurisdictions should note if some income is not included.

Validity checks: If any income unit or household member gets some income, this variable should be greater than zero, unless income is missing for a household member.

Level: Household.

H_SOURCE: Principle source of income for household

Name on merged file H_SOURCE

Data Type: Character *Print format* A5

Field size: Min: 1 Max: 5 *Format on file* A5

Purpose: This variable allows analysis by source of income.

Definition: This variable identifies the main source of income for a household. It is derived by summing individual household member's income by their main source of income. The household's main source of income is then that source of income with the largest (non-missing) dollar value.

Classification

- N No income
- G Government allowances and benefits (including DSS and DVA payments)
- SE Self-employed, own business
- WS Wages and salaries
- OTH Neither government allowances or benefits nor wages and salaries. This includes, for example, income from investments and superannuation.
- 99998 Relevant data not available from the jurisdiction
- 99999 Unknown.

Note: Jurisdictions should indicate if this information is not available. For a household who is paying full rent this should be 99999 not 'N'.

Validity checks: The majority of households should be classified as 'G'.

Level: Household.

H_MRENT Market rent value for household

Name on merged file H_MRENT

Data Type: Numeric *Print format* \$F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: This variable is used to determine the existence and value of rent rebates.

Definition: The market rent for the household is that portion of the dwelling market rent with which the household is associated.

Classification \$ Dollar amount per week (to nearest cent)
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: A method is required for allocating market rent across contributing households.

Validity checks: Rents for GEN and ARHP households should be less than \$300.00 per week.

For multiple income unit households this item should be the sum of market rents (I_MRENT) associated with contributing income units.

Level: Household.

H_RCHARG Rent charged to household

Name on merged file H_RCHARG

Data Type: Numeric *Print format* \$F8.2

Field size: *Min:* 1 *Max:* 6 *Format on file* F8.2

Purpose: This variable is used to determine the existence and value of rent rebates, and to examine housing affordability.

Definition: The rent charged is the actual rent a household is expected to pay after any eligibility for rent rebates/subsidies have been included. The rent charged may or may not have been received. This item reflects the expected and not the actual rent paid as defaults and arrears may reduce or increase the amount received compared to the amount charged.

Classification 0 No rent charged
\$.00 Dollar amount per week (to nearest cent)
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Validity checks: $0 \leq H_RCHARG \leq H_MRENT$.

For a vacant dwelling this should be zero.

For multiple income unit households this item should be the sum of the rent charged (I_RCHARG) charged contributing income units

Level: Household.

H_RSUB Rent subsidy*Name on merged file*

H_RSUB

Data Type:

Numeric

Print format

\$F8.2

*Field size:**Min:* 1*Max:* 8*Format on file*

F8.2

Purpose:

This variable is the value of rent rebates/subsidies received by a household. It is used to examine the distribution of subsidies and the effect of subsidies on housing affordability.

Definition:

This variable is derived as the difference between market rent and rent charged, that is

$$H_MRENT - H_RCHARG.$$

Classification

\$.00

Dollar amount per week (to nearest cent)

99998

Relevant data not available from the jurisdiction

99999

Unknown.

Validity checks:

$$0 \leq H_RSUB \leq H_MRENT$$

For multiple income unit households this item should be the sum of the rent rebates (I_RSUB) received by contributing income units.

Level:

Household.

H_REB Rent subsidy/rebate flag*Name on merged file*

H_REB

Data Type:

Numeric

Print format

F5

*Field size:**Min:* 1*Max:* 5*Format on file*

F5

Purpose:

This variable is used to identify easily those households receiving rent subsidies or rebates.

Definition:

This variable indicates whether or not a household gets a rent subsidy. It is derived from H_RSUB

Classification

0

No subsidy/rebate

1

Receives subsidy/rebate

99998

Relevant data not available from the jurisdiction

99999

Unknown.

Validity checks:

$$\text{If } H_RSUB = 0 \text{ then } H_REB = 0$$

$$\text{If } 0 < H_RSUB < 99998 \text{ then } H_REB = 1$$

$$\text{If } H_RSUB = 99998 \text{ then } H_REB = 99998$$

$$\text{If } H_RSUB = 99999 \text{ then } H_REB = 99999$$

Level:

Household.

H_PAGE: Age of household primary tenant

Name on merged file H_PAGE

Data Type: Numeric Print format F5

Field size: Min: 1 Max: 5 Format on file F5

Purpose: To allow analysis by a tenant age indicator.

Definition: Age in years of the *household primary tenant*. Only completed years are recorded.

Classification Number of completed years (no decimal places)
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: When derive this from date of birth make sure this is coded as 99999 for unknown date of birth.

Validity checks: This variable should generally be greater than 15. Distributions should be checked to ensure that missing values are being recorded correctly, and not as zeros.

Level: Household.

H_PSEX: Sex of household primary tenant

Name on merged file H_PSEX

Data Type: Numeric Print format F5

Field size: Min: 1 Max: 5 Format on file F5

Purpose: To allow analysis by a tenant gender indicator. This is especially useful when examining lone person and single parent households.

Definition: The variable identifies the gender of the *household primary tenant*.

Classification 1 Male
2 Female
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Level: Household.

H_LENGTH: Household length of tenancy

Name on merged file H_LENGTH

Data Type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: To allow analysis of and by length of tenancy.

Definition: This variable is the number of completed months the household has been housed under a particular program. It is generally derived from jurisdictions' data using tenancy start dates.

Classification 0 less than 1 month
Number (no decimal places) of months
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: Length of tenancy may be determined in a number of ways. Problems occur when household members change, for example some members leave and new members join the household. Jurisdictions should indicate how length of tenancy (or tenancy start date) is determined when households change composition or when households move dwellings within a program. For some jurisdictions only length of current tenancy may be able to be derived. Jurisdictions should note if this is the case.

Validity checks: The distribution can be checked within a State/Territory to ensure that length of tenancy has been recorded in months and not years.

Level: Household.

2.2.4 Income unit level items

There are 15 data items specified for income units. For the 1997 collection not all jurisdictions will be able to provide income unit level data.

Income unit level items list

D_ID	Dwelling identifier
H_ID	Household identifier
I_ID:	Income unit identifier
I_TYPE:	Income unit type
I_SIZE:	Income unit size
I_DEPT:	Dependent children in income unit
I_OTH:	Other income unit members
I_TOTINC:	Total income unit weekly income
I_SOURCE:	Income unit principle source of income
I_MRENT	Market rent value for the income unit
I_RCHARG	Rent charged to the income unit
I_RSUB	Rent subsidy
I_PAGE:	Age of income unit reference person
I_PSEX:	Sex of income unit reference person
I_LENGTH:	Income unit length of tenancy

Income unit level items details

I_ID:	Income unit identifier		
<i>Name on merged file</i>	I_ID		
<i>Data Type:</i>	Numeric	<i>Format on file</i>	F10
<i>Field size:</i>	<i>Min:</i> 1 <i>Max:</i> 10	<i>Print format</i>	F10
<i>Purpose:</i>	<p>This variable is used to identify individual <i>income units</i>. The use of such identification is twofold:</p> <ul style="list-style-type: none"> • It allows for follow-up if queries arise concerning data values. Such follow-up may either be through querying the data set as supplied or through returning to the states for clarification. • If there are multiple income units in a household this variable is used to relate income units to households and thence to dwellings. 		
<i>Definition:</i>	The income unit identifier is a code which uniquely identifies each <i>income unit</i> within a household.		
<i>Classification</i>	As provided by jurisdictions.		
	99998	Income unit data not generally available from jurisdiction's data	
<i>Note:</i>	For some jurisdictions, individual income units within households cannot be identified. Jurisdictions should note if this is the case. Also it should be checked that 99998 is not used as an income unit identifier by the State/Territory. If 99998 is used as an income unit identifier by a State/Territory users of the data should be informed to allow appropriate action to be taken.		
<i>Level:</i>	Income unit.		

I_TYPE: Income unit type

Name on merged file I_TYPE

Data Type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: This variable allows analysis to be undertaken by income unit type.

Definition: This variable describes the type of *income unit*.

Classification

- 1 Single only
- 2 Couple only
- 3 Couple with dependent children
- 4 Sole parent with dependent children
- 5 Other
- 99998 Relevant data not available from the jurisdiction
- 99999 Unknown

Note: Income units cannot contain non-dependent children as these are considered to be a separate income unit.

Level: Income unit.

I_SIZE: Income unit size

Name on merged file I_SIZE

Data Type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: This variable allows for analysis by income unit size. It is also useful for examining crowding or for making adjustments for composition when trying to compare the circumstances of different income units and households.

Definition: This variable gives the total count of people in the income unit.

Classification Number, no decimal places.

- 99998 Relevant data not available from the jurisdiction
- 99999 Unknown

Validity checks: If the household is in GEN, household size is generally less than 10.

Level: Income unit.

I_OTH: Other income unit members

Name on merged file I_OTH

Data Type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: This variable allows for analysis by according to the presence or absence of 'other' income unit members. It is also useful for examining crowding and for making adjustments for composition when trying to compare the circumstances of different households/income units.

Definition: This variable gives the total count of people in the income unit who are neither the *income unit reference person* nor their partner nor *dependent children*, but who are not in another income unit. Such people include dependent relatives and others who are not *dependent children*. Those with their own income source, such as students on Youth Allowance, are separate income units.

Classification 0 No 'other' members in the household
Number, no decimal places.
99998 Relevant data not available from the jurisdiction
99999 Unknown

Note: This variable should only count other dependent family/income unit members. For example newly migrated dependent parents or relatives who do not have their own source of income and are not eligible for Social Security benefits.

Level: Income unit.

