



Australian Government

Department of Social Services

National Centre for Longitudinal Data Access and Use Guidelines V2.0

Disclaimer

This document is not a stand-alone document and does not contain the entirety of Authorised Users' obligations. It should be read in conjunction with the terms and conditions set out in the Confidentiality Deed Poll. All capitalised terms in these Guidelines have the same meaning as in the Confidentiality Deed Poll. The purpose of this document is to provide applicants with information about how to apply for access to longitudinal Dataset held by the National Centre for Longitudinal Data and provide an explanation about user responsibilities. This document may be amended by the Department of Social Services from time to time at its absolute discretion.

Table of contents

1. Summary of Authorised Data User Responsibilities	3
2. Background.....	4
Available Datasets.....	4
3. Data applications	5
Storage	5
Usage	6
Publication	7
Notification.....	7
Failure to notify.....	8
DSS Rights.....	8
Commonwealth Legislation	9
Appendix 1 – Applying for access to data	10
Who is eligible for access to the NCLD Datasets?	10
Application – required information	11
Appendix 2 – Selecting a Dataset	14
What is the difference between a wave, a release and an update?	14
Do I need the Restricted Release Dataset?	14
HILDA special purpose datasets; HILDA Cross-National Equivalent File (CNEF)	16
LSAC special purpose Datasets: AEDC linked data file	17
LSAC special purpose Datasets: Child Health CheckPoint data file	17
LSAC special purpose Datasets: Centrelink linked data file	18
Appendix 3 – Understanding your responsibilities	21
Geographic data.....	21
Acknowledgements	21
FLoSse Research.....	22
Replication Materials	23
Appendix 4 – Previous data releases	24
Transition away from organisational licences	24
Individual licences for previous releases of the data.....	25
Appendix 5 – Using the Datasets for training and coursework.....	26
Authorised training activities and coursework	26
Appendix 6 – Definitions of terms.....	26

1. Summary of Authorised Data User Responsibilities

STORAGE:

Store all copies of the Datasets, whether full or partial, in a manner set out in the Guidelines.

Destroy or delete all copies of any Dataset you no longer need.

Understand that you are responsible for the Datasets unless and until all copies of the dataset in your control have been destroyed AND you have notified DSS of their destruction.

USAGE:

Do not attempt to identify individuals in any Dataset

If spontaneous recognition occurs, to not share the identity of any person within the Dataset.

Do not share unit record data with anyone who is not an authorised user of the same Dataset.

Do not match, share, link, integrate or merge any NCLD Dataset with other Datasets without written permission from DSS.

PUBLICATION:

Make all your published or finalised research publicly available in a manner set out in the Guidelines.

Do not publish any Unit Record Data or any research material that could lead to respondent identification.

Include the appropriate Dataset acknowledgement in all your research materials.

Only publish research material that reports at a level of detail (e.g. for geography) as set out in the Guidelines.

NOTIFICATION:

Notify DSS in the following cases:

- Change of personal details, including institutional affiliation
- Change or addition to intended research
- Change in expected completion date of research
- Terminating access to the Dataset
- Loss of the data or other potential breaches
- You are required by law to surrender the data to a third party.

2. Background

The Department of Social Services (DSS), on behalf of the Australian Government, undertakes research and policy development to support the lifetime wellbeing of individuals, families and communities.

To support its objectives, DSS funds four longitudinal studies that collect information about the lives of Australians and the social, economic, housing and family issues that impact them:

- The Household, Income and Labour Dynamics in Australia (HILDA) Survey
- *Growing up in Australia: The Longitudinal Study of Australian Children (LSAC)*
- *Footprints in Time: The Longitudinal Study of Indigenous Children (LSIC)*
- *Building a New Life in Australia (BNLA): The Longitudinal Study of Humanitarian Migrants.*

The National Centre for Longitudinal Data (NCLD) within DSS is responsible for the management of the studies and their output Datasets. Researchers from government, academic institutions and not-for-profit organisations may be granted access to the Datasets at the discretion of the NCLD (refer to Appendix 1). Researchers who have been granted access to DSS longitudinal Datasets are Authorised Users and must agree to and abide by the terms and conditions set out in the Confidentiality Deed Poll. A Deed Poll is a legal document binding one person to an active intention, in this case, keeping and using the Datasets as set out in these Guidelines.

DSS has engaged the Australian Data Archive (ADA) to assist with the management and monitoring of access to the Datasets.

Through ADA, DSS will, provide Authorised Users with:

- longitudinal data in the form of ready to use Datasets in STATA, SAS, SPSS, and tab delimited formats; and
- supporting documentation for use of the Datasets including Data Users Guides, Data Dictionaries and frequencies.

Available Datasets

Datasets available include the following:

- HILDA General Release
- HILDA Restricted Release
- HILDA Cross-National Equivalent File (CNEF)
- LSAC General Release
- LSAC Restricted Release
- LSAC AEDC linked dataset
- LSAC Centrelink General Release
- LSAC Centrelink Restricted Release
- LSAC Checkpoint General releases
- LSAC Checkpoint Restricted Release

- LSIC General Release
- BNLA General Release

Each release of General and Restricted data includes the latest wave in addition to all previous waves. Authorised Users must apply separately for each new release of data. More detailed information about the Datasets and the differences between each General and Restricted Release Dataset is available in Appendix 2 and in the supporting documentation (such as Data Dictionaries and Data Users Manuals) published on the Dataverse website at the time of each new data release.

3. Data applications

Any person wishing to have access to DSS longitudinal Datasets must apply online through the NCLD Dataverse site which is managed by the ADA (<https://dataverse.ada.edu.au>) Detailed information about the application process is available at Appendix 1. Further questions can be addressed to ada@anu.edu.au or NCLDresearch@dss.gov.au.

Access to the Datasets is conditional on agreement to the terms and conditions set out in the **Confidentiality Deed Poll** (the Deed) which must be agreed to by an individual at the time they first apply for a Dataset. The Deed only needs to be signed once unless the Deed is substantially modified in future. To execute the Deed correctly, complete all fields electronically except for the signature fields on the final execution page. Print the entire document and sign the Deed in black or blue pen. Other fields may also be completed in pen after printing. Once complete and signed, scan the entire document and email to ada@anu.edu.au and ncldresearch@dss.gov.au. If you do are unable to scan the document, you can send the original signed hard copy of the Deed to:

Research and Methods
National Centre for Longitudinal Data
PO Box 9820 Canberra ACT 2601

The NCLD must receive the signed Deed before access to any Datasets will be approved. Please note that while a scanned Deed with a wet signature is acceptable, a Deed with an electronic signature is not.

