Australian Government

Australian Institute of Health and Welfare

Developing the National Early Childhood Development Researchable Data Set

Australian Government
Australian Institute of Health and Welfare

Authoritative information and statistics
to promote better health and wellbeing

Developing the National Early Childhood Development Researchable Data Set

Australian Institute of Health and Welfare
Canberra

Cat. no. PHE 179

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 2014

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/>. 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 Digital and Media Communications Unit, Australian Institute of Health and Welfare, GPO Box 570, Canberra ACT 2601.

ISBN 978-1-74249-628-3

Suggested citation

Australian Institute of Health and Welfare 2014. Developing the National Early Childhood Development Researchable Data Set. Cat. no. PHE 179. Canberra: AIHW.

Australian Institute of Health and Welfare

Board Chair	Director
Dr Mukesh C Haikerwal AO	David Kalisch

Any enquiries about or comments on this publication should be directed to:
Digital and Media Communications Unit
Australian Institute of Health and Welfare
GPO Box 570
Canberra ACT 2601
Tel: (02) 6244 1000
Email: info@aihw.gov.au

Published by the Australian Institute of Health and Welfare

This publication is printed in accordance with ISO 14001 (Environmental Management Systems) and ISO 9001 (Quality Management Systems). The paper is sourced from sustainably managed certified forests.

MIX
From responsible sources
FSC® C017269

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> for any amendments.

Contents

Acknowledgments ..... **vi**

Abbreviations ..... **vii**

Summary ..... **ix**

1 Introduction **1**

1.1 Background 1

1.2 Policy context 1

1.3 National agenda for data linkage 2

1.4 Related data linkage projects 3

Building national linkage capability to the AEDC project 4

Proof of Concept project 3 4

Transforming Education and Training Information in Australia 4

National Education Data Linkage Roadmap 5

1.5 Structure of this report 5

2 The National Early Childhood Development Researchable Data Set **6**

2.1 Aim, value and benefits 6

Aim of the NECD RDS 6

Benefits to governments, policy makers and program developers 6

Benefits to the public 7

Questions that the NECD RDS may help answer 7

2.2 Scope and design 8

Research foundations of the NECD RDS 8

Data to be included 9

Documenting available data 10

2.3 Data linkage 14

Practice and protocols 14

Protecting privacy 14

3 Governance framework for the researchable data set **16**

3.1 Privacy considerations 16

AIHW legislative framework—privacy and confidentiality 16

3.2 Ethical considerations 16

Human Research Ethics Committees 17

AIHW Ethics Committee 17
3.3 Stakeholder engagement 18
3.4 NECD RDS Advisory Committee 20
4 Challenges and progress in establishing the researchable data set 21
4.1 Challenges encountered 21
Stakeholder engagement 21
Jurisdictional linkage capabilities 22
Ongoing nature of the NECD RDS 23
Changes to process for generating linkage keys 23
Privacy constraints for preschool data 23
Incomplete data 24
Barriers to obtaining data 24
Other data linkage projects 24
4.2 Obtaining permission to use data for the NECD RDS 25
Commonwealth data sets 25
Jurisdictional data sets 27
Summary of progress 32
Receiving data for inclusion in the NECD RDS 34
Supply and updating of data 36
5 Moving forward 37
5.1 Where are we now? 37
5.2 Next steps 37
5.3 After the NECD RDS is established 38
Stakeholder consultations and meetings 38
Access to the NECD RDS 38
Updating the NECD RDS 39
Data analysis 39
METeOR 40
Appendix A: Childhood development research 41
A.1 The review process 41
A.2 Results of mapping 41
Longitudinal research 42
Appendix B: Data sources to support the NECD RDS 43
B.1 Data included in the NECD RDS 43

B.2 Data to support the NECD RDS....44
Australian Childhood Immunisation Register 44
Medicare data 44
The National Death Index....45
ABS Census 45
Other data obtained through surveys 45
Appendix C: Data items for inclusion 47
Appendix D: Processes to construct the NECD RDS 50
D.1 Linking data to construct the NECD RDS 50
Constructing concordance maps....50
D.2 Merging content data for the NECD RDS 50
Appendix E: Guide for researchers 52
E.1 Access to data for researchers 52
E.2 Access options 53
E.3 AIHW contact details 55
Data Integration Services Centre 55
AIHW Data Linkage Unit 55
Custodian of the NECD RDS....55
AIHW Ethics Committee 55
Glossary....56
References 57
List of tables 60
List of figures 61
List of boxes....62

Acknowledgments

The primary authors of this report were Ilona Brockway, Bernadette Kok, and Louise Wangerek of the Maternal Health, Children and Families Unit of the Australian Institute of Health and Welfare (AIHW). The authors particularly thank Fadwa Al-Yaman, Cathy McNickle, Marianna Stylianou and Conan Liu from the Indigenous and Children's Group, Phil Anderson from the Data Linkage Unit, and the Governance Unit for providing assistance and guidance.

The AIHW would like to acknowledge the funding and valuable input from the Early Childhood Data Sub Group to this project as part of the National Information Agreement on Early Childhood Education and Care work plan.

The South Australian Department of Education and Child Development and the Telethon Institute for Child Health Research collaborated with the AIHW in the developmental phases of this project.

Abbreviations

ABS	Australian Bureau of Statistics
ACIR	Australian Childhood Immunisation Register
AEDC	Australian Early Development Census
AIFS	Australian Institute of Family Studies
AIHW	Australian Institute of Health and Welfare
AIHW Act	*Australian Institute of Health and Welfare Act 1987*
ASGS	Australian Statistical Geography Standard
CCMS	Child Care Management System
CCOPMM	Consultative Council on Obstetric and Paediatric Mortality and Morbidity
CDL	Centre for Data Linkage
Census	ABS Census of Population and Housing
CHeReL	Centre for Health Record Linkage
COAG	Council of Australian Governments
CPSIS	Cross Portfolio Statistical Integration Secretariat
DEEWR	Department of Education, Employment and Workplace Relations, now Department of Education
DISC	Data Integration Services Centre
DLU	Australian Institute of Health and Welfare's Data Linkage Unit
ECD	early childhood development
ECDSG	Early Childhood Data Sub Group
ECEC	early childhood education and care
EthOS	Ethics Online System
EoI	expression of interest
FaHCSIA	Department of Families, Housing, Community Services and Indigenous Affairs, now Department of Social Services
HREC	Human Research Ethics Committee
ID	identification
MBS	Medicare Benefits Scheme

METeOR	Metadata Online Registry
MoU	memorandum of understanding
NAPLAN	National Assessment Program—Literacy and Numeracy
NCRIS	National Collaborative Research Infrastructure Strategy
NDI	National Death Index
NEAF	National Ethics Application Form
NECD RDS	National Early Childhood Development Researchable Data Set
NECD RDS AC	National Early Childhood Development Researchable Data Set Advisory Committee
NIA ECEC	National Information Agreement on Early Childhood Education and Care
NMDS	National Minimum Data Set
NHS	National Health Survey
PHRN	Population Health Research Network
Privacy Act	*Privacy Act 1988*
SA DECD	South Australian Department of Education and Child Development
SEIFA	Socio-Economic Indexes for Areas
SLK	statistical linkage key
SURE	Secure Unified Research Environment
TETIA	Transforming Education and Training Information in Australia
TICHR	Telethon Institute for Child Health Research
URL	unit record level

Summary

This report describes the processes undertaken from 2011–13 towards the development of an ongoing National Early Childhood Development Researchable Data Set (NECD RDS). This data set is intended to foster research across the sectors of early childhood development, education and care, health, and human services. The Australian Institute of Health and Welfare (AIHW) has documented the processes and protocols required to access and link multiple health and education data sets from jurisdictional and national data sources in order to develop an ongoing, linked national data set.

The main aim of the NECD RDS is to provide an evidence base on early childhood development in Australia by providing data on all children in Australia, including those who move between states and territories. The NECD RDS aims to contain data on large enough numbers of children to permit robust analyses of outcomes for particular population groups (such as Indigenous children or children from non-English speaking backgrounds), and to analyse the impacts of particular policies and programs within and between states and territories.

The development of this type of ongoing national linked data set in Australia is relatively new, and the processes for doing so are continuously evolving as new challenges are encountered. This report describes and discusses the complexities and barriers to obtaining data across health and education sectors, at both the national and jurisdictional levels. The AIHW gained further insight into the privacy and legislative constraints. These included the complexities of processes around ongoing receipt of data, how data could be made available to external researchers and how to navigate the multiplicity of approval processes for ethics committees and data custodians for both construction of the NECD RDS and for allowing subsequent access to government and non-government agencies and researchers.

The initial development of this data set revealed differences between jurisdictions, both in their data linkage capabilities and in the processes required to approve data release, including different ethics committee requirements. These issues can create barriers to obtaining cross-sectoral data—in this case, health and education data.

A researchable data set at national level provides a unique challenge for Australia, as well as many opportunities. A data set such as the NECD RDS has the potential to strengthen the evidence base for child development outcomes, allowing the development of sound policies and programs. It would also enhance the current efforts in extending data linkage networks at both jurisdiction and national levels, and ensure that Australia is at the forefront of research developments in this area.

1 Introduction

This chapter provides the background to the project including the policy context and the national agenda for data linkage and related linkage activities.

This report describes the steps undertaken during initial development of the Early Childhood Development Researchable Data Set (NECD RDS). This development project was undertaken from 2011–13 and information in this report dates from this time period. The project involved identifying appropriate data sources and negotiating access to the data. No data linkage has been undertaken and permission to obtain data for inclusion in the NECD RDS has not been granted by all data custodians. Governance arrangements for the NECD RDS have not been agreed.

1.1 Background

Improving the outcomes for Australian children is a high priority for all levels of government in Australia, and research has shown that the early childhood period (birth to primary school age) is a particularly critical time for children's development. Currently, no single collection (or combination of collections) is able to provide national data on the impact of early factors at birth and the types and regularity of care and learning experiences, on children's developmental outcomes. The NECD RDS provides a unique opportunity to examine developmental outcomes as children progress from infancy to school years.

The NECD RDS plans to link unit record level (URL) data across jurisdictions from a number of different data collections including health and education. The proposed data collections include birth, perinatal, pre-school, Australian Early Development Census (AEDC) and National Assessment Program—Literacy and Numeracy (NAPLAN) data. Once established, the NECD RDS would foster research in early childhood development across health, human services, and early childhood education and care across the transition from birth to early years of school education.

This work, undertaken by the AIHW in 2011–13, was funded and overseen by the Early Childhood Data Sub Group (ECDSG) as part of the National Information Agreement on Early Childhood Education and Care (NIA ECEC) work plan (MCEECDYA 2010). The ECDSG comprises members from the Australian Government Department of Education (DoE) (formerly the Department of Education, Employment and Workplace Relations); state and territory governments; Australian Education, Early Childhood Development and Youth Affairs Senior Officials Committee (AEEYSOC); Australian Bureau of Statistics (ABS); the AIHW; and the Secretariat for the Steering Committee for the Review of Government Services Provision.

1.2 Policy context

The Council of Australian Governments (COAG) released *Investing in the Early Years—A National Early Childhood Development Strategy* (the ECD Strategy) in July 2009, to ensure that by 2020, 'all children have the best start in life to create a better future for themselves and for the nation' (COAG 2009). One of the key reform priorities in the ECD Strategy is to build better information and a solid evidence base (COAG 2009). In the context of the ECD Strategy, Australian, state and territory governments, and government information agencies signed the NIA ECEC in early 2010 (MCEECDYA 2010). The NIA ECEC provides a framework for cooperation between the parties to develop the information base for early childhood education and care required for the COAG's early childhood reform agenda. The NECD RDS aims to link health and education data, thus providing a strong evidence base for research in early childhood development. The NECD RDS has been progressed under the work plan of the NIA ECEC, with funding of $305,000 allocated from the $3 million set aside annually over 5 years from the National Partnership Agreement on Early Childhood Education for research, evaluation and data development activities.

Box 1.1 highlights other policy frameworks and initiatives with a focus on early childhood development.

Box 1.1: Policy frameworks and initiatives

National Quality Framework for Early Childhood Education and Care

The National Quality Framework (an initiative under the COAG ECD Strategy) took effect on 1 January 2012. Under this framework, a range of key requirements, including staff qualification levels and educator-to-child ratios, are being phased in between 2012 and 2020.

National Partnership Agreement on Early Childhood Education (2009–13)

The goal under this COAG agreement is universal access (by June 2013) to early childhood education for children in the year before full-time school.

National Partnership Agreement on Universal Access to Early Childhood Education (2013–14)

In April 2013, COAG reaffirmed its commitment to universal access to quality early childhood education, by endorsing a new National Partnership Agreement on Universal Access to Early Childhood Education to the end of 2014.

National Partnership Agreement on Indigenous Early Childhood Development (2009–14)

This agreement supports the Closing the Gap targets through the establishment of 38 Children and Family Centres across Australia by June 2014. The centres will deliver integrated services, including early learning, child care and family support programs.

Source: AIHW 2013:143

1.3 National agenda for data linkage

Data linkage (also known as 'data integration' or 'record linkage') involves bringing together information (from 2 or more sources) about people, places and events in a way that protects individual privacy and confidentiality. Data sets that are constructed using linked information provide public benefits by enabling improved research and supporting good government policy making, program management and service delivery (NSS 2010a). These data sets also create an important opportunity to expand the range of official statistics to better inform Australian society.

In recent years, there has been progress towards developing a national data linkage system for research purposes in Australia, enabling the coordinated linkage of data from state and territory collections as well as from national Commonwealth agencies. Linkage of Commonwealth data for statistical and research purposes is governed by a set of principles endorsed by the Secretaries Board. These principles are supported by a set of governance and institutional arrangements. These arrangements require integrating authorities undertaking linkage with Commonwealth data to:

- be responsible for the sound conduct of the data integration project
- ensure that risks have been assessed, managed and mitigated throughout the duration of the project, including regular reviews of ongoing projects
- ensure that appropriate governance is in place for the data integration project, including using an open approval process, documenting the proposal, considering the privacy impacts, examining the expected costs and benefits of the proposal and considering the access arrangements and dissemination plans
- be responsible for the ongoing management fo the integrated data, ensuring they are kept secure, confidential and fit for the purposes for which they were approved (NSS 2010b).

One of the challenges in developing a national system is to broaden the scope of data that are included to multiple domains, such as health, education and community services. Also, the capability to link data across jurisdictions, and successfully integrate state and territory data with Commonwealth data sets is still being developed, tested and implemented. While significant progress is being made in the health domain, relatively less work has been undertaken in education and social policy.

Substantial work has been undertaken in linking health data for research purposes. In 2006, the Australian Government funded the National Collaborative Research Infrastructure Strategy (NCRIS), which was guided by the *2006 NCRIS Strategic Roadmap.* In 2009, using funding under the NCRIS, the Population Health Research Network (PHRN) was established. It was charged with building a national data linkage research infrastructure by working collaboratively with key stakeholders, including data custodians, researchers and the community. The PHRN, through its members, is building the capability to link data for research purposes between states and territories, as well as between states and territories and the Australian Government. The AIHW and the Centre for Data Linkage have been specifically funded (through the Education Investment Fund) to build national and cross-jurisdictional data linkage research capabilities, and these are now in place.

The development of the NECD RDS would contribute to the overall development of national data linkage infrastructure systems in Australia, capitalising on data that are sourced from both national and jurisdictional collections.

1.4 Related data linkage projects

At the time that work on the NECD RDS started, 2 other Australian linkage projects similar to the NECD RDS were identified. These are:

- the Building national linkage capability to the AEDC project
- the PHRN Perinatal outcomes and child development (risk and protective factors) Proof of Concept project 3.

All three projects use similar methodologies for de-identifying and linking data, and all require similar data sets. The NECD RDS is unique in that it plans to provide an ongoing linked data set that would include linked data for all Australian children as they progress through early childhood into school education. Where possible, all of these projects should complement each other to avoid duplication of effort.

Building national linkage capability to the AEDC project

The Building national linkage capability to the AEDC project uses births and perinatal data, 2009 AEDC data, and education records. It is designed to:

- describe the progress of children from the 2009 AEDC through to the first 2 years of schooling within 4 jurisdiction (SA, NT, WA and NSW)
- explore education system data holdings across jurisdictions.

The project aims to explore what can be learnt for national policy purposes from linking early health and AEDC data to detailed data held within state education departments. This project is managed by the SA DECD on behalf of the Early Childhood Data Sub Group (ECDSG).

