



**Australian Government**

**Australian Institute of  
Health and Welfare**

# **Development of an ongoing national data collection on the educational outcomes of children in child protection services: a working paper**

CHILD WELFARE SERIES NO. 56



**Australian Government**

**Australian Institute of  
Health and Welfare**

*Authoritative information and statistics  
to promote better health and wellbeing*

CHILD WELFARE SERIES  
NUMBER 56

# **Development of an ongoing national data collection on the educational outcomes of children in child protection services: a working paper**

**November 2013**

Australian Institute of Health and Welfare  
Canberra

Cat. no. CWS 46



**The Australian Institute of Health and Welfare is a major national agency which provides reliable, regular and relevant information and statistics on Australia's health and welfare. The Institute's mission is authoritative information and statistics to promote better health and wellbeing.**

© Australian Institute of Health and Welfare 2013



This product, excluding the AIHW logo, Commonwealth Coat of Arms and any material owned by a third party or protected by a trademark, has been released under a Creative Commons BY 3.0 (CC-BY 3.0) licence. Excluded material owned by third parties may include, for example, design and layout, images obtained under licence from third parties and signatures. We have made all reasonable efforts to identify and label material owned by third parties.

You may distribute, remix and build upon this work. However, you must attribute the AIHW as the copyright holder of the work in compliance with our attribution policy available at <[www.aihw.gov.au/copyright/](http://www.aihw.gov.au/copyright/)>. The full terms and conditions of this licence are available at <<http://creativecommons.org/licenses/by/3.0/au/>>.

Enquiries relating to copyright should be addressed to the Head of the Media and Strategic Engagement Unit, Australian Institute of Health and Welfare, GPO Box 570, Canberra ACT 2601.

This publication is part of the Australian Institute of Health and Welfare's series. A complete list of the Institute's publications is available from the Institute's website <[www.aihw.gov.au](http://www.aihw.gov.au)>.

ISSN 1320-081X

ISBN 978-1-74249-509-5

### **Suggested citation**

Australian Institute of Health and Welfare 2013. Development of an ongoing national data collection on the educational outcomes of children in child protection services: a working paper. Child welfare series no. 56. Cat. no. CWS 46. Canberra: AIHW.

### **Australian Institute of Health and Welfare**

Board Chair

Dr Andrew Refshauge

Director

David Kalisch

Any enquiries about or comments on this publication should be directed to:

Media and Strategic Engagement Unit

Australian Institute of Health and Welfare

GPO Box 570

Canberra ACT 2601

Tel: (02) 6244 1032

Email: [info@aihw.gov.au](mailto:info@aihw.gov.au)

Published by the Australian Institute of Health and Welfare

**Please note that there is the potential for minor revisions of data in this report. Please check the online version at <[www.aihw.gov.au](http://www.aihw.gov.au)> for any amendments.**



# Contents

- Acknowledgments..... iv
- Summary ..... v
- 1 Introduction.....1
- 2 Proposed methodology .....4
- 3 Implementation .....14
- Appendix 1: Stakeholder consultation.....18
- Appendix 2: Phase 1 data sources .....19
- Acronyms and abbreviations .....21
- References .....22

# Acknowledgments

The Australian Institute of Health and Welfare (AIHW) was funded by the former Community and Disability Services Ministers' Advisory Council (CDSMAC) – now the Standing Council on Community and Disability Services Advisory Council (SCCDSAC) – to develop, in collaboration with the states and territories, a national methodology for reporting on the educational outcomes of children in child protection services. This working paper summarises the key findings of the final report on this development work, presented to SCCDSAC in late 2012.

The AIHW appreciates the assistance of the various stakeholders involved in the consultation process for the development work, and gratefully acknowledges their cooperation and invaluable contribution. A detailed list of stakeholders is provided in Appendix 1.

The AIHW project team consisted of Nicole Hunter and Tim Beard. The valuable input of the AIHW Data Linkage Unit and Data Integration Services Centre is greatly appreciated.

# Summary

## Background

Improving the educational outcomes of children involved in statutory child protection services has been a high priority for Australian governments in recent years. The inclusion of education-specific national indicators in the *National Framework for Protecting Australia's Children 2009–2020* and the *National Standards for out-of-home care* means the implementation of an ongoing national data collection on the educational outcomes of children in the care of the state has increased in importance and urgency. Such a collection would allow ongoing and longitudinal monitoring of academic progress, to better inform policy, practice and planning of activities to support these children.

This working paper sets out a proposed national methodology for reporting on the educational outcomes of children in child protection services. The former CDSMAC (now SCCDSAC) funded the AIHW to develop this methodology in collaboration with jurisdictions.

## Proposed methodology

National reporting on the educational outcomes of children in care can be best achieved through linking the Child Protection National Minimum Data Set (CP NMDS) with a national set of education data (an 'Education Module', see Section 2). The CP NMDS is the 'base' data set for the Education Module and will be used to identify in-scope children.

In line with the *National Standards for out-of-home care*, the population scope of the Education Module would be children aged 0–17 years whose care arrangements have been ordered through the Children's Court, where parental responsibility for the child or young person has been transferred to the Minister/Chief Executive.

A range of relevant administrative data sets which capture information across the primary and secondary schooling years have been identified, from which data could be sourced for the Education Module (see Section 2 for details).

Undertaking data linkage at the national level will allow the use of nationally-consistent linkage processes to improve match rates and efficiency. The AIHW is a Commonwealth-accredited Data Integration Authority, and therefore well-positioned to undertake this linkage work for the Education Module.

A phased approach to implementation is recommended, commencing with linkage between NAPLAN data and CP NMDS data (Phase 1, further described in Section 3). The Education Module could then be expanded following the successful completion of Phase 1.

## Phase 1 implementation

High-level support from both the child protection and education sectors will be required to implement the Education Module, which would involve national-level data linkage. In-principle support for the implementation of Phase 1 (further described in Section 3) was received from the appropriate national community services and education committees in early 2013—SCCDSAC, and the Australian Education, Early Childhood Development and Youth Affairs Senior Officials Committee (AEEYSOC). The AIHW has received funding from SCCDSAC to roll out Phase 1 over a period of 18 months, commencing in September 2013.



# 1 Introduction

In Australia, state and territory governments are responsible for the welfare of over 35,000 children who are placed in child protection services, such as care and protection orders and out-of-home care (AIHW 2013a). Education is particularly important for children placed in child protection services, as it is integral to their overall development and wellbeing, and provides an important gateway to future employment and life opportunities. However, numerous studies, both local and international, have found that children in the care of the state have poorer educational outcomes than other children, including: poorer school grades; lower scores on standardised tests; developmental delays; higher rates of special education placements and repeating grades; behavioural and disciplinary problems; and higher absenteeism, truancy and drop-out rates (AIHW 2011; Department for Education 2010; Merdinger et al. 2005; Social Exclusion Unit 2003).

## Policy context

Improving the outcomes of children involved in statutory child protection services has been a high priority for Australian governments in recent years. Discussions about the need for a comprehensive child-level (unit record) national child protection data collection, inclusive of outcomes data, have been building over the past decade.