Users of the Datasets under previous licencing arrangements (organisational or individual) must also complete the **NCLD Data Holdings Form** specifying which of the Datasets to which they have been given access they wish to retain or relinquish. The terms of the Confidentiality Deed Poll will apply to all previous Datasets retained by Authorised Users. Send the completed form to to ada@anu.edu.au and ncldresearch@dss.gov.au (See also Appendix 4)

Storage

Authorised Users must store all copies of the data, whether full or partial, using one of the following storage options:

- On a cloud server which has been awarded Australian Signals Directorate (ASD) certification to provide cloud services; or
- On a secure network; or
- On a password protected removable or hard drive; or
- On a local drive on password protected PC or laptop.

When not in use, removable drives and printed material containing Unit Record Data must be kept in a locked drawer or cabinet to which only Authorised Users have access.

The following conditions apply to storage of DSS longitudinal data on a cloud server:

- Storage on an Australian private cloud server is permitted where evidence can be provided that the cloud service provider has completed an Information Security Registered Assessors Program (IRAP) security assessment and been awarded Australian Signals Directorate (ASD) certification to provide cloud services to Australian Government organisations. A list of certified providers is available at https://www.asd.gov.au/infosec/irap/certified_clouds.htm
- Any service agreement or contract entered into between the Authorised User and the cloud service provider must not change or transfer ownership of the Dataset.
- ASD-certified cloud storage complies with Australian Privacy Principle (APP) 11 – Security of personal information. However, ASD-certified cloud storage does not mitigate the need for Authorised Users to adhere to the security requirements outlined in this section of the Protocols and in the Terms and Conditions of Use.

If an Authorised User no longer requires the Datasets they must:

- cancel access to Datasets held on a secure server;
- remove copies from the Cloud server;
- delete all copies of the Dataset held on computer and removable hard drives;
- physically destroy media such as CDs and DVDs; and
- shred printouts of Unit Record Data.

DSS reserves the right to terminate the Deed and remove access to the Datasets and request Authorised Users to destroy or delete all copies of the Datasets in their possession.

Usage

Authorised Users must not attempt to identify individuals in the Datasets. In the unlikely event that this occurs spontaneously or accidentally, Authorised Users must not disclose the identity of the individual to any other person.

Authorised Users must not show or share Unit Record Data with any other person unless that other person is an Authorised User of the same or later release of the same Dataset. Aggregated data may be shown to non-Authorised Users.

Authorised Users must not match, share, merge or link any Dataset with any other dataset without prior written permission from DSS. This includes joining in any way two or more datasets either specifically to identify individuals or to increase the amount of information known about each individual. Authorised Users are asked during the application process

whether they need to merge or link to other datasets. Users wishing to do this will need to provide details of what they propose to do. DSS will consider applications on a case by case basis. If permission is granted and if requested by DSS, Authorised Users may be asked to provide their matched, shared, merged or linked data to DSS for inclusion in future releases of the data.

Authorised Users must not use the ACARA My School institution level data (available in LSAC and LSIC) to calculate or publish material which ranks schools or creates league tables.

The Datasets may be used for training purposes. Both course instructors and course participants must apply for the relevant Datasets. Additional information about the use of DSS longitudinal Datasets for training purposes is set out in Appendix 5.

Publication

Authorised Users must not publish any Unit Record Data or any Research Material that could lead to respondent identification. Methods to limit respondent identification include but are not limited to confidentialisation or cell suppression where the cell count is between one and four and inclusion of small cell sizes would risk identification of individual respondents.

Publication of free text responses is permitted providing any accompanying contextual information (e.g. sex or age of respondent) does not enable identification of any individual

Authorised Users may only report at a level of detail available in the General Release of the Dataset. Authorised Users of the HILDA and LSAC Restricted Datasets must not report, for example, at the postcode level or at the level of rankings of Social-Economic Indexes for Areas (SEIFA). Authorised Users wanting to report at a level of detail other than at a level available in the General Release must seek prior written approval from DSS. Information about the geographic variables in the Datasets is set out at Appendix 3.

Authorised Users must make publicly available all research resulting from the use of the data. Within 30 days of publication or finalisation, Authorised Users are required to upload bibliographic details of published material to FLoSse Research at flosse.dss.gov.au. More information about this is provided at Appendix 3. It is the responsibility of the Authorised User to ensure that any third party who holds the Intellectual Property in the research understands and agrees to this requirement.

Research conducted for the purpose of internal administration of Australian Government agencies or for confidential business purposes of the Australian Government is exempt from this requirement.

Authorised Users are encouraged but not required to upload replication and research material to the NCLD Dataverse site. DSS holds no responsibility for the accuracy of replication material uploaded by Authorised Users to the NCLD Dataverse site (refer to Appendix 3) <https://dataverse.ada.edu.au/dataverse/nclid>

Notification

Authorised Users must immediately notify DSS via email to ada@anu.edu.au in the following situations:

- a. Change of personal details – this includes but is not limited to changes in name, contact details (phone number or email address) or institutional affiliations.
- b. Change to or addition of research project details – Authorised Users must inform DSS of all major research projects for which they intend to use the Datasets.
- c. Change in expected completion date of research.
- d. Access to the data is no longer required – Authorised Users must confirm they have destroyed or deleted all copies of the Datasets no longer required. Authorised Users will remain legally responsible for the Datasets until this is done. For more information about Authorised Users responsibilities for Datasets obtained prior to the introduction of the Dataverse application system, refer to Appendix 4.
- e. Breaches or loss of data – Authorised Users must notify DSS immediately if they lose or are otherwise unable to locate a copy of the Dataset for which they are responsible or if they become aware of any breaches of the terms and conditions of access in relation to their own or other Authorised Users copies of the data.
- f. Legal requirement to surrender the data to a third party – Authorised Users must notify DSS as soon as possible if they are required by law to surrender the data to a third party.

Failure to notify

Failure to comply with these requirements will result in the loss of access to the Datasets and may preclude access to the Datasets in the future.