I_TOTINC: Total income unit weekly income

Name on merged file I_TOTINC

Data Type: Numeric *Print format* \$F8.2

Field size: *Min:* 1 *Max:* 8 *Format on file* F8.2

Purpose: Income unit income is used to examine the affordability of housing costs.

Definition: This variable is the sum of the personal *gross weekly incomes* of each person in the income unit.

Classification

0	No income
\$.00	Dollar amount per week (to nearest cent)
99998	Relevant data not available from the jurisdiction
99999	Unknown.

Note: This variable should include all income, rather than what is considered 'assessable' by jurisdictions. Jurisdictions should note if some income is not included.

Validity checks: If any income unit member gets some income, this variable should be greater than zero, or missing if a member has missing income data.

Level: Income unit.

I_SOURCE: Income unit principle source of income

Name on merged file I_SOURCE

Data Type: Character *Print format* A5

Field size: Min: 1 Max: 5 *Format on file* A5

Purpose: This variable allows analysis by source of income.

Definition: This variable identifies the main source of income for an income unit. It is derived by summing individual income unit member's income by their main source of income. The income unit's main source of income is then that source of income with the largest (non-missing) dollar value.

Classification

- N No income
- G Government allowances and benefits (including DSS and DVA payments)
- SE Self-employed, own business
- WS Wages and salaries
- OTH Neither government allowances or benefits nor wages and salaries. This includes income from investments and superannuation.
- 99998 Relevant data not available from the jurisdiction
- 99999 Unknown.

Note: Jurisdictions should indicate if this information is not available.

Level: Income unit.

I_MRENT Market rent value for the income unit

Name on merged file I_MRENT

Data Type: Numeric *Print format* \$F5

Field size: Min: 1 Max: 5 *Format on file* F5

Purpose: This variable is used to determine the existence and value of rent rebates.

Definition: The market rent for the income unit is that portion of the dwelling market rent with which the income unit is associated.

- Classification*
- \$ Dollar amount per week (to nearest cent)
 - 99998 Relevant data not available from the jurisdiction
 - 99999 Unknown.

Note: A method is required for allocating dwelling and household market rent across contributing income units.

Validity checks: Rents for GEN and ARHP households should be less than \$300.00.

Level: Income unit.

I_RCHARG Rent charged to the income unit

Name on merged file I_RCHARG

Data Type: Numeric *Print format* \$F8.2

Field size: *Min:* 1 *Max:* 8 *Format on file* F8.2

Purpose: This variable is used to determine the existence and value of rent rebates, and to examine housing affordability.

Definition: The rent charged is the actual rent an income unit is expected to pay after any eligibility for rent rebates/subsidies have been included. The rent charged may or may not have been received. This item reflects the expected and not the actual rent paid as defaults and arrears may reduce or increase the amount received compared to the amount charged.

Classification 0 No rent charged
 \$.00 Dollar amount per week (to nearest cent)
 99998 Relevant data not available from the jurisdiction
 99999 Unknown.

Note: A method is required for allocating dwelling and household rent charged across contributing income units

Validity checks: $0 \leq I_RCHARG \leq I_MRENT$.
 For vacant dwellings this should be zero.

Level: Income unit

I_PSEX: Sex of income unit reference person

Name on merged file I_PSEX

Data Type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: To allow analysis by a tenant gender indicator. This is especially useful when examining lone person and single parent income units.

Definition: The variable identifies the gender of the *income unit reference person*.

Classification

1	Male
2	Female
99998	Relevant data not available from the jurisdiction
99999	Unknown.

Level: Income unit.

I_LENGTH: Income unit length of tenancy

Name on merged file I_LENGTH

Data Type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: To allow analysis by and of length of tenancy.

Definition: This variable is the number of completed months that the income unit has been housed under a particular program. It is generally derived from jurisdictions' data using tenancy start dates.

Classification

0	Less than 1 month
Number (no decimal places) of months	
99998	Relevant data not available from the jurisdiction
99999	Unknown.

Note: Length of tenancy may be determined in a number of ways. While not so problematic for income units as households, problems can still occur when income unit members change. Jurisdictions should indicate how length of tenancy (or tenancy start date) is determined when income units change composition or when they move dwellings within a program. For some jurisdictions only length of current tenancy may be able to be derived. Jurisdiction should note if this is the case.

Validity checks: The distribution can be checked within a State/Territory to ensure that length of tenancy has been recorded in months and not years.

Level: Income unit.

2.2.5 Person level data items

There are ten data items specified for each person.

Person data items list

D_ID	Dwelling identifier
H_ID	Household identifier
I_ID	Income unit identifier
P_ID:	Person identifier
P_RELH:	Relationship within household
P_RELI:	Relationship within income unit
P_PAGE:	Age of person
P_PSEX:	Sex of person
P_TOTINC:	Total person weekly income
P_SOURCE:	Person principle source of income

Person data items details

P_ID:	Person identifier		
<i>Name on merged file</i>	P_ID		
<i>Data Type:</i>	Numeric	<i>Format on file</i>	F10
<i>Field size:</i>	<i>Min:</i> 1 <i>Max:</i> 10	<i>Print format</i>	F10
<i>Purpose:</i>	<p>This variable is used to identify individual income units. The use of such identification is twofold:</p> <ul style="list-style-type: none"> • It allows for follow-up if queries arise concerning data values. Such follow-up may either be through querying the data set as supplied or through returning to the states for clarification. • If there are several people in an income unit this variable is used to relate people to income units, and thence to households and dwellings. 		
<i>Definition:</i>	The person identifier is a code which uniquely identifies each person within an income unit.		
<i>Classification</i>	As provided by jurisdictions.		
	99998 Person data not available from jurisdiction		
<i>Note:</i>	For some jurisdictions, person data is not available. This should be noted if this is the case. Also it should be noted if data for particular groups of people, for example children under 5, is not collected.		
<i>Validity checks:</i>	This should not be missing.		
<i>Level:</i>	Person.		

P_RELH: Relationship within household

Name on merged file P_RELH

Data Type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: To allow derivation of income unit and household type.

Definition: This variable describes the relationship between persons within the household.

Classification

1. *Household primary tenant*
 2. *Husband, Wife or Partner of household primary tenant*
 3. *Dependent child*
 4. *Other dependent child (for example, dependent of other resident)*
 5. *Non-dependent child of household primary tenant or partner*
 6. *Other related individual (for example, grandchild or sibling of household primary tenant or partner)*
- 99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note:

Level: Person.

P_RELI: Relationship within income unit

Name on merged file P_RELI

Data Type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: To allow derivation of income unit and household type.

Definition: This variable describes the relationship between persons within income units.

Classification

1. *Income unit reference person*
 2. *Husband, Wife or Partner of income unit reference person*
 3. *Dependent child of income unit reference person or partner.*
 4. *Dependent parent of income unit reference person or partner.*
 5. *Other related individual (for example, grandchild or sibling of income unit reference person or partner)*
 6. *Non-family member (includes unrelated children)*
- 99998 Relevant data not available from the jurisdiction
99999 Unknown.

Level: Person.

P_PAGE: Age of person

Name on merged file P_PAGE

Data Type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: To allow analysis by a tenant age indicator.

Definition: Age in years of person. Only completed years are recorded.

Classification Number of completed years (no decimal places)
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: When derive this from date of birth make sure this is coded as 99999 for unknown date of birth.

Level: Person.

P_PSEX: Sex of person

Name on merged file P_PSEX

Data Type: Numeric *Print format* F5

Field size: Min: 1 Max: 5 *Format on file* F5

Purpose: To allow analysis by a tenant gender indicator.

Definition: The variable identifies the gender of the person.

Classification

1	Male
2	Female
99998	Relevant data not available from the jurisdiction
99999	Unknown.

Level: Person.

P_TOTINC: Total person weekly income

Name on merged file P_TOTINC

Data Type: Numeric *Print format* \$F8.2

Field size: Min: 1 Max: 7 *Format on file* F8.2

Purpose: Person income is used to examine the affordability of housing costs.

Definition: This variable is the usual total personal *gross weekly income* of a person.

Classification

0	No income
\$	Dollar amount per week (to 2 decimal places)
99998	Relevant data not available from the jurisdiction
99999	Unknown.

Note: This variable should include all income, rather than what is considered 'assessable' by jurisdictions. If some income is not included, jurisdictions should inform DSS.

Validity checks: Unknown income should be recorded as 99999 and NOT zero.

Level: Person.

P_SOURCE: Person principle source of income

Name on merged file P_SOURCE

Data Type: Character *Print format* A5

Field size: *Min:* 1 *Max:* 5 *Format on file* A5

Purpose: This variable allows derivation of main source of income for income units and households.

Definition: This variable identifies the main source of income for a person. It is simply that source of income with the largest (non-missing) dollar value for the person.

Classification

- N No income
- G Government allowances and benefits (including DSS and DVA payments)
- SE Self-employed, own business
- WS Wages and salaries
- OTH Neither government allowances or benefits nor wages and salaries. This includes income from investments and superannuation.
- 99998 Relevant data not available from the jurisdiction
- 99999 Unknown.

Note: Jurisdictions should indicate if this information is not available.

Validity checks: Cross-tabulations of P_SOURCE with P_PAGE may indicate if there are any problems. For example, most children under 15 should be 'N'.

Level: Person.

Section 3: Documentation of merged data sets

(User's Guide)

3.1 Introduction

The final merged data sets contain public housing information for four states (NSW, VIC, Qld, and SA) and cover four levels (dwelling, household income unit, and person) of data. All data sets are in SAS data format as required. There are four data sets altogether. Each is a single level data set including up to four states information at the level specified.

List of four single level merged data sets

Dwe.ssd01: containing information relevant to dwellings for NSW, Vic, Qld, and SA.