The *National Framework for the Protection of Australia's Children 2009–2020* (FaHCSIA 2012) and the *National Standards for out-of-home care* (FaHCSIA 2011) include education-specific national indicators including the proportion of:

- children on guardianship and custody orders achieving at or above the national minimum standards for literacy and numeracy (*National Framework indicator 4.5*)
- children and young people achieving national reading and numeracy benchmarks (*National Standards measure 6.1*)
- young people who complete year 10 and the proportion who complete year 12 or equivalent VET training (*National Standards measure 7.1*)

These indicators do not currently have an ongoing national data source. Consequently, the development of an ongoing national data collection on the educational outcomes of children in child protection services is of increasing importance, as there is a strong need to fill these data gaps, to support regular reporting against these national initiatives.

A national linked data set will also allow ongoing and longitudinal monitoring of academic progress, to better inform policy, practice and planning of activities to support these children.

## Previous work

The need for national data on the educational outcomes of children in child protection services was identified in 2004 as part of the former Community Services Ministers' Advisory Council's (CSMAC, later renamed CDSMAC, and currently named SCCDSAC) strategic agenda for the protection and care of children.

## Pilot study

A pilot study was undertaken to progress data development in this area, and to build on previous work by Queensland and South Australian government departments. The AIHW was funded by CSMAC to undertake this work in conjunction with the states and territories. This project was a longitudinal pilot study designed in two stages:

- Stage 1 provided a snapshot of the academic performance of children on guardianship/custody orders who were in Grades 3, 5 or 7 in 2003. These findings were presented in a report (AIHW 2007), which included cross-sectional data from Victoria, Queensland, South Australia, Tasmania and the Australian Capital Territory.
- Stage 2 followed up these children over the next 3 years (2004 to 2006), with the aim of identifying any change in their academic performance over time, as well as providing a second snapshot of the educational achievement of all children on guardianship/custody orders. There were some changes to the participating jurisdictions in Stage 2 – the Australian Capital Territory did not continue due to issues with sample sizes that arose in Stage 1, and Western Australia joined the project, also providing 2003 data. The findings were presented in a second report (AIHW 2011), and included cross-sectional and longitudinal data from Victoria, Queensland, Western Australia, South Australia and Tasmania.

The pilot project confirmed that children on guardianship/custody orders are an academically disadvantaged group, and highlighted the importance of continuing to monitor the academic progress of these children given the many policy and program initiatives in existence that aim to tackle this issue.

However, the pilot data collection was very time-intensive for jurisdictions and a more efficient methodology is required to enable ongoing national collection of these data. The lessons learned and information gathered through the course of the pilot can be used to guide future work and inform national discussions relating to data development issues.

## Development of an Education Module

The former CDSMAC (now SCCDSAC) funded the AIHW to develop, in collaboration with the states and territories, a national methodology for reporting on the educational outcomes of children in child protection services. This project built on the data development work previously undertaken in this area, with the goal of creating an Education Module for linkage with the unit record (child-level) CP NMDS.

The proposed methodology for the Education Module was developed by the AIHW following consultation with a range of stakeholders over August 2011 to July 2012. This included the state/territory departments responsible for child protection and education, along with other data custodians such as the Australian Curriculum and Assessment Reporting Authority (ACARA), the state/territory NAPLAN Test Administration Authorities, the Australian Government Department of Human Services, and the National Centre for Vocational Education Research (NCVER). The Centre for Data Linkage at Curtin University was consulted regarding data linkage options. A desktop review of relevant national indicators from the *Measurement Framework for Schooling in Australia 2010* and relevant education data held by the Australian Bureau of Statistics was also undertaken. A full list of stakeholders consulted is included in Appendix 1.

This working paper summarises the key findings of the final report presented to SCCDSAC in late 2012. In-principle support for the implementation of Phase 1 (further described in

Section 3) was received in early 2013 from the appropriate national community services and education committees (SCCDSAC and AEEYSOC).

## 2 Proposed methodology

### Purpose

The proposed Education Module is set of education-specific data items for linkage with the unit record (child-level) CP NMDS. As the CP NMDS data are required to identify in-scope children, this should be viewed as the 'base' data set for the Education Module.

The main purpose of the Education Module is to provide a national linked data set on the educational activity and outcomes of children while in child protection services, to allow ongoing and longitudinal monitoring of their academic progress, and to better inform policy, practice and planning of activities to support these children. The Education Module will be developed as part of the Phase 2 expansion of the CP NMDS, and involve linkage of state/territory and national data, utilising existing administrative data where possible.

The development of a national linked data set in this field is a high government priority, and the previous pilot project confirmed the value of continuing to monitor the academic progress of children in care (AIHW 2011). Where possible, it is important to compare the progress and outcomes of these children to those of other children (for example, children not in care, and all children).

### Population scope

The scope of the Education Module is children and young people aged 0-17 years whose care arrangements have been ordered through the Children's Court, where parental responsibility for the child or young person has been transferred to the Minister/Chief Executive.

This population scope was selected to align with the agreed scope for the *National Standards for out-of-home care* (FaHCSIA 2011).

### Measurement framework

A broad measurement framework (Table 2.1) has been developed which provides an overview of key issues of interest regarding the education of children in the target population. The framework is intended to guide the inclusion of appropriate data items in the Education Module, and may be useful in identifying options and priorities for future expansion of the Module. Not all items currently have a readily available administrative data source. Data development may be desirable in future to progress these items.

Given the scope of children in the Education Module (see above), the framework focuses on issues related to children aged less than 18 years, for which analysis and reporting may be feasible using administrative data sources. The framework may evolve over time, and non-administrative data may provide a supplementary source of valuable information in future—for example, survey data on school experiences such as bullying, attitudes to school, and relationships with teachers and other students.

The measurement framework focuses on school-based education. Early childhood is also of interest, however due to other national data linkage work under development in that area, it would be prudent to delay the inclusion of these items (including AEDI data) in the

Education Module until that work is further developed, and opportunities for linkage are further explored.

Key issues of interest about the target population of children have been grouped into five broad areas:

- **Participation:** What type of schooling are they enrolled in? Are they remaining in school to years 10 and 12? Are children experiencing multiple changes in schools? Are they attending school? How common are formal actions such as suspensions, exclusions, expulsions, exemptions, and cancellation of enrolments?
- **Progress:** Are they meeting the national minimum standards for literacy and numeracy? Have they repeated a school year? Are they developmentally 'on track' at school entry (as measured by the AEDI)?
- **Attainment:** Have they completed Year 12 or equivalent? What VET qualifications have they attained?
- **Support:** What support/services are they receiving in schools? Are they enrolled in a special education program? Do they have an individual learning plan in place? Are they receiving Youth Allowance?
- **Background characteristics:** What other characteristics may be associated with their educational outcomes (for example, age at first guardianship/custody order, length of time on orders)? How common are disabilities and learning/behavioural difficulties among this group of children?

