

Understanding Australian Migrant Outcomes - Making Better Use of Administrative Data

Jenny Telford (Presenter), Jenny Dobak (Corresponding Author), National Migrant Statistics Group, Australian Bureau of Statistics.

ANZ House, 11 Waymouth Street, Adelaide, SA, 5000, Australia.
Phone: +618 8237 7317 Email: Jenny.Dobak@abs.gov.au

Abstract

Given the diversity of Migrants entering Australia under the Australian Government Migration Program, specific policy responses and programs are often required that reflect the unique circumstances of their arrival. As such, there is a need for relevant, timely data on the socio-economic characteristics of migrants pre- and post-migration to inform policy development, program design and the measurement of settlement outcomes over time. To fill this need, the Australian Bureau of Statistics (ABS) has increasingly looked to data integration techniques which maximise the use of the rich administrative data sources held by other government organisations. This has provided an effective, innovative and timely way of expanding the range of official statistics that can be made available for public benefit without placing extra burden on respondents. The ABS has made significant headway in this area and is now realising the benefits. The release of the Australian Census and Migrants Integrated Dataset, 2011 (ACMID) in 2014, and statistics from the 2009-10 and 2010-11 Personal Income Tax and Migrants Integrated Dataset (PITMID) in 2015, are two examples of innovative ABS data linkage projects which, by utilising administrative data sources, have significantly added to the pool of migrant data in Australia, filling critical data gaps.

This paper provides information about these linked datasets, including an overview of the linkage methods; key findings from the data and plans for possible development of this work into the future. The paper will also share key challenges and discuss the relationship between the ABS and producers and users of the administrative data.

Introduction

The Australian migration program is very carefully managed. Australia is one of the few countries that have formal Migration and Humanitarian programmes to manage the intake of permanent migrants as well as the resettlement of refugees.

Annual planning levels set by Government determine the size and composition of these programs, which are then very closely monitored by the Australian Government Department of

Social Services (DSS), and the Department of Immigration and Border Protection (DIBP). Smith and Smith (2014), states that this monitoring is very well managed by DIBP and DSS and they emphasise the value of timely robust data to provide the evidence base for sound policy formation and evaluation. It is noted that the components of Australia's Migration and Humanitarian Programmes vary significantly. The focus of the Skill stream is to select migrants who have the skills or proven entrepreneurial ability that will contribute to the Australian economy; the Family stream enables an Australian citizen or permanent resident to bring close relatives into Australia; and the Humanitarian Program is one of the ways Australia contributes to the protection of refugees. Clearly, given the difference in the composition of these cohorts, what constitutes successful settlement outcomes for one group can be quite different for another.

The capacity for the Australian Government to understand the settlement outcomes of Australia's recent migrant population by their entry conditions (i.e. visa subclass) has been significantly enhanced through the creation of two new datasets by the Australian Bureau of Statistics (ABS)—the Australian Census and Migrants Integrated Dataset (ACMID) and the Personal Income Tax and Migrants Integrated Dataset (PITMID). These datasets demonstrate the potential of data integration to replace direct collection and to improve the quality of key estimates derived from administrative sources.

Both of these integrated datasets were made possible through access to Australian Government administrative data on permanent migrants (settlement data) from DSS as the data custodian and DIBP as the data collector. The Settlement Database (SDB) has been developed to provide statistical data for government and community agencies involved in the planning and provision of settlement services. It brings together administrative data from various DIBP systems and a number of other sources including Medicare Australia. The SDB is a repository of information on permanent and provisional visas granted to persons who have arrived to settle in Australia since 1 January 1991. The SDB excludes temporary visa holders and non-visa settlers, such as people from New Zealand. PITMID was developed with the additional support of the Australian Taxation Office (ATO). Data on the personal income of migrants was linked to the SDB data to provide more detailed information on the incomes of employees, investors and those with business or other personal income.

The creation of the ACMID and PITMID has greatly enhanced the relevance of the contributing datasets whilst maintaining the confidentiality and privacy of individuals who supplied the data. These new datasets have significantly improved the range and quality of Australian official statistics about recent migrants. The close relationship ABS has with DSS, DIBP and ATO are paramount to the success of both the ACMID and PITMID projects and without the cooperation and commitment from these agencies such initiatives would not be possible.

