Children's Research Digest

Archiving Evaluation Data

December 2017

Children's Research Network

Volume 4
Issue 3
ISSN 209—728X

Editorial

The Prevention and Early Intervention Research Initiative (PEI-RI) is a data archiving project at the Children's Research Network. The central aim of the PEI-RI is to archive research data from a series of evaluations of Prevention and Early Intervention services from around the island of Ireland, often referred to as the Prevention and Early Intervention Initiative (PEII), so that this data is available through the national data archives for further analysis and service development.

Since the Prevention and Early Intervention Initiative began in 2004 (see Birkbeck this volume), practices and experiences of data sharing and re-use have continued to evolve within the ecosystem of social science research, both in Ireland and internationally. In the context of a growing trend towards promoting transparency, rigour and public accountability, the community of researchers, research funders and, increasingly, scholarly journals and other platforms for the dissemination of research findings, are all moving towards requiring or encouraging researchers to make their data, and data analyses, available for others to examine and re-use. Thus, concerns about ensuring the quality and reliability of research findings have been added to earlier ideals of facilitating historical and comparative analysis, maximising the potential and value of data and reducing the burden on research participants.

Since 2017, the EU Framework Programme for Research and Innovation, Horizon 2020, has treated research data as "open by default" with the possibility for researchers to opt out, under the principle that data should be "as open as possible, as closed as necessary." [1]. In Ireland, the Health Research Board (HRB) recently launched an Open Research Platform, including an Open Data Policy [2], which specifies that:

"[O]riginal results should include the source data underlying the results, together with details of any software used to process the results. It is essential that others can see the source data in order to be able to replicate the study and analyse the data, as well as in some circumstances, reuse it."

In this context, support for archiving under the Prevention and Early Intervention Research Initiative led by the Children's Research Network, is leading the way for the development of data sharing practice in Ireland. The articles in this Digest are timely insofar as they reflect many of the challenges and opportunities associated with the new world of open research data.

As Corti and Fielding (2016) argued in a recent article, dedicated academic research data archives and portals play an essential role in meeting the 'FAIR' guidelines that have been adopted both by Horizon 2020 and the HRB. These specify that published data must "embrace the principles of Findability, Accessibility, Interoperability, and Reusability". Meeting these guidelines requires first that digital data are sustainably preserved in a persistent online location. The articles in this digest describe how quantitative and qualitative evaluation data from the PEI initiative have been deposited in the Irish Social Science Data Archive (ISSDA[3]) and the Irish Qualitative Data Archive (IQDA[4]) , both of which have policies oriented towards the preservation of data for the long term. In the case of IQDA, data preservation is secured through its membership of the Digital Repository of Ireland (DRI[5]), a national trusted digital repository for humanities, social sciences and cultural heritage data. DRI adopts the Data Seal of Approval as its policy guideline. In addition, DRI provides persistent citation through Digital Object Identifiers (DOIs) and automatically generated citations. [6,7]

Ensuring accessibility and re-usability also requires that data are appropriately documented using standardised metadata and contextual descriptions. Data documentation facilitates re-use by providing information about context and by enabling search and discovery. As a number of the articles in this Digest reveal (Leavy; Fleming and O'Hare), meeting the requirements for systematic documentation promotes data quality and trust in its provenance. The authors describe their efforts to identify appropriate file versions, standardise naming conventions for variables and anonymise the data in order

to protect participant confidentiality. Trust in data also requires that we can be confident that the data were ethically collected, and that participants have given informed consent for its use and re-use through archiving. Repositories such as IQDA require that data are deposited in compliance with professional norms, including those relating to participant confidentiality and informed consent. A number of the articles in this Digest note the importance of building a plan to share data into the management of the research from the outset, to ensure consent and facilitate data management (Leavy; Fleming and O'Hare; Quinn; Hayes).