**Table 2.1: Measurement framework for the Education Module**

Participation	Progress	Attainment	Support	Background characteristics
Enrolment <sup>(a)</sup> School attendance/absence Formal actions (for example, suspensions, expulsions)	AEDI results (year 1) <sup>(b)</sup> NAPLAN results (years 3, 5, 7 and 9) <sup>(c)</sup> Repeated a school year	Year 12 certificate VET qualifications	Participation in special education program Educational assistance/support received in schools <sup>(d)</sup> Learning plan in place Receipt of education-related income support (for example, Youth Allowance)	Age Sex Ethnicity <sup>(e)</sup> Parent/guardian education Parent/guardian occupation Disability status Learning and behavioural difficulties Remoteness Child protection-related characteristics <sup>(f)</sup>

(a) Enrolment could include: enrolment start/end dates, year level of enrolment, school ID, and type of school/institution (for example, mainstream school, special education school, VET). This should allow calculation of number of schools enrolled in, and retention to years 10/12.

(b) The AEDI is a national progress measure of early childhood development in Australia, and provides an indication of children who are developmentally vulnerable at school entry.

(c) The NAPLAN assesses student performance against national minimum standards for reading, writing, spelling, grammar and punctuation, and numeracy.

(d) May include teacher aide support, specialist behaviour classes, gifted classes, tutoring programs, language support programs, etc.

(e) May include characteristics such as Indigenous status, language background other than English (LBOTE), cultural and linguistic diversity (CALD), and country of birth.

(f) May include characteristics such as length of time on orders, type of living arrangements, stability of living arrangements, and type of abuse. This will be achieved through data linkage with the CP NMDS—these items will not be collected as part of the Education Module.

## Data sources of interest

Data sources of interest for the Education Module are outlined in Table 2.2. Where possible, administrative data sources related to the measurement framework (see Table 2.1) have been identified, with the aim of capturing a variety of data across the primary and secondary schooling years (see Figure 2.1). This is not intended to be an exhaustive list, and it is acknowledged that other relevant data sources may also be available.

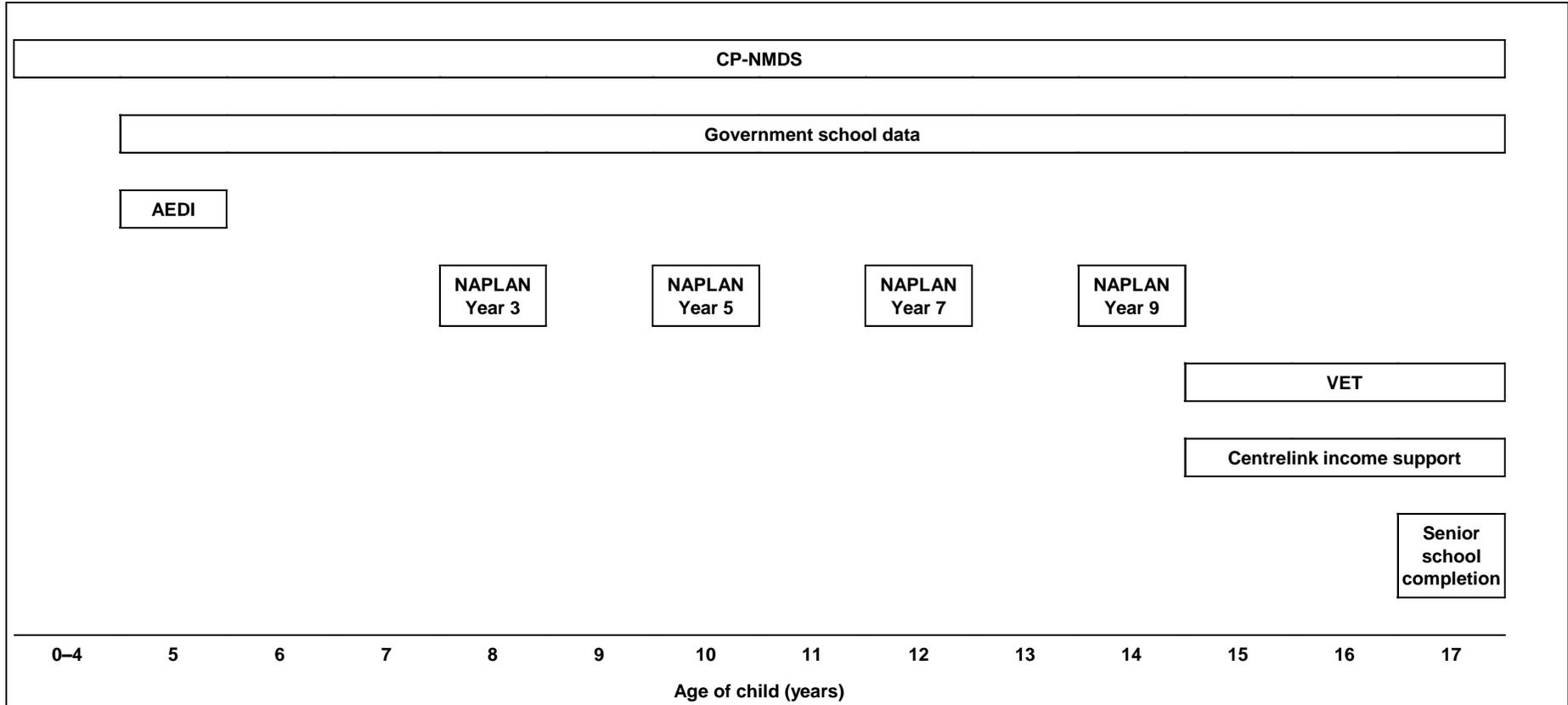
To allow data linkage to the CP NMDS, data for these sources will need to be available in unit record (child-level) format, and include suitable data items to allow linkage. Further consultation with data custodians on data availability and quality is likely to be required for some items.

**Table 2.2: Data sources of interest<sup>(a)</sup>**

Potential data source	Items of interest	Data custodian	Notes
Child Protection National Minimum Data Set (CP NMDS)	<ul style="list-style-type: none"> <li>Orders data (start/end dates).</li> <li>Other child protection-related characteristics (e.g. time on orders, living arrangements)</li> </ul>	AIHW (national data custodian)	<p>The CP NMDS is expected to be operational from the 2012–13 collection (replacing the current aggregate collection).</p> <p>Any project involving CP NMDS data linkage would be subject to approval by each state/territory custodian and the AIHW Ethics Committee.</p>
National Assessment Program—Literacy and Numeracy (NAPLAN)	<ul style="list-style-type: none"> <li>Results (e.g. participation, test scores, etc.) from the 5 assessment domains</li> <li>Student background characteristics (e.g. sex, Indigenous status, etc.)</li> </ul>	NAPLAN Test Administration Authorities (state/territory education department or board/authority)	<p>Each state/territory has its own data release policies and protocols.</p> <p>NAPLAN data are available for government and non-government schools.</p> <p>ACARA holds some national-level unit record NAPLAN data, but the national data sets do not include all the data items required to undertake data linkage with the CP NMDS.</p>
VET provider collection (students and courses) VET in schools collection	<ul style="list-style-type: none"> <li>Participation in VET (e.g. enrolment, qualification type)</li> <li>Qualification completed, particularly completion of Year 12 equivalent (e.g. Australian Qualifications Framework Certificate II)</li> </ul>	National Centre for Vocational Education Research (NCVER)	<p>Unit record data can be requested by external users for the purposes of research.</p> <p>The feasibility of accessing VET data for linkage is to be further investigated. At present, limited data linkage items are available. The introduction of a national unique student identifier in 2014 may allow data linkage in future (IICCSRTE 2013).</p>
Australian Early Development Index (AEDI)	<ul style="list-style-type: none"> <li>Results from the 5 developmental domains</li> <li>Educational experiences in the year before entering school</li> <li>Child demographics</li> </ul>	Australian Government Department of Education (Education)	<p>Education has committed to making AEDI data available for linkage.</p> <p>AEDI data are being explored for inclusion in a number of other national data linkage projects currently under development.</p>
Government school data <sup>(b)(c)</sup>	<ul style="list-style-type: none"> <li>School enrolment</li> <li>School attendance/absence</li> <li>Formal actions (e.g. suspensions, expulsions)</li> <li>Learning plan in place</li> </ul>	State/territory education departments	<p>Each state/territory has its own data release policies and protocols.</p> <p>Data will be available for government schools only. Data availability/quality varies across jurisdictions.</p>
Senior school qualification data	<ul style="list-style-type: none"> <li>Completion of senior secondary school certificate (Year 12)</li> </ul>	State/territory education board/authority (e.g. NSW Board of Studies )	<p>In many states/territories, the education board/authority is an independent statutory authority serving government and non-government schools, and therefore may have data release policies and protocols</p>