Background

The creation of ACMID has enabled the ABS to bring together the breadth of the 2011 Australian Census of Population and Housing data with detailed administrative settlement data on migrants who were granted a visa to permanently live in Australia between 1 January 2000 and 9 August 2011. By effectively attaching 'visa category' and other immigration information to records on the Census file, variation in family formation, labour market and other socio-economic outcomes across different migrant subgroups can be more readily identified and understood.

The ACMID came about as a result of a major ABS initiative in 2006, the Census Data Enhancement (CDE) project. This project aimed to integrate a 5% sample of the Census with various administrative datasets to create new, enhanced data sources for statistical, policy and research purposes in a cost effective way.

The CDE project included a Migrants Quality Study to assess the feasibility of linking the Settlement Database (SDB) data to the 5% Census sample without the use of name and address as linking variables, with a view to creating a longitudinal file. The probabilistic linkage was deemed feasible; see "*Assessing the Quality of Linking Migrant Settlement Records to Census Data*" (Wright, Bishop and Ayre, 2009). Subsequently, the SDB was linked to the 2011 Census dataset (100% file) to create ACMID as well as to the Australian Census Longitudinal Dataset (ACLD).

PITMID also utilises settlement data, integrating it with the ATO Personal Income Tax (PIT) records. The resulting dataset provides detailed information on the sources of personal income that permanent migrants receive including employee, own unincorporated business, investment and foreign income by characteristics such as visa stream, applicant status and location of visa application. It has contributed significantly to meeting a priority data need of understanding the economic wellbeing and financial independence of migrants.

The potential value of the ATO PIT data to provide new data on migrant incomes by visa category had been recognised for some time prior to the linkage work being undertaken. Initially it was hoped that the PIT data file itself may hold comprehensive income and expenditure data coded to a variety of key migrant identifiers, including country of birth, year of arrival and, in some cases, residency status. Preliminary investigations revealed that most of this information was either not collected or was only collected from small sub-populations under certain circumstances. Given this paucity of migrant information, the focus then turned to the potential of being able to link the ATO data with migrant data on the SDB.

In 2013, the ABS was provided with access to the ATO PIT unit record data to assess the feasibility of linking these records with the SDB. The study concluded that linking PIT to the SDB data was feasible and the dataset of sufficient quality, provided name and address information could be utilised as linking variables. The resultant linked dataset provides valuable information on migrant taxpayer's personal income not available from any other source (ABS, 2014b). As well as wage and salary, the PITMID contains income variables such as own unincorporated business income, investment income, superannuation and annuity income. The PITMID has

been well received by stakeholders and researchers and the success of this linkage has resulted in additional funding to support the production of PITMID on an annual basis.

Australian Census Migrants Integrated Dataset (ACMID)

The ACMID contains information on approximately one million persons who responded to the 2011 Census of Population and Housing and had a permanent visa record on the SDB with a date of arrival between 1 January 2000 and 9 August 2011 (Census night). This included permanent settlers from the Skill, Family and Humanitarian visa streams. Exclusions were overseas visitors, permanent departures, non-visa settlers (e.g. New Zealand citizens), deceased persons and those with an "Other Permanent" visa. The linkage of the SDB effectively adds visa sub-class, applicant status (main or secondary) and location (onshore or offshore) of visa application to the already rich Census data items.

Linkage Methodology

The SDB and the 2011 Census datasets were integrated to create ACMID using a probabilistic linking method, without the use of names and addresses. This method aims to link records on two datasets, which belong to the same individual without a unique record identifier. Instead, records from the two files are linked using a number of variables common to individuals on both files.

A key feature of this methodology is the ability to utilise a range of linking variables and record comparison methods to produce a single numeric measure of the likelihood two particular records belonging to the same person. A record pair may be linked in spite of missing or disagreeing values on any given linking variable(s), providing there is sufficient agreement on other linking variables. Records were also weighted, such that estimates produced from the file would match to key SDB benchmarks. For more information see the research paper *"Assessing the Quality of Linking Migrant Settlement Records to 2011 Census Data"* (Richter, Saher and Campbell, 2013).

The quality analysis of the 2011 linked dataset indicated that no major migrant sub-populations were missed in great numbers. However, some subgroups of migrants appeared to be more difficult to link than others. Migrants from the Skill stream and those who applied for a visa offshore were more likely to be under-represented. Migrants on a Humanitarian visa were generally more likely to be over-represented, which can be attributed to their greater degree of engagement with government systems and services e.g. Medicare. Suggested modifications and improvements arising from the 2006 migrant quality study were taken up for the 2011 study resulting in an increase in the linkage rate from 66% to 76%. This increase was primarily due to the identification and removal of individuals who were permanent departures from Australia on Census night (Richter, Saher and Campbell, 2013).