There is a longstanding debate surrounding whether or not the challenges associated with archiving qualitative data are greater than those surrounding quantitative data, aspects of which are mentioned in a number of the contributions to this Digest. As Bishop (2009; 2013) has described, these concerns are both methodological and ethical. Methodological concerns include whether or not it is possible to provide sufficient context for secondary users to be able to analyse the data (Rodriguez), or to avoid the risk of 'misinterpretation.' Some researchers consider that anonymisation of qualitative data may require the removal of so much information that the data becomes unusable (Hayes). Ethical concerns include the idea that qualitative research involves a higher level of moral obligation to participants on the part of the researcher (Hayes). This is not the place to revisit these issues, many of which are addressed in detail in the contribution by Rodriguez. Suffice to say that research carried out by IQDA in collaboration with Tallaght West CDI (see Quinn) showed that, while Irish researchers share similar and understandable concerns with their counterparts in other jurisdictions, there is nevertheless considerable support for the principle of qualitative data archiving (Geraghty 2014).

In summary, the contributions to this Digest testify to the ground-breaking role of the PEI Research Initiative in furthering the practice of social science data sharing and re-use in Ireland.

However, despite the national and international movement towards open data mentioned above, a number of challenges remain. First, as Corti and Fielding (2016) discuss, there is still some misunderstanding about the central importance of dedicated archives situated within international collaborations and infrastructures, for ensuring meaningful preservation and access to data. Ireland has invested in the creation of social science repositories (ISSDA and IQDA/DRI), principally through the Programme for Research in Third Level Institutions, but sustaining and developing these initiatives into the future will require continuing funding and support. In order to secure this, it is vital that their importance for the promotion of high quality, transparent and rigorous primary and secondary research is recognised and acknowledged within the wider Higher Education and Research landscape.

Second, as many of the contributions document, designing and implementing research data management plans to facilitate data archiving is resource intensive. Research funders, (notably including Horizon 2020), increasingly allow costs associated with archiving in research budgets, but this is not consistently true of all funders, especially in relation to the personnel and hours required for managing data for sharing and re-use. Developing a culture of data sharing requires education not only of social science researchers, but also of research funders and commissioners. Finally, while there has been considerable progress towards encouraging and supporting researchers to share their data, work remains to be done to encourage re-use and secondary analysis, especially of archived qualitative data. While there is some international evidence of a pattern of growth in re-use over time (Bishop and Kuula 2017), promoting re-use remains an ongoing challenge in the Irish case. In this context, the inclusion of grants for re-use under CRNINI-PEI research initiative is an exemplar of good practice (see www.childrensresearchnetwork.org).

[1] http://ec.europa.eu/research/openscience/pdf/openaccess/ord_extension_faqs.pdf

[2] https://www.hrbopenresearch.org/

[3] https://www.ucd.ie/issda/

[4] https://content.web.nuim.ie/iqda

[5] http://www.dri.ie/

[6] https://repository.dri.ie/catalog/qz2167463

[7] https://repository.dri.ie/catalog/rx91h486p

References

Bishop, Libby, 2009. Ethical sharing and reuse of qualitative data. *Australian Journal of Social Issues*, 44(3), pp.255-272.

Bishop, Libby, 2014. *Re-using qualitative data: a little evidence, on-going issues and modest reflections.* Studia Socjologiczne, (3).

Bishop, Libby and Kuula-Luumi, Arja, 2017. *Revisiting qualitative data reuse: A decade on.* Sage Open, 7(1), p.2158244016685136.

Corti, Louise and Fielding, Nigel, 2016. *Opportunities From the Digital Revolution: Implications for Researching, Publishing, and Consuming Qualitative Research.* SAGE Open, 6(4), p.2158244016678912.

Geraghty, Ruth, 2014. *Attitudes to Qualitative Archiving in Ireland: Findings from a Consultation with the Irish Social Science Community.* Studia Socjologiczne, (3), p.187.

Jane Gray
Guest Editor

Maja Haals Brosnan
Editor, Children's Research Network

Thank you

We would like to thank all authors and the reviewers for their contributions to this issue. Special thanks are also due to all who helped with proof reading and to AAD for providing the design and layout. We would also like to thank Jane Gray for acting as Guest Editor and Atlantic Philanthropies for their ongoing involvement in the Prevention and Early Intervention Research Initiative.

