

Evaluation of a place-based collective impact initiative through cross-sectoral data linkage

Author

Homel, J, Homel, R, McGee, TR, Zardo, P, Branch, S, Freiberg, K, Manning, M, Wong, G

Published

2020

Journal Title

Australian Journal of Social Issues

Version

Accepted Manuscript (AM)

DOI

[10.1002/ajs4.147](https://doi.org/10.1002/ajs4.147)

Rights statement

© 2020 © 2020 Australian Social Policy Association. This is the peer reviewed version of the following article: Evaluation of a place#based collective impact initiative through cross#sectoral data linkage, Australian Journal of Social Issues, 2020, which has been published in final form at <https://doi.org/10.1002/ajs4.147>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving (<http://olabout.wiley.com/WileyCDA/Section/id-828039.html>)

Downloaded from

<http://hdl.handle.net/10072/400545>

Griffith Research Online

<https://research-repository.griffith.edu.au>

Evaluation of a place-based collective impact initiative through cross-sectoral data linkage

Jacqueline Homel¹, Ross Homel¹, Tara Renae McGee¹, Pauline Zardo², Sara Branch,¹
Kate Freiberg,¹ Matthew Manning³ & Gabriel Wong³

¹ Griffith University

² QLD Department of Youth Justice

³ Australian National University

Dr Jacqueline Homel is a researcher who focuses on individual, family, peer and neighbourhood determinants of successful development during childhood, adolescence, and young adulthood. Of particular interest are early-in-life prevention of criminal offending, the neighbourhood and developmental contexts of substance use in adolescence and young adulthood, and statistical modelling of complex longitudinal processes. Dr Homel is a lecturer in the School of Criminology and Criminal Justice at Griffith University.

Dr Ross Homel, AO is Foundation Professor of Criminology and Criminal Justice at Griffith. He is passionate about the prevention of crime, violence and injuries, and has published approximately 200 books, articles, and high impact government reports. He has, with colleagues, won many awards for his research on the prevention of crime and the promotion of positive development and wellbeing for children and young people in socially disadvantaged communities through family support and family-school-community partnerships.

Dr Tara McGee is an Associate Professor in Criminology at Griffith University. She has worked predominantly with the Mater University Study of Pregnancy, the CREATE project on which is paper is based, and the Cambridge Study in Delinquent Development, to examine the development of antisocial behaviour and offending across the life-course. Dr McGee is the founding co-editor of the Journal of Developmental Life-Course Criminology and the current President of the Australian and New Zealand Society of Criminology.

Dr Pauline Zardo has researched and worked in the area of research engagement, translation and impact, with a focus on strategy, policy and program development, evaluation and impact assessment, in both the higher education and public sectors. Pauline has moved between government and higher education sector over the course of her career, working in both academic and professional roles in the higher education sector. Pauline completed a PhD on translation of research to policy and practice at Monash University in 2013.

Dr Sara Branch has a background in Organisational Psychology and an interest in Workplace Conflict (focus on Workplace Bullying) and Change Management with a broad interest in Social Justice. Her current work focuses on applying organisational theory to enhance processes for collective action within communities and organisations. As part of the CREATE project, Sara has guided the development of the Collective Change Facilitator role.

Dr Kate Freiberg is a Principal Research Fellow at the Griffith Criminology Institute. Her work combines developmental psychology and prevention science. She maintains a strong focus on the development of accessible resources and methodologies to support those who

strive to achieve better outcomes for children and families living in challenging environments.

Dr Matthew Manning is a criminological economist and future crimes scholar. As a future crimes scholar he is committed to advancing knowledge about: (1) why individuals commit future crimes; (2) the processes that can be adopted to make committing crimes more difficult and less rewarding; and (3) the tools that assist government and industry in making strategic economic decisions that create efficiency and enhance return on investment.

Dr Gabriel Wong is a quantitative researcher with experience in evaluation research on crime prevention and health initiatives, cost-benefit analysis, and multi-criteria decision modelling. His research has implications to: (i) the development of targeted institutional and criminal justice policies aimed at mitigating the adverse harms and consequences associated with crime and illicit drug involvement; (ii) the deployment of finite resources that improve the efficiency of policing activities (e.g. arrest, detection); and (iii) understanding the costs and benefits of prevention interventions.

Acknowledgements: We gratefully acknowledge the financial support of the Australian Research Council through LP170100480, *Creating pathways to child wellbeing in disadvantaged communities*, as well as the funding and extensive in-kind support provided by partner organisations: Australian Government Department of Social Services; InVision Media Pty Ltd; Qld Department of Education; Qld Department of Child Safety, Youth & Women; NSW Department of Education; NSW Department of Communities & Justice; Australian Primary Principals Association; Logan Child Friendly Community Ltd Board; The Benevolent Society; The Smith Family; Mission Australia; The Salvation Army. We are especially grateful to Tenniel Guiver for his expertise and feedback on data linkage and integration, especially in regard to Commonwealth Government capabilities and constraints. We also wish to thank Michael Hogan and two anonymous referees for very helpful comments and suggestions on earlier drafts of this paper.

Evaluation of a place-based collective impact initiative through cross-sectoral data linkage

Abstract

Australian governments fund many place-based collective impact initiatives like Communities for Children (CfC); states fund family support services to keep at-risk children out of the child protection system; and schools implement many extracurricular programs. Do these services have a beneficial, cost-effective collective impact on child well-being? This paper describes a proof-of-concept attempt to address this question by linking for one CfC community individual-level data on 5-12 year-olds' participation in programs delivered through these three sectors with child outcomes. This exercise was unsuccessful despite the investigators' prior experience; advice from a data linkage expert and our data custodian partners; five ARC reviews; partners' good will; and ethical safeguards including written, informed parent/carer consent. Obstacles encountered included a lack of data of sufficient quality on children and families' participation in services, lack of data on children's outcomes, and prohibitive costs of linkages within government. We offer for others three key lessons: (1) make assumptions explicit; (2) talk to technical experts in data custodian organisations early in the planning process; and (3) undertake, if possible, an initial scoping exercise. We conclude that despite recent legislative and policy reforms many obstacles we encountered will persist in the absence of a national child wellbeing strategy.

