

Dr. Steven McEachern
Director, Australian Data Archive
ANU Centre for Social Research and
Methods

Overview

- The Five Safes framework
- The Dataverse software system.
- Extensions ADA has made to the Dataverse environment to improve support for sensitive data
- Technical and administrative processes
- Future plans and additional requirements

The Five Safes model (Ritchie, 2017)

https://zenodo.org/record/897821#.W8ekIRMzZE7

Dimension	Question	Approach
Safe projects	Is this use of the data appropriate?	Managerial controls
Safe people	Can the researchers be trusted to use the data in an appropriate manner	
Safe settings	Does the access facility limit unauthorised use?	
Safe data	Is there a disclosure risk in the data itself?	Statistical controls
Safe outputs	Are the statistical results non-disclosive?	

See also:

UK Data Service: https://www.ukdataservice.ac.uk/manage-data/legal-ethical/access-

control/five-safes;

Australian Bureau of Statistics:

http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1160.0Main%20Features4Aug%202017?opendocument&tabname=S

Ritchie, 2017 p.5

- "The key to the model is that the five dimensions severally and jointly contribute to any consideration of whether a data access stratagem meets expectations:
 - The dimensions are designed so that each can be evaluated independently of the others, as far possible
 - All five dimensions need to be considered jointly to evaluate whether a data access system provides an 'acceptable' solution

The Five Safes model in use

- UK Data Service: <u>https://www.ukdataservice.ac.uk/manage-data/legal-ethical/access-control/five-safes</u>;
- Australian Bureau of Statistics: <u>http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1160.0Main%20Features4Aug%202017?opendocument&tabname=S</u>
- South Australia: Public Sector (Data Sharing) Act 2016: https://www.legislation.sa.gov.au/LZ/C/A/PUBLIC%20SECTOR%20(DATA%20SHARING)%20ACT%202016.aspx
- Proposed: Data Sharing and Release Act (per Productivity Commission report): https://www.pc.gov.au/inquiries/completed/data-access/report)

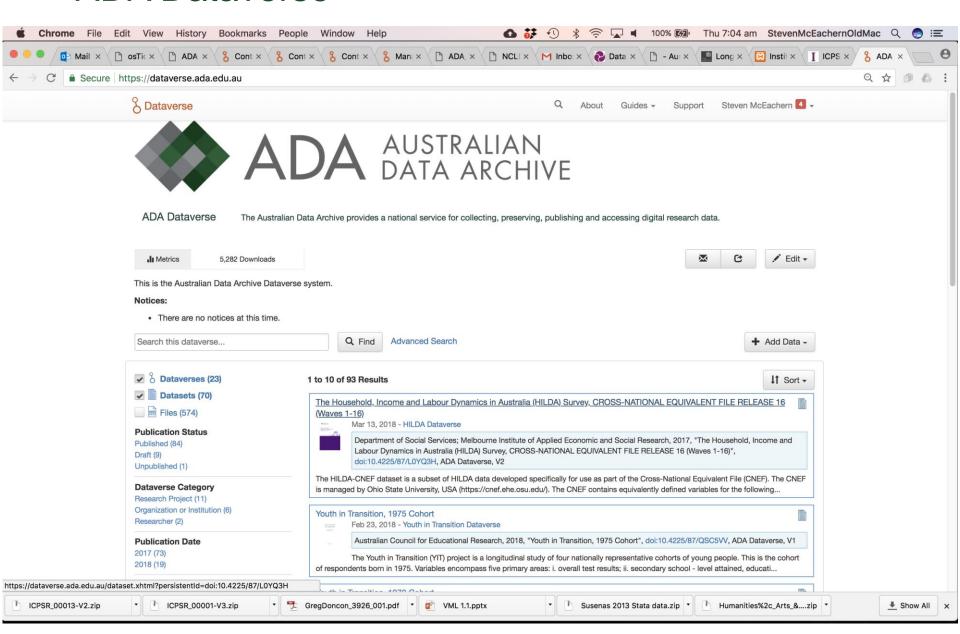
ADA in Brief

- The Social Science Data Archive (now ADA) was set up in 1981, housed in the Research School of Social Sciences at ANU, with a mission to collect and preserve Australian social science data on behalf of the social science research community
- ADA holds over 6000 datasets from around 1500 projects, including national election studies; public opinion polls; social attitudes surveys, censuses, aggregate statistics, administrative data & other sources.
- Data holdings are sourced from academic, government and private sectors.