Archiving data from the Prevention and Early Intervention Initiative

A funder's perspective

Gail Birkbeck

Introduction

The Atlantic Philanthropies (Atlantic) is a global foundation dedicated to advancing opportunity and promoting equity and dignity. Founded in 1982, Atlantic decided in 2002 to fully commit the foundation's assets during founder Chuck Feeney's lifetime and cease operations by 2020. Atlantic, which invested $8 billion over the course of its history, will become the largest foundation to intentionally go out of business. The decision to limit Atlantic's life resulted in a strategic shift and the emergence of new programmes focused on achieving significant outcomes in a relatively short timeframe. This brought a sense of urgency to "get it right" and to bring about change during the life of the founder (Proscio, 2010, p.8). In addition, unlike perpetual foundations which keep grants to a smaller sum, a "culture of big bets prevailed" at Atlantic (Proscio, 2010, p.8). Along with this new outcomes-focused, big bets grantmaking approach, the foundation created a Strategic Learning and Evaluation team, tasked with developing appropriate systems to document the impact of the investments and share the findings. The team also was given responsibility to help grantees measure their progress and learn from their work.

Rigorous evaluation was an integral component of the prevention and early intervention initiative on the island of Ireland. Supporting 52 prevention and early intervention services in areas such as early childhood, learning, child health, child behaviour and parenting, the investments have resulted in substantial knowledge about what works in improving children's lives (see Paulsell and Jewell, 2012, Rochford et al., 2014, The Atlantic Philanthropies, 2015). The extent of data that was generated over a ten-year period is equally impressive. This paper discusses the origins of this programme and the motivation to support the Children's Research Network for Ireland and Northern Ireland to archive and make accessible the data, thereby facilitating further analysis and extending the learning and legacy of this programme.

Atlantic's strategic approach

Once it decided to complete all grant-making by the end of 2016, Atlantic undertook a strategic assessment process to identify how and where – over its remaining years – it could have the greatest impact in improving the life trajectories for disadvantaged and marginalised people, communities and nations. The Children and Youth programme in the Republic of Ireland and Northern Ireland was one such example, with Atlantic investing $172.4 million[1] and $55 million respectively, to change the way children and young people receive services. The investment strategy focused on prevention and early intervention services with the goal of informing government policy and demonstrating a new way of working. To do that required rigorously evaluating the innovative approaches. This focus on promoting evidence-based prevention and early intervention was consistent with international trends that emphasized prevention strategies to cost-effectively address social problems early in a problem cycle, as well as an increased use of programmes and practices with scientific evidence of effectiveness (Paulsell and Jewell, 2013).

A core focus of the strategy Atlantic adopted to achieve its vision was to build capacity in the sectors and fields in which it chose to work. The Prevention and Early Intervention Initiative (PEII) was in keeping with Atlantic's approach of developing capacity and infrastructure for the sector. It made investments in university-based research centres—the Children and Family Research Centre (CFRC), National University of Ireland Galway (NUIG) and the Centre for Effective Education (CEE), Queen's University Belfast. The aim was "to increase capacity to provide service design support and evaluation services on the island of Ireland" (Paulsell and Jewell, 2013. p.20). The Centre for Effective Services, with funding from Atlantic and government, was also established to support evidence-based and evidence-informed practice, translating research from multiple sources in a way that was accessible and relevant for policy makers and practitioners (CES, online).

[1]The then Office of the Minister for Children and Youth Affairs (OMCYA, now the Department of Children and Youth Affairs) was an equal funding partner for three of the community engagement sites in the Republic of Ireland —Northside Partnership Preparing for Life, Tallaght West Child Development Initiatives (CDI), and Youngballymun. These were collectively known as the Prevention and Early Intervention Programme which was established by the Irish government in February 2006 (see Paulsell et al., 2009, p.19).

Capacity building at the practitioner level

In addition to rigorously evaluating innovative approaches in services delivery, Atlantic helped support the development of evaluative capacity at the practitioner level. From Atlantic's perspective it was "important that robust evaluation be incorporated into service delivery in order to ensure that the services delivered are effective. Without including some type of evaluation for each of the prevention and early intervention programs it would have been difficult to know which ones were delivering the intended outcomes and should be continued and which ones were not" (Boyle, 2016, p. 27). To ensure that happened, evaluation budgets were incorporated into the funding. In this way, individual organisations could evaluate their own work and build an evidence base to demonstrate their impact. In practice, this meant that practitioners commissioned, managed, monitored and disseminated their evaluations, while working collaboratively with researchers. The "big bet" culture influenced the size of evaluation expenditures in the Prevention and Early Intervention Initiative. This resulted in extensive budgets to conduct randomised controlled trials and longitudinal studies as well as dissemination campaigns to promote their findings and inform future service provision. While the organizations had full ownership rights to all intellectual property produced with Atlantic's support, the foundation also required that it be granted a royalty free license allowing it to publish and disseminate any of that material for its own knowledge-sharing purposes and for the benefit of future researchers.

