

Confidentialisation & Access

The following outlines a draft strategy that will be further developed to guide data confidentialisation and access.

Classification of users

Due to the relative sensitivity of LSAC data, it has been decided to classify users of the data into various user groups based on the degree of risk associated with their access. This classification system will in turn dictate which datasets the user will have access to, and by which means (i.e. type of contract).

Classification of datasets

The LSAC datasets will be categorised as follows:

1. Moderately confidentialised – postcode data and date of birth removed and some top coding on variables such as income and number of children.
2. Unconfidentialised – postcode data and date of birth will be available and no data will be aggregated.

Researchers

Type I researchers

Academic researchers attached to a university, institute of technology or TAFE will be regarded as Type I researchers. This includes Consortium Advisory Group members, and honours or graduate students under the supervision of an academic. They could be granted access to both datasets and will be required to sign a license with FaCS to use the moderately confidentialised dataset or enter into a fixed term contract for access to the unconfidentialised dataset. FaCS will have responsibility for granting approval to the unconfidentialised dataset.

Type II researchers

Researchers attached to other organisations (such as private research organisations, lobby groups) will be regarded as Type II researchers. Such researchers seeking access to the data will be considered on a case-by-case basis. FaCS will have responsibility for granting this approval.

Overseas researchers

Overseas researchers will be permitted access to the moderately confidentialised dataset only. FaCS will have responsibility for granting this approval, which may include consulting with the FaCS Agency Security Adviser.

Government agencies

Commonwealth agencies

In recognition that all Commonwealth agencies (including Government departments and statutory bodies that operate under the Financial Management Act) are required to comply with Commonwealth privacy and security legislation and policy, greater flexibility will be employed in granting data access to such organisations.



State/Territory agencies

State/Territory agencies will be able to access both datasets under similar conditions as Type I researchers.

Statutory bodies

Statutory bodies whose primary business is research will be able to access both datasets under similar conditions as Type I researchers. Other statutory bodies will be granted access in the same way as Type II researchers; i.e. access considered on a case-by-case basis, with FaCS having responsibility for granting this approval.

Signatories to FaCS' social policy research agreements

FaCS currently has in place Social Policy Research Service agreements with three research providers.¹ These agreements specify the conditions under which FaCS unit record data will be released to these organisations. As such, access to the LSAC datasets may be granted to the relevant parties without the need for a separate contract.

Confidentialisation

We have a duty to respondents to ensure that the data they provide is not identifiable to those accessing the data. A large number of confidentialisation techniques can reduce the identifiability of an individual's data. However, with the implementation of each of these techniques, valuable data are lost. Achieving a balance between the protection of an individual's data and the production of a high quality dataset is always a complex balancing act and the LSAC team expects and welcomes input into fine-tuning the proposed approach. Before the implementation of any confidentialisation methods, the proposal will also need to be shown to meet the security standards required by FaCS.

De-identification

The first and most obvious means of protecting the confidentiality of any data received is for all name and address details to be removed. Upon receipt of name and address information from the Health Insurance Commission (HIC), I-View and NCS Pearson are bound by the same conditions as HIC under the Health Insurance Act; primarily, no personal information is to be passed on to any third party under any circumstances. The fieldwork agency will ensure that no data collected from any individual will be stored directly with contact information. The main datasets will contain identification numbers for each child, which can only be linked to specific personal details by the fieldwork agency. No released datasets will contain this personal information.

¹ The Melbourne Institute of Applied Economic and Social Research (Melbourne Institute) at The University of Melbourne; the Social Policy Evaluation, Analysis and Research (SPEAR) Centre at the Australian National University; and the Social Policy Research Centre (SPRC) at the University of New South Wales.



Even with the deletion of personal contact details, it is, however, difficult (if not impossible) to fully eliminate the possibility of self-identification by a respondent, or identification of another sample member when they are known to have participated (eg a spouse, a child or a neighbour). One method of protecting the confidentiality of de-identified datasets is to provide a restrictive access policy whilst retaining all data intact (i.e. un-manipulated, with no deletion of valuable data).

Thus, it would be possible to allow the final unconfidentialised dataset to be made available for specific use for Type I researchers and government agencies. Before the data could be released, applicants would be required to apply for access (as outlined below) and to sign a contract with FaCS outlining the conditions under which data may be utilised. The purpose of the contract would be to ensure that, in relation to the datasets:

- the datasets are effectively utilised for FACS-approved research projects;
- appropriate data security arrangements are specified and adhered to;
- the parties each act in an effective and efficient manner;
- each party behaves reasonably and in good faith towards the other party; and
- the parties have a mechanism for managing use of the datasets and dealing with any disputes.

Parties granted a deed contract would be required to take all reasonable steps, and do all things that may be reasonably required by FaCS to keep the confidential information, including all documents, and all other things recording, containing, setting out or referred to any confidential information, under effective control of the licensee.