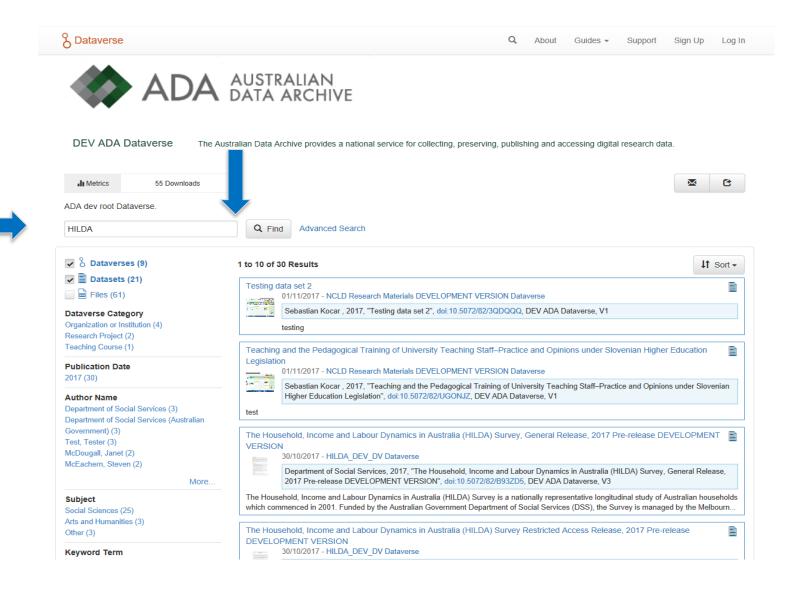
Dataverse at ADA

- Redevelopment of our database and website infrastructure
 - New website: http://sw-dev.ada.edu.au:8085/
 - New data catalogue: http://dataverse.ada.edu.au
- New functionality:
 - Self-deposit of data
 - Open data access
 - API access (both for deposit and access, e.g. through R)
 - Shibboleth authentication
 - ORCID integration
- Currently completing data migration
 - Major studies already migrated

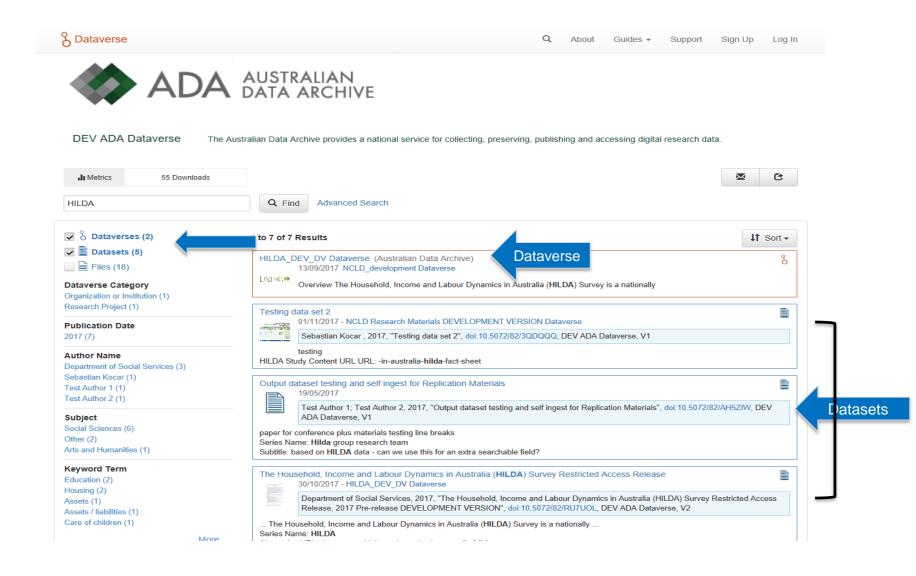
ADA Dataverse



Find the dataset



Search results and limiting



Dataverse landing page



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 Survey is managed by the Melbourne Institute of Applied Economic and Social Research at the University of

HILDA is a panel survey of Australian households and, as such, HILDA surveys all working age members (over 15 years) of the selected households and then re-interviews the same people in subsequent years.