Hhld.ssd01: containing information relevant to households within dwellings for NSW, Vic, Qld, and SA.

Iunit.ssd01: containing information relevant to income units within households for Vic and Qld only.

Pson.ssd01: containing information relevant to persons within income units for the Vic, Qld and SA only.

Three identifiers, D_ID, H_ID, and I_ID can be used to link the four levels of data. More sophisticated analysis can be done using the four levels of data together than can be carried out for any one level on its own. A state indicator is included in each level data set to allow comparisons across the state.

Missing data

For each level a range of data items are available. However, for many variables complete information may not be available for all records. In general, there are two types of missing data:

- Cases where, in general, a jurisdiction's data does not contain the relevant information for derivation of the required variable. This may be true for all or a subset of tenants/dwellings.
- Individual cases where the data is not stated or unknown, but where, in general, a jurisdiction's data contains the relevant information for derivation of the required variable.

These two types of missing data are differentiated and coded correspondingly as thus:

99998 Relevant data not available from the jurisdiction

99999 Unknown.

3.2 Contents of each data set

Dwelling level record content:

Variable	Label	Type	Format
D_ID	Dwelling identifier	Character	A13
D_PGM	Program to which dwelling is assigned	Character	A5
D_NHHL	Number of households	Numeric	F5
D_MRENT	Market rent value for dwelling	Numeric	F5
D_RCHARG	Rent charged to dwelling	Numeric	F8.2
D_RSUB	Rent subsidy for dwelling	Numeric	F8.2
D_VAC	Vacancy flag	Character	A5
D_BED	Number of bedrooms	Numeric	F5
D_TYPE	Dwelling type	Numeric	F5
D_AGE	Age of dwelling	Numeric	F5
D_PCODE	Postcode	Numeric	F5
D_LGA	Local Government Area	Numeric	F5
D_STATE	State or Territory	Character	A3

Household level record content:

Variable	Label	Type	Format
D_ID	Dwelling identifier	Character	A13
H_ID:	Household identifier	Character	A13
H_PGM	Program under which household is assisted.	Character	A5
H_NUNIT:	Number of income units	Numeric	F5
H_TYPE:	Household type	Numeric	F5
H_SIZE:	Household size	Numeric	F5
H_DEPT:	Dependent children in household	Numeric	F5
H_OTH:	Other household members	Numeric	F5
H_TOTINC:	Total household weekly income	Numeric	F8.2
H_SOURCE:	Principle source of income for household	Character	A5
H_MRENT	Market rent value for household	Numeric	F5
H_RCHARG	Rent charged to household	Numeric	F8.2
H_RSUB	Rent subsidy	Numeric	F8.2
H_REB	Rent subsidy/rebate flag	Numeric	F5
H_PAGE:	Age of household primary tenant	Numeric	F5
H_PSEX:	Sex of household primary tenant	Character	A5
H_LENGTH:	Household length of tenancy	Numeric	F5
H_STATE:	State or Territory	Character	A3

Income unit level record content:

Variable	Label	Type	Format
D_ID	Dwelling identifier	Character	A13
H_ID:	Household identifier	Character	A13
I_ID:	Income unit identifier	Numeric	F10
I_TYPE:	Income unit type	Numeric	F5
I_SIZE:	Income unit size	Numeric	F5
I_DEPT:	Dependent children in income unit	Numeric	F5
I_OTH:	Other income unit members	Numeric	F5
I_TOTINC:	Total income unit weekly income	Numeric	F8.2
I_SOURCE:	Income unit principle source of income	Character	A5
I_MRENT	Market rent value for the income unit	Numeric	F5
I_RCHARG	Rent charged to the income unit	Numeric	F8.2
I_RSUB	Rent subsidy	Numeric	F8.2
I_PAGE:	Age of income unit reference person	Numeric	F5
I_PSEX:	Sex of income unit reference person	Character	A5
I_LENGTH:	Income unit length of tenancy	Numeric	F5
I_STATE	State or Territory	Character	A3

Person level record content:

Variable	Label	Type	Format
D_ID	Dwelling identifier	Character	A13
H_ID:	Household identifier	Character	A13
I_ID	Income unit identifier	Numeric	F10
P_ID:	Person identifier	Numeric	F10
P_RELH:	Relationship within household	Numeric	F5
P_RELI:	Relationship within income unit	Character	A5
P_PAGE:	Age of person	Numeric	F5
P_PSEX:	Sex of person	Character	A5
P_TOTINC:	Total person weekly income	Numeric	F8.2
P_SOURCE:	Person principle source of income	Character	A5
P_STATE	State or Territory	Character	A3

3.3 Detailed classifications and explanations

For the current data exchange quite a few data items can not be obtained directly from the state provided data. Therefore where possible derivations based on data provided have been used to produce the required merged data file. This section provides details of the derivations used for each data item, in addition to the detailed classifications and data formats.

3.3.1 Dwelling, or property, level data

There are 13 data items which relate to dwellings.

Dwelling data items list

D_ID	Dwelling identifier
D_PGM	Program to which dwelling is assigned
D_NHHLD:	Number of households
D_MRENT	Market rent value for dwelling
D_RCHARG	Rent charged to dwelling
D_RSUB	Rent subsidy for dwelling
D_VAC:	Vacancy flag
D_BED:	Number of bedrooms
D_TYPE:	Dwelling type
D_AGE:	Age of dwelling
D_PCODE:	Postcode
D_LGA:	Local Government Area
D_STATE:	State or Territory

Dwelling data items details

D_ID	Dwelling identifier
<i>Name on merged file</i>	D_ID
<i>Data type:</i>	Character <i>Print format</i> A13
<i>Field size:</i>	Min: 4 Max: 13 <i>Format on file</i> A13
<i>Definition:</i>	The dwelling identifier is a code which uniquely identifies each <i>dwelling</i> .
<i>Classification</i>	State/Territory identifier three digit alpha code – AAA – followed by the identifying code – NNNNNNNNNNN – as provided by jurisdictions (if available). State/Territory identifiers are: NSW New South Wales VIC Victoria QLD Queensland SA South Australia
<i>Derivation</i>	NSW did not provide its identifying code therefore the default data sequence number is used as the identifying code. For Victoria the household identifier is used as Victoria does not distinguish between households and dwellings. For Queensland the account identifier is used as the identifying code.
<i>Validity checks:</i>	Must be unique within a State/Territory.
<i>Note:</i>	Queensland provided details for both current tenants and old tenants at the account level. Only those accounts attached current tenants have been selected for the merged data. Seven properties with ten and more bedrooms in the SA property file are excluded in the final merged dwelling level data, since these are dwellings which were leased by community organisations from the Housing Trust. Accordingly the organisation is responsible for housing people in these dwellings, not the South Australia Housing trust. For more details, see the reply from SA.

D_MRENT Market rent value for dwelling

Name on merged file D_MRENT

Data type: Numeric *Print format* \$F5

Code range: 0-99999 *Format on file* F5

Definition: The market rent value of the dwelling is the rent the dwelling would receive if it were in the private rental market.

Classification \$ per week, value to be shown to nearest cent

99997	Market rent not relevant, for example for untenable dwellings
99998	Relevant data not available from the jurisdiction
99999	Unknown.

Derivation Only SA provided market rent for dwellings. For the other three states this is same as H_MRENT as they do not distinguish between households and dwellings.

Note: The method of calculating market rent may vary with jurisdiction. This needs to be checked with states.
NSW recorded zero market rent for the new constructed dwellings.

D_RCHARG Rent charged to dwelling

Name on merged file D_RCHARG

Data type: Numeric *Print format* \$F8.2

Code range: 0-99999 *Format on file* F8.2

Definition: The rent charged is what tenants are charged, that is the actual rent they are expected to pay after any eligibility for rent rebates/subsidies have been included. The rent charged to the tenant may or may not have been received. This item reflects the expected and not the actual rent paid as defaults and arrears may reduce or increase the amount received compared to the amount charged.

Classification 99997 No rent charged, for example for untenable/vacant dwellings

\$0.00 per week, value to be shown to nearest cent

99998 Relevant data not available from the jurisdiction

99999 Unknown.

Derivation: For NSW, if a dwelling is vacant then this is assigned as 99997, otherwise this is as it was on the file NSW provided.

For Victoria this is derived from person level data as this is not available at dwelling level.

For SA data at the household level is used for rent charged to dwelling as there is no data available at the dwelling level.

Note: Except NSW, who provided a vacancy indicator, this is same as H_RCHARG as they do not distinguish between households and dwellings.

Validity checks: $0 < D_RCHARG \leq D_MRENT$. For an occupied dwelling this should be market rent minus rent subsidy.

For a vacant dwelling this should be 99997.

D_TYPE: Dwelling type

Name on merged file D_TYPE

Data type: Character *Print format* A5

Code range: 1-5, 99997-99999 *Format on file* A5

Definition: This variable identifies the structure of private dwellings.

Classification

5	Detached house
5	Semi-detached house, townhouse, terrace house, duplex
5	Flat, unit, apartment
5	Boarding house, hostel
5	Other (that is, none of the above, and includes caravans and movable units)
99997	Not known (includes dwellings whose original dwelling type code is based on something other than structure, for example a program)
99998	Relevant data not available from the jurisdiction
99999	Unknown (missing/ not stated).

(Source: DSS specified)

Derivations: Originally four states had their own classification for dwelling type, they are regrouped as above (DSS suggested classification). For details see the table following.