Moderately Confidentialised Dataset

For those researchers or institutions for whom it is not considered appropriate to access deidentified LSAC data, it is proposed that the dataset will be further confidentialised to reduce the risk that individual sample members can be identified. For the creation of anonymised datasets, various techniques could be applied - aggregation techniques, suppressing variables and ensuring there is a delay between collection and reference period of the data and its release as data.

Some possible methods include:

- aggregating geographic coverage;
- rounding;
- grouping or combining categories;
- top- (or bottom-) coding;
- imputing values from a model;
- suppressing fields or cells;
- suppressing variables; and
- time delay.

The implementation of any of these methods leads to a reduction in the data available for research, hence weakening the capacity to address research questions as



powerfully. An example would be the reduction of geographic specificity. To make data available only at the national level (a consideration when aggregating geographic coverage) would be to disallow research at the state/local level. Hence, minimising the use of these techniques is highly desirable, especially as datasets would only be distributed subject to the demands of the deed of license. In creating a confidentialised dataset, it is anticipated that the following processes will be involved:

- i. some variables to be withheld (e.g., postcode);
- ii. others to be aggregated (e.g., occupation is only provided at the two-digit level); and
- iii. others to be top-coded (e.g., age and income variables).

A moderately confidentialised dataset subjected to these methods would then be available for research purposes to approved users through the same data access management regime; namely, through the application and granting of deeds of license.

Managing access to LSAC datasets

Data Managers

The Data Managers in AIFS and FaCS will be responsible for facilitating access to the datasets, monitoring adherence to security requirements, and contributing to a registry of users and completed research papers.

In the case of requests to access the moderately confidentialised datasets, the Data Manager will be a member of the LSAC (AIFS) team (with the exception of Type II and overseas researchers). For all other access requests, the relevant FaCS officer will be the Data Manager.

- FaCS will be responsible for signing the contracts and licenses governing the release of both the confidentialised and unconfidentialised datasets.
- The LSAC (AIFS) team will provide technical assistance to users of both datasets, as well as developing the meta data.

Requests to access the LSAC datasets

All requests to use the LSAC datasets must be made in writing to the Australian Institute of Family Studies (AIFS) using the appropriate form on the AIFS website. A request should include:

- the name of the organisation requesting the data;
- a description of the proposed project or general research purpose;
- the names and qualifications of researchers requiring access to the data under the project; and
- the project methodology and timeframe (necessary only if a specific project is proposed).

Access to datasets will be further subject to a determination, by the relevant Data Manager, that:

- the applicant is able to meet the control measures outlined in this paper; and
- the release is consistent with the purposes for which the data was collected and that there are no impediments to the data being used in this way.

It is acknowledged that researchers may not have a definitive project proposal in mind when applying to access the LSAC datasets. However, in order for FaCS to monitor and record the use of LSAC data, access requests should at minimum provide details of the general purpose of the research, as well as addressing the relevant points listed above.

Access to the datasets will incur an administrative fee of around \$80 per CD and will be charged to each individual wishing to access the data. The fee will apply from April 2005 (Wave 1 public data release) and will be payable to the Australian Institute of Family Studies.

Users must notify their Data Manager if there is any change in their place of employment (including retirement) or in their research topic, as such changes may impact on their access status.

Managing the application process

For the moderately confidentialised dataset, the process will be as follows:

1. prospective users will send an application form and signed deed of confidentiality (downloaded from AIFS website) to AIFS.
2. The Australian Institute of Family Studies will assess suitability of user.
3. If user is approved, AIFS will send the license to FaCS. If user is an overseas or non-academic researcher, AIFS will forward the application form to FaCS.
4. FaCS will sign license and send a copy back to AIFS.
5. AIFS will be responsible for monitoring adherence to security requirements.

For the unconfidentialised dataset, FaCS will be responsible for processing the application, signing the contract and monitoring adherence to security requirements.

Registry of users

Data users will be required to provide the AIFS or FaCS Data Manager with details of any work utilising the LSAC datasets (including advance copies of any papers produced). AIFS will maintain a registry of users and their work (completed and in-progress). FaCS will be responsible for submitting to AIFS details relating to users of the 'unconfidentialised' dataset.

This registry and copies of research findings (including work-in-progress) may be published on the LSAC Internet site and in the LSAC annual report.

Clearance of research papers

Non-Commonwealth government researchers using the unconfidentialised LSAC datasets will be required to clear any papers for publication with FaCS to ensure the privacy of the survey respondents has been maintained. All researchers using the moderately confidentialised datasets will be asked to provide AIFS with an advance



copy of any research to be published. AIFS will pass a copy of the research to FaCS but clearance by FaCS will not be required.

User Support

With the datasets supplied to users, the following documentation will be provided in the form of a user manual which would include:

- Description of the conduct of the survey;
- Details of weighting and imputation procedures;
- Questionnaires;
- Variable listing outlining variable names, labels, response categories;
- Details of derived variables;
- Structure of the datasets; and
- Examples of use of the datasets.

User Training

In conjunction with the public release of the dataset after Wave 1, user training sessions will be offered by AIFS to further develop the information provided in the user manual and to allow users to dynamically interact with the LSAC Data Management team.