Purpose of the study

HILDA provides longitudinal data on the lives of Australian residents. Its primary objective is to support research questions falling within three broad and inter-related areas of income, labour market and family dynamics.

Data sets

There are two versions of the HILDA data available.

- . The General Release means data from which all Personal Information (for the purposes of the Privacy Act 1988) has been removed and any remaining information has been modified by methods such as top coding, aggregation or the use of classification codes at a more general level.
- · The Restricted Release means a release of data from which the names and addresses have been removed but which includes other potentially personal information such as postcodes and data at a more detailed level than the General Release dataset.

Want to know more?

To find out more about the HILDA survey:

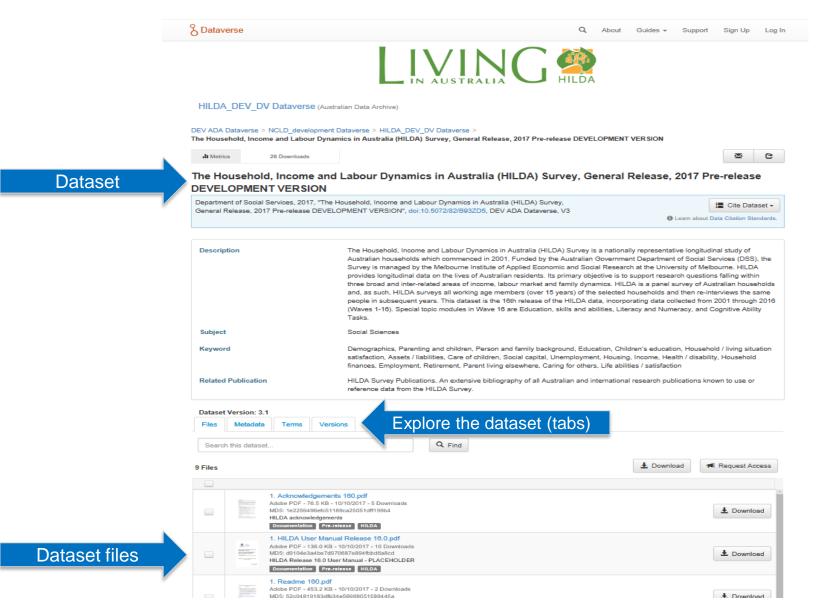
- · View the NCLD HILDA fact sheet.
- · Visit the Melbourne Institute HILDA website.

Details about the National Centre for Longitudinal Data are available from the Department of Social Services