DSS Rights

The Commonwealth owns all intellectual property rights in the data. It does not own the intellectual property rights in any of the Research Material except research agreements where specified by research agreements with the Commonwealth. Authorised Users must include the appropriate acknowledgement and disclaimer in all Research Material in accordance with the requirements set out in Appendix 3.

Initial and ongoing access to the Datasets is at the discretion of the Commonwealth. The Commonwealth may revoke access to the Datasets and require the Authorised user to destroy or delete all documents or copies of the Datasets in their possession.

DSS also has the right to request Authorised Users to provide details relevant to the management of the Datasets. DSS may, through the ADA or independently, seek to verify an applicant's identity and/or confirm the intended use of the Dataset.

For the purposes of data access, or to facilitate ongoing user support, including responses to user queries, Authorised Users' email addresses may be provided to the Authorised User's employer or to organisations contracted by DSS to deliver the Datasets.

- Melbourne Institute of Applied Economic and Social Research (HILDA);
- Australian Institute of Family Studies (LSAC & BNLA);
- Australian Bureau of Statistics (LSAC);

- Australian Data Archive (All Datasets).

The Australian Government and the ADA shall not be held responsible for the accuracy, completeness and fitness for purpose of the data supplied.

DSS may amend these Guidelines from time to time at its absolute discretion and notify Authorised Users.

Commonwealth Legislation

The Authorised User agrees to comply with the law in force in the Commonwealth, including but not limited to:

- *Criminal Code Act 1995*; and
- *Privacy Act 1988* (the Privacy Act).

Appendix 1 – Applying for access to data

Who is eligible for access to the NCLD Datasets?

Access to the Datasets is at the discretion of the NCLD. The majority of users of the Datasets are Australian academics, post-graduate students and employees of government agencies.

The following types of users may be considered for access to NCLD Datasets.

Users in Australia

- Employees of an Australian Government agency
- Employees of an Australian State/Territory Government agency
- Employees of an Australian Local Government agency
- Academics at an Australian University, Institute of Technology or TAFE
- Researchers or research assistants, attached to an Australian University, Institute of Technology or TAFE
- Australian Honours (General Release only), Masters and PhD students at an Australian University, Institute of Technology or TAFE
- Researchers from community based organisations
- Other Australian users whose primary business is research may be given access to the Datasets if it is to be used for non-profit purposes and research outputs are to be placed in the public domain.

Any applicants commissioned by government or another organisation to do research, should note this in their application and identify the commissioning organisation

Users outside Australia

People working outside Australia may apply for the General Release Datasets and the Cross-National Equivalent File (CNEF). All applications for data coming from outside Australia will be considered on a case by case basis.

Access to NCLD Restricted Release Datasets will not be approved for applicants based outside Australia. If you are outside Australia and wish to use the Restricted Release data, the best way to do this is to collaborate with an Australian based researcher who can conduct the data analysis.

Email verification

At the point of downloading files the Authorised User's IP address will be captured as a point of verification of location. Authorised Users in some locations, or unspecified locations, will not be permitted to download data files without further verification. Only organisational email addresses (such as that provided by an agency or university with which you are employed or affiliated) will be accepted. Applicants who register with a personal email address (such as a gmail address) will be asked to provide a valid organisational email address.

Applicants are asked to supply an alternative email address. This secondary email address does not have to be an organisational email address and will only be used if we cannot contact you on your organisational email address.

Agreement to Terms and Conditions

Applicants must send their signed Confidentiality Deed Poll to to ada@anu.edu.au and ncldresearch@dss.gov.au before an initial application for data will be approved. Once signed, this Deed will apply to all DSS longitudinal Datasets and will endure for subsequent applications. Any variation to the terms and conditions in the Deed must be agreed to in writing by DSS and the Authorised User. To the extent there is any inconsistency between these Guidelines and the terms and conditions in the Deed, the terms and conditions in the Deed will prevail.

Authorised Users should retain a copy of their Deed for their records.

Application – required information

You will be asked to provide the following information when you apply for access to NCLD Datasets.

General details for administration and data access approval (All Datasets)

- Your full name *
- Your email address – this must be an organisational email address *
- The name of the institution or organisation with which you are affiliated *
- Your position at the institution or organisation. If you are a student you must specify your current level of study and provide the name and contact details of your supervisor or instructor *
- Type of organisation – for example, Australian Government, State or Territory Government, Local Government, Not-for profit, academia, private sector or private individual
- Location of your organisation – Australia or other
- Secondary email address
- Your telephone number
- Open Researcher and Contributor Identification (ORCID) identifier
- Whether you been an Authorised User of [this] Dataset previously
- The data storage options you intend to use – for example, IRAP assessed cloud server, secure network, password protected removable drive or local drive on password protected PC or laptop. If you store the data on a Cloud Server you must provide the name of the Cloud Service Provider.
- Information about any linkages you propose to undertake with the Datasets

* These fields will be prepopulated from your Dataverse registration. You can update these fields at any time by logging into Dataverse and editing your account information.

Project Details (All Datasets)

- A description of the project(s) you intend to use the data for (max 250 words) +
- Source of funding (if applicable)

- Grant number (if applicable)
- Anticipated project end date +
- Principle purpose of access to or use of the Dataset – for example, academic research, policy/program development or evaluation, professional consulting services (for Government or Not-for profit sectors only - please specify), data user training instructor, data user training student, other (please specify)
- Whether you intend to link the data to other data – if yes, you will be asked for details +
- The type of research outputs you expect to produce from using the data (such as journal article, thesis, report, etc.).

+ Further changes to your data application, such as to your research proposal (including additional research projects) expected end date, and proposed data linkages, can be made by emailing ada@anu.edu.au and DSS can approve and append these to your file.

Additional details for Restricted Release Datasets (HILDA and LSAC only)

- Specify the reasons that the data in the General Release is not sufficient for your research purposes including the Restricted Release variable(s) you are seeking access to.

Details for application for LSAC additional linked Datasets

- Whether you require access to the AEDC linked data file and the reason. If yes, please note your application will be forwarded to the Department of Education through the Social Research Centre (SRC), the Data Management Agency for the AEDC on behalf of the Department of Education, for approval.
- Whether you require access to the Child Health CheckPoint data file and reason. If yes, please note your details will be provided to the Murdoch Children’s Research Institute.
- Whether you require access to the Centrelink linked data file and reason.