NSW	Vic	Qld	SA	Merged data
Detached house	Separate house	Detached house	Single unit	Detached house
Semi-detached house Row house Patio Townhouse Duplex Triplex	Semi detached house Medium density	Cluster Duplex unit Attached house Dual Occupancy	Double unit Attached house	Semi-detached house, townhouse, terrace house, duplex
Maisonette Turner-rigby Multi-unit (walk up) Multi-unit (high rise)	High rise bedsitter High rise flat Low rise bedsitter 1 level Low rise bedsitter 2 levels Low rise flat, 1 level Low rise flat, 2 levels	Apartment Senior unit	Cottage flat Flat	Flat, unit, apartment
Hostel		Community facility Boarding house		Boarding house, hostel
Shop Shop & dwelling	Movable unit Other (shop, office)	Estate office	Other	Other (none of the above, includes caravans and movable units)
			Purchased house AHU	unknown

D_AGE: Age of dwelling

Name on merged file D_AGE

Data type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Purpose: This variable allows analysis by age of dwelling stock.

Definition: Age of dwelling is measured in terms of the number of completed years since the dwelling was built.

Classification Number of months, no decimal places

99998 Relevant data not available from the jurisdiction

99999 Unknown.

Note: This variable does not appear to be available for Victoria and Queensland.

Derivation: For NSW and SA this is derived from the variable 'first tenant date'. The data extracted date is different as below:

NSW: 31 December 1996.

SA: 30 June 1997.

D_PCODE: Postcode

Name on merged file D_PCODE

Data type: Character *Print format* A5

Field size: *Min:* 4 *Max:* 5 *Format on file* A5

Definition: This variable indicates the postcode in which the dwelling is located.

Classification Four digit numeric code

99998 Relevant data not available from the jurisdiction

99999 Unknown.

Note: This is assigned as 99998 for Victoria as this is not available for the jurisdiction.

Validity checks: These values should correspond to official postcodes.

D_LGA: Local Government Area

Name on merged file **D_LGA**

Data type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Definition: This variable indicates the Local Government Area in which the dwelling is located.

Classification As provided by jurisdictions.
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: This is assigned as 99998 for Queensland as this is not available for the jurisdiction.

Validity checks: Each LGA code should have a corresponding name. Missing names suggest problems with the data, for example that LGA codes have not been updated as LGAs have changed boundaries/names.

D_STATE: State or Territory

Name on merged file **D_STATE**

Data type: Character *Format on file* A3

Field size: *Min:* 1 *Max:* 1 *Print format* A3

Definition: This variable indicates the State/Territory in which the dwelling is located.

Classification NSW New South Wales
VIC Victoria
QLD Queensland
SA South Australia

Note: This variable should not be missing.

Household level items

There are 18 data items specified at the household level.

Household data items list

D_ID	Dwelling identifier
H_ID:	Household identifier
H_PGM	Program under which household is assisted.
H_NUNIT:	Number of income units
H_TYPE:	Household type
H_SIZE:	Household size
H_DEPT:	Dependent children in household
H_OTH:	Other household members
H_TOTINC:	Total household weekly income
H_SOURCE:	Principle source of income for household
H_MRENT	Market rent value for household
H_RCHARG	Rent charged to household
H_RSUB	Rent subsidy
H_REB	Rent subsidy/rebate flag
H_PAGE:	Age of household primary tenant
H_PSEX:	Sex of household primary tenant
H_LENGTH:	Household length of tenancy
H_STATE:	State or Territory

Household data items details

H_ID: Household identifier

Name on merged file H_ID

Data type: Character *Format on file* A13

Field size: Min: 1 Max: 10 *Print format* A13

Definition: The *household* identifier is a code which uniquely identifies each *household*.

Classification See D_ID.

Derivation: See D_ID.

Note: NSW included 78 records which are not relevant to households, such as 'place of worship', 'child care centre' etc, therefore these records are excluded in the merged household data. Also since 2195 records are related to vacant dwellings, these are also excluded in the household file.

Validity checks: This should not be missing.

H_PGM Program under which household is assisted.

Name on merged file H_PGM

Data type: Character *Print format* A5

Field size: Min: 1 Max: 5 *Format on file* A5

Definition: This variable identifies the assistance program through which a household is assisted.

Classification

GEN	General public housing (including housing for pensioners)
ARHP	Aboriginal Rental Housing Program
CHP	Community Housing Program
CAP	Crisis Accommodation Program
OTH	Other, including state-specific housing programs
99998	Relevant data not available from the jurisdiction
99999	Unknown.

Derivation: Same as D_PGM.

Note: See D_PGM.

H_NUNIT: Number of income units

Name on merged file H_NUNIT

Data type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Definition: This variable indicates the number of *income units* in a *household*.

Classification Number, no decimal places.

If the number of income units is unknown or cannot be determined the following codes are used:

99997 More than one

99998 Relevant data not available from the jurisdiction

99999 Unknown.

Note: This variable is a derived item for 1997.

Derivation: For NSW and SA this is derived based on household type and number of other non dependent members. Single and couple only and couple with dependent children only and single parent with dependent children only households are assigned as one income unit household, other types of households are coded as 99997.

For Vic and Qld this is derived by counting all income units within household from income unit level data.

H_TYPE: Household type

Name on merged file H_TYPE

Data type: Character *Print format* A5

Field size: Min: 1 Max: 5 *Format on file* A5

Definition: This variable describes the type of household. Households can contain dependent and non-dependent children as well as non-family members.

Classification

- 1 Person living alone
 - 2 Couple only
 - 3 Couple with children
 - 31 Couple with dependent children only
 - 32 Couple with non-dependent children (with or without dependent children)
 - 4 Sole parent with children
 - 41 Sole parent with dependent children only
 - 42 Sole parent with non-dependent children (with or without dependent children)
 - 5 Group (unrelated adults)
- Other, including multiple family
- 99998 Relevant data not available from the jurisdiction
 - 99999 Unknown.

Note: To avoid having a large number of households classified to 'Other' for 1997 a two tier classification is used since SA only provided aggregated categories for household type.

For Vic, categories 32 and 42 could not be distinguished from category 6. In other words some households which were included in 'Other' could have belonged to categories 32 or 42.

Derivation: For SA and Vic this is converted from their own classification to the classification above. For details see table below.

For NSW and QLD this is derived from the household structure. Details of derivation are also listed on the table below.

NSW	Vic	Qld	SA	Merged data
If household size=1	Young single Single Old single Unknown single	If household size=1	Single person	1. Single person
If household size=2 and spouse indicator is 'yes'.	Old couple Couple , no dependent unknown couple	If household size=2 and spouse indicator is 'yes'.	Couple only	2. Couple only
			Couple with children	3. Couple with children
If spouse indicator is 'yes' and H_DEPT>0 and H_oth=0;	Couple , 1 dependent Couple , 2 dependents Couple , 3 dependents Couple , 4 dependents	If spouse indicator is 'yes' and H_DEPT>0 and H_oth=0;		31 Couple with dependent only
				32 Couple with non dependent child
			Single parent	4. Single parent with children
If spouse indicator is 'no' and H_DEPT>0 and H_OTH=0	Single parent , 1 dependent Single parent, 2 dependents Single parent, 3 dependents Single parent, 4 dependents	If spouse indicator is 'no' and H_DEPT>0 and H_OTH=0		41 Single parent with dependent children only
				42 Single parent with non dependent children
			Sharers	5. Group
Those that are left.	Couple , 1 dependent, 1 other Couple , 1 dependent, 2 others Couple , 2 dependent, 1 other Couple , 2 dependent, 2 others Couple , 3 dependent, 1 other Couple , 3 dependent, 2 others Couple , 4 dependent, 1 other Couple , 4 dependent, 2 others Couple , no dependent, 1 other Couple , no dependent, 2 others Couple , no dependent, 3 others Single parent, 1 dependent, 1 other Single parent, 1 dependent, 2 others Single parent, 2 dependent, 1 other Single parent, 2 dependent, 2 others Single parent, 3 dependent, 1 other Single parent, 3 dependent, 2 others Single parent, 4 dependent, 1 other Single parent, 4 dependent, 2 others	Those that are left.		6. Other

			Unknown	99997
--	--	--	---------	-------

H_SIZE: Household size

Name on merged file H_SIZE

Data type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Definition: This variable gives the total count of people in the household.

Classification 0 An 'empty' household
Number, no decimal places.
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Derivation: For NSW this is derived from '1+1 (if have spouse)+number of other household members (includes number of dependents)'.
For QLD this is derived by adding up household members at the person level.

H_DEPT: Dependent children in household

Name on merged file H_DEPT

Data type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Definition: This variable gives the total count of *dependent children* in the household.

Classification 0 No dependent children in the household
Number, no decimal places.
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: Not all jurisdictions identify dependent children the same way. For example, Vic defines children under 18 years old (not 16 which is showed in the Glossary) as dependent children. When using this data item make sure NSW defined Child/Student the same way as given in the Glossary.

Derivation For Vic, SA, Qld this is derived from the person level data. For SA if a person is coded as 'Daughter' or 'Son' and their age is less than 16 years old then he/she is counted as a dependent. For Vic and Qld if a person is coded as 'Dependent' then he/she is counted as a dependent.

H_OTH: Other household members

Name on merged file H_OTH

Data type: Numeric *Print format* F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Definition: This variable gives the total count of people in the household who are neither the *household primary tenant* nor their partner/spouse nor a *dependent child*. Such people include non-dependent children, other relatives and unrelated household members. Dependent children of secondary income units, for example the children of a single mother living with her parents, are counted as *dependent children*.

Classification 0 No 'other' people in the household
Number, no decimal places.
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: See D_DEPT.

Derivation: For Vic this is derived from total number of household members and number of dependent children.
For SA this is derived. If a person is not coded as 'Tenant' or 'Spouse' or 'Dependent' then he/she is counted as an 'other household member'.

Validity checks: This variable is equal to (H_SIZE - H_DEPT).