Proof of Concept project 3

The PHRN's Proof of Concept project 3 titled 'Perinatal outcomes and child development (risk and protective factors)' will link births and perinatal data, 2009 AEDC data, and education records. This project will test the data linkage infrastructure established through the PHRN. This project is being undertaken by the Telethon Institute for Child Health Research. The research aims to explore the relationship between birthweight and developmental outcomes as measured by the AEDC, through to educational outcomes.

More recently, two other data linkage initiatives have been proposed:

- the Transforming Education and Training Information in Australia (TETIA) initiative
- the National Education Data Linkage Roadmap.

Transforming Education and Training Information in Australia

TETIA provides a strategy for measuring educational outcomes over the life-course from early childhood education and development, schooling, higher education and vocational education and training through to employment and other life outcomes. The key difference between TETIA and other projects is that TETIA provides an overall cross-sectoral strategy whilst providing specific projects at the state/territory or national level (depending on data availability) which are intended to provide immediate insights into policy issues. These projects also aim to build capability and provide the foundations for streamlining future data integration and access. In addition, TETIA aims to understand the contextual factors that impact on education and training outcomes, including personal, family and household characteristics such as those available through the Census of Population and Housing and health factors such as those available from a range of health administrative and survey datasets.

Measuring educational outcomes over the life-course

The data integration projects being progressed as part of the TETIA initiative involve the integration of a range of education and training data (including early childhood education and care, AEDC, NAPLAN, school enrolments and Vocational Education and Training data) and Census data.

As part of the TETIA initiative, the ABS is currently undertaking linkage across 2009 and 2012 AEDC and 2011 Census data. The focus of this project is to assess the role of parental and socioeconomic characteristics on child development. The project will also determine whether it is possible to use Census data to assess the role of preschool participation on child development, particularly for children from ‘high risk’ groups.

National Education Data Linkage Roadmap

The National Education Data Linkage Roadmap (the Roadmap) seeks to build on the data linkage capabilities established in the PHRN and to extend these to population data linkage, specifically early childhood and education data. The Roadmap is ambitious and aims to standardise data transfer agreements, develop meta-data for relevant data collections, influence privacy legislation, develop Unique Linkage IDs, streamline ethics approval processes and improve the timeliness and accuracy of linked data.

1.5 Structure of this report

Chapter 1—Introduction: provides the background to the project including the policy context and the national agenda for data linkage and related linkage activities.

Chapter 2—The Early Childhood Development Researchable Data Set: outlines the purpose of the NECD RDS; why the NECD RDS is being developed; the value of the data set; and the scope of the project, including the proposed data sets and how these would be linked and merged. Additionally this chapter outlines the benefits to governments and the wider community and potential research questions.

Chapter 3—Governance framework for the researchable data set: outlines the legislative and ethical considerations in establishing the NECD RDS, as well as stakeholder engagement and the role of the NECD RDS Advisory Committee in ongoing governance.

Chapter 4—Challenges and progress in establishing the researchable data set: provides information on the issues encountered during the developmental stages of the NECD RDS and some of the key lessons learnt in attempting to overcome them. It also provides a description of the progress so far in developing the NECD RDS, including the processes for obtaining data and the processes for data linkage.

Chapter 5—Moving forward: describes the current status of the development of the NECD RDS, some plans if and when data become available for research and recommendations for continued work in this area.

Appendix A—Childhood development research: provides information on a selection of relevant research in childhood development.

Appendix B—Data sources to support the NECD RDS: provides information on other data sources that could be used to enhance information from the NECD RDS.

Appendix C—Data items: provides a draft list of data items proposed for inclusion in the NECD RDS.

Appendix D—Processes to construct the NECD RDS: provides information on how data are to be linked and merged to form the NECD RDS.

Appendix E—Guide for researchers: provides a suggested draft guide to researchers on how the NECD RDS could be accessed if and once established.

2 The National Early Childhood Development Researchable Data Set

The aim of the NECD RDS is to establish a national de-identified and confidentialised linked data set on children from birth to the early years of schooling. For each child, it will contain data on health, development and education.

This chapter outlines the purpose of the NECD RDS; why the NECD RDS is being developed; the value of the data set; and the scope of the project, including the proposed data sets and how these would be linked and merged. Additionally this chapter outlines the benefits to governments and the wider community and potential research questions.

2.1 Aim, value and benefits

Aim of the NECD RDS

Early childhood is a critical time for a child's development. Factors such as children's health at birth, their family environment, where they live, and their access to programs such as high quality preschools, schools and early intervention programs, all combine to affect their overall development, their achievement in school, and their prospects later in life (Elliot 2006, Glascoe & Leew 2010; Kalil et al. 2012; Key et al. 2007). Currently, no single data set (at the national level) can provide a clear understanding of how these different factors interact and how children's outcomes could be improved by examining the impact of experiences early in life on developmental outcomes at the time of entry into school, and on educational outcomes during school.

The NECD RDS would provide a single data set that links ongoing information on children's health and education, allowing a better understanding of how these factors interact.

Benefits to governments, policy makers and program developers

Policy makers and program developers need to know which specific factors affect future development, and the relative importance of risk and protective factors. A solid evidence base will help determine which factors and populations to target in policies and programs, to maximise the Australian Governments' return on investment.

The development of the NECD RDS aligns with 1 of the key reform priorities outlined in the NECD Strategy of developing a better information and evidence base. The NECD RDS would:

- provide longitudinal data on children which enable a picture of the children's characteristics and circumstances to be built up over time
- utilise jurisdictional administrative data sets, as well as relevant Commonwealth data, to understand transition points and factors that influence these transitions
- provide the ability to analyse the impacts of particular policies and programs within and between states
- contain large numbers of population groups that will permit robust analyses of their outcomes (such as Indigenous children or those with a disability)

- provide comprehensive coverage of all children in Australia, including those who move between states and territories
- provide a national de-identified, confidentialised data set on children from birth to the early years of schooling and support research activities throughout Australia that all jurisdictions and researchers could potentially use
- enhance researcher access to a range of nationally linkable data sets that would meet current and emerging research needs in the field of early childhood development.

Given the scope of the data sets proposed for linkage, it is expected that a range of researchers from both academia and government would be interested in the NECD RDS. The creation of the NECD RDS would facilitate existing research studies and provide an impetus for the development of future policy, able to draw upon the best available research.

Benefits to the public

The NECD RDS would provide the public benefit of informing policy development, leading to improved service delivery and integration. This would contribute to the overall goal of ensuring that all Australian children get the best possible start in life.

The NECD RDS would have important community benefits. Indicators of child care quality such as the qualifications and experience of child care workers, the continuity of caregivers, staff/child ratios and group sizes, have been found to relate to child outcomes in the United States (NICHD ECRN 2000). In Australia, experience of formal care has been found to associate with improved socio-emotional and cognitive outcomes (Harrison & Ungerer 1997). However, ongoing debate surrounds the effects of age of entry into child care, the stability of care, and the number of hours per week in care on child development in the long term (Wise et al. 2002). Issues such as these, as well as a wide range of other topics, could be explored using the NECD RDS, and could lead to improved outcomes for Australians on both an individual and national level.

The NECD RDS would also offer the public benefit of enabling Australian, state and territory government policies related to childhood education and care to be monitored and evaluated on a long-term basis. This would:

- increase the transparency and accountability of government programs
- improve the design of policy and delivery of services over time
- inform the broader community of the potential long-term impacts of childhood education and care on children and their future life course.

Questions that the NECD RDS may help answer

The NECD RDS would assist the ability to answer research and policy questions, and identify focus areas where policies or programs may be improved or developed. Although it is difficult to predict or determine all the research questions that could be investigated through the NECD RDS, there are a number of examples of policy-relevant issues that could be explored using the NECD RDS:

- How do the quantity, quality and the starting age of child care impact on early child development and educational outcomes, and do they differ for children from different backgrounds or who live in different areas of Australia?

- For vulnerable children (for example, children with a low birthweight, children who have a disability, or children from disadvantaged backgrounds), does attending an early childhood education program lead to better developmental and educational outcomes at school? How does the impact of these early childhood education programs vary by the number of days attended, the type of service, where the child lives, and other factors?
- What are the long-term outcomes for children who are not participating in formal child care or preschool, compared to those who do? A particular focus may be on children from disadvantaged households.
- What are the key individual, familial, contextual, and policy factors that lead to better developmental and educational outcomes for Indigenous children?
- Do educational outcomes differ for children who move between Australian states and territories?

The results would inform policies aimed at enhancing children's development.

2.2 Scope and design

It is anticipated that the NECD RDS will contain records for children from birth through to approximately 15 years of age. As the NECD RDS would be built using whole-of-population data sets aligning to different stages of development, it would provide extensive coverage of children at different age groups and in sub-groups of the population, such as those born overseas or those who are Indigenous.

Research foundations of the NECD RDS

Childhood development research

In building a data set for research purposes, it is critical to understand the type of research that has already been undertaken, the current and future requirements for research in this area, and where gaps in information exist. These issues were investigated by undertaking a literature search of current research publications in the area of childhood development in general, with a particular focus on its relationship to early childhood education and care and its impact on later outcomes.

The process involved searching for articles that contained discussion about the effects of certain factors—such as early health, disability, socioeconomic background and child care—on early childhood development, preschool, school readiness and educational outcomes in the future.

The results of this literature search indicated that information was available about developmental outcomes for children who were born with a low birthweight and about educational and developmental outcomes for children in relation to child care. There was some information on the educational and behavioural outcomes for those children who had attended preschools, but very little contrasting information about outcomes for children who had not attended preschool (see Appendix A for more detailed information about the literature search and its results).

Other available data sources on children and child development

A number of other data sources are available that could provide additional contextual information about factors impacting on early childhood development. Appendix B provides detailed information about data sources that collect information about all children in the NECD RDS population—for example, the Australian Bureau of Statistics (ABS) Census of Population and Housing (the Census) could provide additional information on socioeconomic factors or unpaid child care arrangements across the Australian population. These data sources are not currently highlighted for inclusion in the NECD RDS, but represent other avenues through which developmental pathways for children could be explored.

It is possible that once preliminary data are loaded into the NECD RDS, other available data sources could be considered for inclusion in the NECD RDS. These data sources include—but are not limited to—Medicare data, the National Death Index (NDI), Australian Childhood Immunisation Register, injury surveillance data from jurisdictional emergency department data collections, family tax benefit data, and disability services data. Linkage to surveys could also be considered, such as the Child Dental Health Survey, ABS National Health Survey, and Australian National Infant Feeding Survey.

Data to be included

Data collections

Advice from collaborating agencies indicated that in order to provide maximum value for government, researchers and the general public, the NECD RDS would need to include health and education data that are regularly collected and updated, but not currently linked at the national level. Such data would be important for describing and analysing outcomes across the early childhood and transition to school period.

The following data sets are proposed for inclusion in the NECD RDS:

- perinatal/midwives data (health outcomes for mothers and babies)
- birth registration data (required for linking data sets only)
- Child Care Management System (CCMS) data (child care data, such as long day care attendance, whether a preschool program is offered by the day care centre, hours in care)
- preschool data that are collected by jurisdictions and contribute to the ABS's National ECEC data collection (preschool enrolment and attendance)
- AEDC data (child development measures at first year of school)
- NAPLAN data (literacy and numeracy in years 3 and 5).

Table 2.1 maps data sets, over time and across cohorts, starting with the cohort that is centred on 2011 ECEC data (the combination of jurisdiction preschool data and Australian Government child care data). URL in the ECEC data collection has been collected since 2010 for some states and territories, and will be available for all states and territories from 2013. This influenced the decision to start the NECD RDS with 2011 data from the ECEC.

The construction of the NECD RDS would start with the 2011 child care and preschool collections (CCMS and ECEC) (see Table 2.1). These children are part of the birth cohort for 2006–07, so data from perinatal data collections would be included for these years. As these children progress through school, education data would be included; that is AEDC data in 2012, NAPLAN Year 3 data in 2015 and NAPLAN Year 5 data in 2017. Table 2.1 shows that each year, data for a new cohort of children would be added to the NECD RDS. For children included in the 2014 data for the CCMS and ECEC, information would be available on their perinatal health, formal child care attendance prior to school, and educational outcomes (AEDC and NAPLAN).

Documenting available data

Once the collections for the NECD RDS were identified, aspects of these collections were documented. These included:

- information about differences between final national data sets and their jurisdictional components
- identifying which variables were required for the construction of linkage keys (identifying variables such as names)
- an assessment of the suitability of variables for the NECD RDS.

A summary of the results is provided in Table 2.2. The lists of variables are not exhaustive and serve only as an example of the types of information that could be included in the NECD RDS (see Appendix C for a more comprehensive list of data items proposed for inclusion in the NECD RDS). In the future, data quality statements for the NECD RDS would be made publically available.

Table 2.1: Proposed data collections to be linked and merged for the NECD RDS

	Approximate age group for the collection									
Birth cohort	**0–1 year**	**1–2 years**	**2–3 years**	**3–5 years**	**4–5 years**	**5–6 years**	**6–7 years**	**7–8 years**	**8–9 years**	**9–10 years**
Birth/perinatal 2006–07				Preschool (CCMS + ECEC) 2011	AEDC 2012			NAPLAN 2015		NAPLAN 2017
Birth/perinatal 2007–08			CCMS 2011	Preschool (CCMS + ECEC) 2012				NAPLAN 2016		NAPLAN 2018
Birth/perinatal 2008–09		CCMS 2011	CCMS 2012	Preschool (CCMS + ECEC) 2013				NAPLAN 2017		NAPLAN 2019
Birth/perinatal 2009–10	CCMS 2011	CCMS 2012	CCMS 2013	Preschool (CCMS + ECEC) 2014	AEDC 2015			NAPLAN 2018		NAPLAN 2020
Birth/perinatal 2010–11	CCMS 2012	CCMS 2013	CCMS 2014	Preschool (CCMS + ECEC) 2015				NAPLAN 2019		NAPLAN 2021

Table 2.2: Data sets considered for inclusion in NECD RDS

Data set	Description of data	Examples of items that may be included	Data custodian	Data owner
Births registration data	These are registrations of all live births in Australia. Data from this collection will be matched with data from the Perinatal Data Collection to obtain the infant's name. This is required to allow data sets to be linked.		ABS	Jurisdictional registrars of births, deaths and marriages
National and state Perinatal Data Collection	The National Perinatal Data Collection is a national population-based cross sectional data collection of pregnancy and childbirth. The data are based on births reported to the perinatal data collection in each state and territory in Australia. Midwives and other staff, using information obtained from mothers and from hospital or other records, complete notification forms for each birth. The data will provide researchers with vital information about children's start in life that will be compared to developmental and academic achievements. The first year of data that will be accessed will be 2003. *Note:* The data will be matched to registered births data for the same year to acquire the name of the infant. This is required to produce the statistical linkage key.	Marital status Mother: Indigenous status Mother: country of birth Area of residence Gestational age Apgar score at 5 minutes Birthweight Admission to special care nursery/neonatal intensive care unit Whether mother died	AIHW for national collection Jurisdictional departments of health for state collections	Jurisdictional departments of health
CCMS (Child care management system)	The CCMS is a child care system that enables child care services to exchange information online with government. One of its aims is to provide information about the supply and usage of child care across Australia. When long-day care data are used in conjunction with preschool program data, a better picture of total preschool program delivery is available. The data will provide an indication of which children are cared for in a child care facility and for how many hours a preschool program is attended by the child.	Maximum preschool program hours available per week Preschool program attendance Preschool program received from a qualified teacher indicator Preschool program hours attended per week Preschool program hours enrolled per week Preschool program repeater indicator Organisation: activity type, early childhood education and care Organisation: delivery setting, early childhood education and care	Australian Government Department of Education	Australian Government Department of Education

(continued)

Table 2.2 (continued): Data sets considered for inclusion in NECD RDS

Data set	Description of data	Examples of items that may be included	Data custodian	Data owner
National Early Childhood Education and Care National Minimum Data Set (NMDS)	The national ECEC collection collects information about the number of children who are enrolled and attending preschool programs across Australia. When used in conjunction with long-day care data, a better picture of total preschool program delivery is available.	Preschool program attendance indicator Preschool program enrolment indicator Preschool program hours attended Preschool program hours enrolled Preschool program received from teacher indicator Preschool program repeater indicator Preschool enrolment fees (per hour) Type of work activity	ABS Jurisdictional departments of health for state collections	Jurisdictional departments of education
AEDC (Australian Early Development Census)	The AEDC (formerly the AEDI) is a national collection of information about young children's development. Data are collected every 3 years and the results provide a snapshot of how our children are developing. The data provide an indication of development during children's first year at school. The first year of data that will be accessed will be for 2009. When analysed against perinatal data, AEDC will show how early life experiences can affect early childhood development.	Language spoken Special needs English proficiency Overall social/emotional development Physical Health and Wellbeing domain score Social Competence domain score Emotional Maturity domain score Developmentally vulnerable on 1 or more domains Developmentally vulnerable on 2 or more domains Communication and General Knowledge domain score Language and Cognitive Skills (school-based) domain score AEDC total score	Australian Government Department of Education	Australian Government Department of Education
NAPLAN National and state level	In 2008, the National Assessment Program—Literacy and Numeracy (NAPLAN) started in Australian schools. Every year, all students in years 3, 5, 7 and 9 are assessed on the same days using national tests in reading, writing, language conventions (spelling, grammar and punctuation) and numeracy. The data will be used to analyse children's education attainment. The first year of data that will be accessed for the RDS will be from 2012.	Year level School sector Participation in reading, language conventions, and writing tests Numeracy scale score Reading scale score Spelling scale score Grammar and punctuation scale score Writing scale score Band scores	Australian Curriculum, Assessment and Reporting Authority (ACARA)	Jurisdictional departments of education

Note: All data sets will need to contain a certain level of identifying information so that linkage can be undertaken, although identifying information would not be available as part of the NECD RDS.