Personal Income Tax Migrants Integrated Dataset (PITMID)

The PITMID has been compiled by linking individual taxpayer personal income tax unit record data from the ATO and migrant settlement records from the SDB. The statistics produced from this linked dataset relate to migrants aged 15 years and over who submitted a tax return, with a permanent or provisional visa who arrived between 1 January 2000 and 6 March 2013. PITMID does not therefore include all migrants in Australia, only more recent migrants. It is important when analysing the data to be mindful that the data represents a subpopulation of recent migrant taxpayers.

For this subpopulation, PITMID provides a wealth of information regarding the sources of personal income that migrant taxpayers received in a financial year. Data includes employee income, own unincorporated business income, investment income, other income and foreign income by selected SDB migrant variables. The ATO provides the ABS with an extract of personal income tax data on an annual basis for a range of purposes, including the production of PITMID. The ATO collects personal income tax information from the lodgement of personal income tax returns as part of their role to manage and shape tax, excise and superannuation systems that fund services for Australians (<http://www.ato.gov.au/>).

Linkage Methodology

The PITMID project began as a feasibility study to ascertain whether the ATO personal income tax (PIT) data could be linked to the SDB utilising only very few available linking variables common to both datasets. The integration method utilised is a combination of deterministic and probabilistic linking to combine the ATO PIT 100% data file for the 2009-10 and 2010-11 financial years with the SDB.

Personal income data are provided by the ATO before the processing of all income tax returns for any given year may have been completed. Data provided to the ABS by the ATO are from returns processed up to 31 October, 16 months after the end of the financial year. It is estimated that approximately an additional 3% of taxpayers lodge their income tax returns in the twelve months after the initial processing cut off each year and approximately 5% of taxpayers lodge their income tax returns in the three years following the processing cut off. This means that data provided to the ABS slightly under-estimates the number of income earners and the total income earned for any given financial year.

Persons who receive an income below the tax-free threshold (\$6,000 in 2009-10) are not necessarily required to lodge a tax return and this can include persons who derive their income from government pensions and allowances. In addition, some Australian Government pension, benefit and allowance payments are exempt from income tax and therefore recipients are not required to include this income in their taxation returns. Consequently, the coverage of low income earners is incomplete in the PIT data and for this reason Government pensions and allowances are excluded from the PITMID dataset (ABS, 2014b)

Despite these characteristics, the quality of the data produced for PITMID is considered to be very high. In order to assess the eligibility of records for linking, the following subpopulation of candidates for linking were identified as those most likely to have submitted a tax return:

- migrants aged 15-64 years;
- Skilled, Family or Humanitarian visa stream;
- Permanent residence start date was prior to the start of the reference period; and
- Departure information showed they did not permanently leave Australia prior to the reference period.

This subpopulation of linked records were assessed against secondary sources such as the Estimated Resident Population (ERP), the 2011 Census and the 2011 Australian Census and Migrants Integrated Dataset (ACMID) in order to estimate the match rate. The match rate was estimated to be almost 90%.

See [Research Paper: Feasibility Study of Linking Migrant Settlement Records to Personal Income Tax Data, Aug 2014](#) (cat. no. 1351.0.55.051) for more information.

Appendix A provides a summary of key findings for both ACMID and PITMID and more detailed information can be found in the links to the ABS publications provided below.

How the Data is Being Used

Data from both ACMID and PITMID has already been well utilised by stakeholders across the Australian Government. As mentioned above, the selection criteria for each of the visa streams enabling migration to Australia vary markedly and are underpinned by different policies and settlement programs. For this reason, migrant characteristics and settlement outcomes are quite different across visa groups. This provides the context for why the analysis of settlement experiences and outcomes by visa entry conditions becomes so important. Compared to the small survey samples available previously, the huge administrative datasets of ACMID and PITMID now allow a valuable level of analysis never before possible.

“The advantage of the ACMID over these other collections is that it provides access to more than 1 million migrant records that have been processed by the ABS to its exemplary standards. This enables us to confidently drill-down to specific migrant sub-groups and... provides the tools needed for sharper policy-thinking.”