Cultural standpoint (LSIC only)

- All LSIC data users will be asked to provide a cultural standpoint. You will also be asked whether you are Aboriginal and/or Torres Strait Islander, whether you will be collaborating with Aboriginal and/or Torres Strait Islander researchers, and whether you have any prior research experience relating to Aboriginal and/or Torres Strait Islander peoples. You will also be asked to describe how your own cultural background may influence your research and how you will address this. In completing your cultural standpoint, you may like to consider such things as your family background and upbringing, your educational background, previous experiences with and knowledge of Aboriginal and Torres Strait Islander peoples and cultures. Your cultural standpoint is requested by the LSIC Steering Committee to ensure that researchers have given some consideration to how their own point of view may influence their research and findings. (For more information refer to “What

is meant by cultural competency and cultural standpoint?” on the LSIC Dataverse site.)

Application for linkage

- Researchers wishing to undertake any linkage with HILDA, LSAC, LSIC or BNLA will need to provide information on
 - the dataset(s) they propose to link to
 - who the custodian of the data is
 - whether the data is publicly available
 - whether there is an associated cost
 - timeframes for when the linked data is required
- Note that some common linkage variables such as geocoding, dates of birth, dates of interview etc., that are not released publicly may be held by DSS or partner agencies, so linkages may be possible as part of a collaboration, even if they are not possible for the applicant to perform.
- Applications for linkage with LSAC are considered on a case by case basis by the LSAC Management Group.
- DSS will assess applications for linkage to HILDA, LSIC and BNLA on a case by case basis, and may contact the applicant to discuss options, such as:
 - Who should integrate the datasets (i.e. DSS or the applicant)
 - Whether linked data could be provided to all data users with subsequent releases of the Dataset.

Submitting the application request

By submitting your request for access you are agreeing to abide by the requirements set out in the Guidelines as well as being bound by the Confidentiality Deed Poll. Applications will not be approved unless DSS has received a signed Deed.

Collaborative research:

The new access arrangements and Deed are tied to approval of access for each individual user. Each researcher collaborating on a research project who needs to use or otherwise have access to the Datasets will need to request access separately. However, it is worth noting that they won't all need to download the data separately, as the Deed does not prohibit providing a copy of the Dataset to another Authorised User of the same Dataset. You can provide access to the Dataset to another Authorised User once they have been approved as long as storage of the Dataset continues to follow the security protocols (e.g. on a secure server). However, please note that if a shared copy of the Dataset is stored on a secure server, the Authorised User responsible for uploading it to the server is responsible for deleting it from the server at the completion of the research or when leaving the organisation.

Students should note that while supervisors are not required to have separate access to the Dataset, you may not share Unit Record Data with your supervisor unless they are an Authorised User for the same Dataset.

Appendix 2 – Selecting a Dataset

When you apply for DSS longitudinal data, you will need to determine exactly which Dataset will meet your research needs. This appendix provides information about the various Datasets for which you can apply.

Researchers may apply for data from any or all of DSS longitudinal studies (HILDA, LSAC, LSIC and BNLA). If applicants want to access more than one Dataset, they must complete a separate application for each Dataset. Some studies issue more than one version of the data and/or supplementary Datasets. You will need to understand what these are and the special requirements attached to these in order to make an informed decision about what you require for your research. In addition to the General Release Dataset for each of the studies, HILDA and LSAC also have a Restricted Release version. LSAC also has three supplementary Datasets which are only supplied if specifically requested; the Australian Early Development Census (AEDC), Checkpoint and Centrelink Datasets (explained in more detail later).

What is the difference between a wave, a release and an update?

A wave of data collection refers to the collection of a particular set of questions from the entire sample. This usually occurs within a defined period, which might, for example be several months or over the course of a year. These are also sometimes called sweeps. In addition to the data collected for specific waves, administrative data is linked to the survey data which may provide continuous data or spells of data, or collection periods that do not relate to a particular wave of the study (e.g. Centrelink income support data, or NAPLAN data).

A release of data occurs when a new wave of data is sent out; and it contains the new wave in addition to all previous waves (e.g. Release 16.0 of HILDA includes wave 16 as well as waves 1-15). When a new release occurs there may be some changes to earlier waves, for example correction of errors, changes to naming or labelling conventions or the addition of derived variables. Because of this, it is important to refer to the data release used when publishing because precise replication may not be possible if using earlier or later releases. Importantly, if you are approved to access a specific release of a Dataset, you will need to reapply if you wish to download a newer release when it becomes available.

An update occurs when edits or additions are made to an existing release. For example, an update to Release 16.0 of HILDA would result in it being reissued as Release 16.1. This is avoided as far as possible, but a major edit or addition can precipitate an update. You do not need to reapply in order to receive an update. If you are an Authorised User, you will receive an email notification and will be able to download the updated Dataset. Information about the new or changed data will be included in the notification.

Do I need the Restricted Release Dataset?

Restricted Release Datasets are available for HILDA and LSAC only. If you wish to apply for a Restricted Release Dataset, you will need to justify your need for the Restricted Release

Dataset in your research. Approval will only be granted if you can show a legitimate reason for seeking access.

Items appearing only in the Restricted Release Datasets generally provide more detailed information on items already in General Release Datasets. Items impacted include geography, specific employment characteristics and occupational codes, date of birth, costs, hours and types of childcare, housing payment amounts, interview outcome codes, and interview start and end times. In the General Release Dataset these items may be top or bottom coded (for example income) or collapsed into larger groups (for example geographic variables, birth dates) or confidentialised in some other way. Importantly, although these variables are available at a more detailed level, there are restrictions on reporting at these levels (refer to the Appendix 3). The most commonly requested items appear in Tables 1 and 2.