H_TOTINC: Total household weekly income

Name on merged file H_TOTINC

Data type: Numeric *Print format* \$F8.2

Field size: *Min:* 1 *Max:* 8 *Format on file* F8.2

Definition: This variable is the sum of the personal *gross weekly incomes* of each member of the household.

Classification 0 No income
\$.00 Dollar amount per week (to nearest cent)
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: This variable includes all income, rather than what is considered 'assessable' by jurisdictions.

Derivation: For Vic and Qld, this is derived by summing up all household member's income in the person level data.

Validity checks: If any income unit or household member gets some income, this variable should be greater than zero, unless income is missing for a household

member.

H_SOURCE: Principle source of income for household

Name on merged file H_SOURCE

Data type: Character *Print format* A5

Field size: *Min:* 1 *Max:* 5 *Format on file* A5

Definition: This variable identifies the main source of income for a household. It is derived by summing individual household member's income. The household's main source of income is then that source of income with the largest (non-missing) dollar value.

Classification

- N No income
- G Government allowances and benefits (including DSS and DVA payments)
- SE Self-employed, own business
- WS Wages and salaries
- OTH Neither government allowances or benefits nor wages and salaries. This includes, for example, income from investments and superannuation.
- 99998 Relevant data not available from the jurisdiction
- 99999 Unknown.

Validity checks: The majority of households should be classified as 'G'.

Derivation: This is derived from the person level data. Household main source of income is the source of income for the household member who had the largest main income in the household.

Note: States had their own classification for this item. Recoding has been used in the final merged data to map state's classification to the above classification. For details of the recoding, see attached SAS code.

H_MRENT Market rent value for household

Name on merged file H_MRENT

Data type: Numeric *Print format* \$F5

Field size: *Min:* 1 *Max:* 6 *Format on file* F5

Definition: The market rent for the household is that portion of the dwelling market rent with which the household is associated.

Classification \$ Dollar amount per week (to nearest cent)
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: A method is required for allocating market rent across contributing households.

Derivation: See D_MRENT.

H_RCHARG Rent charged to household

Name on merged file H_RCHARG

Data type: Numeric *Print format* \$F8.2

Field size: *Min:* 1 *Max:* 8 *Format on file* F8.2

Definition: The rent charged is the actual rent a household is expected to pay after any eligibility for rent rebates/subsidies have been included. The rent charged may or may not have been received. This item reflects the expected and not the actual rent paid as defaults and arrears may reduce or increase the amount received compared to the amount charged.

Classification 0 No rent charged
 \$0.00 Dollar amount per week (to nearest cent)
 99998 Relevant data not available from the jurisdiction
 99999 Unknown.

Derivation: See D_RCHARG

Validity checks: $0 \leq H_RCHARG \leq H_MRENT$.

H_RSUB Rent subsidy

Name on merged file H_RSUB

Data type: Numeric *Print format* \$F8.2

Field size: *Min:* 1 *Max:* 8 *Format on file* F8.2

Definition: This variable is derived as the difference between market rent and rent charged, that is

$$H_MRENT - H_RCHARG.$$

Classification \$0.00 Dollar amount per week (to nearest cent)
 99998 Relevant data not available from the jurisdiction
 99999 Unknown.

Derivation: See D_RSUB.

Validity checks: $0 \leq H_RSUB \leq H_MRENT$

H_PSEX: Sex of household primary tenant

Name on merged file H_PSEX

Data type: Character Print format A5

Field size: Min: 1 Max: 5 Format on file A5

Definition: The variable identifies the gender of the *household primary tenant*.

Classification M Male
F Female
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Derivation: This is the gender for the person who is coded as 'Tenant' under variable 'person relationship in the household'

Note: This is not available in the NSW data.

H_LENGTH: Household length of tenancy

Name on merged file H_LENGTH

Data type: Numeric Print format F5

Field size: Min: 1 Max: 5 Format on file F5

Definition: This variable is the number of completed months the household has been housed under a particular program. It is generally derived from jurisdictions' data using tenancy start dates.

Classification 0 less than 1 month
Number (no decimal places) of months
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: Length of tenancy may be determined in a number of ways. For example, for SA this includes the total length of tenancy within public housing; for Victoria this refers to the length of tenancy in the current dwelling only. The definitions used by NSW and Queensland need to be found out.

Validity checks: The distribution can be checked within a State/Territory to ensure that length of tenancy has been recorded in months and not years.

Derivation: This is derived from the tenancy start date. It should be noted that the data extraction date is not same across the states. For details see P_PAGE.

Income unit level items

There are 16 data items specified for income units. For the 1997 collection only Victoria and Queensland data allow us to derive income unit level data.

Income unit level items list

D_ID	Dwelling identifier
H_ID	Household identifier
I_ID:	Income unit identifier
I_TYPE:	Income unit type
I_SIZE:	Income unit size
I_DEPT:	Dependent children in income unit
I_OTH:	Other income unit members
I_TOTINC:	Total income unit weekly income
I_SOURCE:	Income unit principle source of income
I_MRENT	Market rent value for the income unit
I_RCHARG	Rent charged to the income unit
I_RSUB	Rent subsidy
I_PAGE:	Age of income unit reference person
I_PSEX:	Sex of income unit reference person
I_LENGTH:	Income unit length of tenancy
I_STATE	State or Territory

Income unit level items details

I_ID: Income unit identifier

<i>Name on merged file</i>	I_ID		
<i>Data type:</i>	Numeric	<i>Format on file</i>	F10
<i>Field size:</i>	<i>Min:</i> 1	<i>Max:</i> 10	<i>Print format</i> F10
<i>Definition:</i>	The income unit identifier is a code which uniquely identifies each <i>income unit</i> within a household.		
<i>Classification</i>	As provided by jurisdictions. 99998 Income unit data not generally available from jurisdiction's data		
<i>Derivation:</i>	For both Vic and Qld the family identifier is used as the income unit identifier.		

I_TYPE: Income unit type

Name on merged file I_TYPE

Data type: Numeric *Print format* F5

Field size: Min: 1 Max: 5 *Format on file* F5

Purpose: This variable allows analysis to be undertaken by income unit type.

Definition: This variable describes the type of *income unit*.

Classification 6 Single only
 7 Couple only
 8 Couple with dependent children
 9 Sole parent with dependent children
 10 Other
 99998 Relevant data not available from the jurisdiction
 99999 Unknown

Note: Income units cannot contain non-dependent children as these are considered to be a separate income unit.

Derivation: This is derived from 'person relationship code in income unit'. For example, if I_SIZE=1 then I_TYPE=1; if I_SIZE=2 and spouse indicator is yes then I_TYPE=2; if spouse indicator is yes and I_DEPT>0 then I_TYPE=3; if spouse indicator is no and I_DEPT>0 then I_TYPE=4; otherwise I_TYPE is assigned as 5.

I_SIZE: Income unit size

Name on merged file I_SIZE

Data type: Numeric *Print format* F5

Field size: Min: 1 Max: 5 *Format on file* F5

Definition: This variable gives the total count of people in the income unit.

Classification Number, no decimal places.
 99998 Relevant data not available from the jurisdiction
 99999 Unknown

Derivation: This is derived by summing all income unit member.

I_DEPT: Dependent children in income unit

Name on merged file I_DEPT

Data type: Numeric Print format F5

Field size: Min: 1 Max: 5 Format on file F5

Definition: This variable gives the total count of *dependent children* in the income unit.

Classification 0 No dependent children in the income unit

Number, no decimal places.

99998 Relevant data not available from the jurisdiction

99999 Unknown

Note: Not all jurisdictions identify dependent children the same way. See H_DEPT.

Derivation: For Vic and Qld this is derived from the person level data. If a person is coded as 'Dependent' then he/she is counted as a dependent.

I_OTH: Other income unit members

Name on merged file I_OTH

Data type: Numeric Print format F5

Field size: Min: 1 Max: 5 Format on file F5

Definition: This variable gives the total count of people in the income unit who are neither the *income unit reference person* nor their partner nor *dependent children*, but who are not in another income unit. Such people include dependent relatives and others who are not *dependent children*. Those with their own income source, such as students on Youth Allowance, are separate income units.

Classification 0 No 'other' members in the household

Number, no decimal places.

99998 Relevant data not available from the jurisdiction

99999 Unknown

Note: See H_DEPT.

Validity checks: This variable is equal to (I_SIZE- I_DEPT).

Derivation: For Vic this is derived by subtracting number of dependent children from total number of household members.

For SA this is derived. If a person is not coded as 'Tenant' or 'Spouse' or 'Dependent' then he/she is counted as an 'other household member'.

I_TOTINC: Total income unit weekly income

Name on merged file I_TOTINC

Data type: Numeric *Print format* \$F8.2

Field size: *Min:* 1 *Max:* 8 *Format on file* F8.2

Definition: This variable is the sum of the personal *gross weekly incomes* of each person in the income unit.

Classification

0	No income
\$.00	Dollar amount per week (to nearest cent)
99998	Relevant data not available from the jurisdiction
99999	Unknown.

Note: This variable include all income, rather than what is considered 'assessable' by jurisdictions.

Derivation: This is derived by summing up all income unit member's income in the person level data.

Validity checks: If any income unit member gets some income, this variable should be greater than zero, or missing if a member has missing income data.

I_SOURCE: Income unit principle source of income

Name on merged file I_SOURCE

Data type: Character *Print format* A5

Field size: *Min:* 1 *Max:* 5 *Format on file* A5

Definition: This variable identifies the main source of income for an income unit. The income unit's main source of income is that source of income with the largest (non-missing) dollar value.

Classification

N	No income
G	Government allowances and benefits (including DSS and DVA payments)
SE	Self-employed, own business
WS	Wages and salaries
OTH	Neither government allowances or benefits nor wages and salaries. This includes income from investments and superannuation.
99998	Relevant data not available from the jurisdiction
99999	Unknown.