(Smith and Smith 2014: 221)

The Australian Productivity Commission recently conducted an important inquiry into the Migrant Intake into Australia. The Productivity Commission conducted an investigation into the impact of charges relative to government quotas and selection criteria for migrants coming into Australia. This inquiry relied heavily on the analysis of the ACMID and PITMID microdata. These data sources were considered an invaluable resource enabling the researchers to present

sound evidence based recommendations to Government. A new trusted access framework adopted by the ABS (see below) allowed the Productivity Commission researchers greater access to data, significantly enhancing the 'useability' of these datasets. As a result, a recommendation was made by the Productivity Commission that the Australian Government should continue to invest in the continued production of high quality integrated datasets.

Increased Data Sharing and Making Better use of Administrative Data

In Australia, there is an increasing emphasis on making the most out of existing public sector data holdings. For the ABS, this has been a strategic objective for a number of years. Advances in statistical methods, technology and the availability of administrative data are creating new data sources for evidence based analysis of government programs. In its role as a provider and coordinator of official statistics and related activities, the ABS is vigorously pursuing emerging opportunities to create new datasets, fill data gaps and increase the accessibility of statistics.

The ABS has recently received significant Government funding to transform its infrastructure across a four year period, in order to make the best use of modern technologies. As part of this transformation, the ABS is moving towards a 'statistical solutions' centred data provision model, underpinned by data integration. Transforming population and economic statistics will rely on new approaches to sourcing and creating datasets, and integrating components of these datasets with a wide range of administrative data held by governments and organisations. The ABS will continue to use data integration as a standard statistical tool to increase the depth and breadth of available statistics in Australia.

Partnerships with Data Custodians

Critical to the success of successful data sharing and statistical data integration is the close partnering across the government, research and private sectors. The ABS is building capability through new and existing partnerships with data custodians, to improve accessibility to public information, maximising its use for statistical purposes.

The Australian Government Migrant Statistics Program is a good example of this. Through a partnered migrant statistics work program, DSS, DIBP and ABS have collaborated to take significant steps forward in progressing data integration initiatives for mutual benefit. This work has progressed through a Government commitment for the continual improvement in the collection and collation of data to inform migration and multicultural issues in Australia. For this reason the jointly funded National Migrants Statistics Unit (NMSU - based in the ABS) was established in 2006. Through collaboration and the progression of an agreed work program, the mutual objective is to improve the pool of migrant information available to underpin a sound, objective evidence base for decision making and policy evaluation for Australia. Collectively, the research and statistical outputs provided by the ABS, DIBP and DSS form the core Australian Government Migrant Statistics Program.

The history of collaboration between the ABS, DIBP and DSS, has enabled the development of a true collaborative working partnership built on trust. To facilitate the relationship, ABS officers

are in posted to work within both DSS and DIBP and this allows for a direct conduit between the ABS as a statistical agency and DSS and DIBP as policy agencies. This close working relationship enables the ABS to respond early to emerging data requirements and policy imperatives.

Recent modern reinterpretation of ABS legislation and the implementation of the 'five safes framework' have allowed the ABS to review policies around access to data (see below). Such revised policies allow, subject to approval, the peer review of ABS publications and datasets prior to their public release. This facilitates an improvement in the quality and relevance of data and associated commentary through the early feedback from experts and data custodians prior to the final publication release. The ABS may also approve Government stakeholder agencies to obtain pre-embargo access to data releases to allow them time to prepare briefings for senior management and Ministers.

This more collaborative arrangement contributes to the trust and goodwill between the agencies. The realisation of mutual benefit means that data custodians are more likely to invest the time and resources required to improve their administrative systems – to not only to meet the administrative purpose it was built for, but to provide an improved quality set of statistics. For example, DSS and DIBP work together to maintain the SDB which is recognised as a critical set of information which informs the Migrant Statistics Program, bringing in sources from DIBP and other agencies (e.g. Medicare Australia data), to keep it contemporary from an information and geospatial perspective.

Greater Data Access and the 'Five Safes' Framework

A critical aspect underpinning the sharing of data between agencies is ensuring data is stored, accessed and disseminated in a way that ensures the privacy, confidentiality and legislative obligations of each agency are met.

The ABS recognises that flexible and wide-ranging microdata access is critical to meeting data needs of researchers and policy-makers. As already discussed, allowing stakeholder agencies greater access to integrated datasets; particularly where they are 'part custodians', is critical if a truly mutually beneficial relationship is to be realised. The challenge is to maintain the strict privacy and confidentiality conditions which ABS legislation demands while making data accessible. Fortunately, technological advances and modern statistical techniques continue to open up new frontiers in data integration, facilitating greater access to microdata while safeguarding the privacy of individuals.