An independent evaluation of the programme confirmed that it had "led to significant changes in grantee organisational capacity" (Paulsell and Jewell, 2012, p.16). Specifically, "grantees have gained substantial experience in implementing evidence-based prevention and early intervention programmes in real-world"... "Participation in the rigorous evaluations also stimulated growth in... capacity (p.20). It had also introduced grantees to "a new way of thinking about how to identify needs, design services and...use evaluation methods" (The Atlantic Philanthropies, 2015, p.3). The significance of these evaluations on the island of Ireland is in no doubt as Boyle (2016) notes, "A level of analysis has been undertaken that wouldn't have happened otherwise. The creation and existence of Irish cases where there is strong evidence of what works and what doesn't is viewed very positively by policymakers" (p.4).

Findings from the Prevention and Early Intervention Initiative

Much has been written about the impact of the programme (Paulsell et al., 2009; Paulsell and Jewell, 2013; Rochford et al., 2014) as well as the outcomes of specific interventions (see https://www.atlanticphilanthropies.org/subtheme/prevention-early-intervention for summative information and links to the individual projects funded). In all, 39 initiatives were funded to deliver 52 evidence-based services to children and young people on the island of Ireland. "Almost all of these programmes have been evaluated positively – under demanding testing" (The Atlantic Philanthropies, 2015, p.3). At a minimum this translates to about 90,000 children and young people, 24,000 parents or caregivers, and 4,000 professionals benefitting from the programme in areas such as early childhood, learning, child health, child behaviour and parenting. Furthermore, this has resulted in substantial knowledge about what works in improving children's lives as well as significant datasets providing rich and detailed information on all aspects of childhood, family life and service provision. In addition, new networks such as the Prevention and Early Intervention Network and the Children's Research Network for Ireland and

Northern Ireland (CRNINI) emerged organically from the work to share the results of evaluation with government. These networks brought together a wide range of professionals with an interest in research on child and family issues across the island of Ireland, respectively building on the work to date.

Building on the learning of the Prevention and Early Intervention Initiative

After ten years of funding, extensive and wide-ranging datasets were created and collated as part of the Prevention and Early Intervention initiative. While much of this data has been analysed as part of the evaluations of the projects, there is still much to be gained from further exploration. In 2014 The Atlantic Philanthropies funded a learning initiative, managed by the Children's Research Network for Ireland and Northern Ireland (CRNINI), for archiving and further analysis of the data to embed the legacy and learning. To date funding has been made available to prepare and archive datasets in the Irish Social Science Data Archive (ISSDA), the Digital Repository of Ireland and The National Archives in the UK. A small, competitive grants scheme for further interrogation of the data is also underway for secondary analysis of a PEII dataset, meta-analysis of a number of PEII datasets, comparative analysis of PEII data and any other relevant data, development of training in evaluation research using PEII data and disseminating and/or presenting research outputs nationally and internationally. In this special issue we learn about some of the practical, ethical and methodological issues and challenges encountered in archiving and analysing the PEII datasets. Especially relevant are the processes and requirements involved in archiving data retrospectively as well as preserving qualitative data for further analysis. There is small infrastructure for archiving on the island of Ireland and the implications for funders who, more and more, require data be archived, are useful. These include: the need to plan for data preservation from project inception as well as to ensure there is capacity in the sector to manage data throughout lifecycle of a project. In addition, CRNINI's work on this project has helped garner knowledge on archiving and data preservation more generally. The findings from this initiative will be disseminated throughout 2018 while the datasets can be accessed into the future.