2.3 Data linkage

Practice and protocols

The AIHW adheres to best practice standards for linking data and is an accredited Commonwealth data integrating authority, meaning that the AIHW has met stringent criteria covering project governance, capability, data management, and the protection of privacy and confidentiality (NSS 2010b).

A key feature of best practice and protocols is that data containing identifying information (for example, names and dates of birth) are separated from the content data (the subject related data), and that identifying data are only used for the initial linkage stage, from which a concordance file (sometimes known as a 'linkage key') to project IDs is produced. These processes:

- maximise the protection of individual privacy
- provide linked data files only to nominated researchers involved in specific approved research projects
- provide the researcher with no more than the data sets required for their specific project
- assure data custodians that those data that are their responsibility will be used appropriately and that security obligations will be met (Kelman et al. 2002).

More detailed information about the processes proposed to link data for the NECD RDS is provided in Appendix D.

Protecting privacy

It will be essential to protect the privacy of individuals whose data may be included in the NECD RDS, both during the linking and merging of data sets to form the NECD RDS and if the NECD RDS is established.

One important strategy for protecting privacy during linkage and merging is known as the 'data linkage separation principle'. This principle separates the data linkage step from the data merging step. The data linkage separation principle is illustrated in Figure 2.1. This shows how the data files supplied by data custodians are protected by keeping identifying information separate from content information. The AIHW Data Integration Services Centre infrastructure, operational procedures and security controls keep these files separate within the AIHW, so that data linkage staff do not have access to content data and data merging staff do not have access to identifying information.

The data linkage separation principle ensures that identifying information is only seen by data linkers, whose role is to match or link the same individual in different data sets. This linkage process is based on comparing identifying information such as full name, sex, full date of birth and residential address between data sets. A pair of records is considered a match if the 2 sets of information have a high probability of belonging to the same person. Only the identifying information is seen by the data linkers; they do not see any of the content data (Figure 2.1). For more information about the processes for data linkage, refer to Appendix D.

The other aspect of the data linkage separation principle is that data mergers do not see any identifying information; they have access to de-identified content data only. The data mergers join information about the same person from different data sets together, based on the de-identified concordance file provided by the data linkers. This file contains information about the project IDs in each data set that have been identified as belonging to the same individual, and therefore the content data that should be merged (Figure 2.1).

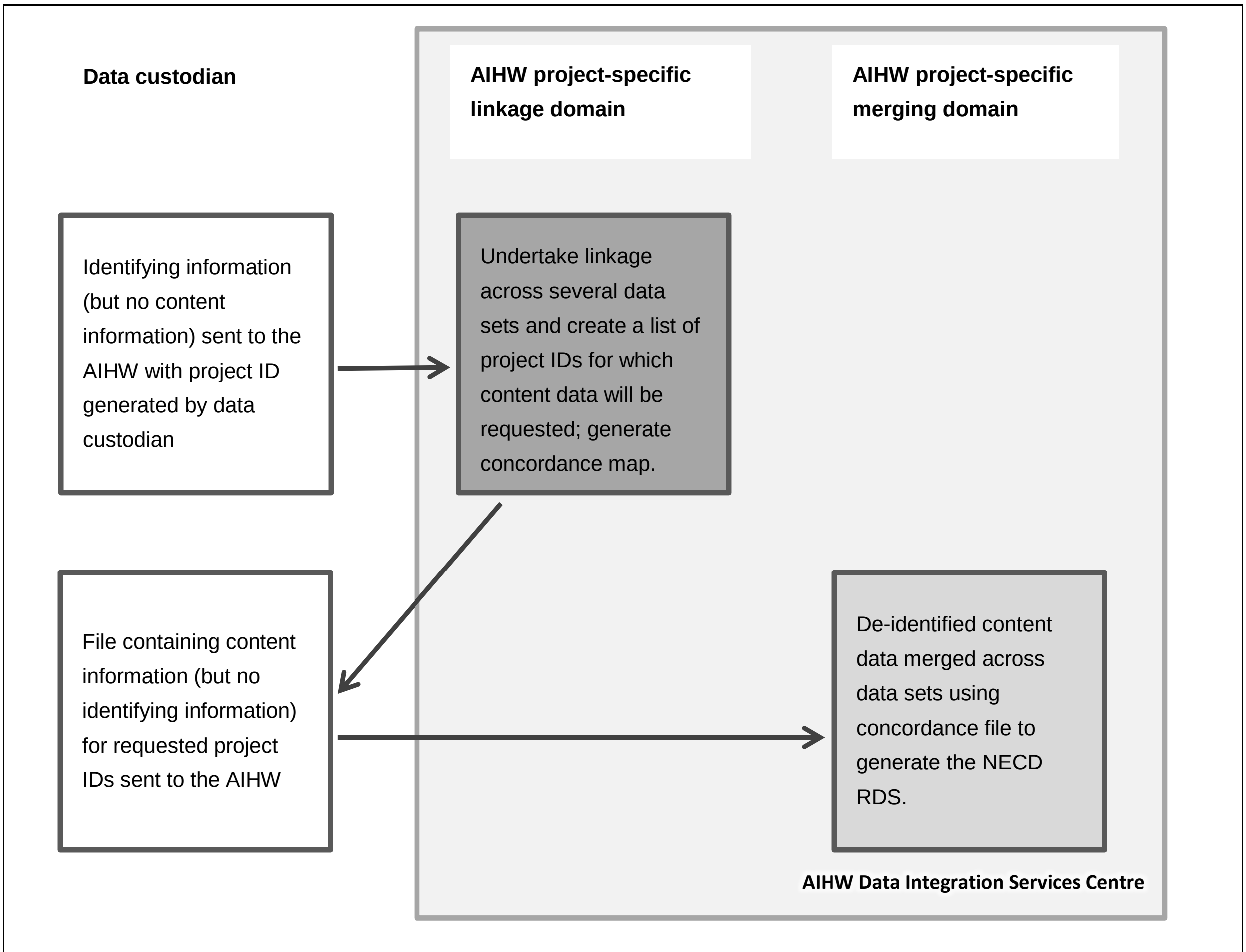


Figure 2.1: Establishing the NECD RDS—proposed flow of data from data custodians to the AIHW

If the NECD RDS is established, it will be essential to maintain the privacy of children whose data are included. Privacy can be protected by ensuring confidentiality undertakings, removing all IDs from the merged data prior to supplying them to researchers, providing the data through secure data access mechanisms and ensuring that identifying information does not form part of the NECD RDS. In addition, data can be confidentialised as required to minimise the risk of indirect identification. Data released from the NECD RDS would comply with the *High level principles for data integration involving Commonwealth data for statistical and research purposes* (NSS 2010a) and data requests would be assessed on a case-by-case basis. At this early stage in development of the NECD RDS, protocols for this assessment have not been developed, however, the NECD RDS Advisory Committee would provide advice.

3 Governance framework for the researchable data set

This chapter outlines the legislative and ethical considerations in establishing the NECD RDS, as well as stakeholder engagement and the role of the NECD RDS Advisory Committee in ongoing governance.

3.1 Privacy considerations

Maintaining the privacy of individuals whose data are included in the NECD RDS is essential. Whilst the AIHW must comply with Australian Government privacy legislation, states’ and territories’ privacy legislation must also be complied with when obtaining data. This includes:

- *Privacy Act 1988* (Cwlth)
- *Privacy and Personal Information Protection Act 1998* (NSW)
- *Information Privacy Act 2000* (Vic)
- *Information Privacy Act 2009* (Qld)
- *Personal Information Protection Act 2004* (Tas)
- *Privacy Act 1988* (as modified by the Australian Capital Territory Government Service (Consequential Provisions) Act 1994) (ACT)
- *Information Act 2002* (NT).

WA and SA do not have specific privacy legislation (note that in South Australia *Premier and Cabinet Circular No 12–Information Privacy Principles Instruction* regulates the manner in which SA government agencies handle personal information). Changes to state and territory privacy legislation may impact on the NECD RDS in the future.

AIHW legislative framework—privacy and confidentiality

As well as complying with its obligations under the *Privacy Act 1988* (the Privacy Act), the AIHW is also subject to requirements under the *Australian Institute of Health and Welfare Act 1987* (the AIHW Act). The confidentiality provisions in s. 29 of the AIHW Act apply to information obtained and/or held by the AIHW ‘concerning another person’. The definition of 'person' under s. 29 of the AIHW Act is broader than that in the Privacy Act as it includes, for example, bodies politic and deceased persons. It is an offence to release information concerning a person other than in accordance with the AIHW Act. The prohibition against release of such information applies even to exclude divulging or communicating it to a court of law, under s. 29(1)(e).

3.2 Ethical considerations

There are many complex and ethically significant issues related to research involving humans (NHMRC 2013a). Ethical oversight of such research is provided by ethics committees established under the National Health and Medical Research (NHMRC) *National Statement on Ethical Conduct in Human Research* (2007) (updated 2013) (the National Statement).

Ethics committees are usually established by organisations (public, not-for-profit or private) which conduct research involving humans. Universities and hospitals are the most common of these organisations (NHMRC 2013b).

Human Research Ethics Committees

Human Research Ethics Committees (HRECs) play a central role in the Australian system of ethical oversight of research involving humans. HRECs review research proposals involving human participants to ensure that they are ethically acceptable and in accordance with relevant standards and guidelines, including the National Statement. The National Statement sets out the relevant ethical principles and values by which research should be designed and conducted and to which HRECs should refer when reviewing research proposals (NHMRC 2013b).

Application forms

To simplify the process of obtaining HREC approval, a National Ethics Application Form (NEAF) has been developed. This form has been designed to address all the relevant ethical considerations.

Box 3.1: What is the National Ethics Application Form?

The NEAF is a web-based tool that enables researchers of all disciplines to complete research ethics proposals for submission to HRECs and assists HRECs to consistently and efficiently assess these proposals. It has been designed to meet the requirements of relevant guidelines with the aim of increasing the efficiency and quality of the ethical review process for all parties involved (NHMRC 2013c).

The amount of information required by HRECs varies across individual committees. Some HRECs have their own application forms that need to be completed, either in addition to or in place of the NEAF.

AIHW Ethics Committee

The Australian Institute of Health and Welfare Ethics Committee is established under s.16(1) of the *Australian Institute of Health and Welfare Act 1987*. The Australian Institute of Health and Welfare (Ethics Committee) Regulations 1987 set out the committee's functions and role. The regulations give the committee the power to make determinations on the ethical acceptability of AIHW activities, and activities engaged in by external parties under arrangements with the AIHW. They also allow the committee to make decisions on the acceptability of the release of identifiable information to researchers for research purposes. The committee reviews all projects, including internal proposals, that involve the release of identifiable information or the creation of new data sets or data linkage. In forming its opinion, it considers relevant guidelines, such as the National Statement.

As part of initial processes to test the feasibility of the NECD RDS, the AIHW Ethics Committee approved the pilot testing of some ECEC data in September 2011.

Further approval was sought from the AIHW Ethics Committee to begin the process for developing the NECD RDS and this was granted on 5 June 2012. With any significant changes to the project, a variation outlining these must be submitted to the AIHW Ethics

Committee. The project will continue to be monitored for progress throughout its lifetime. A number of variations have been submitted and approved due to the changing nature of the project. The last variation was approved in October 2013.

As part of the approval for the development of the NECD RDS, the AIHW NECD RDS project has to report progress annually to the AIHW Ethics Committee. It is a legislative and ethical requirement that once a research project receives approval from the AIHW Ethics Committee, it continues to be monitored until completion (completion includes the time taken to disseminate the results of the research, for example, publication as a report or journal paper). Researchers wanting to access the NECD RDS in the future may need to seek approval from the AIHW Ethics Committee, depending on the data requested. It is envisaged that secondary use of post-linkage anonymised data would facilitate greater access for researchers.

3.3 Stakeholder engagement

Good stakeholder engagement promotes better and consistent sharing of information, and builds trusting and professional relationships between parties. Stakeholder engagement has been an essential component of the NECD RDS project. As part of the development process for the NECD RDS (and as a deliverable under the Agreement), a stakeholder communication strategy was developed, outlining who the stakeholders are, and how they would be informed about the project.

The stakeholders to the project are varied, and range from government agencies to researchers, to those who implement policies and programs around early childhood. There were 16 stakeholder groups identified (see Table 3.1). Organisations that are likely to provide data for the NECD RDS are important stakeholders, as are parents and carers, and children themselves, who may indirectly benefit (through changes in health, education or early child care systems) from the wealth of information that would be collected and analysed.

Stakeholder engagement is essential to building confidence in the project and is considered a key to the success of the NECD RDS. Engaging with stakeholders has been vital in allowing the AIHW to understand their views, concerns and expectations, particularly those of data custodians, who have been or may be asked to provide data.

An important stakeholder engagement activity was a series of road shows to jurisdictions undertaken in November and December 2012 with the Department of Education, Employment and Workplace Relations, now Department of Education, South Australian Department of Education and Child Development and the Telethon Institute for Child Health Research . Representatives from departments of education and health, as well as jurisdictional data linkage units were invited to attend these meetings, where information about the NECD RDS was presented and information sought about the processes required to access data. These road shows provided the opportunity to outline the purpose of the NECD RDS, and illustrate how it fits in with other data linkage projects being developed. Since the road shows, 2 other projects that relate to linking education data have emerged: the National Education Data Linkage Roadmap and the Transforming Education and Training Information in Australia (TETIA) (see section 1.4 for more information).

Table 3.1: Stakeholders for the NECD RDS and their role in the process

The following stakeholders can be categorised into 3 groups: policy, data custodians and ethics committees.

Stakeholders	Relationship to/ role in the NECD RDS
Early Childhood Data Sub Group (ECDSG) (policy and data custodians)	The ECDSG manage structures and processes under the NIA ECEC. The AIHW reports progress in the development of the researchable data set to this group. This group also funds the Building national linkage capability to the AEDC project, which aims to describe the progress of children from the 2009 Australian Early Development Census (AEDC) cohort through their first 2 years of schooling.
Australian Government Department of Education (DoE), previously Department of Education, Employment and Workplace Relations (DEEWR) (policy and data custodians)	DoE oversee the activities undertaken by the AIHW in the development of the researchable data set, and manage the funding from ECDSG. DoE are also data custodians for the preschool component of the Child Care Management System (CCMS) data, and the AEDC data.
State and territory departments of education (policy and data custodians)	These departments are the data custodians for preschool data and National Assessment Program—Literacy and Numeracy (NAPLAN) data.
South Australian Department of Education and Child Development (SA DECD) (policy and data custodians)	The AIHW collaborates with the SA DECD to exchange information and experience about best practices in data linkage.
State and territory, Registrars of Births, Deaths and Marriages (BDMs) (data custodians)	The state and territory BDMs are the data custodians of births registration information in Australia.
State and territory health departments (data custodians)	State and territory health departments are the data custodians of perinatal data.
ACT Chief Minister’s Office (CMO) (policy and data custodians)	The ACT CMO is taking the lead in data integration projects that cross portfolios. As such the AIHW has liaised with them (in conjunction with the Health and Education Directorates) so that everyone is aware of current activities.
AIHW Ethics Committee (ethics)	The AIHW’s Ethics committee have approved the development of the NECD RDS. In the future, the AIHW Ethics Committee would have to approve research proposals from researchers wishing to access the data.
Human Research Ethics Committees (HRECs) (ethics)	For the data sets identified for inclusion in the NECD RDS, the HRECs are predominantly within the state and territory health departments and responsible for approving uses of the data within that portfolio, that is perinatal data. These committees provide in-principle approval and determine the specific requirements for that jurisdiction and are responsible for approval of the projects. In some states, additional approval is required from Aboriginal HRECs.
Other approval bodies such as the AEDC National Committee	This committee oversees the AEDC collection, processing and dissemination of findings and data, considers any complex or exceptional requests and must see results of publications using AEDC data 1 month prior to intended release.