Derivation: This is derived from the person level data. Income unit main source of income is the source of income for the income unit member who had the largest income within the income unit.

Note: See H_SOURCE.

I_MRENT Market rent value for the income unit

Name on merged file I_MRENT

Data type: Numeric *Print format* \$F5

Field size: *Min:* 1 *Max:* 5 *Format on file* F5

Definition: The market rent for the income unit is that portion of the dwelling market rent with which the income unit is associated.

Classification \$ Dollar amount per week (to nearest cent)
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Derivation: This is derived by dividing market rent for the household by the number of income units in that household.

I_RCHARG Rent charged to the income unit

Name on merged file I_RCHARG

Data type: Numeric *Print format* \$F8.2

Field size: *Min:* 1 *Max:* 8 *Format on file* F8.2

Definition: The rent charged is the actual rent an income unit is expected to pay after any eligibility for rent rebates/subsidies have been included. The rent charged may or may not have been received. This item reflects the expected and not the actual rent paid as defaults and arrears may reduce or increase the amount received compared to the amount charged.

Classification 0 No rent charged
\$.00 Dollar amount per week (to nearest cent)
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Validity checks: $0 \leq I_RCHARG \leq I_MRENT$.

Derivation: This is derived by dividing the rent charged for the household by the number of income unit in that household.

I_PSEX: Sex of income unit reference person

Name on merged file I_PSEX

Data type: Character Print format A5

Field size: Min: 1 Max: 5 Format on file A5

Definition: The variable identifies the gender of the *income unit reference person*.

Classification M Male
F Female
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: See P_PSEX.

Derivation: This is the gender for the person who is coded as 'SELF' under variable 'person relationship in the income unit/family'.

I_LENGTH: Income unit length of tenancy

Name on merged file I_LENGTH

Data type: Numeric Print format F5

Field size: Min: 1 Max: 5 Format on file F5

Definition: This variable is the number of completed months that the income unit has been housed under a particular program. It is generally derived from jurisdictions' data using tenancy start dates.

Classification 0 Less than 1 month
Number (no decimal places) of months
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: See H_LENGTH.

Derivation: See H_LENGTH.

Validity checks: The distribution can be checked within a State/Territory to ensure that length of tenancy has been recorded in months and not years.

Person level data

There are 11 data items specified for each person. For the 1997 collection only Victoria, Queensland and SA data allow us to derive person level data.

Person data items list

D_ID	Dwelling identifier
H_ID	Household identifier
I_ID	Income unit identifier
P_ID:	Person identifier
P_RELH:	Relationship within household
P_RELI:	Relationship within income unit
P_PAGE:	Age of person
P_PSEX:	Sex of person
P_TOTINC:	Total person weekly income
P_SOURCE:	Person principle source of income
P_STATE:	State or Territory

Person data items details

P_ID:	Person identifier		
<i>Name on merged file</i>	P_ID		
<i>Data type:</i>	Numeric	<i>Format on file</i>	F10
<i>Field size:</i>	Min: 1	Max: 10	<i>Print format</i> F10
<i>Definition:</i>	The person identifier is a code which uniquely identifies each person within an income unit.		
<i>Classification</i>	As provided by jurisdictions. 99998 Person data not available from jurisdiction		
<i>Validity checks:</i>	This should not be missing.		
<i>Note:</i>	In the SA person file persons with a zero id are full rent payer or rent frozen (see SA documentation).		

P_RELH: Relationship within household

Name on merged file P_RELH

Data type: Numeric Print format F5

Field size: Min: 1 Max: 5 Format on file F5

Definition: This variable describes the relationship between persons within the household.

Classification

- 1 Household primary tenant
- 2 Husband, Wife or Partner of household primary tenant
- 3 Dependent children
- 4 Other dependents
- 5 Non-dependent child of household primary tenant or partner
- 6 Other related individual (for example, grandchild or sibling of household primary tenant or partner)
- 99998 Relevant data not available from the jurisdiction
- 99999 Unknown.

Note: For Qld and Vic, 'non-dependent children' can not be separated from 'other related individual'. For SA, other dependents (that is, not children) are included in other related individual. For details see table below.

Vic	Qld	SA	Merged data
Tenant	Tenant	Tenant	1. Tenant
Spouse	Spouse	Spouse, wife, husband, De facto	2. Spouse
Dependent	Dependent	Daughter, Son , age under 16	3. Dependant child
	Dependent of residents		4. Other Dependant
			5. Non-dependent child
Resident	Resident	All other, including sister, brother, cousin, grandson, etc	6. Other related individual
		Caretaker, extra person	8. Other non-related individual
'MULND'			99999

P_RELI: Relationship within income unit

Name on merged file P_RELI

Data type: Character *Print format* A5

Field size: Min: 1 Max: 5 *Format on file* A5

Definition: This variable describes the relationship between persons within income units.

Classification

SELF: Income unit reference person

SP: Husband, Wife or Partner of income unit reference person

DEPT: Dependent child

99998 Relevant data not available from the jurisdiction

99999 Unknown.

P_PAGE: Age of person

Name on merged file P_PAGE

Data type: Numeric *Print format* F5

Field size: Min: 1 Max: 5 *Format on file* F5

Definition: Age in years of person. Only completed years are recorded.

Classification Number of completed years (no decimal places)

99998 Relevant data not available from the jurisdiction

99999 Unknown.

Note: The extract data date for states is different as below.

NSW: 31 December 96.

VIC: 6 August 97.

SA: 30 June 97.

QLD: 31 May 97.

Derivation: This is calculated from the person's date of birth and the data extraction date for NSW, Vic, and Qld.

P_PSEX: Sex of person

Name on merged file P_PSEX

Data type: Character *Print format* A5

Field size: Min: 1 Max: 5 *Format on file* A5

Purpose: To allow analysis by a tenant gender indicator.

Definition: The variable identifies the gender of the person.

Classification M Male
F Female
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: Code 'U' is recoded to '99999'.

P_TOTINC: Total person weekly income

Name on merged file P_TOTINC

Data type: Numeric *Print format* \$F8.2

Field size: Min: 1 Max: 8 *Format on file* F8.2

Definition: This variable is the usual total personal gross weekly income of a person.

Classification 0 No income
\$.00 Dollar amount per week (to 2 decimal places)
99998 Relevant data not available from the jurisdiction
99999 Unknown.

Note: This variable includes all income, rather than what is considered 'assessable' by jurisdictions.
For those paid full rent their income are usually recorded as 0, this is changed to 99999.

Derivation: This is derived by summing all sources of income for the person.

P_SOURCE: Person principle source of income

Name on merged file P_SOURCE

Data type: Character *Print format* A5

Field size: Min: 1 Max: 5 *Format on file* A5

Definition: This variable identifies the main source of income for a person. It is simply that source of income with the largest (non-missing) dollar value for the person.

Classification

- N No income
- G Government allowances and benefits (including DSS and DVA payments)
- SE Self-employed, own business
- WS Wages and salaries
- OTH Neither government allowances or benefits nor wages and salaries. This includes income from investments and superannuation.
- 99998 Relevant data not available from the jurisdiction
- 99999 Unknown.

Derivation: For a person who had more than one source of income this is taken as the source of income which associated with the largest dollar value.

Validity checks: Cross-tabulations of P_SOURCE with P_PAGE may indicate if there are any problems. For example, most children under 15 should be 'N'.

Note: States did not record income related data for persons who are full rent payer. Therefore this is coded as '99999' for full rent payer.
See H_SOURCE.

Glossary

<i>Child</i>	A <i>child</i> is a related or unrelated person aged under 16 years who forms a parent-child relationship with one person over 16 years of age in the dwelling/household/income unit. (Source: 1997-98 Public Housing Data manual).
<i>Child- dependent</i>	A <i>dependent</i> child is a <i>child</i> aged under 16 years of age; or a dependent full time student aged 16 to 24. (Source: 1997-98 Public Housing Data manual).
<i>Child-non-dependent</i>	A <i>non-dependent</i> child is a <i>child</i> who is aged 16 or more and is not a full-time student. (Source: 1997-98 Public Housing Data manual)
<i>Dwelling</i>	A dwelling is a building or structure in which people live. This can be a building, such as a house; part of a building, such as a flat; or it could be a caravan or demountable unit. (Source: based on ABS Census 1996 Data Dictionary)
<i>Dwellings – occupied</i>	These are <i>dwellings</i> occupied by households with a current 'tenancy agreement'. (Source: 1997-98 Public Housing Data manual)
<i>Dwellings – tenantable</i>	Tenantable dwellings are <i>dwellings</i> currently occupied by tenants as well as those vacant properties where maintenance has been completed. All tenantable dwellings whether occupied or vacant have a market rent value. (Source: 1997-98 Public Housing Data Manual)
<i>Dwellings – untenantable</i>	Untenantable dwellings are defined as <i>dwellings</i> not currently occupied by a tenant where maintenance has either been deferred or not been completed. (Source: 1997-98 Public Housing Data Manual)
<i>Dwellings – vacant</i>	Vacant dwellings are <i>dwellings</i> not currently occupied by tenants. Such dwellings may be tenantable or untenantable. Includes newly purchased and constructed dwellings that are awaiting tenancy and dwellings under repair. (Source: 1997-98 Public Housing Data Manual)
<i>First income unit</i>	The <i>first income unit</i> is that with I_INUM=1 within a household.
<i>Gross weekly income</i>	Gross weekly income is the income before tax, superannuation, health insurance, or other deductions are made. Gross income includes family allowance, family allowance supplement, pensions, unemployment benefits, student allowances, maintenance (child support), superannuation, wages, salary, overtime, dividends, rents received, interest received, business or farm income (less operation expenses) and worker's compensation received. (Source: ABS Census 1996 Data Dictionary)
<i>Household</i>	<i>Household</i> equates to tenancy, and consists of the usual persons in a dwelling covered by a single 'tenancy agreement' with the housing authority. Persons living in the same dwelling but covered by different 'tenancy agreements' are considered to be in different households. Persons who are co-tenants, that is who have co-signed the same tenancy agreement, belong to the same household. (Based on the 1997-98 Public Housing Data manual)

<i>Household primary tenant</i>	The <i>household primary tenant</i> is the household member responsible for the tenancy.
<i>Income unit</i>	Income units can be considered to be analogous to family units with the distinction that <i>non-dependent</i> children and other adults living in the same household are treated as separate income units. (Source: ABS 1994 Australian Housing Survey definition)
<i>Income unit reference person</i>	The <i>income unit reference person</i> is the income unit member responsible for the income unit.