For Atlantic, support to archive and preserve the data from the Prevention and Early Intervention Initiative served as a mechanism to prolong the life of the datasets and to add to the knowledge base. There is a growing recognition among funders of the value of preserving and sharing data and Atlantic is no exception as it prepares to archive and make available to researchers and others its more than 30-year history so that it can continue to "...inform, influence and inspire current funders, emerging philanthropists and the public (Florino, 2017)" even after the foundation ceases to exist.

References

Boyle, P. (2016) *Philanthropy Working with Government: A Case Study of The Atlantic Philanthropies' Partnership with the Irish Government.* Dublin: Institute of Public Administration [Online] Available at: https://www.atlanticphilanthropies.org/case-studies/philanthropy-working-with-government-a-case-study (Accessed: 9 September, 2017)

Florino, J.V. (2017) *The Atlantic Philanthropies and its archives: limited life, enduring Legacy* [Online] Available at: www.atlanticphilanthropies.org/news/the-atlantic-philanthropies-and-its-archives-limited-life-enduring-legacy (Accessed: 9 September, 2017)

Paulsell, D. and Jewell, C.P. (2013) *The Atlantic Philanthropies' Children and Youth Programme in Ireland and Northern Ireland: Programme Evaluation Findings – Final Report.* Princeton:

Mathematica Policy Research, Inc. [Online] Available at: https://www.atlanticphilanthropies.org/evaluations/evaluation-prevention-and-early-intervention-programme-ireland-and-northern-ireland (Accessed: 10 September, 2017).

Paulsell. D., Del Grosso, P. and Dynarski, M. (2009) T*he Atlantic Philanthropies' Disadvantaged Children and Youth Program in Ireland and Northern Ireland: Overview of Program Evaluation Findings.* Princeton: Mathematica Policy Research, Inc. [Online] Available at: https://www.atlanticphilanthropies.org/evaluations/children-youth-programme-ireland-and-northern-ireland-overview-programme-evaluation-finding (Accessed: 10 September, 2017).

Proscio, T. (2010) *Winding Down the Atlantic Philanthropies – 2001-2009: The First Eight Years. North Carolina: Duke University* [Online] Available at: https://www.atlanticphilanthropies.org/research-reports/report-winding-down-atlantic-philanthropies-first-eight-years-2001-2009 (Accessed: 9 September, 2017)

Rochford, S., Doherty, N. and Owens, S. (2014) *Prevention and Early Intervention in Children's and Young People's Services: Ten years of Learning.* Dublin: Centre for Effective Services [Online] Available at: http://www.effectiveservices.org/resources/article/prevention-and-early-intervention-in-children-and-young-peoples-services-te (Accessed: 9 September, 2017).

The Atlantic Philanthropies (2015) *Prevention and Early Intervention in Ireland and Northern Ireland: Making a Real Difference in the Lives of Children and Young People.* New York: The Atlantic Philanthropies [Online] Available at: https://www.atlanticphilanthropies.org/app/uploads/2016/03/Prevention-Early-Intervention-Ireland-NI-Report.pdf (Accessed: 10 September, 2017).

Author information

Gail Birkbeck is the former Head of Strategic Learning & Evaluation at The Atlantic Philanthropies (2004–2017)

Gail Birkbeck has spent 13 years working in foundation monitoring, evaluation and learning where she supported staff and grantees to use evaluation as a learning practice, putting processes in place to inform their work. Working in close collaboration with programme staff, she determined strategic areas for evaluation in the Children and Youth, Ageing, Reconciliation and Human Rights and Population Health programmes on the island of Ireland and in South Africa. As Head of Strategic Learning and Evaluation (EuroAfrica, 2015–2017) the end of the foundation grantmaking brought an urgency to maximise the contribution of evaluation to Atlantic's influence and legacy. This entailed working closely with communications colleagues, curating and archiving materials to solidify institutional memory, capture knowledge and facilitate future learning.

Gail's academic background is in psychology, research and statistics. More recently she has completed managerial studies in strategy and innovation and is an MSc candidate in Data Business studying how organisations collect, evaluate, manage and use data for maximum value.