Table 3.2: Other groups involved with the NECD RDS and their roles

Group	Involvement with the NECD RDS
Population Health Research Network (PHRN)	The Population Health Research Network is an Australian Government initiative established as part of the National Collaborative Research Infrastructure Strategy. This has helped fund the establishment of the Centre for Data Linkage (CDL) at Curtin University in Western Australia and the Secure Unified Research Environment (SURE) facility at the Sax Institute.
SAX Institute	It is planned that Government and academic researchers would be able to access the NECD RDS remotely via the SURE facility at the Sax Institute using a secure electronic login. Researchers wanting to access the NECD RDS to explore specific research questions would need to apply for ethics approval from the AIHW Ethics Committee.
AIHW units	Other AIHW units that may assist with the development of the NECD RDS include: • Data Integration Services Centre, which provides the infrastructure to undertake secure data linkage in line with Commonwealth and PHRN requirements • Data Linkage Unit, which undertakes the data linkage and would manage the NECD RDS • Maternal Health, Children and Families Unit • Governance Unit • Technology and Transformation Unit and ICT Operations Unit • METeOR (Metadata Online Registry) Unit • Online Communications Unit (in the future).
The public, including parents and carers, and children	Would be the eventual beneficiaries of new policies and programs that are developed using data from the researchable data set.
Telethon Institute for Child Health Research (TICHR).	The Telethon Institute is experienced in the creation and analysis of linked data, about children's health.

3.4 NECD RDS Advisory Committee

As noted previously, the NECD RDS involves many stakeholders, including Australian Government and state and territory health and education departments, as well as HRECs. A key challenge for this project will be establishing an advisory committee with appropriate representation.

It is expected that the NECD RDS Advisory Committee (AC) will comprise staff from the AIHW, (including a representative from the Data Integration Services Centre [DISC]), representatives from the Australian Government departments of education and health, and representatives from departments of education and health for each jurisdiction.

The operational framework of the NECD RDS AC has not been finalised and will require further development. It is proposed that the NECD RDS AC will guide and advise on the continuing development of the NECD RDS. In addition, the advisory committee will also provide advice on proposed research projects for both internal and external researchers.

4 Challenges and progress in establishing the researchable data set

The ECD Strategy (COAG 2009) includes a key reform priority of building better information and a solid evidence base. This reflects the support at a national level for projects such as the NECD RDS. In principle, many jurisdictions also support the concept of the NECD RDS; however, there have been a number of challenges in the initial development of the NECD RDS.

This chapter provides information on the issues encountered during the developmental stages of the NECD RDS and some of the key lessons learnt in attempting to overcome them. It also provides a description of the progress so far in developing the NECD RDS, including the processes for obtaining data and the processes for data linkage.

4.1 Challenges encountered

A summary of the major challenges encountered during initial development of the NECD RDS is provided in Table 4.1 and more detail is included in the sections below.

Table 4.1: Summary of challenges encountered during development of the NECD RDS

Short term	Medium term	Long term
Access to education (preschool) data—privacy issues	Access to NECD RDS—how will researchers access data and who will be involved in the project approval process	Ongoing provision of data—conditions for release of data; resources required to supply data on an ongoing basis
Ongoing nature of NECD RDS—data custodian concerns	Other linked data sets and national data linkage strategies—the role of the NECD RDS compared with other linkage projects	
Jurisdictional linkage experience—less experienced have more concerns about the project	Incomplete data—data not collected at the unit record level (URL); data not collected for all children (for example, non-government preschools in some jurisdictions)	
Access to Commonwealth and jurisdictional datasets—difficulties in obtaining permission		

Stakeholder engagement

Good stakeholder engagement takes time. To progress a large project such as the NECD RDS, a substantial amount of time and effort was spent on meeting with data custodians, data managers, collaborators and ethics committee managers, clarifying information and following up with stakeholders. As the stakeholders for the project are such a diverse group of people, this resulted in significant time spent on coordinating meetings, road shows and presentations, and often collaborators were reliant on each other's availability. It was however, valuable in terms of building rapport and trust between parties.

Jurisdictional linkage capabilities

Data linkage activities are developing throughout Australia, and each jurisdiction is at a different stage of contributing to the progress of these projects (Table 4.2). In turn, this means that there are different processes and requirements for obtaining data for the NECD RDS.

Three states (New South Wales, South Australia and Western Australia) have comparatively advanced data linkage units with a large number of data sets included or planned for inclusion in their master linkage. They also have well-established processes for accessing data and information about these processes is readily available. Other jurisdictions do not have such advanced linkage systems. For the AIHW, this means additional effort has been and will continue to be required to progress arrangements for data linkage, and to provide data custodians in those jurisdictions with enough information to satisfy their requirements.

As there is a well-established history of linking health data in Australia, the initial development of the NECD RDS could have benefited from convening a group of representatives from health agencies to promote the benefits of the project, and allay concerns about linking to education data. The potential benefits of this type of group were shown during the road show meetings in 2012 where data custodians from both health and education domains attended (often with representatives from PHRN linkage units) and could share information about processes and linkage experiences in their jurisdiction.

Table 4.2: Summary of linkage capabilities and examples of linked data in states and territories, as at December 2013

Jurisdiction	Data linkage node in place	Ethics committee forms available from website	Education data linked on system	Health data linked on system
NSW	Yes, CHeReL	Yes	No	Hospital, cancer, death, perinatal, diabetes, maternal/infant health
Vic	Yes, Victorian Data Linkages (limited)	No	No	A list is not available on the website
Qld	Yes, Health LinQ	No	No	Hospital admissions, perinatal, births/deaths/marriage
WA	Yes, Data linkage WA	Yes	Education data specific details not listed	Child protection, health, hospital, cancer, accident, mortality
SA	Yes, SANT DataLink	Yes	AEDC, NAPLAN	Perinatal, cancer, hospital, emergency/injury
Tas	No	No	No	No
ACT	Yes, via CHeReL (limited)	Yes	No	As for NSW
NT	Yes, SANT DataLink	Yes	AEDC, NAPLAN	As for SA

Note: The above information on the education and health data sets that each of these linkage nodes uses is provided as an example only. WA has been linking data since 1995 and has linked a vast number of data sets over the years using numerous data sets.

Ongoing nature of the NECD RDS

One issue that has been an obstacle for data providers is that the NECD RDS is planned as an ongoing linked data set. Although the concept of a longitudinal data set is not new, it does pose significant barriers, as it challenges the current processes and practices of data custodians. Traditionally, data linkage projects are undertaken for one-off specific projects, and this is what data custodians, ethics committees and other groups that review projects and approve data for research are generally familiar with. The proposed ongoing nature of the NECD RDS is not common and, for some agencies, quite challenging. The AIHW has liaised with data custodians (and others) to reassure them about the safety, security, benefits and feasibility of this type of project. In addition, AIHW has endeavoured to accommodate requests when developing governance arrangements for the NECD RDS, such as giving data owners the chance to review research findings based on their own data prior to publication.

The planned ongoing provision of data has caused concern for some data custodians and jurisdictional ethics committees, in terms of conditions of data release and the resources required to provide data on an ongoing basis. Addressing these issues has taken considerable time. For example, it has taken over 6 months since initial communication with some data custodians about the data, and in many cases these types of issues are still not resolved.

One of the most profound lessons learnt from the development of this data set to date is underestimating the time and complexities involved in obtaining data within the health and education sectors for an ongoing linked data set such as the NECD RDS.

Changes to process for generating linkage keys

It was initially proposed that data containing identifying information would be sent to the CDL in Western Australia for the construction of linkage keys and concordance maps. However, it became apparent that it would not be possible to send Commonwealth data to the CDL.

In June 2012, the Cross Portfolio Data Integration Oversight Body accredited the AIHW as an integrating authority. To achieve this accreditation, the AIHW was required to be fully compliant against all criteria outlined in the accreditation process. As an accredited integrating authority, AIHW can now work with and link Commonwealth data.

Privacy constraints for preschool data

When negotiating with data custodians to include data in the NECD RDS, it became apparent that some jurisdictions were not in the position to provide preschool data to the AIHW due to their own jurisdiction’s privacy arrangements. This is because during the collection phase, some departments of education had not gained consent from the parents/guardians of the children to pass information on to a third party. This is not an issue for collections where consent is obtained, or for health collections, where the AIHW is able to gain a waiver under s. 95 of the Privacy Act, where data are to be used for health or medical research.

AIHW legal representatives confirmed that using the information collected about children in preschools would also be in breach of Australian Privacy Principles (under which the AIHW operates), if consent was not obtained from parents or guardians.

This problem was not anticipated during the formative stages of the NECD RDS. To overcome this in most jurisdictions, the AIHW requested that all preschool data be provided using the Statistical Linkage Key 581 (SLK 581) rather than identifying information such as full names. This emulates the way data are transferred between jurisdictions and the ABS for the Early Childhood Education and Care (ECEC) collection. For jurisdictions where this solution has not adequately addressed the issue, the AIHW would continue to work with them to find alternative solutions.

Incomplete data

In order for the NECD RDS to provide information for all children, it is essential that all collections provide data that relate to each individual child, that is, at URL. In 2011, ECEC collection did not collect URL data for parts of Queensland, South Australia and Western Australia. In addition, for the Australian Capital Territory, data were only collected from Australian Government preschools, which means that information about children attending Catholic or Independent preschools is not available. The result of missing URL information is that the cohort of children who do not go to preschool is more difficult (if not impossible) to define for 2011.

Due to the proposed longitudinal nature of the NECD RDS, not all data would be available from the start. For example, the Australian Early Development Census data were collected in 2012 however would not be available for inclusion in the NECD RDS until mid-2013. Similarly, any NAPLAN data for children who were in preschool in 2011 would not be collected until those children reach Year 3 in 2015.

Barriers to obtaining data

The NECD RDS was first conceived under the premise of agency-committed resources being readily available and useable. The AIHW endeavoured to work closely with other agencies that undertake projects using data linkage and need to follow similar pathways to obtain data. Although the sharing of experiences and knowledge between agencies has been invaluable to the AIHW, it has become apparent that individual agencies arrive at different points in the development of their data sets at different rates. For agencies that are able to develop more quickly, this can result in periods of little or no progress.

One example is the CCMS data, which contains information that, when combined with jurisdictional preschool data, forms the Early Childhood Education and Care (ECEC) data set. The Department of Education originally indicated that they could supply these data to the AIHW. These data were requested in July 2012 but at the time of writing this report, CCMS data had not been provided to the AIHW.

Other data linkage projects

Currently there are a number of data linkage projects being developed using much of the same data as the NECD RDS, linking education and/or health sector data for individual states or across more than 1 jurisdiction (see section 1.4). Data custodians have sought clarification about the differences and the benefits of the NECD RDS compared with the other linkage projects. Furthermore some stakeholders have questioned why it would be beneficial to have a national data set (as an ongoing linked data set) when information may already exist in some jurisdictions. The AIHW acknowledges that there are similar projects (or projects that may be perceived as such), and as part of a collaborative road show with the other agencies undertaking these projects, consulted with data custodians and other stakeholders to explain the differences between the projects, and the benefits of the NECD RDS.

4.2 Obtaining permission to use data for the NECD RDS

As noted above, the AIHW encountered significant challenges in the initial development of the NECD RDS, with the work undertaken revealing differences between jurisdictions in their data linkage capabilities and in the processes required to approve, release and provide data. These differences are sometimes significant and can create barriers to the provision of data across health and education sectors. Other disparities included:

- the multiplicity of ethics committees (and other approving agencies) and approvals processes required by data custodians
- the complexities of retaining data over a long period of time (since the NECD RDS is proposed as an ongoing linked data set)
- privacy and legislative constraints
- provisions for making data available to external researchers.

The issues encountered in obtaining permission to use data for the NECD RDS are discussed below for both Commonwealth and jurisdictional data sets.

Commonwealth data sets

The CCMS and the AEDC are two data sets that have been requested as part of the first round of requests. These data sets are managed by different areas within the DoE.

CCMS data, 2011 and 2012

In July 2012, 2011 CCMS data were requested from DoE. Initial advice suggested that the DoE would have to obtain a Public Interest Certificate, but further legal advice indicated that only approval from the Secretary would be necessary. The mandatory requirements to obtain CCMS data are shown in Table 4.3.

The ABS provides DoE with a ‘return-to-source’ file containing the data DoE supplied for the ECEC collection and as well as some data items derived by the ABS (for example, geographic codes). These files include the statistical linkage and match keys, which are codes that enable two or more records belonging to the same individual to be brought together. However, the DoE identified that the return-to-source file was not suitable for the NECD RDS.

DoE has proceeded to explore the use of the original source data with the option of AIHW applying Socio-Economic Indexes for Areas (SEIFA) and Australian Statistical Geography Standard (ASGS) codes using the geographic variables on the file.

Since the original request for data, the scope of the request has changed to include 2012 CCMS. At the time of writing this report, CCMS data have not yet been provided.

AEDC data, 2009 and 2012

Data from the 2012 AEDC became available for researchers in April 2013. The 2012 data represent the first AEDC data that can be linked for children whose records were collected through the 2011 preschool or CCMS collections. The AIHW met with DoE, the custodians of the AEDC, in April 2013 to discuss pathways for obtaining data. The required steps are summarised in Table 4.3.

AIHW was advised that a memorandum of understanding (MoU) would be necessary between DoE and AIHW to provide both the 2012 and 2009 AEDC data sets, and the identifying information required to link data for the NECD RDS. To access these data, a Deed of Agreement was signed and the AIHW has been supplied with data.

The next AEDC collection will occur in 2015, with data available in 2016.

AEDC data are historically used for linkage purposes, and there are a number of established processes for the acquisition of data which are outlined on the AEDC website. Complex requests involving unpublished AEDC data will be considered by the AEDC National Committee.

AEDC National Committee

The AEDC National Committee was established to guide the national implementation of the AEDC and to help ensure that the potential value of the AEDC in contributing to improved early childhood outcomes is realised (AEDC 2013). The AIHW is a member of the AEDC National Committee.

The Committee oversees the AEDC implementation, including the collection, processing and dissemination of findings and data, and provides high-level guidance relating to access and appropriate use of AEDC data.

Requests for access to unpublished AEDC data will be approved by the Department of Education, as Secretariat to the AEDC National Committee, and the AEDC National Committee will consider complex or exceptional requests, as necessary.

Before publishing any results using AEDC data, AEDC National Committee must see the results at least 1 month before their intended release.

Table 4.3: Requirements to obtain access to Commonwealth data for the NECD RDS

	Mandatory requirements						
Collection	**Review by human research ethics committee**	**Request for data to data custodian**	**Data custodian approval (in-principle)**	**Approval to be provided by senior delegates**	**Contract or MoU required**	**PHRN linkage unit**	**Brokerage available by linkage unit**
Project approval—AIHW Ethics Committee	✓						
Child Care Management System data		✓	✓	✓	*		
Australian Early Development Census data		✓	✓	✓	✓		

✓ Denotes that this is a requirement for accessing Commonwealth data. It does not signify that approval has been obtained.

* Denotes requirement for an MoU or contract yet to be determined.

Jurisdictional data sets

Obtaining most jurisdictional data for the NECD RDS requires the approval in-principle from each data custodian of the collection; however, this does not automatically guarantee access to the data. Often the proposal for a project will also need approval from another committee that will consider the ethical implications for the project and the project’s scientific validity. Depending on the organisations, these can be ethics, privacy or other review committees.

Jurisdictions hold several data sets identified for inclusion in the NECD RDS. These include births data (used only for linking data files), perinatal data and preschool data, as well as NAPLAN data, which would be requested in the future. A common requirement of data custodians, particularly of education data, is the development of an MoU between jurisdictions and the AIHW, to establish arrangements for the transfer of data for the NECD RDS.

Based on the initial requirements of 1 jurisdiction, an MoU was drafted that forms the basis of an agreement between the AIHW and jurisdictions. This draft MoU contains an anticipated timetable for the provision of data (for up to 3 years), as well as a list of the data items required.