Keywords: child wellbeing, place-based initiatives, collective impact, family support services, data linkage and integration

1 | INTRODUCTION

Capacity-building initiatives in socially disadvantaged areas mounted by community and government agencies aspiring to achieve a collective impact on human welfare are a booming industry in Australia. Since at least the end of the Second World War Australian government policies for national economic and social development have been informed by an understanding that success depends at least in part on incorporating a regional and community perspective. However, recent years have seen a proliferation of what have come to be termed ‘place-based’ initiatives which take as their primary mission the improvement of social conditions and population wellbeing in a defined, but often quite large, geographical area through the combined efforts of mostly non-government and local organisations funded by the Commonwealth or state governments.

A prominent place-based program is *Communities for Children (CfC)*, a program funded and administered by the Australian Department of Social Services. CfC has been operating successfully for 15 years in 52 disadvantaged communities (Edwards et al., 2014). The program involves partnerships of community agencies funded through a variety of non-government organisations (NGO) called Facilitating Partners. Much has been learned by practitioners, policy people, and researchers through the array of imaginative CfC initiatives that have been mounted over the years. Perhaps because CfC pre-dated the collective impact movement and is not usually viewed through this prism its potential for informing the field is not as appreciated as it should be, despite the fact that the program is in effect striving for a collective impact on child wellbeing.

Sadly, the failure to build an empirically well-grounded body of knowledge from years of practice experience is a feature of many government programs in Australia. Perhaps the most significant single aspect of this failure is that there has been insufficient investment in routine data collection to support the rigorous measurement, planning, and evaluation

needed to demonstrate collective impact. For example, the most recent published evaluation of CfC (Edwards et al., 2014) was forced to rely on specially commissioned surveys of parents in CfC and comparison communities to assess possible impacts because until that time services had not been funded to record program activities or to measure outcomes at the individual child or family level. The reality is that the vast bulk of what government programs may have achieved over the past few decades through community agencies and schools remains, especially from a preventive perspective, unexamined, unmeasured, and unevaluated.

The aims of this paper are (1) to present a research design intended to overcome some of the limitations in the data infrastructure of CfC and allow evaluation of the collective impact of place-based initiatives using data linkage and integration at the child level, and (2) to discuss the practical problems that arose during the implementation of this design.

Although we were not able to implement our ambitious design for a number of reasons, our failure was instructive. This paper sheds light on some of the data deficiencies, technical barriers, and privacy considerations that need to be navigated or rectified to enable the construction of big data to address important policy issues related to place-based initiatives.

2 | WHAT IS NEEDED TO DEMONSTRATE COLLECTIVE IMPACT FOR PLACE-BASED INITIATIVES?

The collective impact movement takes its name from an influential paper by Kania and Kramer in the *Stanford Social Innovation Review*. According to Kania and Kramer (2011), the defining features of a collective impact approach comprise:

. . . long-term commitments by a group of important actors from different sectors to a common agenda for solving a specific social problem. Their actions are supported by a shared measurement system, mutually reinforcing activities, and ongoing communications, and are staffed by an independent backbone organization. (p. 39)

The commitment to “a shared measurement system” in the collective impact rubric is of particular importance for our purposes in this paper. Shared measurement implies at the very least the sharing of data from different organisations and sectors, data that can be integrated or combined in some way and which throws light on both project inputs and project outputs. By ‘project inputs’ we mean data on what was done with individuals or families, by which organisation, for what purposes, using what methodologies, and for how long. ‘Project outputs’ require data on immediate and mid-term effects defined in a program logic or theory of change, as well as long term project effects on outcomes for children, families, or the whole community.

If we follow the logic of collective *impact* we need data that can be used to construct valid measures of outcomes that are agreed by the whole partnership or collective, and their funders, to reflect the fundamental purposes of the initiative. If improving child or youth wellbeing is an agreed goal, which it is for the great majority of prominent Australian collective impact projects described above, then valid, reliable, and believable measures of wellbeing are absolutely critical. These measures should preferably be based on data drawn from a number of sources (schools, families, the children themselves) and should be able to be easily connected with the data on project inputs and outputs.

However, the logic of collective impact requires yet more: if the project collective wishes to generate plausible evidence that all the money they have spent and all their heroic efforts have indeed had their intended impact, then attention must be paid to the *counterfactual*. What would have happened to children and families if the project had never happened and business as usual had prevailed? What happened in other communities without collective impact funding, or in parts of target communities where people were not exposed to project activities?

Thus, to demonstrate collective impact for place-based initiatives, data are needed about inputs (e.g. child participation in programs and services), child outcomes (e.g. wellbeing), and the costs of services. In addition, a control group is required to fulfil all the conditions of a high quality quasi-experimental or experimental design. These conditions would increase understanding of what interventions are most effective for maximising outcomes, are most cost-effective, and the optimal times and durations for these interventions.

These questions, including cost-effectiveness and cost benefit, were precisely those addressed by our data linkage proposal.

3 | CONTEXT FOR THE RESEARCH: THE *CREATE* PROJECT.