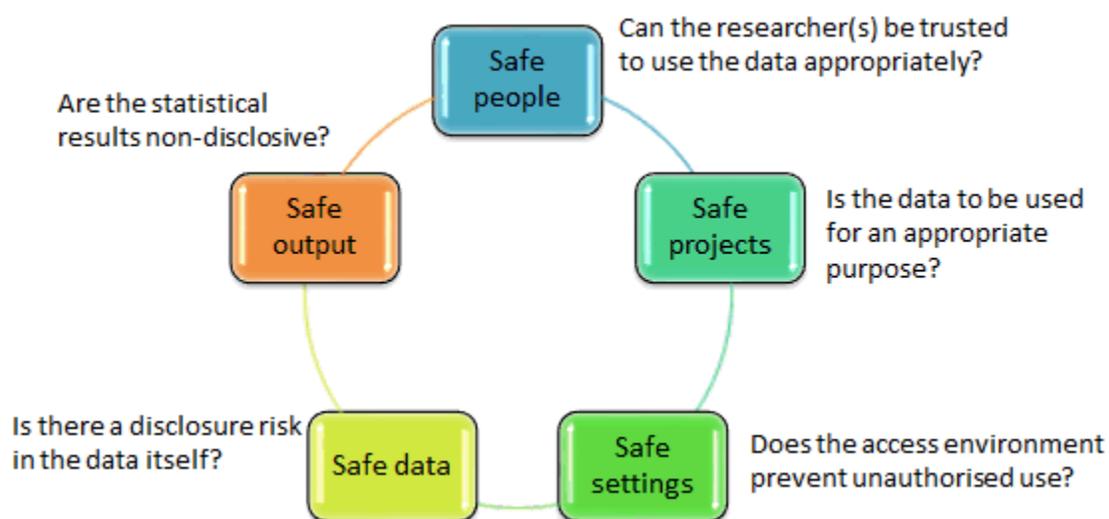
The ABS has recently reviewed microdata access arrangements of integrated datasets to support in-depth research, with a view to increasing its availability. Over the past six months ABS has been transforming its approach to managing disclosure risk. The focus has been on finding a balance between providing access to data for analysis and research purposes, while also protecting privacy and confidentiality.

The ABS has adopted the *Five Safes Framework*, first developed in the United Kingdom and now widely used in Europe and New Zealand, for comprehensive assessment of disclosure risk.

The five safes (safe people, safe project, safe settings, safe data and safe output), (see box below), can be applied to assess all data releases, but they have a special application when it comes to detailed microdata. Historically, the 'safe data' lever has been used to strongly safeguard confidentiality. This new approach to microdata access is also making greater use of the 'safe people' lever. A stronger focus on 'safe people' creates opportunities for access to detailed microdata without compromising confidentiality. 'Safe settings' (e.g. controlled IT access environment) and 'safe output' (e.g. checking output before it is delivered to the researcher) provide further safeguards.

What are the five safes?

The five safes are: safe people, safe projects, safe settings, safe data and safe output. They are controls or levers, each of which is considered independently on a sliding scale, as well as jointly, to evaluate whether or not a particular method of data access meets confidentiality and privacy requirements.



Four detailed microdata files have recently been released through the ABS DataLab environment, after each was assessed using the five safes framework. The DataLab allows Government agencies greater *access to microdata from a safe external environment*.

Under this model, the ABS is also making greater use of in-posting arrangements to allow employees from 'trusted' stakeholder agencies greater access to data in a 'safe' setting for their own purposes as well as to provide feedback and guidance regarding possible improvements to these data sets. It was under this model that employees from the Australian Productivity Commission were able to be seconded into the ABS to analyse PITMID and ACMID data for input into their Migrant Intake draft report (link provided below). This proved to be extremely successful and beneficial for both the ABS and the Productivity Commission.

“It has provided Productivity Commission employees with direct contact with the microdata and an opportunity to apply analytical approaches that would not otherwise have been possible”

Matthew Forbes, Productivity Commission

Future Directions

With support from DSS and DIBP, the ABS is producing the next iterations of PITMID and ACMID, as well as creating a new dataset linking temporary migrant data (sourced from DIBP) to the Census. If successful this will fill another significant data gap.