In some cases, data custodians may have unique requirements, for example mandatory reporting requirements for monitoring purposes. These requirements will be incorporated into the MoU on a case-by-case basis.

Another specific requirement of some data custodians, who provide identifying information for data linkage, is that the AIHW notify them of any records that may be multiple records for an individual or any other issues relating to the quality of the data.

The requirements for accessing jurisdictional data sets and including them in the NECD RDS are summarised in Table 4.4. Additional challenges identified are detailed below.

Table 4.4: Requirements to obtain access to state and territory data for inclusion in the NECD RDS

Collection	NSW	Vic	Qld	WA	SA[(a)]	Tas	ACT	NT
Births registration data								
Mandatory requirements								
Review by Human Research Ethics Committee	✓	✓	✓	✓	✓	✓	✓	✓
Request for data to data custodian	✓	✓	✓	✓	✓	✓	✓	✓
Data custodian approval (in-principle)	✓	✓	✓	✓	✓	✓	✓	✓
Contract or MoU required	✓	*	*	*	*	*	*	*
Data held by jurisdictional linkage unit	✓	✓	✓	✓	✓	✓	✓	✓
Brokerage available by linkage unit	✓			✓	✓			✓
Perinatal data[(b)]								
Mandatory requirements								
Review by Human Research Ethics Committee	✓	✓	✓	✓	✓	✓	✓	✓
Request for data to data custodian	✓	✓	✓	✓	✓	✓	✓	✓
Data custodian approval (in-principle)	✓	✓	✓	✓	✓	✓	✓	✓
Contract or MoU required	✓	*	*	*	*	*	*	*
Data held by jurisdictional linkage unit	✓	✓	✓	✓	✓	✓	✓	✓
Brokerage available by linkage unit	✓			✓	✓			✓
Preschool collection[(c)]								
Mandatory requirements								
Review by Human Research Ethics Committee								
Request for data to data custodian	✓	✓	✓	✓	✓	✓	✓	✓
Data custodian approval (in-principle)	✓	✓	✓	✓	✓	✓	✓	✓
Contract or MoU required	*	✓	✓	*	✓	✓	*	✓
Data held by jurisdictional linkage unit								
Brokerage available by linkage unit								

✓ Denotes that this is a requirement for the state or territory. It does not signify that approval has been obtained.
* Denotes requirement for an MoU or contract yet to be determined.
(a) For perinatal data, South Australia also requires review by the Aboriginal Human Research Ethics Committee.
(b) Name of ethics committees for perinatal data:
NSW—NSW Population and Health Service Research Ethics Committee
Vic—Department of Health and Department of Human Services Human Research Ethics Committee
QLD—Queensland Health Office of Health & Medical Research—Human Research Ethics Committee
WA—Department of Health WA Human Research Ethics Committee
SA—Aboriginal Health Research Ethics Committee; SA Department for Health and Ageing Human Research Ethics Committee
TAS—Tasmania Health and Medical Human Research Ethics Committee
ACT—ACT Human Research Ethics Committee
NT—Human Research Ethics Committee for the Northern Territory Department of Health and Menzies School of Health Research.
(c) For preschool data, there is also the requirement that approval be provided by senior delegates in all jurisdictions.
Note: In the ACT, the Chief Minister's Directorate is coordinating data linkage processes that occur between different directorates.

Births data

This information is collected by the Registrars of Births, Deaths and Marriages for each state and territory, which are housed within different Departments and Directorates within those jurisdictions. Births data would be required to undertake linkage but would not form part of the NECD RDS.

In late 2012, AIHW officers contacted the Registrar-General of Births, Deaths and Marriages, and provided information about the project. Currently, Australian Coordinating Registries (ACR) have been established or are planned to oversee data access for the individual elements of births, deaths and marriages data. It is likely that with the new processes, accessibility to data will be streamlined. However, an ACR process to obtain birth registration data is yet to be established. Therefore, at the time of writing this report, it is planned that these data will be requested from individual state and territory registrars, once their jurisdiction's HREC approves the project.

In September 2013, the Centre for Health Record Linkage (CHeReL) advised the AIHW that approval from the New South Wales Registrar of Births, Deaths and Marriages might be required before an HREC would review the application. Originally, this was not considered necessary, as no birth data were to be included in the NECD RDS, and these data would only be used for linkage. At the time of writing this report, this had not been considered a necessary step for the Australian Capital Territory or South Australia.

Perinatal data

For perinatal health data, most jurisdictions required approval for the project to be granted by a state-based HREC. Some states also indicated a possible need for approval from an Aboriginal health ethics committee. However, only 1 jurisdiction mandated this requirement.

Most HRECs have processes other than a single application form that must be completed before an application is formally reviewed. These usually involve completion of documents or the provision of more detailed lists of data items, and most processes differ across jurisdictions. What became clear through application processes is that most projects (studies) that are reviewed by HRECs are clinical in nature, so many of these forms (including the National Ethics Application Forms—see Box 3.1) adapt poorly to data and information projects like the NECD RDS, although the same forms and procedures apply. This can result in HRECs requesting additional information to address specific concerns in their jurisdictions. Table 4.5 shows the variety of forms that have been requested across each state and territory.

Table 4.5: Types of application forms required when obtaining permission to use health (perinatal) data in the NECD RDS, by jurisdiction, as at September 2013

Form type	NSW	Vic	Qld	WA	SA	Tas	ACT	NT
PHRN linkage unit application form	✓				✓		✓	✓
NEAF*	✓		✓	✓	✓	✓[(a)]	✓	✓
HREC application form						✓	✓	
Research protocol form	✓							
Request for data form		✓	✓	✓[(b)]				
Low risk application form							✓	
Privacy form	✓					✓		
Public Health Act 2005 form			✓					
Core application form					✓			
Declaration of Interests		✓						
Finance and administration forms						✓		
CCOPMM[(c)] data and research request form*		✓						

✓ Denotes that this is a requirement for the state or territory. It does not signify that approval has been obtained.

*Forms are filled out and submitted online.

(a) The NEAF used by this jurisdiction may be one that was specifically used and approved by another jurisdiction.

(b) Initially used as an expression of interest.

(c) Consultative Council on Obstetric and Paediatric Mortality and Morbidity.

For obtaining perinatal data, advice was requested from the perinatal data custodian on whether approval should be sought through the PHRN or state health department and whether in-principle approval was required prior to submitting the HREC application for each state and territory (also as part of births).

Specific challenges encountered when applying to use perinatal data included:

- Although standardised application forms for HRECs have been developed, these are designed for one-off research projects and often have a more clinical focus, so are not well suited for an ongoing project, as planned for the NECD RDS. This meant that additional forms and information often had to be supplied. Some HRECs also felt that they should review future applications to use data.
- Application processes differed between jurisdictions, for example, 1 jurisdiction requested in-principle approval from data custodians prior to an HREC application, whereas another jurisdiction requested HREC approval prior to data custodian in-principle approval.
- The planned ongoing nature of the NECD RDS concerned some jurisdictions and additional information on governance and management was requested. In addition, concerns about the resources required for ongoing supply of data were raised.
- Other data linkage projects, largely state-based, are using the same data sources. This created some confusion about the role of the NECD RDS, concerns about duplication and concerns that data custodians would lose control over data access and use.
- Jurisdictions are at different stages of establishing protocols and procedures for data linkage projects, making it more difficult to obtain permission to use data for linkage in some jurisdictions.

Preschool and education data

For education-related (preschool) data, all the jurisdictions that gave their in-principle approval (to date 5/8) also required that an MoU be developed between the AIHW and their respective departments of education. Draft MoUs were sent to these jurisdictions for comment. The MoUs cover the processes for transmitting data, annual reporting of data usage, and annual requests to update data and timeframes, and outline specific requirements of the states, such as the need to review researchers' projects or the provision to provide data on an ongoing basis. Some data custodians required that any results be provided to them before publication. This allows data custodians to prepare briefs for senior ministers or to check data for inappropriate use. As previously noted, there is considerable divergence in the process for data provision across jurisdictions.

Another theme that became evident through liaising with jurisdiction education departments was their wish for access to linked data. In response to these requests, the AIHW has now agreed that each data provider will have access to their state's linked health and education data (if and when the NECD RDS is established), with the provision that all data custodians in that state agree to this, and that no cross-border records are included. At the time of writing this report, the AIHW was considering the implications of including Commonwealth data (relative to that state). It is likely that for this to happen, each jurisdiction would have to gain in-principle approval from other jurisdictions as well as the AIHW Ethics Committee. In the case of AEDC data, the AEDC National Committee would also need to review and approve applications.

The proposed process for obtaining education data was to make contact with ECDSG representatives during the road shows and via teleconferences/meetings. This would be followed up with letters to senior officials within the state and territory education departments requesting data. Any queries would be addressed and contracts or MoUs drafted before the final step of transferring the data. Transfer of data would require agreement between the parties about the best method to maintain the state-, territory- and AIHW-specific confidentiality and privacy requirements, while accessing the best quality data for linkage.

Some specific issues encountered with preschool data are detailed below:

- Privacy legislation varies across jurisdictions. In some jurisdictions, privacy legislation prevents disclosure of personal information to third parties without individual consent, and for education data, this consent was not always obtained.
- The majority of jurisdictions would require an MoU with the AIHW prior to transfer of any education data, with some requiring additional documentation such as an Information Access and Disclosure *pro forma*.
- A number of other projects are currently being developed that link education data (the National Education Data Linkage Roadmap and the TETIA), and the NECD RDS was seen as replicating these projects.
- Jurisdictions requested that they have access to linked data from their own jurisdiction.

Summary of progress

During the development of the NECD RDS, significant progress was made in obtaining permission to use a range of data sets in the NECD RDS. This is summarised in Table 4.6.

Table 4.6: Progress towards obtaining data for the NECD RDS, as at October 2013

Data collection and jurisdiction	Date applied for	Response	Current status
Child Care Management System —Commonwealth	Letter requesting data sent in July 2012	Data not available to date	Data not available to date
Australian Early Development Census—Commonwealth	Letter sent in January 2013 Meeting in April 2013	Requirement of an MOU	Deed of Agreement MoU signed and AIHW has received the data.
Births registrations			
NSW			Approved in September 2013 by Registrar of Births, Deaths and Marriages (BDM).Request for data currently being reviewed by the NSW Population & Health Services Research Ethics Committee (PHSREC)
Vic			Not yet requested
Qld	Preliminary discussions and letter outlining project to Registrar-General in November 2012	As the Australian Coordinating Registries (ACR) is not yet established to oversee data access to births, deaths and marriages data, AIHW will pursue data from individual state and territory registrars, once their jurisdiction's Ethics Committee approves the project	Data yet to be formally requested
WA			Not yet requested
SA			SANT DataLink will request on AIHW's behalf. Currently being reviewed by the South Australian Health Human Research Ethics Committee.
Tas			Not yet requested
ACT			Not yet requested
NT			Not yet requested

(continued)

Table 4.6 (continued): Progress towards obtaining data for the NECD RDS, as at October 2013

Data collection and jurisdiction	Date applied for	Response	Current status
Perinatal data			
NSW	CHeReL contacted to establish processes in February 2013 Perinatal contacts/data custodian contacted in February 2013	Applications submitted to CHeReL in March 2013. Data custodian gave in-principle approval in March 2013 NSW NEAF submitted in September 2013. Additional information provided in October 2013.	In-principle approval by data custodian. Application reviewed by the NSW PHSREC and approval granted.
Vic	Contacted Vic Health in March 2013 to discuss processes for obtaining data. Provided additional information in May 2013. Contacted Data Linkage Unit in April 2013.	Submitted on-line applications in July 2013, including CCOPMM Data and research request form and Declaration of interest	Application form submitted and awaiting comments and advice from CCOPM. A change of staff slowed the process.
Qld	Contacted HealthLinQ in February 2013 and submitted NEAF and associated applications for comment.	Contacted the perinatal data custodian in March 2013 and applications submitted. Advised in April 2013 that due to the on-going requirement of perinatal data that our request would require further consideration from Director-General	Application forms submitted and awaiting advice from Queensland Department of Health with advice from Director-General.
WA	Expression of interest applications submitted to WA for comment in February 2013	It was decided to wait until a working model for data provision and linkage was in use.	Application forms submitted but not being progressed currently.
SA	Contacted SA-NT DataLink in January 2013. Applications submitted for comment in February 2013. Contacted perinatal data custodians in April 2013 and they provided in-principle support.	NEAFs and associated applications submitted to SA-NT DataLink to progress in March 2013. Support from data custodian revoked in September 2013 due to the number of data linkage projects that will be storing data over a period of time.	HREC requested additional information in May 2013, including peer review of the project, which was undertaken in July 2013. Additional information was requested, including AHREC Approval and letters of support, in August 2013. Data custodian is developing strategies for projects that hold data over a period of time.
Tas	Initial contact with Data Linkage Unit in January 2013 to discuss the processes to obtain data.	Advised in February 2013 that approval by an HREC was required before submitting applications.	Applications submitted to University of Tasmania HREC in September 2013.
ACT	Contacted ACT in May 2013 and confirmed that HREC approval is required as first step	HREC application submitted in July 2013 and approved in August 2013. Perinatal data can be applied for now.	Further contact with the ACT Directorate has taken place.

(continued)

Table 4.6 (continued): Progress towards obtaining data for the NECD RDS, as at October 2013

Data collection and jurisdiction	Date applied for	Response	Current status
NT	NT Health contacted in May 2013 regarding perinatal data and discussion with NT Health and perinatal data custodian.	NEAF and information provided to NT data custodian in May 2014. After further discussion with NT perinatal data custodian it was agreed that the preferred option for obtaining NT data is through the SA-NT DataLink.	Requested through SA-NT DataLink and awaiting advice.
Preschool collection			
NSW	Letter requesting data sent April 2013	Reply in June	Problems with privacy yet to be resolved
Vic	Letter requesting data sent April 2013	In-principle approval	Draft MoU provided for comment
Qld	Letter requesting data sent April 2013	Reply in May	Draft MoU provided for comment
WA	Letter requesting data sent April 2013	Reply in May	Not progressing at this time
SA	Letter requesting data sent May 2013	Reply in June	Draft MoU provided for comment
Tas	Letter requesting data sent April 2013	Reply in May	Draft MoU provided for comment
ACT	Letter requesting data sent April 2013		Meeting with Chief Minister's Office to be convened to discuss queries.
NT	Letter requesting data sent May 2013	Reply in May	Draft MoU provided for comment

Note: When contacting stakeholders for permission to obtain data there was a need to contact a number of stakeholders. For example, for perinatal data, ethics committee, data custodian, PHRN and in some jurisdictions the data linkage unit all needed to be contacted.

Receiving data for inclusion in the NECD RDS

Once permission has been granted to include data in the NECD RDS, the data would need to be securely transferred to the AIHW. Data for linkage could be transferred to the AIHW either directly or via the PHRN linkage node/unit within the specific state or territory. If the data custodian prefers using the state-based PHRN node, the AIHW would work with state data linkage units to facilitate data transfer.

The AIHW Data Linkage Unit (or state/territory data linkage unit) would use identifying information to produce a concordance map and a list of project IDs for which content data would be requested (see Figure 2.1). The data custodian or the AIHW would then use the concordance map to merge de-identified content data. The data linkage and data merging would be undertaken in accordance with the strict governance and institutional arrangements that apply to Commonwealth integrating authorities (NSS 2010b).

Table 4.7 outlines the data linkage steps required for each specific type of data collection (births, perinatal and preschool).

Table 4.7: Steps for obtaining data from birth registration, perinatal and preschool data collections

Birth registration data	Perinatal data	Preschool data
Birth registration data would be required for the linkage step only. In jurisdictions where the baby's name is not available in the perinatal data, birth registration data would be matched with perinatal data to provide sufficient identifying information to undertake linkage. 1. Contact registrar for in-principle approval as the data custodian—through the head registrar (Qld) 2. Apply HREC in each state and territory (also as part of perinatal) 3. Births data linked with perinatal data 4. Refer to perinatal data for remaining steps	1. Contact perinatal data custodian for advice on going through the PHRN/ state health department 2. Contact perinatal data custodian (in-principle approval) 3. Submit HREC application for each state and territory (also as part of births) 4. Transfer of linked births and perinatal data (identifying variables only) for creation of concordance map and request content data for subset of project IDs from perinatal data custodian 5. Perinatal data custodian transfers de-identified content data for inclusion in NECD RDS	1. Initial contact with ECDSG representatives and road shows 2. Letter requesting data to head of education department 3. Response from each education department 4. Respond to address any queries/issues raised based on response from head of education department 5. Draft and execute an MoU where applicable 6. Transfer of data (identifying variables only) for creation of concordance map and request content data for subset of project IDs from preschool data custodian 7. Preschool data custodian transfers de-identified content data for inclusion in NECD RDS

Supply and updating of data

Data for the NECD RDS would be sourced from a number of different collections, and therefore have varying timeframes around their collection cycles and availability for inclusion in the NECD RDS (see Table 4.8). Some initial data supplies may include data for a number of years before settling into a regular pattern of supply. MoUs to facilitate ongoing supply of data and obligations from all parties regarding the maintenance and updating of data are being developed.