The research design described in this paper was embedded within a much larger project focused on improving child wellbeing in disadvantaged communities: *Creating Pathways to Child Wellbeing in Disadvantaged Communities* (CREATE). The project was funded by the Australian Research Council and a wide range of partner organisations (Homel, Freiberg & Branch, 2015). Adopting the 22 CfC sites in New South Wales and Queensland as its service delivery framework, the program sought to *go wide* by fostering the formation of coalitions of schools and community agencies (moving from coordination to collaboration to achieve effective engagement with families and children), and to *go deep* by putting collective impact aspirations onto a firm scientific footing. Findings and methods from the interdisciplinary field of prevention science (e.g., Van Ryzin et al., 2018) were introduced through the use of field-tested electronic measures and resources, as well as through the services of an emerging profession, that of *Collective Change Facilitators* (Branch et al., 2019; Homel et al., 2019). These human and electronic resources were underpinned by six practice principles encapsulated in the acronym CREATE (Collaborative; Relationships-driven; Early in the pathway; Accountable; Training-focused; Evidence-

based), and were designed to be implemented through an iterative cycle of collaborative learning and changed practices: Deciding Together, Planning Together, Doing Together, and Reviewing Together. The electronic and human resources, organised according to the CREATE principles and the change cycle, were termed the *Griffith Prevention Translation and Support System* (G-PTSS) to distinguish it from the infrastructure and supports provided by the Department of Social Services (DSS)ⁱ.

4 | THE DATA LINKAGE AND INTEGRATION DESIGN

The data linkage and integration research design was a proof-of-concept project in one of the Queensland CfC sites involved in the CREATE project. The design reflected both the logic of what is needed to demonstrate collective impact in terms of inputs, outcomes and costs, as well as the directions recommended in recent government reports on data integration in the child and family field (Australian Government Department of Education and Training, 2018). An amount of \$50,000 was reserved for the data linkage exercise in the CREATE project budget but was not expended for reasons that are explained in Section 5. The main resource expenditure was the time of two Chief Investigators.

The focus was children aged 5 to 12 years old. Subject to written informed parental consent, we sought to relate data on a range of outcomes for children (including their self-reported wellbeing) to patterns and costs of participation in programs or services designed to improve their wellbeing. These services or programs could be delivered by schools, or by community agencies funded by CfC, or by the state government. Child participation could be direct, or indirect through parents or other family members (but child outcomes had to be available). Several of the outcome measures were intended to be derived directly from children or teachers at the school, capitalising on the fact that primary schools constitute a near-universal platform for data collection.

Linking multiple datasets (schools/education department records; state-funded services; and CfC coalition-level data) over time would produce a comprehensive (although not complete) picture of children and families in CfC communities and the extent and nature of service provision across agencies. In addition to quantifying the collective impact of programs and services this information could be used to identify where services are being unnecessarily duplicated or indeed where there are children in need who are not receiving adequate services. The linkage process was intended to be repeated at least once in the life of the project so that changes in patterns of service participation could be related to changes in outcomes. Ideally the linkage would be repeated several times for a cohort of children, making the data truly longitudinal.

Given our interest in new ways of evaluating the collective impact of place-based interventions, our specific research questions were:

1. Do CfC, school-based, and state-funded services have a collective beneficial effect on children?
2. Do benefits outweigh costs?
3. What are the effects of different program mixes or types, their duration, and their intensity, on a range of outcomes for various kinds of children?
4. What evidence is there for the separate impact on children of school-based, CfC, or state funded family support services?

In order to be able to answer these questions it is necessary to be able to construct control groups of children not receiving any services beyond normal school or community activities. Provided enough parents give consent for school-based data collection and for linkage to other datasets, it would be relatively easy to select samples for a range of control groups matched on key factors such as age, gender, race, and classroom behaviour. This methodology was pioneered successfully in one Australian community through the *Pathways*

to *Prevention Project* (Homel, Freiberg, Branch & Le, 2015), which led to the construction of a longitudinal database of nearly 5,000 children from seven schools over a 10-year period. In particular, analyses of these data have allowed methods of matching intervention and non-intervention children to be explored and refined.

4.1 | Sources of data

Details of data measuring inputs and outcomes that were envisaged to be included in the child-centric linked database are summarised in Table 1. For inputs, the project sought to include *participation data* (nature of program, program duration, frequency of contact) from three sources:

1. *Communities for Children* services delivered through the local partnership of service providers, coordinated by the Facilitating Partner. The DSS Data Exchange System (DEX) is able to provide partnership-level service delivery data (with some complications described later in this paper); we did not seek to evaluate the programs offered by individual service providers.
2. *State government-funded family support services* designed to help prevent children at risk from falling into the child protection system. Data was believed to be available from the NGOs delivering these services via the Queensland Department of Child Safety, Youth and Women.
3. *Schools* implement a range of programs for children or families, directed at behaviour, wellbeing, or poverty-related issues. Little is known about these programs since generally neither schools nor departments of education record or publish information about them.

Measures of child outcomes (ages 5-12) were also designed to be collected through schools and included social-emotional wellbeing reported by children and classroom behaviour reported by teacher. To measure children's social and emotional wellbeing we proposed using [*Rumble's Quest*](#), which is a new 57-item measure which takes the form of a

computer/tablet game. The child playing the game selects an avatar and has an adventure in a magical land with their pet creature. Through the conversation between the creature and the child and the tasks that must be attempted, the software generates valid and reliable measures for children aged 6-12 years of their overall social-emotional wellbeing; attachment to school; emotional and behavioural self-regulation; social confidence with respect to peer relationships; supportive environments and relationships with caring adults; and five measures of executive function (Day et al., 2019). The Queensland Department of Education would provide NAPLAN numeracy and literacy scores; and school attendance and suspensions/exclusions.

