

# **FACT SHEET 5**

## **Specialist Datasets**

## Specialist Datasets

In addition to the HILDA, LSAC LSIC and BNLA General Release and HILDA LSAC and BNLA Unconfidentialised datasets there are three other datasets available.

1. HILDA-CNEF Dataset
2. Beta Datasets
3. HILDA Training Dataset

These are all covered under the Organisational Deed of Licence or MOU in terms of the data security requirements but other rules apply in terms of access limitations, application procedures or payment. People wanting access to these datasets should read the following instructions for the relevant datasets. Note that the HILDA-CNEF and the HILDA Training Dataset do not have LSAC, LSIC or BNLA equivalents, and there is no Unconfidentialised version of the LSIC dataset.

### HILDA-CNEF Datasets

The HILDA-CNEF dataset is a subset of HILDA data developed specifically for use as part of the Cross National Equivalence File (CNEF). 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, and
- Household, Income and Labour Dynamics in Australia Survey (HILDA).
- Swiss Household Panel
- Korea Labour and Income Panel Study
- The Russia Longitudinal Monitoring Survey-Higher School of Economics

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

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 approving access to the HILDA-CNEF. Details on how to order the CNEF data for other surveys can be viewed on the [Ohio State University CNEF website](#).

Access to the HILDA-CNEF data is by completing an Individual Deed of License HILDA-CNEF application, even for those who are eligible for access to other datasets under an Organisational Deed of Licence or MOU. DSS will email the researcher to let them know whether permission for access to the dataset has been granted. DSS will also inform the CNEF team at Ohio State University. It is the responsibility of the researcher to contact the CNEF team at Ohio State University ([CNEF@osu.edu](mailto:CNEF@osu.edu)) to organise to be sent the CNEF datasets or Tel: +1-614-292-4561 Columbus, Ohio USA.

In the case of users who are at an organisation with an Organisational Deed of Licence or MOU, DSS will also inform the Data Manager of the individual's permission for access to the HILDA-CNEF. This is done so that the Data Manager is aware of all datasets held by users of the organisation. However, the organisation has no responsibilities under the Organisational Deed of Licence or MOU for the HILDA-CNEF dataset held by its users.

Ohio State University provides users with a set of CDs containing all of the CNEF datasets. Each of the datasets in the CNEF may only be accessed with a password. The passwords for each dataset will only be provided once Ohio State University has been notified that the applicant has met the access requirements of the specific dataset.

DSS has been informed that the cost for obtaining the set of CDs with all of the CNEF datasets is US\$125 but advises users to check this. This fee is payable to CNEF team. This cost is NOT covered under the Organisational Deed of Licence or MOU.

Users are permitted to use the HILDA-CNEF Dataset on CD ROM / DVD off organisational premises but must comply with the security requirements in their Deed.

Users should note that because HILDA-CNEF is not on a separate CD from the other CNEF datasets, it is not possible to relinquish responsibility of the HILDA-CNEF without relinquishing access to all of the CNEF datasets.

Unlike other datasets, the Data Manager will not hold a copy of the HILDA-CNEF dataset and each applicant will be supplied with a copy directly by Ohio State University.

## **Beta Datasets**

A beta dataset is one that precedes the public release of that dataset. Much of the data in the beta dataset will remain unchanged in the final version. However, some information such as weights, imputations and the coding framework may be updated prior to the final release or may not be complete in the beta dataset.

DSS's longitudinal beta datasets are released to a small number of users for the purposes of testing to identify such problems as documentation errors, illogical data, unusual national estimates and missing files.

Access to the beta datasets may be provided under the provisions of the Organisational Deed of Licence, MOU or under an Individual Deed of Licence. However, access is highly restricted and is generally at DSS's invitation. As with all other datasets, users must sign an appropriate deed before being given access. Users of the beta datasets agree to report any problems or errors they identify to the relevant Survey Management Contractor.

No analysis from the beta dataset may be reported except with the express written permission of DSS. The beta dataset may be used for preliminary analysis but any

reported research must be based on the final version of the dataset once it becomes available.

Unit record data from the beta dataset must not be shown to anyone who is not an Authorised User of the beta dataset.

There is no administration fee charged for beta datasets.

## **HILDA Training Dataset**

The HILDA Training Datasets are sub-sets of the HILDA data that have been created as a tool for teaching about the use and analysis of longitudinal data. There are three training datasets; one with a balanced three year panel (waves one to three), one with a balanced five year panel (waves one to five) and one with a balanced seven year panel (waves one to seven). They contain information across 77 fields. They do not contain household information.

The HILDA Training Datasets are only available under the organisational licensing arrangements. They will not be supplied under individual licensing arrangements.

The access policy for the Training Datasets is dependant upon the type of course the data is to be used for; a short course or a long course.

### **Short Courses**

Short courses may last anywhere between a few hours and five days. However, the primary distinguishing factor is that participants of a short course do not have access to the dataset outside the classroom. If participants are required to use the dataset outside the classroom, the rules for the long courses will apply regardless of the length of the course.