Table 1: Examples of data differences between HILDA General and Restricted Releases

General Release HILDA	Restricted Release HILDA
Year of birth	Date of birth
Income, wealth and age - highest values are top-coded	Income and wealth - all values
State, Remoteness Area, Greater Capital City Statistical Area, Major Statistical Region, Section of State, SEIFA deciles, distance moved since last wave	Additionally- postcode, census collection district, Local Government Area, Statistical Division, Statistical Subdivision, Statistical Local Area, SEIFA rankings, SA1, SA2, SA3, SA4
Occupation (ANZSCO) – coded to two digit level	Occupation (ANZSCO) – coded to four digit level
Childcare data aggregated to household level	Costs, hours & types of childcare – all variables at individual child level

Table 2: Examples of data differences between LSAC General and Restricted Releases

General Release LSAC	Restricted Release LSAC
Year of birth	Date of birth
Income, housing costs, child support paid by parent 2, children' and parents' current height, weight and waist circumference,	Income, housing costs, child support paid by parent 2, children' and parents' current height, weight and waist circumference, number of hours spent in child care

number of hours spent in child care (highest values are top-coded)	
State, Remoteness Area, Greater Capital City Statistical Area, Major statistical region, Section of state, SEIFA index rounded to nearest 10, distance to coast	Additionally- postcode, census collection district, local government area, statistical division, statistical subdivision, statistical local area, SEIFA rankings, SA1, SA2, SA3, SA4.
Occupation (ANZSCO) – coded to two digit level	Occupation (ANZSCO) – coded to four digit level

The HILDA data dictionary only includes variables available in the General Release Datasets. The LSAC data dictionary includes both General and Restricted Release variables. It is recommended that you check the following website links to determine whether the General Release Dataset will provide the level of detail you need:

- <https://www.online.fbe.unimelb.edu.au/HILDAodd/Default.aspx> and
- <http://data.growingupinaustralia.gov.au/data/datadict/index.html>

Additional information about the Restricted Release Datasets may be found in user manuals for the relevant studies. Section 3.12 “Confidentialisation” of the HILDA manual (<https://melbourneinstitute.unimelb.edu.au/hilda/for-data-users/user-manuals>) provides information about top-coding and section 4.3 “Geography” provides information about derived geographic variables (many of which only appear in the Restricted Release files). The LSAC user manual is available at <http://data.growingupinaustralia.gov.au/data/docs/userguide/index.html>

If you are still unsure, you can email NCLDresearch@dss.gov.au and explain what you need for your research and we will provide advice on which Dataset most appropriately fits your research requirements.

Authorised Users approved for access to the Restricted Release Datasets will need to put in a separate application if they also require access to the General Release Datasets.

HILDA special purpose datasets; HILDA Cross-National Equivalent File (CNEF)

The HILDA Cross-National Equivalent File is automatically included within the HILDA General Release and Restricted Release Datasets.

The HILDA-CNEF Dataset is a subset of HILDA data developed specifically for use as part of the Cross-National Equivalent File (CNEF) which is managed by Ohio State University, USA (<https://cnef.ehe.osu.edu/>). The CNEF contains equivalently defined variables for the following panel studies:

- Panel Study of Income Dynamics (USA),
- German Socio-Economic Panel,

- British Household Panel Study,
- Canadian Survey of Labour and Income Dynamics,
- The Household, Income and Labour Dynamics in Australia (HILDA) Survey,
- Swiss Household Panel,
- Korea Labour and Income Panel Study, and
- The Russia Longitudinal Monitoring Survey-Higher School of Economics.

The Datasets are designed to allow cross-national researchers with less experience in panel data analysis to have access to a simplified version of these panel data, while providing experienced panel data users with guidelines for formulating equivalent variables across countries. Most importantly, the CNEF provides a set of constructed variables (for example, pre-and post-government income and international household equivalence weights) that are not directly available in the original Datasets.

Access policies for the Datasets within the CNEF vary depending upon the administrative requirements of the individual surveys from which the Datasets are taken. DSS is only responsible for and can only approve and provide access to the HILDA-CNEF. Access to the HILDA-CNEF data is obtained by applying via HILDA Dataverse for either the General or Restricted Release.

Details on how to order all other CNEF data files can be viewed on the Ohio State University CNEF website: <http://cnef.ehe.osu.edu/>.

LSAC special purpose Datasets: AEDC linked data file

AEDC linked data is not automatically provided with either the General or Restricted LSAC Releases, but can be requested as an additional file accompanying either one.

Applicants wishing to access the AEDC linked data file must specify so separately and provide details of how the data will be used. All applications for the AEDC linked data will be forwarded to the Social Research Centre (SRC) for approval. Access to the AEDC linked data is contingent upon this approval. However, approval for access to LSAC is not contingent upon a successful application for the AEDC linked data.

Use of the AEDC linked data is subject to Authorised Users following guidelines on conduct and publication of research as outlined in the [AEDC Data Guidelines](#). The guidelines require data users to notify Department of Education and Training of intent to publish material using AEDC data (email support@aedc.gov.au)

LSAC special purpose Datasets: Child Health CheckPoint data file

The Child Health CheckPoint file contains data collected in 2014-2015 from LSAC children in the B cohort to help address gaps in the evidence on the antecedents of adult disease. Linking LSAC's existing early-life data with a suite of measures allows researchers to explore early-life mechanisms underlying health gradients of the children now and the adults they become. Further information can be found at <https://www.mcri.edu.au/research/projects/longitudinal-study-australian-childrens-child-health-checkpoint>

Child Health CheckPoint data is not routinely provided with either the General or Restricted Release LSAC data, but can be requested as an additional file accompanying either one. Child Health CheckPoint data is available in General or Restricted Release formats and data users should justify the reason for needing the Restricted Release. Refer to the CheckPoint Data User Guide available on LSAC Dataverse for information about the difference between the two release formats.

Applicants wishing to access the Child Health CheckPoint data file must explain which version they need and why they need it in the LSAC application *and* agree to have their details provided to the Murdoch Children's Research Institute. Approval is at the discretion of the NCLD and on a case by case basis.

LSAC special purpose Datasets: Centrelink linked data file

Centrelink linked data is not automatically provided with either the General or Restricted LSAC data, but can be requested as an additional file accompanying either one. Centrelink data is available in General or Restricted Release formats and data users should justify the reason for needing the Restricted Release.

Applicants wishing to access the Centrelink linked data file must explain why they need it in the LSAC application. Approval is at the discretion of the NCLD and on a case by case basis.