Table 4.8: Proposed supply and updating of data for the NECD RDS

Collection	State or Commonwealth based	Collection frequency	Years required for the NECD RDS	Update schedule	Comments
Perinatal data	State	Ongoing	2006 onwards	Yearly	For first delivery requesting all data from 2006 onwards
Births data	State	Ongoing	2006 onwards	Yearly	Data have not yet been requested, but it is expected that years requested will match those requested for perinatal data
Preschool data	State	Yearly	2011 onwards	Yearly, after submission to the ABS ECEC data collection	
CCMS	Commonwealth	Ongoing	2011 onwards	Yearly	To align with data requested for preschool years
AEDC	Commonwealth	3 yearly	2012	3 yearly. Once data become available, usually the year after the collection	For first delivery, 2009 and 2012 data requested
School Census	State	Twice per year	To be determined	To be determined	
NAPLAN	Commonwealth	Yearly for certain grades	2015 onwards	Once data become available	Data not yet requested, as the 2011 preschool children have not yet done the NAPLAN

5 Moving forward

This chapter describes the current status of the development of the NECD RDS, some plans if and when data become available for research and recommendations for continued work in this area. It must be recognised that funding to continue development of the NECD RDS has not been secured and that progress would be dependent on the project receiving adequate funding.

5.1 Where are we now?

At the time of writing this report, applications had been submitted to all jurisdictions for access to perinatal and education data and to the DoE for access to CCMS and AEDC data. Each of these applications is at a different stage in the approval process. An exception is Western Australia, where a decision was made to wait until a working model for data provision and linkage was in use.

To date, the AIHW has HREC approval from one state to transfer some jurisdictional data to the AIHW. An MoU with the DoE for the transfer of AEDC data to the AIHW has been developed and data received. AIHW has continued to build upon previous work. For example, the AIHW continues to build and maintain relationships with data custodians, engage with other researchers who are interested in data linkage, and build national data linkage infrastructure.

5.2 Next steps

In order to continue development of the NECD RDS, funding for a project manager and a small team of project officers would need to be secured. This team would continue to work with jurisdictions to develop MoUs and to obtain required ethics approval for each data collection. If development of the NECD RDS is continued, the next steps would include:

- negotiating with data custodians
- progressing approvals
- liaising with HRECs and other groups
- establishing governance and access processes for researchers.

Upon the arrival of data, several activities would need to be undertaken, including:

- identification of project IDs for which content data would be requested
- return of project IDs to data providers, who would then send de-identified content data
- construction of concordance maps
- merging of de-identified content data
- analysis of linked data to establish linkage rates and assess quality of the linked data
- development of documentation for data custodians and users (for example, data guides, data quality statements and data dictionaries).

5.3 After the NECD RDS is established

The efforts of the AIHW, its collaborators and data custodians have been substantial, and there has been significant progress towards the development of the NECD RDS.

Assuming that the NECD RDS project continues, the following sections address future aspects of the project, including access to the NECD RDS, continuing development of the NECD RDS, updating data in the NECD RDS and communication with stakeholders.

Stakeholder consultations and meetings

Consultation with stakeholders will continue and be enhanced to incorporate new stakeholders as required. As with current consultation processes, the project officers will report on the status of the NECD RDS, review processes for data delivery and linkage, and once enough data are available, describe how researchers can access data for their own analysis.

The AIHW has also agreed that it will provide updates on development of the NECD RDS to committees such as the Early Childhood Development Sub Group and the Children's Services Working Group. These types of committees provide a forum for reporting on the progress of the NECD RDS and promoting its benefits.

Access to the NECD RDS

A common theme arising through consultations with data custodians and other stakeholders was the need to know how the NECD RDS will be used and accessed. The concerns mainly focused on the arrangements in place for data custodians to gain access to data, the processes researchers will have to follow to gain approval to access the NECD RDS and, once approved, the means by which they will actually access the data.

For jurisdictional data custodians

At July 2013, the AIHW had decided to make a copy of a state's linked health and education data (perinatal and preschool) available to the jurisdiction, providing agreement from all data custodians in that jurisdiction is obtained, and that no records from another state are included in that file.

If jurisdictions want other data included in that file, for example Commonwealth or data from other states or territories, a separate process will have to be undertaken. This will include seeking in-principle approval from other data custodians as well as receiving approval from the data owners, NECD RDS AC and the AIHW Ethics Committee.

For researchers

As some jurisdictions were concerned about the processes by which researchers could access the NECD RDS, suggested draft guidelines for access were developed (see Appendix E). These were largely based on experience with other data sets that the AIHW manages and processes used by other organisations that make data available for research purposes. Readers should note the status of these and use them only as an indication of how processes to access the NECD RDS may operate. Once the NECD RDS is ready for research, the AIHW will make additional documentation, such as user guides and data quality statements, available to researchers.

Researchers will only be able to access the NECD RDS for research purposes. As part of the terms and conditions for data provision to the AIHW, some data custodians have requested to review and approve research proposals and publications. This may occur separately from, or as part of membership to, the NECD RDS AC. Data custodians will review research proposals to ensure:

- the relevance of the data being sought and that the way they will be used meets the desired outcomes of the applicant
- relevant expertise is sought to determine the appropriateness of the purpose and use of the requested data
- information being sought will not breach confidentiality (AEDC 2013).

As described in Appendix E, the NECD RDS AC will produce a guideline for researchers to access data in the NECD RDS. Access options include:

- the Secure Unified Research Environment (SURE) facility
- the AIHW data laboratory
- the DISC mobile data laboratory
- as aggregates or outputs from analysis undertaken at the AIHW
- through specialised data requests to the AIHW.

Researchers will be encouraged to contact the AIHW to discuss which options suit them best.

Researchers will be required to gain approval from data custodians before submitting an application to the AIHW Ethics Committee. Some data custodians may request to review the results of any research before publication. This option enables the early development of briefing papers and responses to research if required.

Updating the NECD RDS

The NECD RDS will be updated as additional data become available. This will include new information about the initial cohort of children, such as Year 3 NAPLAN data in 2015, as well as adding new cohorts to the NECD RDS. Although initial documentation (Table 2.1) shows that Year 5 NAPLAN data have been considered for inclusion, the AIHW will seek advice from the NECD RDS Advisory Committee before deciding to progress this.

Other education-related data that will be considered for inclusion are School Census data, sometimes referred to by jurisdictions as school enrolment data. These data sets will allow further demographic information about the children (for example country of birth of parents) to be included. In some jurisdictions, School Census data also provide indicators of disability, which will allow further analysis of preschool experience and educational outcomes for children with disabilities. At the time of writing this report, the intervals for obtaining School Census data sets had not yet been decided.

Data analysis

The first series of analysis using NECD RDS data will focus on describing the number of records and the proportion of successful matchings between records for multiple collections.

Once sufficient data exist in the NECD RDS for analytical purposes, the AIHW may produce statistical reports describing the educational or developmental outcomes for Australian children. In line with the reporting policies of the AIHW, all reports would be freely available on the AIHW website.

Over time, the AIHW may undertake more specific projects and analysis. AIHW will have to apply for approval from data owners and ethics committees prior to undertaking any research.

METeOR

METeOR is Australia's ongoing linked data set for national metadata standards for health, housing and community services statistics and information. Information about the NECD RDS, including the data items it contains and their supporting information will be included in METeOR.

Information about METeOR can be found at <http://meteor.aihw.gov.au/content/index.phtml/itemId/181414>.

Information about the NECD RDS, the data that comprise it and statements of data quality will be loaded onto METeOR once the NECD RDS contains data from a number of sources, thereby providing this information to a wide audience.

Appendix A: Childhood development research

When building a new data set to be used for research purposes, it is important to understand what types of research have already been undertaken; what are the current and perhaps future requirements for research; and where gaps in information exist. The AIHW investigated the research knowledge about childhood development. This appendix provides information on a selection of relevant research in childhood development.

A.1 The review process

The first step was to search for articles that contained discussion about the effects of certain factors—such as early health, disability, socioeconomic background and child care—on early childhood development, preschool, school readiness and educational outcomes in the future.

The search included academic articles, reports, policy documents, and conference papers. The scope of searching was constrained to the last 10–15 years and encompassed databases (Science Daily, Proquest, Informit, ScienceDirect and PubMed) as well as websites—Australian Institute of Family Studies (AIFS), Australian Early Development Census (AEDC), AIHW, Department of Social Services (formerly FaHCSIA) and Centre for Community and Child Health, and Google Scholar. A random sample of the articles identified was used in further analyses.

Using 181 randomly selected publications, a matrix was constructed to map the results of this research. It plotted factors by outcomes, for example low birthweight to future health. The themes discovered in the search process were used as the column and row headings in the matrix.

This type of mapping highlighted which topics were covered by research activity and indicated areas where more or less of this research had occurred. Mapping the types and amounts of research in a matrix format also highlighted where potential data gaps exist.

A.2 Results of mapping

The mapping indicated that research findings were available about developmental outcomes for children who were born with a low birthweight and about educational and developmental outcomes for children in relation to child care.

There was some research on the educational and behavioural outcomes for those children who had attended preschools; but, very little contrasting information about outcomes for children who had not attended preschool. Internationally, there were some smaller studies undertaken about children who either attended or did not attend preschool programs in the 1960s. These studies were quite specific in their focus (for example, the Perry Preschool project collected information about African-American children who lived in poverty), but revealed positive aspects for the futures of children who were exposed to quality preschool programs (Berrueta-Clement 1984). More recently, longitudinal studies have been undertaken in England, which included children with no preschool experience, and investigated the relationships between preschool quality and several outcomes, such as educational outcomes at age 11 (Sylva et al. 2011) and cognitive development (Halle et al. 2009).

Research was available (but limited) linking socioeconomic status, cultural background or Indigenous status, as identified through parental variables, and future outcomes for children, as was research about the children's Indigenous or cultural background and the links to future outcomes. Articles based on parental characteristics (for example, mother's country of birth) focused on educational outcomes for reading, numeracy and language/literacy, but not on outcomes for intellectual and behavioural development, or social competence. Some small-scale studies were identified, although these were beyond the 10–15 years of the original search criteria and were based on small numbers of children. The results of these studies did indicate that there were correlations between selected parental characteristics and behavioural and mental development of children (Dodge 1994; Bayley 1965).

There was very little research found about the effects of injuries, trauma or disability, and future health, developmental or educational outcomes. The scope of articles related to these topics was more likely to be prevention based, particularly in relation to child abuse, or more clinical in nature, for example the effects of head trauma on memory (Petersen & Bell 1996).

Analysis of research themes confirmed that many of the research questions identified in Section 2.1 will contribute to filling gaps in current knowledge about child development. It must be noted that the analysis conducted here was not exhaustive and that gaps identified may reflect a smaller body of research rather than a lack of research.

Longitudinal research

The NECD RDS will enable analysis of data over a number of years, which will benefit early childhood development research in Australia. Much of the research identified through the exercise relates to studies undertaken at 1 point in time. Data from the NECD RDS will also show how changes in factors, such as socioeconomic status of parents, can translate to different outcomes for children, as was found in the United Kingdom (Bowers & Strelitz 2012).

Appendix B: Data sources to support the NECD RDS

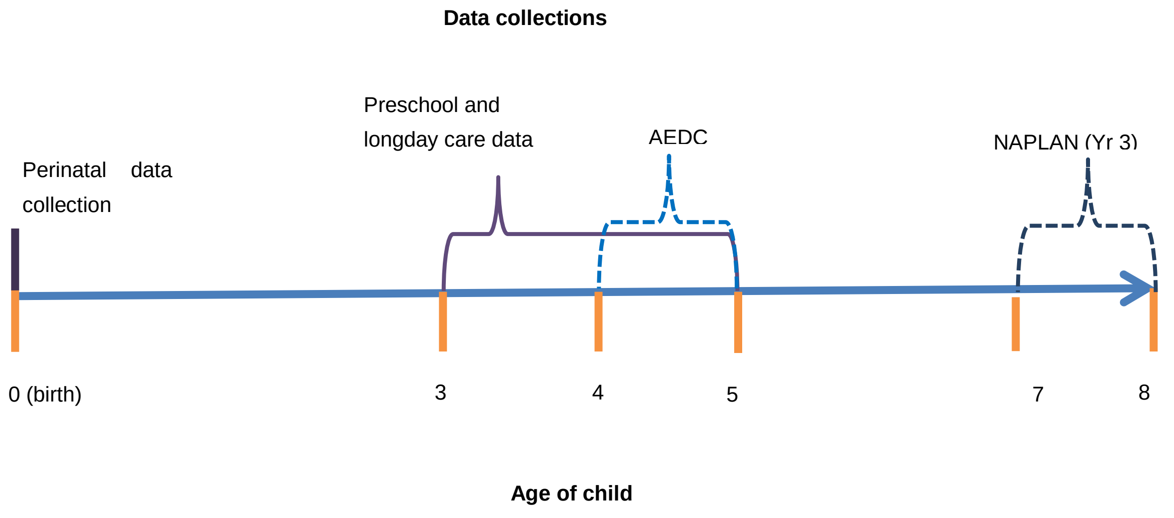
This appendix provides information on other data sources that could be used to enhance information from the NECD RDS.

B.1 Data included in the NECD RDS

The most rapid phase of children's development occurs during the early years of life. During this time, children develop in many ways, for example physically, emotionally and intellectually. They learn how to communicate and how to interact socially. Vital to these processes is good brain development, and early childhood is the most intensive period of brain development during the lifespan (WHO 2013). Adequate stimulation and nutrition are essential for development during the first 3 years of life. It is during these years that a child's brain is most sensitive to the influences of the external environment. Rapid brain development affects cognitive, social and emotional growth. Such development helps to ensure that each child reaches his or her potential and is a productive part of a rapidly changing, global society.

Many things can affect the cognitive and non-cognitive development of children, including families and schools, and differences by family income and family background emerge early and persist (Cunha & Heckman 2006). Research indicates that 'If anything, schooling in the early grades widens these differences. However, most of the gaps in these skills that are found in adulthood emerge before schooling begins.' (Cunha & Heckman 2006).

A review of Australian data showed that there are very few data sources available for inclusion in the NECD RDS that contain information about all Australian children from just after the birth to around the age of 3 years (or until they go to preschool) (Figure B.1). This gap in data sources limits information about the development of children in that age group, as well as limiting opportunities for the NECD RDS to access and use data for this age group.



Note: Perinatal, AEDC and NAPLAN (at Year 3) are collected once for each child. Preschool and long day care data are collected for each year the child is attending that service and attendance at multiple services in the same year are possible.

Figure B.1: Core NECD RDS data sources and at what ages they are collected

B.2 Data to support the NECD RDS

There are a number of data sources that could provide contextual information for the NECD RDS. Some data sets could potentially be included in the NECD RDS, while others can provide only general information about Australian children. More information on available data sources to support the NECD RDS is provided below.

Australian Childhood Immunisation Register

The Australian Childhood Immunisation Register (ACIR) is a national register that records vaccinations given to children under 7 years. It was established in 1996 in response to a decline in childhood immunisation in Australia and an increase in vaccine-preventable childhood disease. The proportion of infants and children assessed as fully immunised was identified as a key indicator of positive childhood development (AIHW 2011a). This data set could be considered for inclusion in the NECD RDS.

Medicare data

Medicare statistics comprise a number of collections, one of them being the ACIR which is discussed above. Other collections include:

- Medicare Benefits Scheme (MBS)
- Pharmaceutical Benefits Scheme (PBS) and Repatriation Pharmaceutical Benefits Scheme (RPBS)
- Australian Organ Donor Register
- Practice Incentives Program (Medicare 2012).

The two most relevant collections for statistics about children are the MBS, the PBS and the RPBS. The MBS is based on the items and groups in the Medicare Benefits Schedule, and the

PBS and RPBS are based on the items and Anatomical Therapeutic Chemical (ATC) groups in the Pharmaceutical Benefits Schedule. MBS data record the use of private medical services, for example single doctor consultations, pathology tests and services from a practice nurse (AIHW 2012). Although these data would give information about the services provided, they do not give information about reasons for the service usage. PBS data can show prescriptions for those medications which are subsidised by the Australian Government. However, they do not show the reasons for that prescription. Linking MBS and PBS information could reveal patterns in health service use or prescription use for children, which may contribute to knowledge about their health and their outcomes in health and development over time.