### **Short Course Instructors**

Instructors of the course will sign a HILDA Training Dataset Deed of Confidentiality for Course Instructors in which they undertake to restrict participant access to the dataset and take steps to ensure participant adherence to security requirements. If there is more than one instructor for a course, all instructors must sign a copy of the HILDA Training Dataset Deed of Confidentiality for Course Instructors. Instructors who have previously signed a Deed of Confidentiality for another HILDA dataset such as the General Release or Unconfidentialised datasets will still need to sign the HILDA Training Dataset Deed of Confidentiality for Course Instructors as it contains details of additional responsibilities in relation to the provision of the dataset to course participants.

Instructors will be responsible for providing copies of the dataset to the participants in a nominated secure environment and must be in attendance at all times that the participants have access to the dataset. The dataset will preferably be on a password-protected server. Instructors must ensure that the file containing the dataset is closed down at the end of each session and is not accessible by unauthorised persons.

If the dataset cannot be provided on a secure server, it may be provided on a CD ROM / DVD. The Data Manager will provide the instructor with the required number of copies. The cover of each CD ROM / DVD as well as each disc must be clearly labelled with the name of the dataset, the name of the instructor to whom it has been issued, the name of

the organisation and a serial number for tracking purposes. These details must be recorded in the register kept by the Data Manager. The Instructor is responsible for all copies of the dataset and must collect them at the end of each session. If the dataset has been made available on the computer hard drives, the instructor must make sure that the dataset is removed from all hard drives prior to the computers being used for other purposes or at the conclusion of the course.

A list of course participants should be forwarded to DSS a week prior to the commencement of the course either by the instructor or through the Data Manager. Instructors must witness the HILDA Training Dataset (Short Courses) Deed Polls which must be completed and signed by all participants before they are given access to the dataset.

Instructors may retain copies of the training dataset at the conclusion of the course for use in future or subsequent training purposes. However, DSS needs to be informed of each intended use. All copies in the possession of the instructor must be kept secure in accordance with the security requirements set out in the HILDA Training Dataset Deed of Confidentiality for Course Instructors. If the instructor wishes to relinquish responsibility for the dataset, they must return all copies to the Data Manager or destroy them in accordance with the Deed of Confidentiality and inform the Data Manager. The Data Manager will inform DSS when this has been done.

## **Short Course Participants**

Course participants are required to sign a HILDA Training Dataset (Short Courses) Deed Poll in which they undertake not to use the dataset outside the venue of the course and to comply with the instructions given by the instructor including the security arrangements. The Deed Polls do not need to be signed in advance of the course, as they do not need to be counter signed by DSS. However, they must be signed by the participants and witnessed by an instructor who has signed a HILDA Training Dataset Deed of Confidentiality for Course Instructors for the course, before the participants are given access to the data. The Deed Polls must be forwarded to DSS once signed.

Participants who already have access to the latest release of the HILDA dataset (the General Release or the Unconfidentialised) do not need to sign the Deed Poll as its provisions are already met under the Deed of Confidentiality for the other datasets. They do however need to inform DSS that they will be using the training dataset as this constitutes another purpose for which they wish to use the data. This can be done by the participants contacting DSS ([longitudinalsurveys@dss.gov.au](mailto:longitudinalsurveys@dss.gov.au)) directly or through the Data Manager or through the course instructor indicating on the participant list sent to DSS. These participants may not retain the training dataset or use it outside the classroom.

Before giving access to the participants who have a previous Deed of Licence or MOU with DSS, the instructor must seek proof either by sighting the signed copy of the participant's deed or through a confirmation email from the Data Manager or DSS.

## **Long Courses**

Long courses are primarily indicated by the fact that course participants will need access to the dataset outside the class environment.

Instructors must sign a HILDA Training Dataset Deed of Confidentiality for Course Instructors and will be provided with access to the dataset by the Data Manager.

Participants of long courses will need to complete a HILDA Training Dataset (Long Course) Deed of Confidentiality for Participants which must be approved by DSS before being given access to the dataset by the Data Manager.

Participants will need to comply with the destruction requirements for all copies of the training dataset that they hold within one week of the completion of the course. If participants wish to retain the dataset after the completion of the course, they should contact DSS prior to the end of the course, stating the reasons they wish to retain the dataset. Alternatively, they may apply for the General Release Dataset.

While use of a secure server is the preferred method of access to the dataset, participants are permitted to use the HILDA Training Dataset on CD ROM / DVD off organisational premises as long as they comply with the security requirements in the Deed of Confidentiality.

Users must not publish research and analysis based on the HILDA Training Dataset.

## **Training Material**

Instructors may use unit record data or analysis based on the Training Dataset in their training material only but must abide by the following conditions:

1. no more than twenty lines of unit record data may be included per example;
2. lines of data must be randomly selected where possible;
3. training material that includes unit record data may not be published or otherwise made publicly available;
4. training material that includes the unit record data may be provided to course participants.