Given our interest in economic analyses such as cost-benefit and cost effectiveness, we also planned to ask participating schools and agencies to record cost data about their programs using another tool developed for the G-PTSS, the *Economic Support and Reporting Tool* (ESRT; Manning et al., 2018). Since these cost data pertain to programs not children, they would not have been part of the child-centric linked database, hence they are not listed in Table 1. We planned to use the ESRT after the data linkage was implemented at least once to undertake a cost-benefit analysis of the collective effects of all the linked programs, and also to guide schools and agencies to undertake self-administered economic evaluations of their own programs. We discuss this tool in more detail in our exploration of future directions in the last section of this paper.

4.2 | Minimum viable data and sample size

Although we envisaged the integration of all the data shown in Table 1, we considered that collective impact could be demonstrated with a subset of the data. This would include at least two measure of inputs (e.g. data on participation in CfC services and data on participation in school-based initiatives) and at least one child outcome (e.g. social-emotional wellbeing reported by children or classroom behaviour reported by teacher).

A sample size was planned that was large enough to test out the data linkage and integration design. To achieve this, we aimed to seek consent from parents of all children aged 5-12 in the government primary schools in the community. Even allowing for a consent rate of 30%, this would have yielded a sample size of about 6,000 children. At a minimum, we aimed for at least 500 children in both the treatment and comparison samples.

4.3 | Proposed data linkage procedure

The procedure we originally proposed for the data linkage is shown in Figure 1. In Step 1 of the procedure, custodians for each of the data sources measuring inputs and child outcomes (see Table 1) would send a data integrating authority a full child-level database including identifiable dataⁱⁱ. These identifiable data would include name, date of birth, gender, and school attended for data relating to schools and education (e.g. participation in school-based initiatives; suspensions and exclusions). Additional identifiable data, such as First Peoples status, would be subject to negotiation. In Step 2, the data integration authority would use the identifiable data to link children across datasets. Once linked, the identifying information would be removed. In Step 3, the data integration authority would supply the full integrated, de-identified dataset to the research team. Finally, in Step 4, value would be added for data suppliers by the project research team (or possibly by the data integration authority) through data analyses that would, for example relate patterns of participation in CfC services to child outcomes. The value of such analyses could be enhanced if in addition enough local schools had supplied data on all their children to allow a matched control group of children to be selected, as outlined earlier.

It is important to note that this procedure was an initial proposal only, which we expected to be considerably refined. However, as outlined in Section 5 below, the data linkage exercise did not proceed far enough for this to happen.

5 | BARRIERS TO IMPLEMENTATION

As we have intimated, unfortunately our plans for data linkage in the demonstration community did not eventuate. There were multiple issues that, individually, were not fatal for the exercise. We might have been able to pursue linkage of a smaller number of datasets. However, collectively the problems meant that the data linkage exercise was not feasible. There are lessons to be learned from each point of failure.

5.1 | Data on state government-funded NGO services was not accessible

In our discussions with the Queensland Department of Child Safety, Youth and Womenⁱⁱⁱ prior to submitting the grant application there were positive indications that some useful data for individual data linkage could be provided. However, on commencing the project we found that the NGO providers were not required to provide data on individual program participants to the department, only aggregated data for the purposes of contract management. This lacuna reflects both constraints imposed by government and the sensitive nature of service delivery to vulnerable families. The reality is that the department was never funded to provide a shared client information platform or data sharing facility, especially broken down to individual family members, and at the time there were a number of legislative barriers that any endeavour along these lines would have had to navigate.

A workaround would have been to approach the providers themselves, but we received advice that it was very unlikely that providers were keeping client lists where children and families would be identified, and also unlikely that information was collected on frequency and duration of participation. The fundamental problem is that any form of effective service delivery to vulnerable children and their families must be founded on trust, which is not easily achieved but can easily be lost. NGO services generally take the view, based on long experience, that the provision to government (or researchers funded by government) of even apparently innocuous individual-level data such as age or gender can jeopardise these trusting relationships which they put huge efforts into fostering with

vulnerable families. These sensitivities are particularly acute for First Nations families, where the Stolen Generations is not just a memory but a present reality (Funston & Herring, 2016; Homel et al., 1999).

We know from our own experience in working with a national community agency in the Pathways to Prevention project that there are ways in which these difficulties can be overcome using strength-based measures and co-creation processes (e.g., Freiberg et al., 2014). We discuss these approaches in more detail later in this paper in the context of exploring how data linkage strategies might be made more acceptable to families and community agencies, but for present purposes it is sufficient to state that we realised that we could not pursue our data linkage idea using data from agencies funded by the Queensland government. However, we remained hopeful about the rest of the data linkage exercise.

5.2 | Difficulties involving schools in data collection

A few schools in the demonstration community enthusiastically embraced Rumble's Quest as an innovative and school-friendly child wellbeing measure as soon as we made it available. Generally, however uptake was slow because we simultaneously asked schools to participate in a number of other CREATE project activities that might have appeared daunting. This was the case even though we had already abandoned plans to ask classroom teachers to assess classroom behaviour data using the Strengths and Difficulties Questionnaire (SDQ) after feedback from our departmental partners. At the time of writing, 30% of the primary schools in the demonstration community had registered for Rumble's Quest, and only 13% had implemented it for any students. Moreover, in some schools it was only implemented for a small number of students, possibly those about whom the school was particularly concerned.