The National Death Index

The NDI is a file specifically maintained for data linkage. It contains registrations of all deaths in Australia, supplemented with cause-of-death information as it becomes available. If linked to the NECD RDS, the NDI would allow for adjustments in the NECD RDS population, thereby allowing for more accurate analysis. Information for the NDI is collected by the Registrars of Births, Deaths and Marriages in each state and territory, and transferred to the AIHW.

ABS Census

The Census of Population and Housing (the Census) is a descriptive count of everyone who is in Australia on 1 night, and of their dwellings (ABS 2012). Its objective is to accurately measure the number and key characteristics of people who are in Australia on Census Night, and of the dwellings in which they live. The most recent Census was conducted on 9 August 2011.

The types of data collected by the Census are vast and varied, and range from demographic economic information to statistics about how people travel to work, unpaid child care and disability.

Although organisations outside the ABS will not be given data about an individual person or individual household, the ABS can undertake data linkage using the Census. The confidentiality of Census data is protected by the *Census and Statistics Act 1905* and the Privacy Act. Both these Acts ensure that data are not provided to anyone where those data can be used to identify an individual. Census data do however provide further information about Australian children and their families.

Other data obtained through surveys

Data about children are also available from a number of surveys in Australia, and information about a selection of these is presented next. Each of these surveys has its strengths and limitations. Because they are usually sample surveys (not all children in the population are surveyed) they cannot be included in the NECD RDS. However, results from these surveys could provide additional information about the health and wellbeing of children in Australia.

LSAC and LSIC

In Australia, there are 2 main data sources covering the early childhood years: *Growing Up in Australia: The Longitudinal Study of Australian Children* (LSAC) and *Footprints in Time – the Longitudinal Study of Indigenous Children* (LSIC) (AIFS 2013; DSS 2014). Both these studies comprise 2 cohorts of children who are re-surveyed over time. At the start of the surveys, the children in the cohorts were aged 3–19 months and 3½–5 years. Both these collections aim to provide researchers with insight into the effects of early years on child development.

Child Dental Health survey

The Child Dental Health survey is an annual survey that monitors the dental health of children and young people enrolled in school dental services. Data are derived from routine examinations of children, and include information on dental health status, such as decay experience, immediate treatment needs and fissure sealants. Data have been collected annually since 1990.

ABS National Health Survey

The series of ABS National Health Surveys (NHSs) collects information about health status, health-related actions (for example, visits to doctors) and risk factors (for example, diet and exercise) of Australians. The survey collects information about people in all age groups; however not all aspects of the survey relate to all age groups. For example, some questions may be asked of adults only. The ABS has conducted the NHS since 1977 at different intervals. The current ABS NHS program has scheduled surveys every 3 years.

Australian National Infant Feeding Survey

The 2010 Australian National Infant Feeding Survey is the first specialised national survey of infant feeding practices in Australia (AIHW 2011b). The survey also collected information on attitudes towards, enablers for and barriers to breastfeeding.

Appendix C: Data items for inclusion

Once the collections proposed for inclusion in the NECD RDS were identified, information about each of them was documented, including the data items contained, the data items likely to be—or not to be—relevant to include in the NECD RDS, and those data items required for linkage (that is, identifying variables such as names). This appendix provides a draft list of data items proposed for inclusion in the NECD RDS.

As noted previously, identifying information from the Births, Deaths and Marriages collections will be used for linkage purposes but not included as part of the NECD RDS. Data items from the AEDC will be determined once the list of data items is released by the DoE.

The tables below list the data items from the Perinatal Data Collection, the ECEC Collection and the CCMS proposed for inclusion in the NECD RDS.

Table C.1: Data items from the Perinatal Data Collection proposed for inclusion in the NECD RDS, 2006 to present

Mother’s data items	Mother’s data items	Mother’s data items	Baby’s data items
Mother’s age	Gravidity	Complications of Labour	Baby’s date of birth
Area of residence	Alcohol in pregnancy	Pre-existing diabetes	Baby’s Indigenous status
Mother’s marital status	Gestation at first antenatal visit	Pre-existing hypertension	Plurality
Mother’s Indigenous status	No. of antenatal visits	Pre-existing epilepsy	Birth status
Mother’s country of birth	Assisted Reproductive Technology	Obstetric complications	Gestational age
Smoking during pregnancy	Actual place of birth	Gestational hypertension	Apgar score at 5 minutes
No. of cigarettes daily	Hospital sector	Gestational diabetes	Birthweight
State or territory of birth	Accommodation status	Ante-partum haemorrhage	Admission to SCN/NICU
Alcohol in pregnancy	Date of admission	Placenta praevia	
Maternal height	Maternal medical conditions	Placental abruption	
Maternal weight	Foetal distress in labour	Mortality (mother died)	
Maternal body mass index	Method/type of birth		
Mother’s mental health	Substance use		
Screening for domestic violence	Whether interpreter required	Maternal education	

Note: Although not currently collected for all jurisdictions, the shaded variables could be included if available.

Table C.2: Data items from the Early Childhood Education and Care Collection proposed for inclusion in the NECD RDS, 2011 to present

Child level data items	Early childhood education worker level data items	Service provider level data items
Maximum preschool program hours available	Person employed series number	Postcode—Australian
Preschool program attendance indicator	Paid employment indicator	Suburb/town/locality name
Preschool program enrolment indicator	Field of highest qualification relevant to early childhood education and care	Geographic location of organisation
Fees charged (early childhood education and care)	Level of highest qualification relevant to early childhood education and care	Management type (early childhood education and care)
Preschool program hours attended (total)	Delivery of preschool program indicator	Preschool program service operation weeks (calendar year)
Preschool program hours enrolled (total)	Hours worked in preschool program delivery	Organisation identifier (early childhood education and care)
Preschool program received from teacher indicator		Service activity type (early childhood education and care)
Preschool program repeater indicator		Service delivery setting (early childhood education and care)
Postcode—Australian		State or territory funding indicator
Suburb/town/locality name		Type of work activity
Area of usual residence		Role of early childhood education and care worker
Australian state/territory identifier		
Age at collection time		
Indigenous status		
Sex		
Preschool enrolment fees (per hour)		

Note: Due to issues discovered by some jurisdictions on the ABS return-to-source file, the AIHW is reliant on the data custodian's discretion as to whether they provide the return-to-source or the original data file for inclusion in the NECD RDS.

Table C.3: Data items from the Child Care Management System proposed for inclusion in the NECD RDS, 2011 to present

Child level data items	Early childhood education worker level data items	Service provider level data items
Maximum preschool program hours available	Person employed series number	Postcode—Australian
Preschool program attendance indicator	Paid employment indicator	Suburb/town/locality name
Preschool program enrolment indicator	Type of work activity	Geographic location of organisation
Fees charged (early childhood education and care)	Role of early childhood education and care worker	Management type (early childhood education and care)
Preschool program hours attended (total)	Field of highest qualification relevant to early childhood education and care	Preschool program service operation weeks (calendar year)
Preschool program hours enrolled (total)	Level of highest qualification relevant to early childhood education and care	
Preschool program received from teacher indicator	Delivery of preschool program indicator	Service activity type (early childhood education and care)
Preschool program repeater indicator	Hours worked in preschool program delivery	Service delivery setting (early childhood education and care)
Postcode—Australian		State or territory funding indicator
Suburb/town/locality name		
Area of usual residence		
Australian state/territory identifier		
Age at collection time		
Indigenous status		
Sex		
Preschool enrolment fees (per hour)		

Appendix D: Processes to construct the NECD RDS

This appendix provides information on how data are to be linked and merged to form the NECD RDS.

D.1 Linking data to construct the NECD RDS

As an accredited Commonwealth data integrating authority, the AIHW meets strict requirements when undertaking data linkage, including linkage to build the NECD RDS. This includes separating the processes of data linkage and data merging.

With the changing environment of data linkage in Australia, the processes considered at the start of the NECD RDS project have changed as the project has developed. The process, as described below, has evolved over time, and may change again as data linkage processes evolve.

Constructing concordance maps

Figure 2.1 shows the current proposed process for the construction of concordance maps and merging of records using state and territory data to construct the NECD RDS.

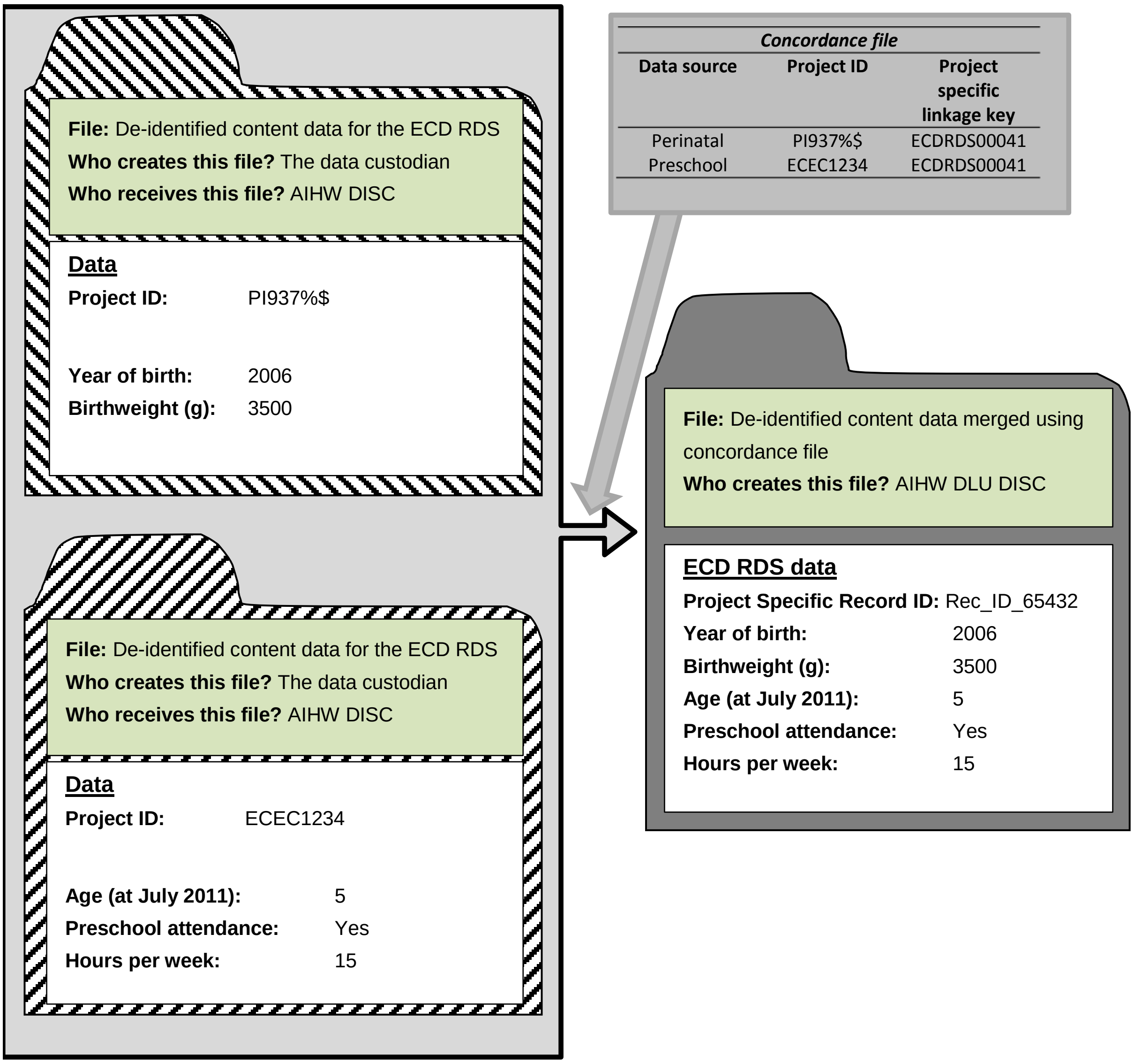
The AIHW is a member of the PHRN and is able to work with state linkage units. State and territory data custodians have the option of passing their information via state and territory linkage units, rather than transmitting data directly to the AIHW. One of the benefits of using the state and territory linkage units is that they may have already linked identifying information from a number of collections and produced linkage maps, therefore eliminating the requirement to repeat this process.

In general, if the jurisdictional linkage units have linked the required data, it is quicker and simpler to follow this path. However, if the linkage has not occurred, passing data directly to the AIHW is a simpler and quicker process.

It is recognised that different data custodians have different capacity to prepare files for transfer and to de-identify information. The AIHW is able to assist data custodians with any of these processes.

D.2 Merging content data for the NECD RDS

Figure D.1 presents a hypothetical example of 1 child’s record, which shows the proposed process for how data are obtained and de-identified from individual data sources (perinatal and preschool data) and merged to form a record in the NECD RDS.



Concordance file

Data source	Project ID	Project specific linkage key
Perinatal	PI937%$	ECDRDS00041
Preschool	ECEC1234	ECDRDS00041

File: De-identified content data for the ECD RDS
Who creates this file? The data custodian
Who receives this file? AIHW DISC

Data
Project ID: PI937%$
Year of birth: 2006
Birthweight (g): 3500

File: De-identified content data for the ECD RDS
Who creates this file? The data custodian
Who receives this file? AIHW DISC

Data
Project ID: ECEC1234
Age (at July 2011): 5
Preschool attendance: Yes
Hours per week: 15

File: De-identified content data merged using concordance file
Who creates this file? AIHW DLU DISC

ECD RDS data
Project Specific Record ID: Rec_ID_65432
Year of birth: 2006
Birthweight (g): 3500
Age (at July 2011): 5
Preschool attendance: Yes
Hours per week: 15

Figure D.1: Proposed method for merging records for the NECD RDS

In this example, a concordance file is constructed in the AIHW Data Linkage Unit (DLU) by matching identifying data (such as name, data of birth, SLK581 and postcode) across perinatal and preschool data collections. A project specific linkage key is generated for each matched individual. The DLU then sends a list of Project IDs to data custodians, requesting content data for these Project IDs.

Data custodians then create a file with de-identified content data to be included in the NECD RDS, by attaching Project IDs to content data. Files containing de-identified content data from different data collections are merged in the AIHW Data Integration Services Centre (DISC) to form the NECD RDS.

Appendix E: Guide for researchers

This appendix provides a suggested draft guide to researchers on how the NECD RDS could be accessed if and once established. This draft has been prepared in response to queries from data custodians regarding future access to the NECD RDS. It is largely based on AIHW experience with other data sets and processes used by other organisations to make data available for research purposes. This draft guide has not been through an approval process and therefore provides only an indication of how access to the NECD RDS may occur.

E.1 Access to data for researchers

Once data are considered robust enough (for example, having adequate numbers of records) for analysis, sections of the NECD RDS will be released for research and statistical purposes. In general, it would be preferable to have 1 complete cohort of data, for example, complete URL records for children who attended preschool in a particular year, however this may not be possible, as provision of data for some jurisdictions is more difficult.

Information about data availability from the NECD RDS will be provided to researchers and other users through the *Guide to the NECD RDS* (not yet published). This will be an online product that will be updated regularly.

All data will be released under the specifications and conditions imposed by the AIHW Ethics Committee and data custodians.

How will researchers access data from the NECD RDS?

There are a number of secure data dissemination options that will be available to researchers, and most of these will require approval from the AIHW Ethics Committee. Data custodians may also impose restrictions to access data; therefore, their approval may also be needed. The following is a general stepwise guide to accessing data.

1. Contact the AIHW Maternal Health, Children and Families Unit (see AIHW contact details below).
2. Obtain approval from custodians that contribute data to the NECD RDS. This step may be incorporated with Step 3, if custodians are members of the NECD RDS Advisory Committee (AC).
   a. This may be a requirement for some of the data custodians who provide data for the NECD RDS.
   b. Data custodians may also request to see results before they are published. This will allow them to prepare briefs for their senior officers and anticipate any questions that may arise.
3. Obtain approval from the NECD RDS AC.
4. Obtain approval from AIHW Ethics Committee, if appropriate.
5. Arrange the mode of access to the data.

Examination by the NECD RDS Advisory Committee

The NECD RDS AC will review applications for their value regarding research (Step 3).

This process may be iterative between the advisory committee and the researcher to fine-tune the applications, if required.