Archiving the Preparing for Life data

Motivation and historical context

Orla Doyle

Introduction

Preparing for Life (*PFL*) is a prevention and early intervention programme which aims to improve the life outcomes of disadvantaged children in Dublin, Ireland. *PFL* was designed and implemented by the Northside Partnership and was subject to an extensive evaluation conducted by the UCD Geary Institute for Public Policy between 2008 and 2015 using a randomised control trial design. The evaluation found that the *PFL* programme had a significant impact on children's skills by raising cognitive ability, reducing behavioural problems, and improving health. Please see Doyle (2017) and Doyle and *PFL* Evaluation Team (2016) for a description of the final results. The programme was one of 52 programmes funded by The Atlantic Philanthropies and the Department of Children and Youth Affairs as part of the Prevention and Early Intervention Initiative. In mid-2017 almost all the quantitative data collected as part of the *PFL* evaluation were placed in the Irish Social Science Data Archive (ISSDA). The decision to archive the data was made prospectively during the design of the study. The aim of this article is to describe the motivation for archiving the *PFL* data and the processes involved in prospectively designing, collecting, and storing data which was destined for a national archive.

The *PFL* Study

The goal of the *PFL* programme was to reduce social inequalities in children's skills by working with parents from pregnancy and until school entry. Families were recruited during pregnancy and randomly assigned to a high (n=115) or low (n=118) treatment group. The high treatment group received 1) bi-monthly home visits from a trained mentor to support parenting and child development using Tip Sheets, 2) baby massage classes to support reciprocal communication, and 3) the Triple P Positive Parenting Program to support positive, effective parenting practices. Both groups also received developmental toys, access to preschool and public health workshops, and a support worker. A 'services as usual' comparison group (n=99) from another community was also recruited.

The impact evaluation investigated the impact of the programme at frequent time points (baseline, 6, 12, 18, 24, 36, 48, and 51 months) using parent-report interviews, observations, and direct assessments. Families also gave consent to access their maternity and child hospital records, and teachers completed online surveys about the children's school readiness skills. Qualitative interviews with *PFL* mothers, fathers, children, and the *PFL* staff were also conducted.

Motivation for Archiving the *PFL* Data

Unlike most studies of early intervention programmes, the evaluation of *PFL* was led by a group of economists. Traditionally economists, particularly those who study human development, conduct secondary analysis of publicly available cohort or registry data. Thus the decision to archive the *PFL* data was driven by a strong belief among the investigators that any data collected as part of this publically funded study should be made available as a public good to be used by other researchers. The decision was also influenced by the location of ISSDA, which at the time, was housed within the UCD Geary Institute. The investigators, and in particular Professor Colm Harmon the then Institute director, had in-depth knowledge about the value of archiving and disseminating quantitative social science data. Within economic journals in particular, authors were increasingly required to make their datasets and code publicly available.

The decision was also motivated by one member of the study's advisory group, Professor James Heckman, who was seeking access to data from some of the landmark U.S. early intervention studies. Prior to this, almost all evaluation data were privately held, often by the researchers who conducted the original study. By making these historical data available, the data could be reanalysed and reinterpreted using new methods and different theoretical perspectives. In particular, Heckman and his team at the

University of Chicago accessed data from the Perry Preschool Program, the Carolina Abecedarian Program, and the Nurse Family Partnership Program. As a result, several new papers emerged offering new insights into these important studies (e.g. Heckman, Moon, Pinto, Savelyev, and Yavitz, 2010; Heckman, Pinto, and Savelyev, 2013; Gertler et al., 2014; Campbell et al., 2014).

Thus, the resolution to archive the *PFL* data, which we believed would be another landmark study in the early intervention field, was embedded into the design of the study from its inception.

Impact of Archiving the *PFL* Data on Study Design

The decision to archive the data had a number of implications for the study design which can be broadly grouped into four main categories: consent and ethics process, survey content, data quality and protection, and data documentation.

The first step in prospectively archiving the *PFL* data was to design an information and consent form which would provide the *PFL* participants with the necessary information to make an informed decision about joining the study. The form included consent both to join the *PFL* programme and the evaluation. As we were seeking consent to deposit the evaluation data into ISSDA, the form included a detailed section describing what would happen to the participant's data when the study ended. The information sheet explained that an anonymised dataset would be placed in ISSDA and could be used by other researchers. We reiterated that this dataset would not contain any personal details and that all names would be replaced by numbers to ensure that no-one could identify any individual responses. The consent form then explicitly asked participants whether or not they would permit an anonymised version of their data to be used in other research studies and publications. Of the 332 participants recruited, only one did not consent for their data to be used. While archiving social science data is slowly becoming standard practice, when we applied for ethical approval to conduct the *PFL* study in 2007, there was little precedence of prospectively archiving data. Despite this, none of the three ethics committees from whom permission was sought (UCD Human Research Ethics Committee, Rotunda Hospital's ethics committee, and National Maternity Hospital's ethics committee), raised any concerns with this aspect of the proposal.