			that differ from those of the education departments. Data are available for government and non-government schools.
Centrelink income support	<ul style="list-style-type: none"> <li>• Youth Allowance recipients</li> <li>• ABSTUDY recipients</li> </ul>	Australian Government Department of Human Services (DHS)	Data access to be negotiated with DHS and Education. Unit record data can be requested by external users for research purposes in specific circumstances.

- (a) This table provides an overview. Further detail on the CP NMDS and NAPLAN data sources is provided in Appendix 2.
- (b) Initial consultation with state/territory education departments suggests that data related to disability, learning and behavioural difficulties, and relevant supports for individual students may be less readily available on their electronic departmental data systems than other items of interest. As such, these items have not been included in Table 2.2, but may be further explored in future.
- (c) Non-government school data have not been explored. Non-government schools maintain their own data systems and have their own reporting arrangements. Selected items are reported for national data collections.



*Notes*

1. Age of child is indicative. Age at various stages of schooling will vary for individual students and across jurisdictions. Although formal education does not usually commence until 4–5 years of age, a research question of interest is whether age at first guardianship/custody order is associated with educational outcomes.
2. Further information on these data sources is provided in Table 2.2.

**Figure 2.1: Expected reference age for data sources of interest**

## Longitudinal data

The in-scope population will be captured over multiple waves of data. This population is not a stable one, as children tend to move in and out of child protection services over the years, so a fixed cohort approach (for example, birth cohort) will result in high attrition rates, thereby limiting the longitudinal sample sizes. As such, a similar approach to Stage 2 of the pilot (see Section 1; AIHW 2011) is proposed, where each wave includes cross-sectional data on all children falling into the in-scope population in that year (regardless of whether they appeared in earlier waves)—this allows the total sample to be ‘topped up’ regularly (see Figure 2.2). Longitudinal data will be provided by tracking the children in each wave over the remaining waves. It is acknowledged that longitudinal data are complex, and appropriate methodologies will need to be developed to handle and analyse these data. Advice received from state/territory child protection departments during the pilot indicated there are limitations as to which data can be accessed once a child has exited the jurisdiction’s guardianship.

It should be noted that not all data sources will provide annual data for each child. Over a four-year period, most children will have no more than 2 years of NAPLAN test results—for example, they may have sat the Grade 3 test in 2013 and the Grade 5 test two years later in 2015. Similarly, not all data described in Figure 2.1 will be available for each child in the in-scope population—a child would need to remain continuously in care from the age of 5 to 17, and the Education Module would need to include 13 years of data (commencing 2012–13, when CP NMDS data are expected to first become available).

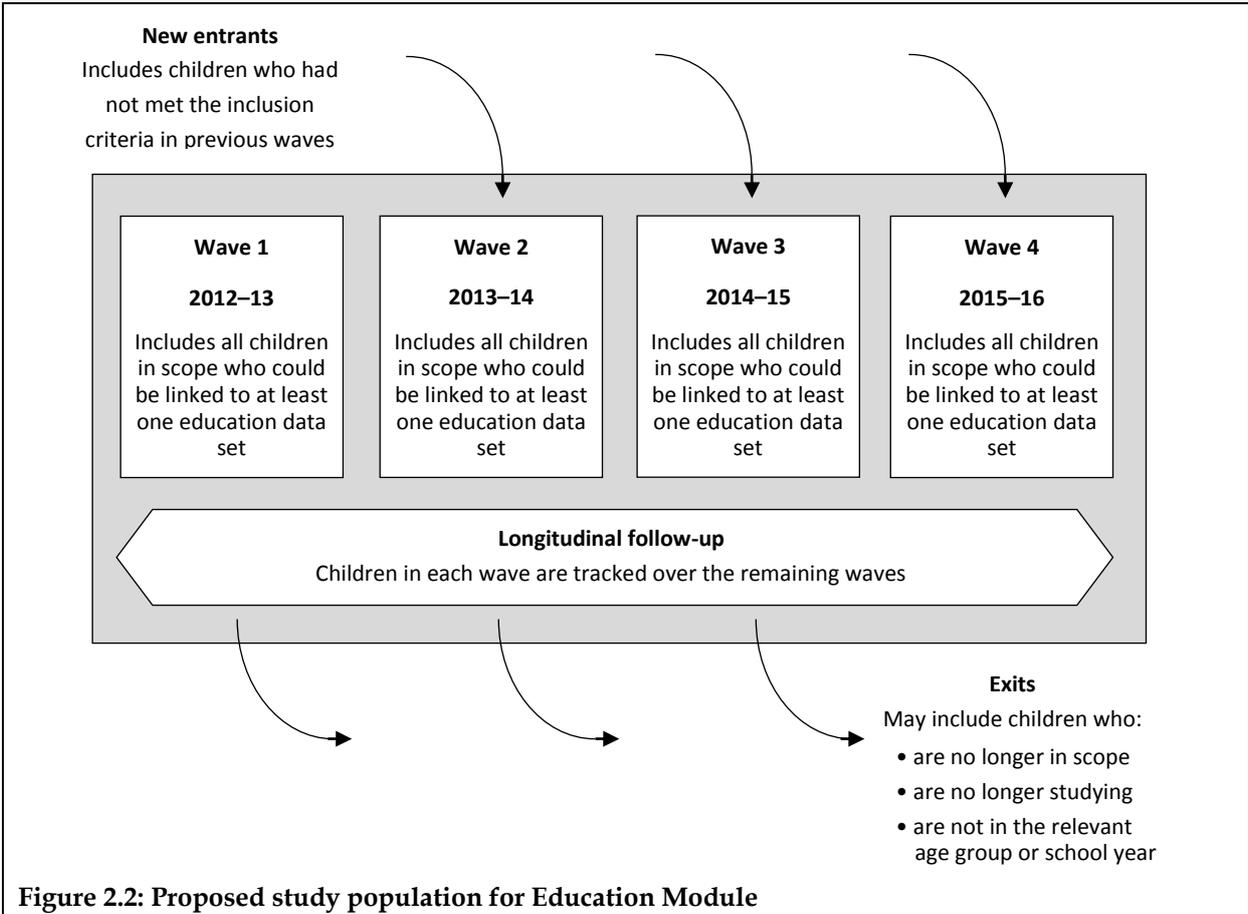


Figure 2.2: Proposed study population for Education Module

## Data access

The process of accessing data from the various custodians is one of the most time-consuming components in establishing a linkage project. Given the national scale of the Education Module, it is expected to be a complex and lengthy process to negotiate the parameters of the project, and to establish the necessary agreements and processes to access the data for linkage. The proposed data linkage process is described below.