Appendix 1: The input data sets

The main purpose of this Appendix is to summarise the work done in the first stage of the project. While raw data for the four states were read in successfully, some issues need to be discussed.

The discussion is divided into three parts. Part 1 describes the raw data as provided by each state. Data structure, data records, and data items are described. Part 2 addresses the problems we encountered while reading the data into SAS datasets. Part 3 contains recommendations on standards to be used in data exchange and the information that should be included in data documentation.

Part 1: Description of public housing data files from four states

A description of the raw data sets provided by four states (New South Wales, Victoria, Queensland, South Australia) is given below.

New South Wales

New South Wales data structure is shown below:

NSW Data Structure

The NSW data seems to be primarily a property based system and contains

1. Property type and management data, data items include:

Tenancy start date (end date)	Dwelling type	Number of bedrooms
Local Government Area	Suburb or Location	Post code
Date first occupied	Date last vacated	Reason of last vacation
Department of Housing zone	Building material	Current tenancy start date
2. Program type of dwelling
3. Rent data including

Market rent	Rent arrears	How rent paid
Rent payable		
4. Broad household composition data, such as,

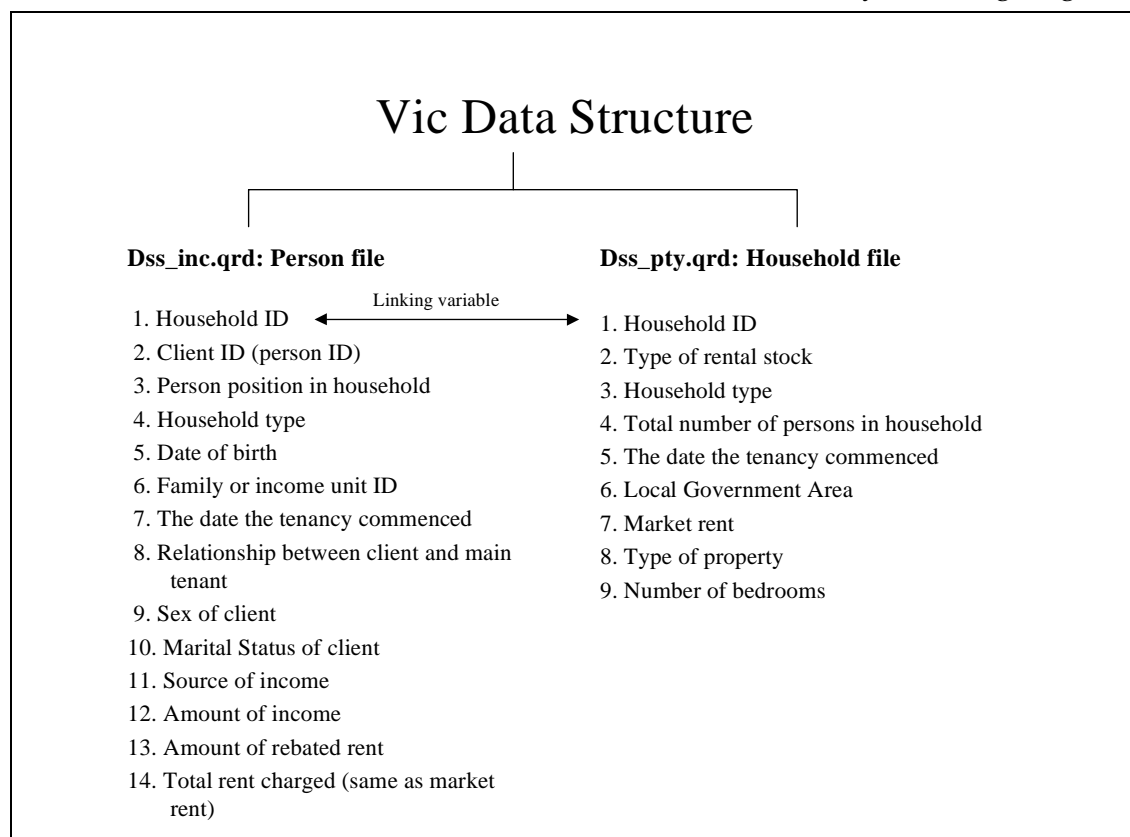
Number of children/students
Number of other household members (excluding spouse and tenant)
Flag for tenant's spouse (Y/N)
Date of birth of principal tenant
5. Income by household member (type and amount for tenant, spouse and up to four others, includes FAS specifically)

Note: A complete list of data items is at Attachment 1.

The NSW data was provided in a single large file which contains 130,975 records and 48 data items. The data is primarily attached to properties rather than tenants. It contains information on rent, income by household member (type and amount for tenant, spouse and up to 4 others), type of dwelling, property type, and broad household composition.

Victoria

The data structure for the data sets for Victoria can be described by following diagram.



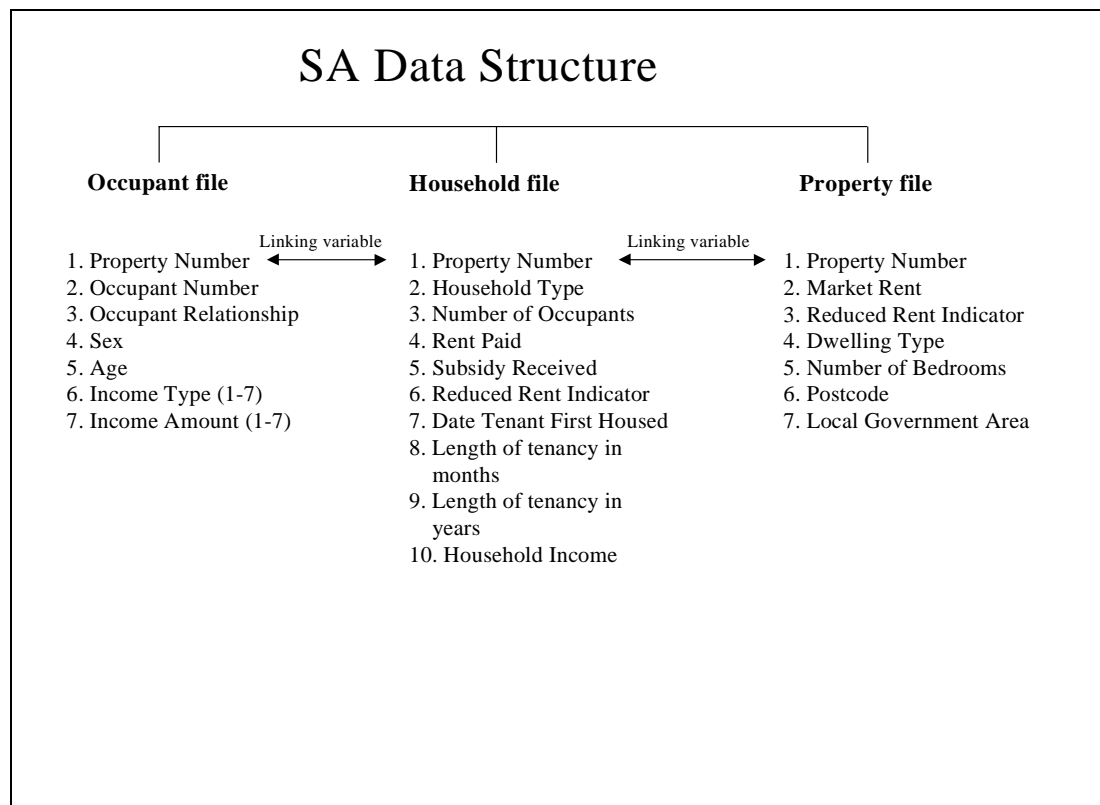
The Victorian data contains two data files, with Household ID linking records in the two files. The Person file contains 162,293 records and 15 data items. It has one record for each person in each household. Data includes identifiers for household, income unit and client (person), client characteristics (age, sex, marital status, role in household, relationship to main tenant) and their income source and amount, household type, rebated rent and market rent.

The household file contains 60,735 records and 9 data items. It includes household ID, stock type, dwelling type and size (number of bedrooms), household type, number of persons in household, date tenancy commenced, Local Government Area and market rent.

Aboriginal Rental Housing Program clients are included on the persons file but there is no matching data in the household file for ARHP households. Therefore there are person files with no matching household file in the Victoria data.