Background information on Centrelink linkage files

During wave 7 enumeration, consent was collected from the K cohort study child's parents (P1 and P2) to link their Centrelink welfare benefits data back to 1 January 1999 and from the K cohort study child to link back to their 16th birthday. The data includes information on income support payments, Family Tax Benefit, Carer Allowance and concession cards. The data released with wave 7 is extracted up until the end of the 2016-17 financial year (30 June 2017) apart from the Family Tax Benefit data which is only extracted up until 30 June 2015 because it is based on the entitlement calculated after reconciliation with tax data.

The linked Centrelink data is provided in separate Datasets from the main LSAC data files and there are both General Release and Restricted Release versions. Neither of these versions are supplied automatically with the LSAC data files and must be explicitly requested. Access to either (or both) versions of the Centrelink data can be requested by an applicant when applying for the General the Restricted Release LSAC data. Applicants can also apply for the Restricted Release of the Centrelink data when applying for the General Release LSAC file, and applicants applying for Restricted Release LSAC data can choose to apply for the General Release Centrelink data.

Applicants for the restricted Centrelink files are required to present a project rationale for access to the Restricted Release data making it clear why this data is essential for their research. Applicants must specify why particular data items are required or why the research questions require access to income support data in episodic form

General Release Centrelink Files

File Name: ISP_Summary

Contains data for income support payments receipt (ISP) aggregated at financial year level. For each participant who has received an income support payment in a particular year there will be a single observation. Information provided in the summary file includes:

- Income support benefit type received by the participant for the greatest duration during the year
- Number of days that the participant received an income support payment and duration they received the primary benefit type
- Duration in receipt of rent assistance, home ownership status and rent type
- Number of days the participant received other income while in receipt of an income support payment
- Number of days the participant was partnered
- Indicators for receipt of carer allowance payment and low income card

File Name: FTB_Summary

Contains data for Family Tax Benefit (FTB) summarised aggregated at financial year level based on a participant's reconciled eligibility and entitlement determined after receipt of their taxable income provided by the Australian Taxation Office (ATO). Information is only provided up to two years prior to the extraction date at which point the data is considered 'mature' (i.e. the vast majority have tax data against which their entitlement can be reconciled). Information provided includes:

- Number of days the participant was eligible for FTB (in total), FTB-A and FTB-B
- Number of days the participant was eligible for an ISP while eligible for FTB
- Number of days customer was partnered with a primary partner whilst eligible for FTB
- Number of days the participant was partnered with ex-partners whilst eligible for FTB
- Number of children assessed as FTB children
- Total validated adjusted taxable income (customer + primary partner + ex-partners)

File Name: Concession_cards

Contains episodes of concession cards data for participants where a participant held a concession card. As a participant can have multiple concession cards during the same time period, in some cases this file has overlapping episodes of concession cards for a participant. Information includes the benefit type which qualified them for a concession, the concession card type, eligibility start and end dates, and the number of dependent children.

Restricted Release Centrelink Files

In addition to the files provided with the General Release Centrelink data, the Restricted Release Centrelink data includes the following files.

File Name: ISP_Episodic

Holds the information for each episode of ISP receipt. In addition to the variables in the summary file the following information is provided:

- Entitlement rate
- Activity requirements
- Reason for end of payment
- Earnings amount and work hours
- Educational details – student status, course level and type, highest educational level before episode
- Rent amount
- Homelessness
- Medical conditions (currently a binary indicator pending confidentialisation) and impairment rating
- Vulnerability indicator

File Name: FTB_Customer_Reconciled

Has the same structure as the ISP_Summary file. Additional information provided includes:

- Age, citizenship, Indigenous indicator, overseas indicator, preferred written language, remoteness area
- Number of days eligible for FTB-A (by rate type)
- Number of days eligible for of FTB-B
- FTB-A and FTB-B pre-reconciliation eligibility amounts (paid and notional)
- FTB-A and FTB-B post-reconciliation entitlement amounts
- Maintenance Income and amount of FTB-A not paid due to Maintenance Income test
- Number of days overseas
- Count of FTB shared care children
- Number of days also eligible for an ISP
- Adjusted taxable income broken down by components

File Name: FTB_Child_Reconciled

This file holds the reconciled data for the FTB children for which a participant received FTB payments in an entitlement year. The data contains one observation for each FTB customer/FTB child combination for each entitlement year during which the participant/customer received FTB payment for the corresponding child. Details for children aged 16 or over not included due to privacy considerations. Information includes:

- Age, gender, overseas indicator and duration
- Post-reconciliation durations for FTB-A and FTB-B
- Regular and shared care durations
- FTB-A supplement amount

Appendix 3 – Understanding your responsibilities

Geographic data

As a general rule, researchers may publish research findings at any level of geographic detail that is available in the General Release Datasets. Geographic variables in the General Release Datasets¹ have been included on the basis that they are unlikely to enable identification of individuals in the data. However, if restricting the reported data by combining with other variables, researchers need to ensure that cell sizes reported are not low enough to make identification of individuals possible.

As LSIC and BNLA have small samples and both used a cluster sampling technique, a variable showing the state or territory in which a respondent lives would significantly increase the likelihood of identification and is therefore not available for these two studies.

Lower level geography such as postcode and local government area is available in Restricted Release Datasets but Authorised Users with access to the Restricted Release Datasets may not publish research at these lower levels and should note that the Datasets are not representative at this level. LSAC General Release data includes a postcode indicator variable which identifies respondents in the same postcode with a randomly assigned number so that the actual postcode is not identified.

Researchers wishing to report at a level of geographic detail other than that which is available in the General Release data will need to request permission from DSS. Requests will be considered on a case by case basis.

Researchers may use detailed geographical data to identify respondents living in areas with particular characteristics (e.g. respondents living in areas affected by adverse weather conditions during a certain period). This will generally involve linkage with other datasets. Authorised users wishing to link geospatial data must provide additional information about how they propose to undertake the linkage. Researchers may not use linked data to report at a level of geography more detailed than available with the General Release without written permission from DSS.

Acknowledgements

The Australian Government owns all intellectual property rights in the Datasets. Authorised Users must acknowledge DSS in all research material using the following acknowledgement:

This paper uses unit record data from [insert full name of study/ies and abbreviation here]. [Abbreviated name/s] conducted by the Australian Government Department of Social Services (DSS). The findings and views reported in this paper, however, are those of the author[s] and should not be attributed to the Australian Government, DSS, or any of DSS' contractors or partners. DOI: #####

¹ Details of differences between General and Restricted Release Datasets are provided in Appendix 2.