Opportunities for improving and expanding existing integrated datasets are also being considered including linking PITMID to ACMID and developing a longitudinal ACMID and/or PITMID. In the longer term, provision of a longitudinal ACMID and PITMID would reveal the movement patterns of migrants over time provided that their initial location (first address) can be identified. In the long term, data would allow for analyses of regional migration programs to assess their effectiveness in building skill capability in particular areas and generating regional population stability or growth.

There is also strong potential value in linking to other available datasets, such as:

- linking PITMID with the newly created Linked Employer-Employee Dataset (LEED) to provide information regarding the business characteristics of where migrants work
- linking ACMID to Overseas Arrival and Departures data (OAD) which would inform analysis against the priority policy area of migrant mobility and circular migration.

Conclusion

To achieve great outcomes such as the new ACMID and PITMID, the collaboration and ongoing support of the data custodians is paramount. The ‘seed’ projects underpinning the design and production of integrated datasets such as ACMID and PITMID only came to fruition because of the development of effective relationships, built on trust and cooperation between the data custodians and the integrating authority to achieve a mutually beneficial outcome.

Data issues encountered along the way often require expertise and resources from both sides to resolve. A collaborative relationship can deliver new and innovative ways of enhancing the data or resolving data issues. These projects also require high levels of commitment from the senior levels of all stakeholder organisations to work through the governance, access arrangements and agreements that are a necessary part of these data integration projects.

References

Australian Bureau of Statistics (2013) Understanding Migrant Outcomes - Enhancing the Value of Census Data cat. no. 3417.0, ABS, Canberra.

— (2014a) *Microdata: Australian Census and Migrants Integrated Dataset, 2011* cat. no. 3417.0.55.001, ABS, Canberra.

— (2014b) *Research Paper: Feasibility Study of Linking Migrant Settlement Records to Personal Income Tax Data* cat. no. 1351.0.55.051, ABS, Canberra.

— (2015) *Personal Income of Migrants, Australia, Experimental, 2010-11* cat. No. 3418.0, ABS, Canberra.

— (2016) *Information Paper: Construction of Experimental Statistics on Employee Earnings and Jobs from Administrative Data, Australia, 2011-12* cat. no. 6311.0.55.001, ABS, Canberra.

Department of Immigration and Border Protection (2015), Historical Migration Statistics, Table 3.2 and 4.1, Canberra.

Viewed at <https://www.border.gov.au/about/reports-publications/research-statistics/statistics/live-in-australia/historical-migration-statistics>

Richter, K., Saher, G and Campbell, P. (2013) "Assessing the Quality of Linking Migrant Settlement Records to 2011 Census Data", *Methodology Research Papers*, cat. no. 1351.0.55.043, Australian Bureau of Statistics, Canberra.

Smith, D. and Smith, T. (2014) "Using Linking to Sharpen Policy-Thinking: Early Findings from the Australian Census and Migrants Integrated Dataset", *Australian Economic Review*, Vol. 47, Issue 2. Viewed at <http://onlinelibrary.wiley.com/doi/10.1111/1467-8462.12069/pdf>

Wright, J., Bishop, G. and Ayre, T. (2009) "Assessing the Quality of Linking Migrant Settlement Records to Census Data", *Methodology Research Papers*, cat. no. 1351.0.55.027, Australian Bureau of Statistics, Canberra.

Links to ABS publications

[Assessing the Quality of Linking Migrant Settlement Records to 2011 Census Data](#) (ABS cat. no. 1351.0.55.043)

[Assessing the Quality of Linking Migrant Settlement Records to Census Data](#) (ABS cat. no. 1351.0.55.027)

[Microdata: Australian Census and Migrants Integrated Dataset, 2011](#) (ABS cat. no. 3417.0.55.001)

[Personal Income of Migrants, Australia, Experimental, 2009-10](#) (ABS cat.no. 3418.0)

[Personal Income of Migrants, Australia, Experimental, 2010-11](#) (ABS cat.no. 3418.0)

[Research Paper: Feasibility Study of Linking Migrant Settlement Records to Personal Income Tax Data](#) (ABS cat. no. 1351.0.55.051).

[Understanding Migrant Outcomes - Enhancing the Value of Census Data](#) (ABS cat. no. 3417.0)

Other links of Interest

[Australian Productivity Commission Draft Intake Report, November 2015](#)