South Australia

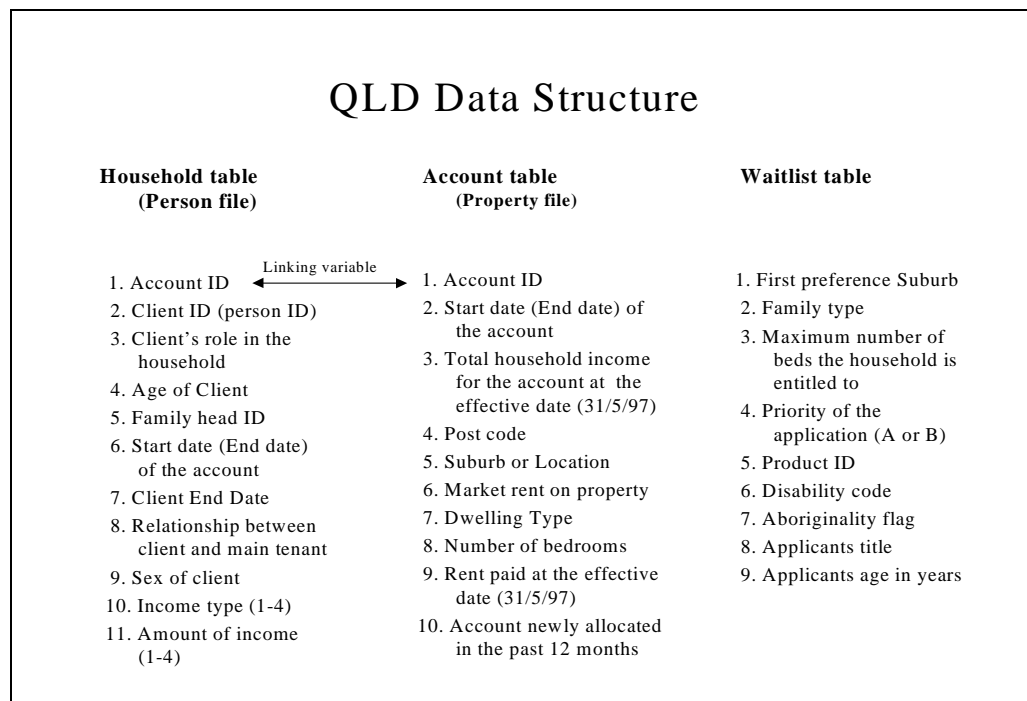
The following is a diagram of the South Australia public housing data structure:



As we can see from the above data structure diagram, SA data has three files (Occupant file, Household file and Property file) which are linked by the key variable 'Property Number'. The Occupant file contains 105,374 records, one record for each person living in a public dwelling. It provides each occupant's age, sex, relationship in household, and income by type of income (up to 7 are allowed). The Household file (56,780 records) covers all households in public dwellings and has information on household type, number of occupants in household, rent paid and rent subsidy received, household income and information on date of tenant first housed and length of tenancy. The Property file (59,352 records) contains information on public dwellings including dwelling type, if tenants are on reduced rent, Local Government Area, number of bedrooms and market rent.

Queensland

The following diagram shows the Queensland data structure.



Queensland provided three files (Account table, Household table, and Waitlist table). Account ID is used to link the first two of the three files.

The Account table includes 59,659 records and 11 data items. It gives details at the account level for current tenants and those tenants who have come and gone in the 12 months prior to the effective data date. Data include start date, end date, location, dwelling size (number of bedrooms) and type, total household income and rent paid at chosen date, market rent, and new allocations indicator. It appears that an 'account' refers to a particular property although this needs to be checked.

The Household table gives details about current individuals relating to an account (property) over last 12 months prior to effective data date. It includes 171,102 records and 19 data items. Of the 19 data items, 3 are identifiers for the account, client (person) and family head. Other data items include account start/end dates, client end date, household end date, client details (age, sex, household position, tenancy position, income and its source for up to 4 sources).

The Waitlist table contains 25,999 records and 10 data items. It gives details of households currently on the public housing waiting list. Data includes preferred location, family type, size entitled to, priority, product type (general stock, aged etc.), age, disability and Aboriginal and Torres Strait Islander codes. This data does not relate to the household or account table, persons/households currently in public housing may be on this list for some reason but cannot be identified as such in these data.

Summary

The following table summarises the size and formats of the public housing raw data files from the above four states.

Table 1: Summary of public housing data files for four states.

State	File	File format	Data records	Data items	Effective data date *
NSW	Property	Tab delimited free field file with variable name in first line, and no end of line mark	130,975	48	Not known
VIC	Person	Tab delimited free field file with variable name in first line	162,293	15	13/6/97
	Property	Comma delimited free field file with variable name in first line	60,735	9	13/6/97
QLD	Property	Fixed column text file	59,659	11	31/5/97
	Person	Same as above	171,102	19	31/5/97
	Waiting list	Same as above	25,999	10	31/5/97
SA	Household	Fixed column text file	56,780	10	30/6/97
	Occupant	Same as above	105,374	19	30/6/97
	Property	Same as above	59,352	7	30/6/97

* Effective date is the date on which data were extracted from state public housing databases.

Part 2: Problems in reading in the data

Each state provided documentation on the data sets to be used when reading the data in. However, the format and quality of the documentation varied across States. The problems we had when reading in the data mainly related to shortcomings or inaccuracies in the data documentation. Details of problems are listed below:

New South Wales

- The documentation for New South Wales data did not tell us that the data file was a comma delimited free field file. This was determined by checking the data using VI editor.
- With the date variable, it did not clearly state which date format was used.
- It did not tell how many records were in the data file, so that this could not be used to check that at least the correct number of records had been read in.
- The variable name was included in the first line of the data file but the accompanying documentation did not mention this.

Victoria

- The documentation for the Victoria data stated the data file type and records. It said that the two files were tab delimited free field files. However, the property file was a comma delimited free field file.
- The documentation stated that the person file included 165,175 records, when it actually only contained 162,293 records.
- There was one extra variable on the file which was not listed in the documentation.

- The order of two variables was not correct. This was discovered by checking variable labels including in first line of data file.
- The data documentation did not mention that the data files included variable labels in the first line.

Queensland

Two errors occurred in the documentation. They are:

- two variables 'IncAmnt5' and 'IncType5' were listed in the documentation, but were not on the data file; and
- position number was wrong for variable 'IncType4'. It should be '196' not '796'.

Apart from these two errors, Queensland provided a very good documentation. It listed clearly variable names, type, position, width and description. It also provided an electronic copy of all the codes used in the data file. However, it did not say how many records were on the files.

South Australia

SA provided the most comprehensive documentation. It stated clearly variable names, description, format used and position.

Part 3: Recommendation

As we can see from Part 1, the data structure of the public housing data files varied across state. Data items held in state databases are different for each state.

Developing a national public housing tenants database would obviously be much easier if a standard or common data form was sent by each state. If states cannot send data in the same form, then standard documentation practices would greatly assist the development of a national database.

Therefore, developing a pro forma for documentation for each state to follow is highly recommended. The main information which should be included in this pro forma is listed below.

Brief data file description

First, the documentation should briefly describe the files provided. Information included in this description of each file should be:

- file name;
- what the data in the file is concerned with (persons, dwellings, households,...);
- how many records are included in the file;
- what kind of data format is used in the data set (for example, comma or tab delimited or fixed column).

A list of data items

Second, the documentation should list information about all the data items or variables, including:

- variable name,
- description of variable,

- type (number, text, date, currency etc),
- position (if applicable),
- width (if applicable),
- format to be used to read in the variable,
- valid range.

Codes used for value of data

Codes used for each data item are also very important, so that invalid values can be identified and the data can be interpreted. Providing an electronic copy of codes used in the data file for relevant variables saves time required to key in code value and avoids typing errors. More importantly, it avoids code value errors and allows for more meaningful reports to be printed.

Data example

An example of data on the file is very useful for validating data after reading it in. For example, listing the first 10 records by the providing state would allow simple checks to be made.

Missing data

Documentation should include a description for each variable on how missing values are handled. If there are missing values in the data, it is preferable for these to be assigned an artificial number (eg, 99, or 999) rather than leaving them blank. This is especially important for free field files as SAS reads two commas or two tabs as a single delimiter. In other words, two commas or tabs together in the raw data are not recognised as representing missing data and lead to errors when reading the data in.

Common terms

Common terms need to be used in all state files. Currently, each state uses their own terminology to describe similar things. For example, to describe a file containing information about each person within each household, Queensland had a 'Household Table', SA had an 'Occupant File', and Victoria had a 'DSS_inc.qrd' file.

Attachment 1: List of data items on New South Wales public housing tenants and properties

Property vacant or occupied

Building/cladding material (eg, brick veneer, timber framed)

Dwelling type (eg, detached/semi-detached house)

Housing program (eg, GHS-general housing, state funded)

Department of Housing Administrative Region

Suburb from current address

Postcode from current address

Australian Bureau of Statistics Local Government Area

Department of Housing allocation zone

Department of Housing maintenance zone

Department of Housing client service team/zone

Tenancy start date

Market Rent, \$pw

Flag for tenant's spouse (Y/N)

Number of children/students

Number of other household members, (exc. Spouse and tenant)

Total household income, \$pw

Total Family Allowance, \$pw

Rent payable, \$pw

Whether rent automatically deducted by DSS

Whether rent automatically deducted by banks/building societies

Spouse income, \$pw

Income type, spouse

% spouse income assessed as rent payable

Tenant Income, including FAS if applicable

Income type, tenant

% tenant income assessed as rent payable

Income, other household member, number 1

Type of income, other household member, number 1

% tenant income assessed as rent payable, other household member number 1

Income, other household member, number 2

Type of income, other household member, number 2

% tenant income assessed as rent payable, other household member number 2

Income, other household member, number 3

Type of income, other household member, number 3

% tenant income assessed as rent payable, other household member number 3

Income, other household member, number 4

Type of income, other household member, number 4

% tenant income assessed as rent payable, other household member number 4

Total other household member's income

Number of bedrooms

Date property last vacated

Reason property last vacation

Status of property (eg, owned, leased)

Type of income (eg, statutory, with no additional income)

Date property first tenanted

Appendix 2: The SAS code for reading raw data and producing the final merged nationally consistent data