Given the many reporting requirements on schools and principals from different quarters, it is not surprising that a non-required initiative from an outside organisation should

fall by the wayside in many busy schools. These difficulties are common in undertaking research in schools and are certainly not unique to data linkage projects. On balance, despite the slow uptake we believed that we would over time receive enough Rumble's Quest data for some useful data linkage to take place, subject to the consent of the Queensland Department of Education, the schools, and parents. We were also optimistic that we could obtain NAPLAN data and data on attendance, suspensions and exclusions from the Queensland Department of Education (see Table 1). Thus, we were reasonably confident that we would have some outcome data at the individual child level, and we therefore started to progress our plans to link these data with information about federally funded service participation via the Commonwealth Data Exchange System.

5.3 | Data collected on program participation in the DSS Data Exchange did not identify children who are beneficiaries of programs/services

Since July 2014, organisations that are funded by grants from the Commonwealth Department of Social Services to work with families and children must use the Data Exchange (DEX) to report service use, performance, and client outcomes to government (Department of Social Services, 2019). The Data Exchange provides a range of secure, web-based tools for organisations to report data. Each organisation must report a small set of mandatory items (the *priority requirements*) and can elect to report a larger number of items (the *partnership approach*). The mandatory priority items are generally client-level unit record items. The first set includes client identifying information (e.g. name, date of birth, address), consent to collection and storage of personal information, and consent to participate in research and evaluation (but not, notably, consent to link to other data sources). A statistical linkage key is created from several of these items. The second set of mandatory items includes details of the service, such as service type, date of each service contact, and outlet.

In early 2018 we started to discuss arrangements for data linkage involving DEX with DSS. A major problem became apparent: clients whose data are uploaded to DEX are usually adults. For example, a parent who accesses family support services is usually the client in DEX. Details of that parent's children are not necessarily collected, even though the primary focus of the CfC program is on prevention and early intervention for children aged 0-12. Children may be entered as clients, but there is no item to capture relationships between clients (which clients are the parents, offspring, and siblings). Thus, there was no way of identifying children who were the intended beneficiaries of Commonwealth-funded services, and therefore no way of linking their outcome data to information about service participation. Another consequence of this was that, even at the community level, it would not be possible to quantify the number of children in a community who were the direct or indirect beneficiaries of funded services.

At this point we feared that the data linkage design might be completely impossible. However, our DSS partners suggested that we might explore the use of DOMINO to identify relationships between parents, carers and children.

5.4 | Cost of using DOMINO to link children and families

[DOMINO](#) stands for Data Over Multiple INdividual Occurrences. DOMINO contains administrative data generated by individuals as they interact with DSS payments and programs over the course of their life. It is not a single dataset but a DataMart consisting of multiple datasets made to be linked together through a centralised linkage key. These datasets incorporate data on both payments and life events, as well as demographic details. Importantly for our project, it would be possible to link a DEX client (via a statistical linkage key, SLK) to their records in DOMINO. That client would then become a reference person within DOMINO, and their relationships with other family members identified in the

administrative data could be mapped. The quality of this relationship mapping is especially high from parent to child.

While this seemed like a good solution to our woes, we soon realised that issues relating to privacy and consent meant that cost was a major barrier to accessing DOMINO. At that time, the default arrangements for access to DOMINO were through AIHW via the [Sax Institute's Secure Unified Research Environment](#) (SURE). As we planned to seek consent from parents to link their data, we had anticipated that a sufficiently deidentified dataset could be released to the research team for analysis, with appropriate secure storage and access arrangements in place. However, we were advised that the data custodian could stipulate the conditions of data release, effectively overriding parental consent. Thus, access to and analysis of data would have to be done in SURE, for which we had not budgeted. Moreover, data integration would also have been required to take place in SURE, and this would have incurred substantial costs. Thus, it became quite clear that our data linkage proposal would not have been affordable even if it had been feasible.

5.5 | Data quality might have made linkage impossible anyway

After realising the costs involved, we considered whether we should try to raise extra funds. We decided that this could be done if the whole exercise were looking feasible. Unfortunately, there were two more major problems relating to data quality.

The most serious of these was that most of the clients (78%) recorded in DEX for our demonstration community were unidentified. This occurs when the partner organisation runs services for groups of participants, and does not collect data for each individual, but instead records in DEX the number of people who attended and some other statistical details at the session level. Sometimes this is because sessions are run with large groups and it is not practical to collect data from each individual. However, as we noted earlier in the discussion of data from state-funded services, many organisations have serious concerns about

collecting data from individuals, at least in some settings, because it can be a barrier to the development of trusting relationships and participation. One service provider in particular aimed to provide soft entry points and soft entry activities that provided a non-stigmatising, neutral environment that parents would choose to attend, with no expectation of further engagement. The service providers took the view that focusing on data collection in these settings might alienate these clients (Cortis et al., 2009).

The second major problem relating to data quality revolved around linkage and consent. As described, we intended to seek consent from parents to access and link child data, and this would have extended to consent to access records in DOMINO. Having consent might have allowed deterministic linkage to take place, providing very high-quality linkages. However, this would require pristine identifying data, including date of birth and Centrelink customer reference number (CRN). However, if dates of birth and CRNs were not present or inaccurate, many links would be missed. An alternative would have been probabilistic linkage with a SLK. However, for this to be effective for our purposes, we would have needed high quality data from both children and parents to construct the SLKs. If we had progressed to the point of negotiating an agreement with the Queensland Department of Education to seek parental consent via schools, we would have had to ensure that we obtained this high-quality identifying data from parents during the consent process. Even if this were achieved however, we would have returned to the issues with DEX: potential problems with data quality, and a large number of unidentified clients. Finally, although whole families can be identified in DOMINO, the complexities of real families – changing names, separations – can make matching parents and children in families messy. With a relatively small sample such as ours, this could have been a significant problem.