Datasets

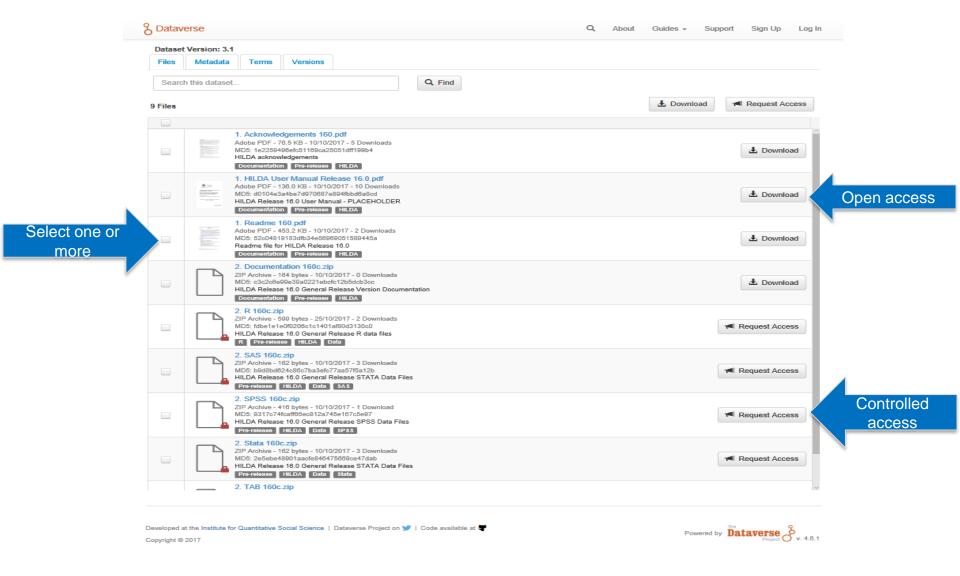
Dataset landing page



Applying the Five Safes model at ADA

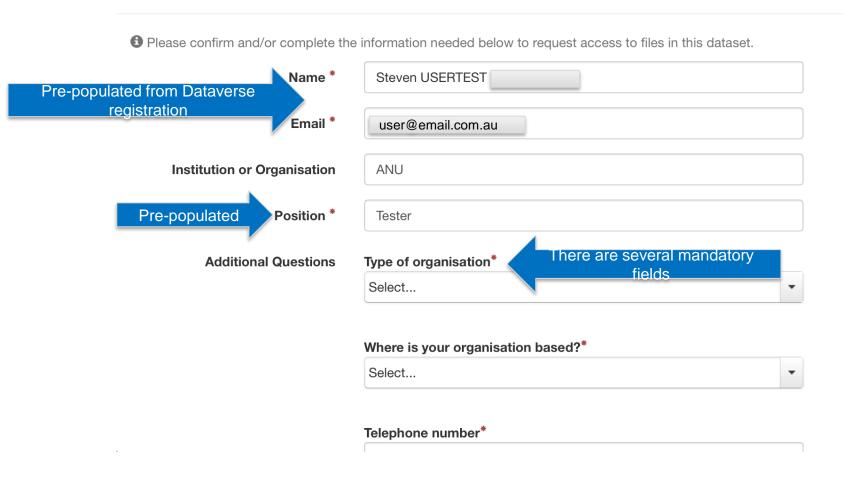
Dimension	Implementation	Assessment
Safe projects	Project proposal	Strong
Safe people	Institutional email	Moderate- Strong
Safe settings	User agreement not to share data	Weak
Safe data	Data is anonymised and deidentified	Strong
Safe outputs	Not assessable	Weak

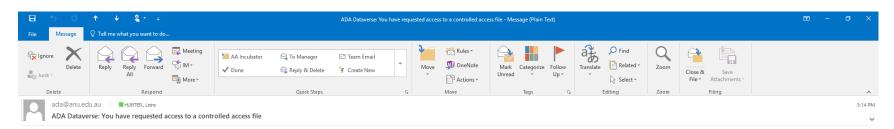
Request access to data



Request access to the data (2)

Request File Access





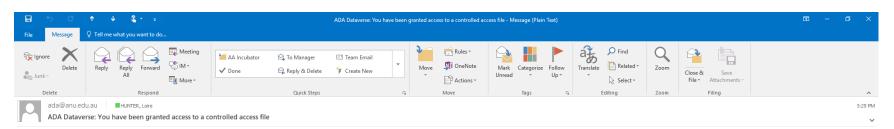
Hello.

Thank you for your recent application for access to controlled access files from the dataset The Household, Income and Labour Dynamics in Australia (HILDA) Survey, General Release, 2017 Pre-release DEVELOPMENT VERSION. Your application is currently being reviewed. (view at https://dataverse-dev.ada.edu.au/dataset.xhtml?persistentid=doi:10.5072/82/893ZD5).

ADA or the Data Owner may need to follow up further with you to clarify details of your request. Should this be required, we will contact you within three (3) business days.

Thank you,

The Australian Data Archive on behalf of the data owner.



Hello.

You recently applied for access to controlled access files in dataset: The Household, Income and Labour Dynamics in Australia (HILDA) Survey, General Release, 2017 Pre-release DEVELOPMENT VERSION. We are pleased to advise that your application for access has been approved.

Your obligations as an Authorised Data User are contained in the Terms and Conditions of Use.