Insert the appropriate name or names of the Datasets you used in a single acknowledgement as per the following:

Full name of study	Abbreviation
Household Income and Labour Dynamics in Australia	HILDA
Growing Up in Australia: the Longitudinal Study of Australian Children	LSAC
Footprints in Time: the Longitudinal Study of Indigenous Children	LSIC
Building a New Life in Australia: the Longitudinal Study of Humanitarian Migrants	BNLA

The DOI (digital object identifier) references the Dataset(s) used and can be found on the download page of the NCLD Dataverse site. The DOI is circled in the example below.

The Household, Income and Labour Dynamics in Australia (HILDA) Survey, RESTRICTED RELEASE 16 (Waves 1-16)
 Feb 7, 2018 - HILDA Dataverse

Department of Social Services; Melbourne Institute of Applied Economic and Social Research. 2017. The Household, Income and Labour Dynamics in Australia (HILDA) Survey, RESTRICTED RELEASE 16 (Waves 1-16). doi:10.4225/87/QFUIBM. HILDA Dataverse, V3

The Household, Income and Labour Dynamics in Australia (HILDA) Survey is a nationally representative longitudinal study of Australian households which commenced in 2001. Funded by the Australian Government Department of Social Services (DSS), the HILDA Survey is managed by the Me...
 Series Information: on the survey is available at the NCLD website.

FLoSse Research

DSS longitudinal studies are funded by the Australian Government so that research resulting from their use can form part of the evidence base for policy development. As such, research resulting from the use of data from the studies is required to be made publicly available².

When they publish or otherwise finalise their research, researchers must upload bibliographic details of their research onto the FLoSse Research website. FLoSse Research is a publicly available searchable repository of research which uses one or more of DSS longitudinal studies. Details of research must be uploaded onto FLoSse Research (flosse.dss.gov.au) within 30 days of completion or publication.

Types of research that should be uploaded to FLoSse include but are not limited to:

- Annual reports
- Audio and visual presentations
- Books and book chapters
- Information Booklets, brochures and newsletters
- Journal article
- Presentations and conference papers
- Research papers
- Technical, working papers and reports
- Theses and student dissertations

² The use of the data for government purposes is exempt from this requirement.

Replication Materials

Please note that this functionality is not available at this time. Authorised users will be notified when it becomes available.

DSS encourages Authorised Users to consider their ethical obligation to facilitate analytical transparency so that their work can be tested or replicated. To this end, Authorised Users are invited to upload replication materials to the NCLD Dataverse site. Replication materials should include a full account of the procedures used to manipulate the data and a full account of how the research drew their analytical conclusions from the data (i.e. publishing replication code).

Authorised Users uploading information to this site are asked to provide sufficient information to replicate their results – noting full disclosure may increase the utility of the research, leading to increased citation, which also elevates scientific rigour.

No Unit Record Data should be provided in the replication materials.

DSS is not responsible for the accuracy of any replication material placed on the NCLD Dataverse site.

Upload guidelines for replication materials

1. In summary, a replication package should contain:

- a. A brief README.txt file that summarises the materials that are part of the replication package (see detail below).
- b. Well-documented and well-named code for producing the publicly available results.
- c. Documentation so that replicators know how to use the data and code to reproduce the publicly available results.

2. To facilitate ease of replication, the README.txt file should include, at a minimum:

- a. *Table of contents*: a brief description of every file in the replication folder.
- b. *Notes for each table and figure*: a short list of where replicators will find the code needed to replicate all parts of the results.
- c. *Base software dependencies*: a list of all software required for replication, including the version of software used by the author (e.g. Stata 11.1, R 2.15.3, 32 bit Windows 7, OSX 10.9.4), and references to specialised software packages, modules, or routines that are not a standard component of public-release, off-the-shelf software.
- d. *Additional dependencies*: a list of all libraries or added functions required for replication, as well as the versions of the libraries and functions that were used and the location from which those libraries and functions were obtained.

e. *Seed locations*: Please set seeds in your code for any analyses that employ randomness (i.e. simulations or bootstrapped standard errors). The README.txt file should include a list of locations where seeds are set in the analyses so that replicators can find and change the seeds to check the robustness of the results.

Appendix 4 – Previous data releases

Transition away from organisational licences

The new data access model through Dataverse will progressively phase out organisational licences and move to individual ones. Organisational licences will continue to apply until all data users within the institution have formally relinquished their responsibility under the organisational licencing arrangements.

However, to expedite the process, DSS is asking all Authorised Users of Datasets under previous licensing arrangements to make a choice about whether they wish to retain the Datasets and to either relinquish their responsibility or agree to the new arrangements by signing the Confidentiality Deed Poll (available on Dataverse).

Authorised Users who no longer wish to retain access can fully relinquish their responsibility for all Datasets by:

- cancelling access to all copies of the Datasets held on a secure server;
- removing all copies of the Dataset from the Cloud server;
- deleting all copies of the Datasets held on fixed or removable computer storage
- physically destroying all media such as CDs and DVDs that hold copies of the Datasets; and
- shredding all printouts of Unit Record Data from the Datasets.

Once this has been done, Authorised Users will need to inform their Data Manager and DSS by email that the relevant steps listed above have been taken in respect of all copies, either full or partial, of the Datasets the Authorised Users have obtained or generated.

Authorised Users who wish to retain access to the Datasets initially obtained under organisational licensing arrangements can do so by signing the Confidentiality Deed Poll and notifying DSS of which Datasets they hold. To do this, Authorised Users must complete the NCLD Data Holdings form listing the names and release of the Datasets they wish to retain and then sign the new Confidentiality Deed Poll for access to the Datasets and email or post the signed Deed to DSS. The Deed will also cover subsequent applications for Datasets through NCLD Dataverse. The Deed and the NCLD Data Holdings form are available on the main NCLD Dataverse page. It is not necessary to complete an application for a Dataset- unless you wish to apply for one.

Authorised Users who wish to relinquish responsibility for some but not all of the Datasets they currently have access to will need to undertake a combination of both processes; relinquishing responsibility for Datasets no longer required and sign the Deed for the Datasets still required.

A copy of the signed Deed will be attached to the Authorised User's record. ADA will be informed of all researchers who have signed the deed regardless of whether they have applied for access to the Datasets through Dataverse or through direct contact with DSS.