6 | LESSONS LEARNED

In summary, there were multiple points of failure in our attempt to do cross-sectoral linkage. Moreover, each possible solution was contingent on other aspects of data with different problems, rendering these solutions impractical. To illustrate, under different funding arrangements we might have been able to afford to pursue data linkage through DOMINO. But this would not have solved the problem of the lack of identified client data in the Data Exchange. However, even if all clients were identified, data quality may not have been sufficient for a statistical linkage key. Even if data were of high quality, we may not have obtained consent from enough parents to create a large enough sample to be useful. And finally, if consent from sufficient parents were obtained, the quality of identifying data obtained may not have been high enough for linkage.

This catalogue of failures perhaps suggests that the investigators were inexperienced and did not seek expert advice before attempting to implement the design. This was in fact not the case. First, we spent considerable time consulting with one of Australia's experts in this field, a person highly experienced in working with a number of data linkage projects and authorities. Second, the proposed design was thoroughly discussed with all the partner organisations who were data custodians. None of these organisations intimated that the design was not feasible, although they indicated the need for department-specific data access procedures. Thirdly, three of the investigators on the CREATE project had previously constructed a longitudinal database integrating service participation data from a large NGO, education department records, and child- and teacher-reported child wellbeing. Another investigator also had experience with other data linkage projects. Finally, none of the five reviewers of the ARC linkage proposal raised serious objections to the data linkage component, although one asked for more details about the procedures we had engaged in with custodians to that point.

For all these reasons we were surprised to encounter as many obstacles as we did. There were three key lessons that we learned from the data linkage and integration attempt. The first, and probably the most important, is that *it is critically important to make assumptions explicit*. Our partners in the project were enthusiastic about participating in the data linkage exercise and supportive of the objectives of the project, and we generally assumed that this indicated that data would be available, and that linkage would be feasible. Taking the example of the DEX data, our taken-for-granted and hence unstated assumption was that it was obvious that the child was the unit of analysis, and so it didn't occur to us to determine the availability of these data early on. Similarly, we assumed that because we planned to seek informed parental consent to access and link child data, this would facilitate linkage between different datasets. As discussed, this was a faulty assumption for both technical and legal reasons.

Second, researchers should *talk to data custodians very early in the project planning stages*. While we had close relationships with project partners, we learned that it is important to talk to the knowledge holders in data custodian organisations and in data linkage organisations who understand what the rows and columns look like in the datasets of interest. While this may require a research team member with some technical skill, it is the only way to match expectations about what a researcher needs with what is available in the data, assess whether it is accessible, and start to determine whether linkage would be technically feasible.

Third, many problems could be avoided if as much information as possible about data quality, linkage procedures and costs, access and storage requirements, and consent requirements for each dataset were obtained through a *thorough scoping exercise well ahead of time*, for instance prior to writing a funding application. However, in practice this may be difficult to achieve. Many of the barriers we encountered - technical, legal, and ethical - only became apparent once the attempt to start accessing data began. For instance, for some

datasets it was difficult to find out who to talk to about the data without making a data access application. One consequence of this for our project was that we found it difficult to get an accurate picture of the costs of linkage ahead of time. In addition, the regulatory and privacy frameworks are complicated and difficult to interpret. Mitchell and colleagues (2015) described similar issues in a proof-of-concept project linking health and hospital data across Australian jurisdictions.

Two practical steps that researchers can take to implement these lessons and avoid the problems we experienced are first, to request from custodians metadata and data dictionaries for review. This can reveal the structure and content of the data. Moreover, if such resources cannot be provided, it suggests that the data may not be available or suitable for linkage. Second, written data sharing agreements could be established with data custodians. This process could identify many potential problems. However, although we did not pursue formal written data sharing agreements, there were extensive discussions with custodians about the research design, and letters of support were obtained. Nonetheless, these steps are best practice for data linkage projects.

7 | FUTURE DIRECTIONS

The lessons learned in this proof-of-concept project will be valuable for future projects involving data linkage. However, we are no further progressed in our original objective: evaluating the collective effects on child wellbeing of national and state programs delivered by community agencies, as well as school-based initiatives. At present, it is not possible to achieve this objective in Australia at the level of children and families. Nonetheless, it remains an important goal to pursue because despite the millions spent on initiatives to improve child wellbeing in disadvantaged communities, we still have very little evidence of what works, for which children, and under what circumstances. We offer two broad recommendations that could help achieve this goal:

1. Community agencies and schools should be properly funded to measure both the wellbeing of children and young people, and the programs and services that seek to improve it. This should include collecting the high-quality data that is needed to both capture program costs and assess program logic or theories of change, so that schools and agencies can engage in fast cycle innovation, constantly improving their practices in light of timely and relevant data (Shonkoff & Fisher, 2013).

The CREATE project has developed some technologies that can support and enhance this process. For instance, Rumble's Quest is not simply a computer game that gives children a voice and validly and reliably measures their social-emotional wellbeing, it is a comprehensive system which provides schools and agencies real-time aggregated summary reports of data designed to support data-driven decisions about priorities for action and the strategies with the best evidence. A second example is the ESRT app, described earlier (Manning et al., 2018), which provides a logical and user-friendly interface to support schools and community agencies to record details of the services they provide for children and families. Complementing Rumble's Quest, it is designed to assist schools and communities in: (i) setting priorities and plans; (ii) identifying the most cost-effective way of achieving strategic objectives; (iii) providing information on alternatives that have the lowest impact on external parties while maximising benefits to program participants and their families; (iv) providing evidence on return on investment; and (v) documenting the decision process.

Putting the means for data collection and summary reports of that data into the hands of schools and community agencies in these ways could result in higher quality data for both data linkage purposes and for agency use in a continuous cycle of reflection, evaluation, and improvement of services and data protocols^{iv}.