The development of a broad Education Module will require access to data from a range of stakeholders. Essentially, approval will be required from each of the data custodians, whether at the national level, or from the individual custodian in each jurisdiction. In some jurisdictions, state-level linkage across administrative data sets is well developed, and relevant agreements and processes are already in place. Where possible and appropriate, it would be beneficial to build on existing work in jurisdictions, while acknowledging that a nationally-consistent approach to data linkage is highly desirable.

There may be several options to be explored in relation to data access. This will largely hinge on the agreed data linkage process, the resulting partnership arrangements that will need to be established, and resolving the relevant legislation and ethics issues. A brief commentary based on the current status of each data set is provided above in Table 2.2.

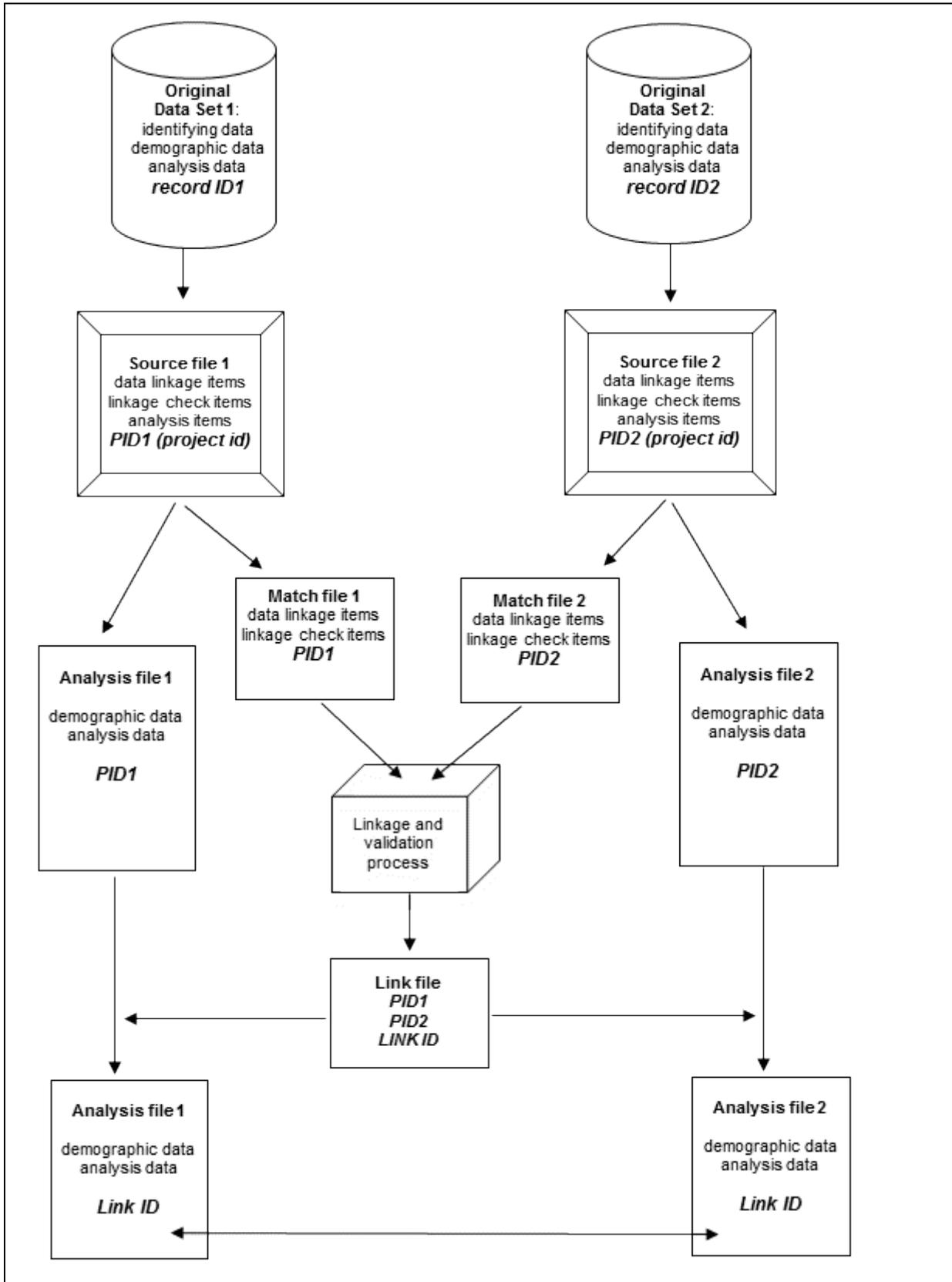
## Data linkage

Given the multiple data sources of interest for the Education Module, it is likely that data linkage will be conducted in stages. Several linkage models were considered, but the preferred model would involve the AIHW creating data linkage keys and integrating all data, with data custodians providing data directly to the AIHW. These data custodians would separate the data items required for linkage from those items required for analysis, and send them separately to the AIHW. This would ensure separation of the linkage process from any processes that extract and deliver content data for analysis. The linkage and integration would be undertaken by the AIHW Data Integration Services Centre (DISC) using the separation principle, whereby staff only have access to those data sets required for the particular operation they are undertaking at any point in time. Staff undertaking linkage can only access identifying items (such as names and dates of birth) and staff undertaking integration can only access analysis items (see Figure 2.3).

External separation by the data custodians and internal merging (in conjunction with controlled data access mode) provides the strictest control over privacy and data access in accordance with the guidelines for *Data Integration Involving Commonwealth Data for Statistical and Research Purposes* (NSS 2010).

This approach will use nationally-consistent linkage processes (particularly for state/territory-level data sets), ensuring uniform linkage results across states. The approach allows best practise linkage methods to be applied to enhance match rates and operational efficiency. The variety of data sources and existing (or developing) data linkage mechanisms were also taken into consideration.

It may be desirable to provide analysis data following the linkage process (known as 'post linkage content provision'). This would allow data custodians to provide only the analysis data required for in-scope clients to the AIHW. In this process, only the linkage file is initially provided. After linkage, a set of project specific identifiers are returned to the data custodian, allowing them to provide only those analysis records for linked clients, thereby



Source: AIHW 2013b.