You can now view and download files from the dataset at this link: https://dataverse-dev.ada.edu.au/dataset.xhtml?persistentid=doi:10.5072/82/B93ZD5

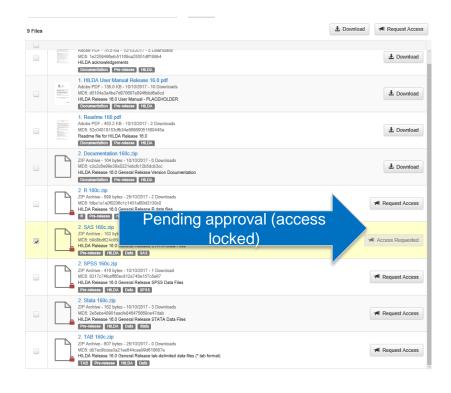
If you have any queries in relation to the data please refer to the The Household, Income and Labour Dynamics in Australia (HILDA) Survey, General Release, 2017 Pre-release DEVELOPMENT VERSION supporting documentation or click on the email icon from anywhere in Dataverse to submit a query.

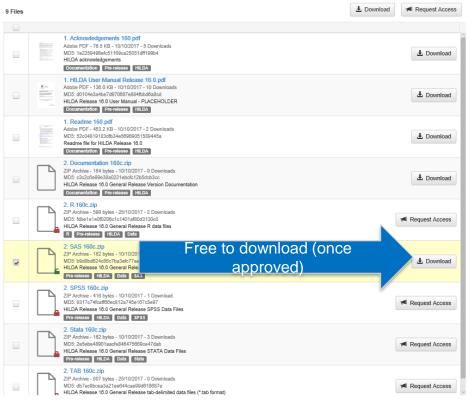
Thank you,

The Australian Data Archive

on behalf of the data owner.

Status changed to controlled file: download data









Approving requests

Business rules

- ADA approves "standard" requests, subject to business rules (based on the contents of the user request)
- ADA refers "non-standard" requests (e.g. from commercial organisations) to data owner for review
- Data owner reviews and approves/rejects nonstandard requests
- Approvals are managed in Dataverse
 - Established a new "Approver" role in DV to support this
- Communications between users, ADA and data owner regarding requests is managed via email, and logged in the ADA ticketing system

Email notification of a new data request

New Ticket Alert - ADA Dataverse: Access has been requested for a restricted file [#502003] Australian Data Archive Support <ada@anu.edu.au> Today, 8:50 AM Steven McEachern > Inhox Hi Steven McEachern, A new ticket #502003 has been created **Department: NCLD** Hello, File access requested for dataset: The Household, Income and Labour Dynamics in Ticket Assigned to you - Re: Ticket Assigned (Waves 1-16) DEVELOPMENT VERSION. Manage permissions at https://dataverse-dev.ad.to you - HILDA guery hildadvtest@gmail.com - Gmail files.xhtml?id=198. Thank you, The Australian Data Archive

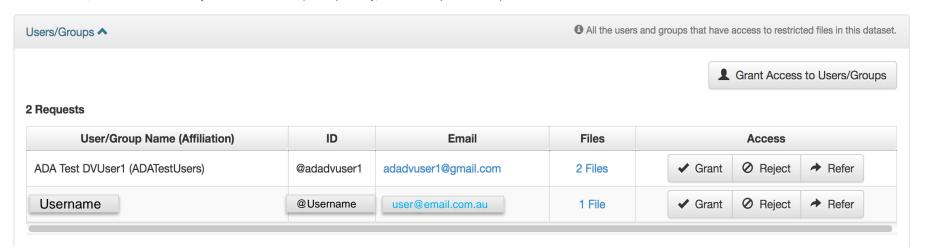
Also comes into the notifications in Dataverse

Data request permissions



HILDA_DEV_DV Dataverse (Australian Data Archive)

DEV ADA Dataverse > NCLD_development Dataverse > HILDA_DEV_DV Dataverse >
The Household, Income and Labour Dynamics in Australia (HILDA) Survey, Release 16 (Waves 1-16) DEVELOPMENT VERSION > Restricted File Permissions







Future directions

Future directions

- Safe settings: PHRN, ABS Data Lab
- Safe people: Shibboleth (AAF)
- Safe outputs??
- Integration with other programs and requirements:
 Data Sharing and Release Act, HREC
- Integration with other systems: virtual environments (e.g. HASS DEVL), DMPs, RAIDs





Questions?

Steven McEachern
steven.mceachern@anu.edu.au
ada@anu.edu.au
http://ada.edu.au