2. Data custodians and government gatekeepers should work with consumers of services (residents, children, parents) as well as with researchers, schools, and community agencies to find creative solutions to privacy concerns. This should take place within a co-creation framework (Pfitzer et al., 2013) that encourages respectful dialogue and meaningful partnerships that recognise and correct for power imbalances. Such an approach is completely consistent with the radical reforms for “policy partnerships” and “place-based partnerships” in the new National Agreement for Closing the Gap (Commonwealth of Australia, 2020). Of particular relevance to this paper is the objective that “Aboriginal and Torres Strait Islander people have access to, and the capability to use, locally-relevant data and information to set and monitor the implementation of efforts to close the gap, their priorities and drive their own development.” (p. 3) In our view, the principles, objectives, and practices in this Agreement should apply in all disadvantaged communities and with all ethnic groups.

When explicit and informed parent or carer consent is obtained through respectful dialogue, data linkage and the release of data could be greatly facilitated. However, we also recognise that even with consent a balance needs to be sought between the benefit to individuals and communities of using data to improve services and wellbeing, and harms that could occur. For instance, if we had been able to construct our child-centric dataset, there was a chance the data would have been re-identifiable due to small cell sizes, even with the removal of identifying information. There is also the risk of identifying vulnerable groups of children and families, which could lead to stigmatisation and further marginalisation (Broadley, 2018).

There are ways of accessing data, however, that minimise risks. Using secure facilities for data storage and analysis such as SURE and Griffith University’s Social Analytics Lab is one way of doing this, and there are other technical solutions to minimise re-identification risk that can be explored (Simon et al., 2019). More broadly, we argue that the

risks and benefits of data linkage should be identified and considered on a case-by-case basis and in a collaborative and respectful process between data custodians, communities, and researchers. Establishing trust would be fundamental to this process.

New technologies and legislative change will continue to alter the landscape of big data and data linkage in Australia in the coming years. Particularly important in this regard is the new Data Availability and Transparency Bill (Office of the National Data Commissioner, 2020). The Bill emphasises the importance of making the best use of existing government data, especially through sharing or linking across departments and levels of government, and “streamlines sharing by overriding secrecy provisions and non-disclosure prohibitions ...” (p. 8).

The new Bill however will not immediately solve the problem of incorporating data held by NGOs or schools into data linkage arrangements. Cross-sectoral data linkages of the kind we attempted will remain out of reach without other committed strategies to update and align data standards and systems. An exercise such as this would likely need a guiding framework such as a national child wellbeing strategy agreed between the Commonwealth and states and territories and the community, health, and education systems^v. While ambitious, this would be a significant step forward in improving outcomes for Australia’s most vulnerable children.

References

- Australian Government Department of Education and Training. (2018). *Better evidence for children's policy. Towards improving children's developmental outcomes: Scoping an approach to build better evidence for policy.*
https://docs.education.gov.au/system/files/doc/other/childrens_policy_evidence.pdf
- Branch, S., Freiberg, K. and Homel, R. (2019). Strengthening the prevention delivery system for children in disadvantaged communities through infrastructure development. *Developing Practice: The Child, Youth and Family Work Journal*, 53, 54-73.
- Broadley, K. (2018). What can surveillance data and risk factor research contribute to a public health approach to preventing child sexual abuse? *Australian Journal of Social Issues*, 53(4), 372-385. <https://doi.org/10.1002/ajs4.46>
- Cortis, N., Katz, I., and Patulny, R. (2009). *Engaging hard-to-reach families and children. Stronger Families and Communities Strategy 2004–2009* (Occasional Paper No. 26). Canberra: Australian Government Department of Families, Housing, Community Services and Indigenous Affairs.
- Day, J., Freiberg, K., Hayes, A. and Homel, R. (2019). Towards Scalable, Integrative Assessment of Children's Self-Regulatory Capabilities: New Applications of Digital Technology. *Clinical Child and Family Psychology Review*, 22(1), 90-103.
<https://doi.org/10.1007/s10567-019-00282-4>
- Department of Social Services. (2019). *The Data Exchange Framework. Streamlined performance reporting.*
- Edwards, B., Mullan, K., Katz, I., and Higgins, D. (2014). *The Stronger Families in Australia (SFIA) Study: Phase 2* (Research Report No. 29). Melbourne: Australian Institute of Family Studies

- Freiberg, K., Homel, R. and Branch, S. (2014). The Parent Empowerment and Efficacy Measure (PEEM): A Tool for Strengthening the Accountability and Effectiveness of Family Support Services. *Australian Social Work*, 67(3), 405-418.
<https://doi.org/10.1080/0312407X.2014.902980>
- Funston, L., and Herring, S. (2016). When will the stolen generations end?: A qualitative critical exploration of contemporary 'child protection' practices in Aboriginal and Torres Strait Islander communities. *Sexual Abuse in Australia and New Zealand*, 7(1), 51-58.
- Homel, R., Branch, S. and Freiberg, K. (2019). Implementation Through Community Coalitions: The Power of Technology and of Community-Based Intermediaries. *The Journal of Primary Prevention*, 40(1), 143-148. <https://doi.org/10.1007/s10935-019-00541-8>
- Homel, R., Freiberg, K. and Branch, S. (2015). CREATE-ing capacity to take developmental crime prevention to scale: A community-based approach within a national framework. *Australian and New Zealand Journal of Criminology*, 48(3), 367-385.
<https://doi.org/10.1177/0004865815589826>
- Homel, R., Freiberg, K., Branch, S. & Le, H. (2015). Preventing the onset of youth offending: The impact of the Pathways to Prevention Project on child behaviour and wellbeing. *Trends and Issues in Crime and Justice*, 481
- Homel, R., Lincoln, R., and Herd, B. (1999). Risk and resilience: Crime and violence prevention in Aboriginal communities. *Australian and New Zealand Journal of Criminology*, 32(2), 182-196. <https://doi.org/10.1177/000486589903200207>
- Kania, J., and Kramer, M. (2011). Collective Impact. *Stanford Social Innovation Review*, Winter.