Figure 2.3: Schematic example illustrating the AIHW data linkage principles

streamlining the data extraction and integration process. Post-linkage content provision is not appropriate if the integrated data comprises both linked and unlinked records.

Under any model, the data linkage will need a mix of different processes including probabilistic linkage where names are available, and linkage using statistical linkage keys (SLK) where identifiers are not readily available. An SLK is a string of characters and numbers that contains sufficient information to link records for statistical analysis, but does not contain sufficient information to identify individuals. They are typically used to link data sets where information such as full name is not available (for example, the SLK-581 includes selected letters of name, date of birth, and sex). In particular, the CP NMDS has an SLK-581 which allows it to be linked to other data sets that include matching SLK information, reducing the need for identifiers. The AIHW has developed an enhanced stepwise deterministic method to maximise the quality of data linkage using the SLK-581, supplemented by other suitable items as available.

It is proposed that the AIHW undertake the data linkage process. The AIHW is a Commonwealth-accredited Data Integration Authority, thus allowing the AIHW to successfully manage data linkage projects which are considered to pose a high systemic risk. Commonwealth accreditation provides assurance that the AIHW has been audited, and demonstrated a rigorous standard of security, confidentiality, privacy, staff skills, IT infrastructure, and a highly secure working environment (NSS 2013). Additionally, this allows the entire set of integrated data to be covered by the AIHW legislative arrangements regarding data provision (AIHW 2012). Following accreditation by the Cross Portfolio Integration Oversight Board, the AIHW is further developing its integration facility. The AIHW is also a collaborating participant of the Population Health Research Network which has been established to build nationwide data integration infrastructure.

## Privacy and confidentiality

Privacy and confidentiality must be considered whenever data about individuals, service provider organisations or funding departments are collected or disseminated. Consultation with data custodians is required to ensure privacy and confidentiality issues are addressed appropriately, particularly given the complex and sensitive nature of the proposed data set.

The *Privacy Act 1988* (Privacy Act) protects personal information. Section 6 of the Privacy Act defines personal information as information or an opinion about an individual whose identity is apparent or can be reasonably ascertained from that information or opinion. An individual is a natural person only, and the Act does not protect the information of deceased individuals or other entities. As a government agency, the AIHW is bound by the requirements of the Information Privacy Principles (IPPs) contained within section 14 of the Act, which regulate how government agencies deal with personal information. Specifically, the IPPs lay down rules around how agencies collect, store and release personal information.

The AIHW is also bound by the *Australian Institute of Health and Welfare Act 1987* (AIHW Act). Section 29 of the AIHW Act applies criminal penalties to any individual who releases 'information concerning a person' outside of the limited permitted exceptions provided in the Act. For the purposes of section 29, 'person' is defined broadly to include other entities such as bodies-corporate and deceased individuals. The penalties for breach of section 29 extend to employees, contractors, researchers who have obtained identifiable information through approved ethics applications and individuals who have obtained information concerning a person under an agreement with the Institute.

The AIHW has an Ethics Committee that assists in ensuring compliance with relevant ethical guidelines. The Australian Institute of Health and Welfare Ethics Committee Regulations 1989 (the Regulations) provide that the committee has the power to form an opinion on ethical grounds about the acceptability of any Institute activity. As a matter of policy, the Ethics Committee requires that all AIHW projects involving data linkage, release of identifiable information or the creation of new data sets are presented to the committee for approval.

The committee forms its opinion on the acceptability of a project with reference to National Health and Medical Research Council guidelines such as the *National Statement on Ethical Conduct in Research Involving Humans*, relevant legislative obligations such as the Privacy Act, AIHW Act, the regulations and relevant scientific standards.

In June 2012, the AIHW was fully accredited as an Integrating Authority by the Commonwealth Data Integration Oversight Board, thus allowing the AIHW to successfully manage data linkage projects which are considered to pose a high systemic risk (see 'Data linkage' above).

### 3 Implementation

Implementation of the Education Module is likely to be complex and take a number of years. Based on investigation throughout the development project (see Section 1), the AIHW proposes that a phased approach will be most manageable.

Although a range of data sources of interest have been identified, addressing the current national reporting gaps under the *National Framework for the Protection of Australia's Children 2009–2020* and *National Standards for out-of-home care* (see Introduction) is a key priority. As such, it is proposed that the linkage of CP NMDS and NAPLAN data be the initial focus of the Education Module, to support national reporting under these frameworks on the achievement of national minimum standards for reading and numeracy by children in care.

The AIHW proposes a phased approach to progressively build a longitudinal linked data set. Proposed phases are outlined in Table 3.1. Linkage of CP NMDS and NAPLAN data has been prioritised as Phase 1, to enable reporting to the Council of Australian Governments (COAG) within the 2009–2020 lifetime of the *National Framework*. The Phase 1 data sets are also expected to include nationally-consistent data and adequate items for data linkage. The NAPLAN data also have a clearly specified timeframe for data linkage – NAPLAN is administered over 3 national testing days in May each year – whereas other data sets are expected to have more complex and variable timeframes for linkage purposes (for example, VET data are collected for the calendar year, CP NMDS data are collected for the financial year, and children may move in and out of both collections during these periods, thereby complicating the data linkage process).

The various data sources have different reference periods. The CP NMDS is the 'base' data set for the Education Module, and will be used to identify in-scope children. Additional data sets and years of data may be added over time as the relevant CP NMDS data become available (expected to commence in 2012–13).

Based on the detailed consultation undertaken, the best approach would be for the Education Module to initially comprise the data sets which have a greater degree of national standardisation (for example, CP NMDS, NAPLAN, Centrelink), with state-specific data sets added later on. This will allow the data linkage process to be refined for the larger national data sets first, and will allow more time for the development of national data specifications for seeking data from the state-specific collections.

Furthermore, the AIHW is aware that the AEDI and a range of school-based data are being explored for inclusion in the other national data linkage projects currently under development. It is expected that some of these data will meet the data needs of the Education Module. To minimise duplication of any national data linkage work, it would be prudent to delay the inclusion of these items in the Education Module until those projects are further developed, and opportunities to link to those data sets are further explored.

**Table 3.1: Proposed phased approach to building a longitudinal linked Education Module**

Potential data source	Proposed phases			
	1	2	3	4
CP NMDS <sup>(a)</sup>	2012–13	2013–14	2014–15	2015–16
NAPLAN <sup>(b)</sup>	2013	2014	2015	2016
Centrelink income support <sup>(c)</sup>		2013–14	2014–15	2015–16
VET <sup>(d)</sup>			2014	2015
Senior school qualifications <sup>(e)</sup>			2014	2015
Government school data <sup>(e)</sup>			2014	2015
AEDI <sup>(f)</sup>		2012 <sup>(f)</sup>		2015
<i>Potential timing of implementation<sup>(g)</sup></i>	<i>Mid 2013— late 2014</i>	<i>Mid 2014— late 2015</i>	<i>Mid 2015— late 2016</i>	<i>Mid 2016— late 2017</i>

(a) CP NMDS data are for the financial year (1 July to 30 June).

(b) NAPLAN testing occurs in May each year.

(c) Data are expected to be available for the financial year (1 July to 30 June).