Examination by the AIHW Ethics Committee

After approval has been given by the NECD RDS AC and any data custodians, researchers who wish to obtain approval from the AIHW Ethics Committee should lodge an application using the online EthOS (Ethics Online System).

EthOS allows researchers to enter information and upload documents such as the NEAF and peer reviews directly into the online application form. Importantly, EthOS gives researchers the ability to:

- track the progress of their application online
- lodge annual and final monitoring reports online.

The AIHW Ethics Committee looks at each proposal on its merits, and ensures there is a good understanding of the proposal by all members. Each application is presented by an assigned member of the AIHW Ethics Committee to ensure this understanding and that a thoughtful perspective is placed before the AIHW Ethics Committee. AIHW staff attend Ethics Committee meetings on request to clarify issues and expedite response to the Committee's questions.

Before submission to the AIHW Ethics Committee, each application is examined by the relevant Data Custodian and signed off by the responsible Group Head.

Is there a cost involved?

An administration fee is charged for applications from external researchers. Please contact the AIHW Ethics Manager (see AIHW contact details below).

E.2 Access options

There are a number of secure ways that researchers may be able to access data from the NECD RDS for their own research purposes.

The SURE facility

The SURE is a secure computing environment developed as part of the PHRN, and managed by the Sax Institute. It provides a consistent level of security so that all researchers using it to work with data follow the same security protocols. The AIHW has negotiated a contract with the Sax Institute for the provision of services.

Researchers will use data that are stored by SURE by logging in remotely and working on a remote virtual computer desktop. Researchers who want to use the SURE must apply to SURE, and have approval from the AIHW Ethics Committee. All users must complete training on privacy, ethics, information security and statistical disclosure control and sign a deed that sets out the terms and conditions for using SURE. The training is provided online through the SURE website <https://www.sure.org.au/> and ensures that:

- Researchers have an awareness of legal and ethical issues and considerations relating to information security and data management relevant to undertaking linked data research.

- Researchers are given detailed guidance on the features of SURE, including accessing and using the system.

The modules presented in the training program are:

- **Data linkage research in Australia**: an overview of the data linkage process and the development of linked data research in Australia
- **Responsibilities of a researcher**: an overview of the legal and ethical considerations of researchers when conducting linked data research
- **Information security**: the current approach to information security in linked data research and a description of the security model of SURE
- **Statistical disclosure analysis**: an introduction to principles and methods of statistical disclosure analysis and statistical disclosure control.

When a study is approved by both the AIHW Ethics Committee and the NECD RDS AC, the researcher will have to establish a project workspace in SURE. Instructions on how to do this are available on the SURE website at <https://www.sure.org.au/accessing-sure/establishing-a-project-in-sure>.

Each set of study data inside SURE is confined within its own security perimeters, so there is no possibility of data exchange between studies. No data are stored on a researcher's local computer or in their institutions' computing environment, and complete audit trails of all information entering or leaving SURE are maintained. The SURE facility is protected by multiple firewalls and data are stored centrally on servers located in a high-level secure data centre with strict access controls and 24-hour security surveillance.

Costs to access SURE are calculated based on project duration, user numbers and computing resources required by users for the project. Researchers are advised to build the costs of accessing data through SURE into their research applications. Information about access costs can be found at <https://www.sure.org.au/accessing-sure/access-charges-new>.

The AIHW data laboratory

There is a data laboratory option available to researchers who do not want to use the SURE facility. This involves a researcher working on site at the AIHW, within a secure environment that is monitored by staff from the DISC. Sessions to use the AIHW's data laboratory must be booked in advance, and a cost may be incurred for this.

To use this option, researchers must contact the AIHW DISC (see AIHW contact details below).

DISC mobile data laboratory

Where approved by the data custodians, de-identified, confidentialised data may also be migrated to the DISC mobile data laboratory.

The DISC mobile data laboratory is a secure laptop with all contents encrypted using very strong encryption. Two-factor authentication is required to access the mobile data laboratory. All input and outputs, including floppy disc drives, optical disc drives, input and output ports (including serial, parallel, USB, video), except for keyboard, mouse and monitor are disabled for all user accounts excluding those persons with a DISC system manager role.

The mobile data laboratory can be transported by DISC to staff to an offsite location for use by approved researchers.

Researchers must contact the AIHW DISC (see AIHW contact details below) to organise access to the mobile data laboratory, and a cost may be incurred for this.

Other options for access to de-identified data

The AIHW Act provides for other ways to access de-identified, confidentialised data from the NECD RDS. Providing that necessary approvals are met, in line with conditions specified by the AIHW Ethics Committee and data custodians, the AIHW may release data in aggregated forms, or if approvals exist, as URL data that have been further confidentialised as required. Each of these options would incur a cost, which will offset the resources used to provide the data.

Researchers wanting to obtain NECD RDS data in this way should initially contact the Maternal Health, Children and Families Unit (see AIHW contact details below) to discuss the request and organise a quote.

Specialised data requests

The AIHW can make statistics from the NECD RDS available through external projects, and usually provides results from such projects as tabulated information. These requests are subject to confidentialisation of the tabulated data

Requests for data are charged on a cost-recovery basis. The length of time it takes to complete a request or linkage depends on a number of factors such as the complexity of the work, competing projects, the ethics application process, additional approvals (for example from the NECD RDS AC) and correspondence with data custodians where necessary.

E.3 AIHW contact details

Data Integration Services Centre

The Manager
Data Integration Services Centre
Email: disc@aihw.gov.au

AIHW Data Linkage Unit

Head, Data Linkage Unit
Email: linkage@aihw.gov.au

Custodian of the NECD RDS

Head, Maternal Health, Children and Families Unit
Email: conan.liu@aihw.gov.au

AIHW Ethics Committee

AIHW Ethics Manager
Email: ethicssec@aihw.gov.au

Glossary

Concordance map: Shows the link between specific project IDs in a collection. Data custodians or AIHW can apply to merge content data using the map as a reference.

Confidentialised data: De-identified data which have undergone a further step, such as combining categories or limiting the number of variables in a data set, to reduce the risk of identification.

Content data: The data that is subject specific and not identifying information. For example, preschool content data may include the number of hours that child X attended preschool.

Data linkage/integration: The process of bringing together information from 2 or more sources about people, places and events in a way that protects individual privacy and confidentiality. The NECD RDS will be the outcome of data being linked from a number of different data sets such as births, perinatal, preschool and long day care.

Data linkage separation principle: Separates the data linkage step from the data merging step to maintain the privacy of the individual whose information is linked and merged.

De-identified data: Data from which identifying information such as name and address has been removed.

Identifying information: Data from which a person’s identity may reasonably be ascertained. They may include the full name, date of birth or address of a child, but do not include any information about their preschool attendance. These data can be used to construct concordance maps and allow linkage across multiple data sets. Often these data are common across data sets, however the degree of completeness can differ. For example, some administrative data sets do not collect full names, therefore partial names are used in linkage.

Longitudinal data set: Collects information about a group of people (a cohort) over a period of time.

Project identifier: Also known as a project ID or project specific ID, it is a unique identifier that data custodians or managers apply to each record in a data collection.

Population Health Research Network (PHRN) data linkage units: The PHRN was established to build a nationwide data linkage infrastructure capable of securely and safely managing health information from around Australia. The PHRN is a national network comprising a Program Office located in Perth, Western Australia, a Centre for Data Linkage located at Curtin University in Western Australia, a remote access laboratory located at the Sax Institute in New South Wales and a network of project participants and data linkage units located in each Australian state/territory. The AIHW is a member of the PHRN.

Record linkage: See **Data linkage/integration**.

SURE: A secure computing environment developed as part of the Population Health Research Network (PHRN) and managed by the Sax Institute.

References

ABS (Australian Bureau of Statistics) 2012. What is the Census? Viewed 1 July 2013, <http://www.abs.gov.au/websitedbs/censushome.nsf/home/what?opendocument&navpos=110>.

AEDC (Australian Early Development Census) 2013. AEDI Data protocol 2012–14. Viewed 9 September 2013, <http://www.aedidata.com.au/docs/default-source/public-library/aedi-data-protocol-2012-2014.pdf?sfvrsn=2>.

AIFS (Australian Institute of Family Studies) 2013. Growing up in Australia: The Longitudinal Study of Australian Children. Viewed 28 February 2013, <http://www.growingupinaustralia.gov.au/index.html>.

AIHW (Australian Institute of Health and Welfare) 2011a. Headline indicators for children's health, development and wellbeing, 2011. Cat. no. PHE 144. Canberra: AIHW.

AIHW 2011b. 2010 Australian national infant feeding survey: indicator results. Cat. no. PHE 156. Canberra: AIHW.

AIHW 2012. Australia’s Health 2012. Australia’s health series no. 13. Cat. no. AUS 156. Canberra: AIHW.

AIHW 2013. Australia's welfare 2013. Australia's welfare no. 11. Cat. no. AUS 174. Canberra: AIHW.

Bayley N 1965. Comparisons of mental and motor test scores for ages 1–15 months by sex, birth order, race, geographical location, and education of parents. Child development: 379–411.

Berrueta-Clement JR 1984. Changed Lives: The Effects of the Perry Preschool Program on Youths through Age 19. Monographs of the High/Scope Educational Research Foundation, No. 8. ERIC.

Bowers A & Strelitz J 2012. An Equal Start Improving Outcomes in Children’s Centres: An Evidence Review. Viewed 10 September 2012, <http://www.instituteofhealthequity.org/>.

COAG (Council of Australian Governments) 2009. Investing in the Early Years: A National Early Childhood Development Strategy. Canberra: COAG. Viewed 11 September 2012, <http://www.coag.gov.au/sites/default/files/national_ECD_strategy.pdf>.

Cunha F & Heckman J 2006. Investing in Our Young People, Unpublished manuscript, University of Chicago, Department of Economics.

Dodge KA, Pettit GS & Bates JE 1994. Socialization mediators of the relation between socioeconomic status and child conduct problems. Child Development 65:649–65.

DSS (Department of Social Services) 2014. Footprints in Time—The Longitudinal Study of Indigenous Children (LSIC). Viewed 07 August 2014, <http://www.dss.gov.au/about-the-department/publications-articles/research-publications/longitudinal-data-initiatives/footprints-in-time-the-longitudinal-study-of-indigenous-children-lsic>.

Elliot A 2006. Early childhood education: pathways to quality and equity for all children. Australian Education Review no. 50. Melbourne: Australian Council for Educational Research.

Glascoe F & Leew S 2010. Parenting behaviors, perceptions, and psychosocial risk: impacts on young children's development. Pediatrics 125:313–9.

Halle T, Forry N, Hair E, Perper K, Wandner L, Wessek J, Vick J, 2009. Disparities in early learning and development: lessons from the Early Childhood Longitudinal Study—Birth Cohort (ECLS-B). Washington DC, Child Trends.

Harrison, LJ & Ungerer, JA 1997. Child care predictors of infant mother attachment security at 12 months. Early Child Development and Care 137: 31–46.

Kalil A, Ryan R & Corey M 2012. Diverging destinies: maternal education and the developmental gradient in time with children. Demography 49:1361–83.

Kelman CW, Bass AJ & Holman CDJ 2002. Research use of linked health data—a best practice protocol. Australian and New Zealand journal of public health 26(3):251–5.

Key A, Ferguson M, Molfese D, Peach K, Lehman C & Molfese V 2007. Smoking during pregnancy affects speech-processing ability in newborn infants. Environmental Health perspectives 115:623.

MCEECDYA (Ministerial Council for Education, Early Childhood Development and Youth Affairs) 2010. National Information Agreement on Early Childhood Education and Care. Viewed 8 May 2013, <http://www.mceecdya.edu.au/verve/_resources/FINAL_National_Information_Agreement_on_Early_Childhood_Education_and_Care_-_Signed.pdf>.

Medicare 2012. Statistics. Viewed 1 March 2013, <http://www.medicareaustralia.gov.au/about/stats/index.jsp>.

NCRIS (National Collaborative Research Infrastructure Strategy) 2006. National Collaborative Research Infrastructure Strategy Strategic Roadmap. Canberra: NCRIS.

NHMRC (National Health and Medical Research Council) 2013a. Human research ethics. Viewed 10 April 2014, <http://nhmrc.gov.au/health-ethics/human-research-ethics>.

NHMRC 2013b. Overview of Human Research Ethics Committees (HRECs). Viewed 4 April 2014, <http://nhmrc.gov.au/health-ethics/human-research-ethics-committees-hrecs/overview-human-research-ethics-committees-hrecs>.

NHMRC 2013c. Welcome to NEAF. Viewed 1 March 2013, <https://www.neaf.gov.au/>.

NHMRC, ARC (Australian Research Council) & AVCC (Australian Vice-Chancellors' Committee) 2007. National Statement on Ethical Conduct in Human Research 2007 (Updated 2013). Commonwealth of Australia, Canberra.

NICHD ECRN (National Institute of Child Health and Human Development Early Childhood Research Network) 2000. The relation of child care to cognitive and language development. Child Development 71: 960–980.

NSS (National Statistical Service) 2010a. High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes. Viewed 23 November 2012, <http://www.nss.gov.au/nss/home.nsf/NSS/7AFDD165E21F34FDCA2577E400195826?opendocument>.

NSS 2010b. Governance and Institutional Arrangements. 2. Integrating Authorities. Viewed 24 April 2014, <http://nss.gov.au/nss/home.nsf/NSS/0E887A88A9224F8BCA2577F20016FE5D?opendocument>.

NSS 2013. Statistical integration involving Commonwealth data: Best practice guidelines. Viewed 13 February 2013, <http://www.nss.gov.au/nss/home.nsf/NSS/085988C74EAFAB58CA2577F20017118A?opendocument>.

Petersen C & Bell M 1996. Children's memory for traumatic injury. Child development 67: 3045–70.

Sylva K, Melhuish E, Sammons P, Siraj-Blatchford I & Taggart B 2004. The Effective Provision of Pre-School Education (EPPE) project: findings from pre-school to end of Key Stage 1. London: Institute of Education, University of London.

Sylva K, Melhuish E, Sammons P, Siraj-Blatchford I & Taggart B 2011. Pre-school quality and educational outcomes at age 11: low quality has little benefit. Journal of Early Childhood Research 9(2):109–24.

WHO (World Health Organization) 2013. Early child development. Viewed 1 March 2013, <http://www.who.int/mediacentre/factsheets/fs332/en/index.html>.

Wise S, Ungerer J & Sanson A 2002. Childcare Policy to Promote Child Wellbeing. Australian Economic Review 35(2):180–7.

List of tables

Table 2.1: Proposed data collections to be linked and merged for the NECD RDS 11

Table 2.2: Data sets considered for inclusion in NECD RDS....12

Table 3.1: Stakeholders for the NECD RDS and their role in the process 19

Table 3.2: Other groups involved with the NECD RDS and their roles 20

Table 4.1: Summary of challenges encountered during development of the NECD RDS 21

Table 4.2: Summary of linkage capabilities and examples of linked data in states and territories, as at December 2013....22

Table 4.3: Requirements to obtain access to Commonwealth data for the NECD RDS....26

Table 4.4: Requirements to obtain access to state and territory data for inclusion in the NECD RDS....28

Table 4.5: Types of application forms required when obtaining permission to use health (perinatal) data in the NECD RDS, by jurisdiction, as at September 2013....30

Table 4.6: Progress towards obtaining data for the NECD RDS, as at October 2013....32

Table 4.7: Steps for obtaining data from birth registration, perinatal and preschool data collections....35

Table 4.8: Proposed supply and updating of data for the NECD RDS....36

Table C.1: Data items from the Perinatal Data Collection proposed for inclusion in the NECD RDS, 2006 to present 47

Table C.2: Data items from the Early Childhood Education and Care Collection proposed for inclusion in the NECD RDS, 2011 to present....48

Table C.3: Data items from the Child Care Management System proposed for inclusion in the NECD RDS, 2011 to present....49

List of figures

Figure 2.1: Establishing the NECD RDS—proposed flow of data from data custodians to the AIHW.....15

Figure B.1: Core NECD RDS data sources and at what ages they are collected.....44

Figure D.1: Proposed method for merging records for the NECD RDS.....51

List of boxes

Box 1.1: Policy frameworks and initiatives 2
Box 3.1: What is the National Ethics Application Form? 17

This information paper outlines the processes undertaken towards establishing the National Early Childhood Development Researchable Data Set. This data set aims to link health and education data, using both jurisdictional and national data sources, which would provide a valuable resource to researchers and policy-makers. The paper documents the privacy, legislative and data custodianship and supply hurdles encountered during the initial stages of establishing this data set. The paper concludes with a pragmatic approach for the next steps and way forward.