- Manning, M., Wong, G. T. W., Graham, T., Ranbaduge, T., Christen, P., Taylor, K., Wortley, R., Makkai, T., and Skorich, P. (2018). Towards a 'smart' cost-benefit tool: using machine learning to predict the costs of criminal justice policy interventions. *Crime Science*, 7(1), 12. <https://doi.org/10.1186/s40163-018-0086-4>
- Mitchell, R. J., Cameron, C. M., McClure, R. J., and Williamson, A. M. (2015). Data linkage capabilities in Australia: practical issues identified by a Population Health Research Network 'Proof of Concept project'. *Australian and New Zealand Journal of Public Health*, 39(4), 319-325. <https://doi.org/10.1111/1753-6405.12310>
- Moore, H. C., Guiver, T., Woollacott, A., de Klerk, N., and Gidding, H. F. (2016). Establishing a process for conducting cross-jurisdictional record linkage in Australia. *Australian and New Zealand Journal of Public Health*, 40(2), 159-164. <https://doi.org/10.1111/1753-6405.12481>
- Office of the National Data Commissioner. (2020). *Data Availability and Transparency Bill 2020 Exposure Draft Consultation Paper*. Canberra: Department of the Prime Minister and Cabinet.
- Pfitzer, M. W., Bockstette, V., and Stamp, M. (2013). Innovating for shared value. *Harvard Business Review*, September. <https://hbr.org/2013/09/innovating-for-shared-value>
- Shonkoff, J. P., and Fisher, P. A. (2013). Rethinking evidence-based practice and two-generation programs to create the future of early childhood policy. *Development and Psychopathology*, 25(4 Pt 2), 1635-1653. <https://doi.org/10.1017/S0954579413000813>
- Simon, G. E., Shortreed, S. M., Coley, R. Y., Penfold, R. B., Rossom, R. C., Waitzfelder, B. E., Sanchez, K., and Lynch, F. L. (2019). Assessing and Minimizing Re-identification Risk in Research Data Derived from Health Care Records. *EGEMS (Washington, DC)*, 7(1), 6-6. <https://doi.org/10.5334/egems.270>

Van Ryzin, M. J., Fishbein, D., and Biglan, A. (2018). The promise of prevention science for addressing intergenerational poverty. *Psychology, Public Policy, and Law*, 24(1), 128-143. <https://doi.org/10.1037/law0000138>

ⁱ DSS supports include Child Family Community Australia (CFCA) resources, Data Exchange (DEX), and funding for CfC partnerships to access evaluation and other expertise.

ⁱⁱ In Australia, when high risk Commonwealth data are involved (e.g. data on children), data linkage can only be carried out by a national [Accredited Data Integrating Authority](#) as these have the capacity and legislative protections to securely manage Commonwealth data linked to data from other sources (Moore et al., 2016). Such authorities have proliferated in recent years, but at the time the data linkage experiment was being designed (2018) there were only three: the Australian Bureau of Statistics, the Australian Institute of Health and Welfare, and the Australian Institute of Family Studies.

ⁱⁱⁱ In 2018 this was the Qld Department of Communities, Child Safety and Disability Services; it is now the Qld Department of Child Safety, Youth and Women.

^{iv} Not even the best funded initiatives, such as the NSW Government's [Their Futures Matter](#), have yet managed to link government with NGO data.

^v This could build on the 'National Thriving Children Early Years Strategy' that is being advocated by the [Australian Research Alliance for Children and Youth](#), and the [National Action Plan for the Health of Children and Young People: 2020-2030](#)

Table 1. Details of data measuring inputs and child outcomes

Dataset	Source	Details	Consent process
<u>Inputs: Child and family participation in programs and services</u>			
School-based initiatives	School principals	Participation in programs intended to improve social-emotional wellbeing or behaviour	Principals seek consent from parents to share children's participation data
Communities for Children Services	Commonwealth Department of Social Services	Data on the participation of children in CfC services where children were the intended beneficiaries supplied via the Data Exchange (DEX)	Obtained from parents/carers via the Facilitating Partner; approval from DSS
State-funded NGO services	QLD Dept of Child Safety, Youth and Women	Child participation in family support services delivered by state-funded NGOs	NGOs facilitate parental consent
<u>Outcomes for children age 5-12 years old</u>			
Rumble's Quest computer game (6-12 years)	Schools	Five dimensions of social-emotional wellbeing and five executive function scores. Data collected in class groups in schools.	Schools seek parental consent and research team seek principals' consent for de-identified but linkable data
Strengths and Difficulties Questionnaire	Schools/classroom teachers	Screening scale used to score behavioural problems and prosocial behaviours. Classroom teachers score each student	Schools seek parental consent
NAPLAN	ACARA or QLD Dept. of Education	Grades 3, 5 & 7: Numeracy and literacy	Schools seek parental consent
School attendance, suspensions and exclusions	QLD Department of Education		Schools seek parental consent

Note: NGO = Non-government organisation. NAPLAN = National Assessment Program - Literacy and Numeracy. ACARA = Australian Curriculum, Assessment and Reporting Authority. DSS = Australian Government Department of Social Services. CfC = Communities for Children. DEX = Data Exchange.

Figure 1: Proposed data linkage procedures

Note: DSS = Australian Government Department of Social Services, NGO = non-government organisation, QLD = Queensland