(d) At present, limited data linkage items are available for national VET data. However, the introduction of a national unique student identifier in 2014 may allow data linkage in future (IICCSRTE 2013). VET data are for the calendar year (1 January to 31 December). As such, two years of CP NMDS data (for example, 2013–14 and 2014–15) will be required for linkage to one year of VET data (for example, 2014).

(e) Data are expected to be available for the calendar year (1 January to 31 December).

(f) AEDI data are available every 3 years (2009, 2012, and 2015). For the 2009 AEDI, data were collected between 1 May and 31 July. As such, two years of CP NMDS data (for example, 2014–15 and 2015–16) will be required for linkage to one year of AEDI data (for example, 2015). If deemed appropriate, 2012 AEDI data could be pursued in Phase 2, pending progress with other national linkage projects; however data for 1 May to 30 June 2012 (part of the AEDI collection period) will not be available from the 2012–13 CP NMDS.

(g) Implementation is likely to include development of technical specifications, attaining relevant data access permissions and ethics approvals, data linkage, data analysis, and reporting on results. The suggested timeframes are estimates, pending support from key stakeholders and available resources.

## Phase 1 implementation

Phase 1 will involve linkage of CP NMDS and NAPLAN data – further details on these data sources are provided in Appendix 2. It is expected that Phase 1 implementation could be rolled out over a period of 18 months, commencing in September 2013 (Table 3.2).

**Table 3.2: Potential workplan for Phase 1 Education Module implementation**

Activity/ Key milestone	Expected completion date
Develop data collection specifications	End September 2013
Secure data-access permissions, ethics approvals, data linkage arrangements	End February 2014
AIHW to receive data for linkage	End May 2014
Complete data linkage and analysis	End September 2014
Complete project report, including recommendations for future phases	End November 2014
Circulate project report to relevant committees and key stakeholders	End February 2015

*Note:* The suggested timeframes are estimates, pending support from key stakeholders, available resources, and the availability/quality of 2012–13 CP NMDS data.

Appropriate governance arrangements will need to be established. High-level support from both the child protection and education sectors will be required for national-level data linkage. In-principle support for implementation of Phase 1 was received in early 2013 from

the appropriate national community services and education committees – SCCDSAC and AEEYSOC.

The AIHW has received funding from SCCDSAC to implement Phase 1. It is vital that appropriate additional resources are allocated for other organisations contributing to Phase 1 – in particular, the state/territory NAPLAN data custodians. It will not be possible to progress the Education Module within existing resources.

A project working group should also be established to oversee ongoing administration and implementation of Phase 1 of the Education Module. This group could include suitable representatives from state/territory child protection departments, state/territory NAPLAN data custodians, the non-government school sectors and the AIHW. The working group membership can be expanded as appropriate in subsequent phases to include additional data custodians.

It is expected that minimal data development work will be required for Phase 1, as both data sources are national-level collections with established data specifications. A set of detailed technical specifications will need to be developed for Phase 1 of the Education Module and agreed by key stakeholders (including the AIHW and state/territory data custodians). These specifications should include: governance arrangements including roles and responsibilities; the specific data items required, including those needed for data linkage; data clearance, provision and linkage processes; and how the data will be used (for example, for analysis, reporting, access to the linked data set). The information in this report will serve as a useful starting point for the technical specifications.

Arranging data access is likely to be a complex process. NAPLAN data will be sought from the state/territory data custodians – the Test Administration Authorities in each jurisdiction, which hold data for government and non-government school students. Each jurisdiction has its own data access protocols and requirements for releasing NAPLAN data. As such, permissions and arrangements for the Education Module would need to be negotiated on a state-by-state basis, and may involve the non-government school sectors. It is likely that formal agreements will need to be developed between the AIHW and each state/territory data custodian to allow access to NAPLAN data for linkage with CP NMDS data on an annual basis.

Consultation with data custodians will be required to ensure privacy and confidentiality issues are addressed appropriately, particularly given the complex and sensitive nature of the proposed data set. The AIHW will need to seek approval from the AIHW Ethics Committee to ensure the Education Module complies with relevant ethical legislation and guidelines. Each jurisdiction may also have its own ethics approval process which will need to be completed.

## **Implementation of additional phases**

Following Phase 1, the progress of other national education-related data linkage projects should be reviewed, to assess opportunities to use those data sets and minimise duplication.

The implementation of additional phases of the Education Module will involve similar tasks to those described above, including: seeking funding for implementation, establishment of governance and data procurement arrangements, development of appropriate technical specifications, and ethics clearances.

Possible timeframes for implementation of Phases 2–4 are provided in Table 3.2. These timeframes are estimates, pending support from key stakeholders, available resources, and the success of Phase 1 implementation. It may be feasible to run components of multiple phases in parallel.

Data sets which have a lower degree of national standardisation are likely to require more data development because common data items and national data specifications are not readily available (for example, government school data vary across states and territories, see Table 2.2).

# Appendix 1: Stakeholder consultation

The proposed methodology was developed by the AIHW following consultation with a range of stakeholders over August 2011 to July 2012. Further details are provided in Section 1.

**Table 1: Stakeholders consulted during development of the proposed data collection**

<b>Jurisdiction</b>	<b>Department/organisation</b>
NSW	Department of Family and Community Services
	Department of Education and Communities
	Office of the Board of Studies
Vic	Department of Human Services
	Department of Education and Early Childhood Development
	Victorian Curriculum and Assessment Authority
Qld	Department of Communities
	Department of Education and Training
	Queensland Studies Authority
WA	Department for Child Protection
	Department of Education
	School Curriculum and Standards Authority
SA	Department for Education and Child Development
	South Australian Certificate of Education Board of South Australia
Tas	Department of Health and Human Services
	Department of Education
	Tasmanian Qualifications Authority
ACT	Community Services Directorate
	Education and Training Directorate
	Office of the Board of Senior Secondary Studies
NT	Office of Children and Families
	Department of Education and Training
Commonwealth	Australian Government Department of Human Services
	Australian Curriculum and Assessment Reporting Authority
Other	National Centre for Vocational Education Research
	Centre for Data Linkage

## Appendix 2: Phase 1 data sources

### Child Protection National Minimum Data Set (CP NMDS)

Child protection data are 'owned' by the state/territory departments responsible for child protection, but the AIHW is the data custodian for the national minimum data set. Utilising the CP NMDS for the Education Module will minimise duplication and reduce the effort required from the state/territory departments.

The unit record (child-level) CP NMDS is expected to fully replace the existing aggregate data collection from 2013. As such, unit record data from the CP NMDS are expected to be operational from the 2012–13 collection, pending data quality assessment. As the CP NMDS is the 'base' data set for the Education Module (and will be used to identify in-scope children), any issues with the availability or quality of CP NMDS data may impact the capacity to undertake the data linkage required for the Education Module.

CP NMDS data items of interest include:

- data linkage items (for example, letters of name, person IDs, date of birth, sex)
- items required to identify in-scope children (for example, National Standards for out-of-home care flag, order start/end dates)
- various child protection-related characteristics (for example, time on orders, living arrangements)

Any project involving data linkage with the CP NMDS would be subject to approval by each state/territory custodian and the AIHW Ethics Committee.