Once all Authorised Users at the organisation have either relinquished responsibility for all Datasets or entered into the Confidentiality Deed Poll, the Data Manager will be required to destroy any remaining copies of the Dataset held by them for administrative purposes (i.e. remove from a server or destroy CDs or DVD ROMs). The Organisational Delegate will then be informed that the organisational licence has been terminated.

Individual licences for previous releases of the data

Authorised Users of DSS Datasets under individual licensing arrangements will be asked to decide whether they wish to retain or relinquish access to Datasets obtained under previous individual licencing arrangements.

Authorised Users who no longer wish to retain access can fully relinquish their responsibility for all Datasets by:

- cancelling access to all copies of the Datasets held on a secure server;
- removing all copies of the Dataset from the Cloud server;
- deleting all copies of the Datasets held on fixed or removable computer storage;
- physically destroying all media such as CDs and DVDs that hold copies of the Datasets; and
- shredding all printouts of Unit Record Data from the Datasets.

Once this has been done, Authorised Users will need to inform DSS by email that the relevant steps listed above have been taken in respect of all copies, either full or partial, of the Datasets the Authorised Users have obtained or generated.

Authorised Users who wish to retain access to the Datasets initially obtained under previous individual licensing arrangements can do so by signing the Confidentiality Deed Poll and notifying DSS of which Datasets they hold. To do this, Authorised Users must complete the NCLD Data Holdings form listing the names and release of the Datasets they wish to retain and then sign the new Deed for access to the Datasets and email or post the signed Deed to DSS by post. The Deed will also cover subsequent applications for Datasets through NCLD Dataverse. The Deed and the NCLD Data Holdings form are available on the main NCLD Dataverse page.

Authorised Users who wish to relinquish responsibility for some but not all of the Datasets they currently have access to will need to undertake a combination of both processes; relinquishing responsibility for Datasets no longer required and sign the Confidentiality Deed Poll for the Datasets still required.

A copy of the signed deed will be attached to the Authorised User's record. ADA will be informed of all researchers who have signed the deed regardless of whether they have applied for access to the Datasets through Dataverse or through direct contact with DSS.

Appendix 5 – Using the Datasets for training and coursework

Authorised training activities and coursework

NCLD Datasets can be used for university coursework or as training Datasets (i.e. for data user training) with the appropriate approvals. Data access requirements vary according to the course delivery methods. Instructors wishing to use any of the Datasets for training or coursework should email the ADA with details of the proposed training.

Appendix 6 – Definitions of terms

The definitions used in these Guidelines are the same as those in the Confidentiality Deed Poll:

ADA means the Australian Data Archive.

Authorised User means a person who has signed a Deed of Confidentiality substantially in the same form as this Deed and has been given permission by the Commonwealth to access and use the Datasets.

Business Day means a weekday other than a public holiday in the Australian Capital Territory.

Commonwealth means the Commonwealth of Australia as represented by the Department of Social Services.

Confidant means the signatory of this Deed.

Confidential Information means information that:

- a) is by its nature confidential;
- b) is designated in writing by the Commonwealth as confidential;
- c) is Personal Information under the Privacy Act; or
- d) is contained in the Datasets.

Datasets means any or all data from:

- a) General or Restricted Release of the household and person level Unit Record Data from The Household, Income and Labour Dynamics in Australia (HILDA) survey, also known as Living in Australia;
- b) The household and person level Unit Record Data from the HILDA Cross-National Equivalent File (CNEF);
- c) General or Restricted Release of the household and person level Unit Record Data from The Longitudinal Study of Australian Children (LSAC), also known as Growing up in Australia;
- d) General Release of the household and person level Unit Record Data from The Longitudinal Study of Indigenous Children (LSIC), also known as Footprints in Time;

- e) General Release of the household and person level Unit Record Data from The Longitudinal Study of Humanitarian Migrants, also known as Building a New Life in Australia (BNLA);
- f) Any linked datasets associated with any or all of the above.

Deed means this Confidentiality Deed Poll.

Document includes:

- a) any paper or other material on which there is writing;
- b) any paper or other material on which there are marks, figures, symbols or perforations having a meaning to persons qualified to interpret them; and
- c) any article, material or media from which sounds, images or writings are capable of being reproduced with or without the aid of any other article or device.

Electronic Communication has the same meaning as in the *Electronic Transactions Act 1999* (Cth).

Information System has the same meaning as in the *Electronic Transactions Act 1999* (Cth).

General Release means a release of data from which Personal Information has been removed and other information has been modified by various methods such as top coding and the application of classification codes as set out in the Guidelines.

Guidelines means the National Centre for Longitudinal Data Access and Use Guidelines as amended by the Commonwealth from time to time, published on the Dataverse website available at <https://dataverse.ada.edu.au/dataverse/nclid>

Intellectual Property means copyright (and all associated rights, including moral rights), and all rights in relation to inventions, registered and unregistered trade marks (including service marks), registered and unregistered designs, and circuit layouts, and any other rights resulting from intellectual activity in the industrial, scientific, literary or artistic fields.

NCLD means the National Centre for Longitudinal Data in the Australian Government Department of Social Services.

Personal Information has the meaning given in section 6 of the Privacy Act.

Privacy Act means the *Privacy Act 1988* (Cth).

Privacy Commissioner means the person appointed under section 14 of the *Australian Information Commissioner Act 2010* (Cth) as the Privacy Commissioner.

Registered Australian Privacy Principle code (APP code or APPC) has the meaning given in section 6 of the Privacy Act.

Release means a dataset that differs from another dataset from the same survey in that it contains additional information based on new responses from survey respondents. For the purposes of this Deed, a Release does not include a new

version of the same release of the data in which changes have been made to the previously released information from respondents.

Research Material means any final research findings based on the analysis of the Dataset created by the Confidant.

Restricted Release means a release of data from which the names and addresses of respondents have been removed but includes other information at a more granular level than the General Release and access and use of which is governed by stricter requirements.

Term means the term of this Deed, which is from the date of execution of this Deed until:

- a) the Confidant deletes or destroys all copies of the Datasets it holds in accordance with clause 7; or
- b) the Deed is terminated in accordance with clause 14.

Unit Record Data means records about individual respondents from the Datasets, as opposed to aggregated data.