### National Assessment Program—Literacy and Numeracy (NAPLAN) data

In Australia, national minimum standards have been developed for reading, writing, spelling, language conventions (grammar and punctuation) and numeracy for students in Years 3, 5, 7 and 9. Achievement against these standards is assessed on an annual basis through the NAPLAN. Students who achieve the minimum standards have demonstrated at least the basic understanding required for their year level.

NAPLAN data for government and non-government school students are held by the Test Administration Authorities (TAAs) in each state and territory (ACARA 2013):

- the NSW Board of Studies
- the Victorian Curriculum and Assessment Authority
- the Queensland Studies Authority
- the WA School Curriculum and Standards Authority
- the SA Department for Education and Children's Services
- the Tasmanian Department of Education
- the ACT Education and Training Directorate
- the NT Department of Education and Children's Services.

The TAAs hold identifying NAPLAN data for their state/territory (including the student's full name and date of birth). Each jurisdiction has its own data access protocols and requirements for releasing NAPLAN data; permissions and arrangements for the Education Module would need to be negotiated on a state-by-state basis, and may involve the non-government school sectors.

Data items of interest from the NAPLAN collection include:

- data linkage items (for example, letters of name, student IDs, date of birth, sex)
- results from the 5 NAPLAN tests, including:
  - name of test (for example, reading)
  - year level (for example, Year 3)
  - participation (for example, sat test, exempt, absent, withdrawn)
  - test achievement score
  - percentile
  - national achievement band
- child demographics, including:
  - Indigenous status
  - language background other than English (LBOTE) status
  - geolocation (for example, metro, provincial, remote, very remote)
  - parental education
  - parental occupation (ACARA 2012; MCEECDYA 2009).

It is expected that these data items would be available from the state/territory Test Administration Authorities. Permissions and arrangements to access these data for inclusion in the Education Module would need to be negotiated on a state-by-state basis, including all parties involved in the local arrangements.

ACARA is the data custodian of national-level de-identified NAPLAN data. However, the data held by ACARA does not contain all the information required to create a statistical linkage key – specifically, it does not contain student IDs or any name information, although date of birth, sex and Indigenous status are included. As such, the ACARA-held NAPLAN data is not considered a feasible data source for the linkage requirements of the Education Module.

# Acronyms and abbreviations

ACARA	Australian Curriculum, Assessment and Reporting Authority
ACT	Australian Capital Territory
AEDI	Australian Early Development Index
AEEYSOC	Australian Education, Early Childhood Development and Youth Affairs Senior Officials Committee
AIHW	Australian Institute of Health and Welfare
CDSMAC	Community and Disability Services Ministers' Advisory Council
COAG	Council of Australian Governments
CP NMDS	Child Protection National Minimum Data Set
CSMAC	Community Services Ministers' Advisory Council
DEEWR	Australian Government Department of Education, Employment and Workplace Relations
DHS	Australian Government Department of Human Services
DISC	Australian Institute of Health and Welfare Data Integration Services Centre
Education	Australian Government Department of Education (formerly DEEWR)
FaHCSIA	Australian Government Department of Families, Housing, Community Services and Indigenous Affairs
NAPLAN	National Assessment Program—Literacy and Numeracy
NCVER	National Centre for Vocational Education Research
NSW	New South Wales
NT	Northern Territory
Qld	Queensland
SA	South Australia
SCCDSAC	Standing Council on Community and Disability Services Advisory Council
SLK	statistical linkage key
TAA	Test Administration Authority (for NAPLAN)
Tas	Tasmania
VET	vocational education and training
Vic	Victoria
WA	Western Australia

# References

- ACARA (Australian Curriculum, Assessment and Reporting Authority) 2013. Test Administration Authorities. Viewed 13 September 2013, <<http://www.nap.edu.au/contacts/contact-test-administration-authorities.html>>.
- ACARA 2012. 2012 National Assessment Program: Literacy and Numeracy: achievement in reading, persuasive writing, language conventions and numeracy. Melbourne: ACARA.
- AIHW (Australian Institute of Health and Welfare) 2007. Educational outcomes of children on guardianship or custody orders: a pilot study. Cat. no. CWS 30. Canberra: AIHW.
- AIHW 2011. Educational outcomes of children on guardianship or custody orders: a pilot study, Stage 2. Cat. no. CWS 37. Canberra: AIHW.
- AIHW 2012. Privacy of data. Viewed 4 October 2012, <<http://www.aihw.gov.au/privacy-of-data/>>.
- AIHW 2013a. Child protection Australia 2011–12. Child welfare series no. 55. Cat. no. CWS 43. Canberra: AIHW.
- AIHW 2013b. Data linkage and protecting privacy: a protocol for linking between two or more data sets held within the Australian Institute of Health and Welfare. Viewed 1 October 2013, <<http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=6442468416&libID=6442468414>>.
- Department for Education 2010. Outcomes for children looked after by local authorities in England, as at 31 March 2010. London: UK Department for Education.
- FaHCSIA (Australian Government Department of Families, Housing, Community Services and Indigenous Affairs) 2011. An outline of National Standards for out-of-home care. A priority project under the National Framework for Protecting Australia's Children 2009–2020. Canberra: Commonwealth of Australia.
- FaHCSIA 2012. Protecting children is everyone's business. National Framework for Protecting Australia's Children 2009–2020: second three-year action plan 2012–2015. Canberra: FaHCSIA.
- IICCSRTE (Australian Government Department of Industry, Innovation, Climate Change, Science, Research and Tertiary Education) 2013. Unique Student Identifier (USI) initiative for VET. Viewed 29 May 2013, <<http://www.innovation.gov.au/SKILLS/NATIONAL/UNIQUESTUDENTIDENTIFIERFORVET/Pages/default.aspx>>.
- MCEECDYA (Ministerial Council for Education, Early Childhood Development and Youth Affairs) 2009. 2010 data standards manual: student background characteristics. Fifth edition. Melbourne: MCEECDYA.
- Merdinger JM, Hines AM, Osterling KL & Wyatt P 2005. Pathways to college for former foster youth: understanding factors that contribute to educational success. *Child welfare* 84(6):867–96.
- NSS (National Statistical Service) 2010. High level principles for data integration involving Commonwealth data for statistical and research purposes. Viewed 4 October 2012, <<http://www.nss.gov.au/nss/home.nsf/NSS/7AFDD165E21F34FDCA2577E400195826?opendocument>>.

NSS 2013. Accredited Integrating Authorities. Viewed 14 January 2013,  
<<http://www.nss.gov.au/nss/home.nsf/pages/Data%20Integration:%20Accredited%20Integrating%20Authorities>>.

Social Exclusion Unit 2003. A better education for children in care. London: Office of the Deputy Prime Minister.

Improving the educational outcomes of children involved in statutory child protection services has been a high priority for Australian governments in recent years. This working paper provides an overview of a proposed national linked dataset on the educational activity and outcomes of children while in child protection services, to allow ongoing and longitudinal monitoring of the academic progress, and to better inform policy, practice and planning of activities to support these children. National linkage of multiple administrative data sources is proposed, with the aim of capturing a variety of data across the primary and secondary schooling years.