In terms of survey content, to ensure the usefulness of the *PFL* collection as a panel dataset, the same instruments were used in multiple waves to allow future researchers to model changes over time in child and parent outcomes. We also ensured that each survey could be utilised as a stand-alone dataset to facilitate cross-sectional analysis. As archive users may wish to compare the *PFL* data to other national and international datasets, we also included commonly used instruments in the field, such as the Child Behavioral Checklist and the Home Observation Measurement of the Environment scores.

In terms of data quality and protection, as these data would be a publically available resource, the highest possible standards were maintained throughout the study to ensure that quality data were collected and stored appropriately. As the data would eventually be archived in electronic format, all the research interviews were conducted using tablet laptops to record responses directly. This served to reduce administrative burden, as well as increase the reliability of the data by minimising imputing errors. To guarantee data protection, we developed a *PFL* Data Management and Protection Protocol, alongside a Data Confidentiality Agreement, which everyone involved in the study signed. This document detailed the security procedures to be followed regarding the collection, storage, and analysis of the data. As the study was conducted over an extended period of time, with over 30 researchers working on the project, a *PFL* Research Training Manual was developed to facilitate staff turnover and preserve institutional memory.

In terms of data documentation, archived data requires clear and detailed documentation to ensure that researchers not involved in the original study can effectively re-use the data. Therefore, a number of standardised procedures were put in place to capture key information at each data collection wave. These included maintaining detailed codebooks and instrument descriptions, using a standardised variable naming convention, and recording information on the sample population, attrition, and missing data. After each research assessment was complete, we produced an evaluation report documenting this information, alongside the impact results for that assessment point. These procedures helped to ensure that the data archival process conducted at the end of the study was less onerous. Please see Wong (2017) in this edition for details on the practical steps involved in preparing the *PFL* data for archival.

Potential Uses of the *PFL* Data

The ISSDA website provides detailed information on all the *PFL* quantitative data that are available for analysis. Broadly, these include the eight research interviews conducted with families and the directly assessed measures of children's cognitive development. The birth and hospital records are not available due to the sensitive nature of these data, and the qualitative data will be archived in the Irish Qualitative Data Archive (IQDA).

The entire quantitative collection includes over 14,000 variables collected from ~300 families. Thus, as well as providing detailed information about the effectiveness of the *PFL* programme, they also provide comprehensive information on a population that is often under-represented in social surveys. The archived data has many potential uses. For example, it could be used to reproduce the impact results derived in the original evaluation. The issue of reproducibility of RCTs has received much attention in recent years in both the medical and social sciences (see special editions of Science, December 2011 and the American Economic Review, May 2017) and it is argued that making research data publicly available may help to reduce the dissemination of incorrect results, as well as prevent scientific fraud.

Regarding the *PFL* data, archive users could test the sensitivity of the original results to different statistical methods. It is also possible for archive users to examine outcomes that were not the primary focus of the original evaluation. For example, while academic papers have been published on the impact of the programme on child outcomes (e.g. Doyle, Harmon, Heckman, Logue, and Moon, 2017; Doyle, Fitzpatrick, Rawdon, and Lovett, 2015; Doyle, Delaney, O'Farrelly, Fitzpatrick, and Daly, 2017), less research has been published on parent outcomes. There is also potential to examine the mechanisms underlying the treatment effects and the longitudinal nature of the data could be exploited by modelling changes in outcomes over time. Finally, the presence of such a large amount of data on child development, health, parenting, social support systems, childcare, service use, as well as detailed socio-economic profiles, allows a thorough investigation of the lives of disadvantaged families in Ireland during a period of economic recession and recovery.

Conclusion

The archival of the *PFL* data capitalises on the substantial time and financial investment made in its original collection. In an era when the replicability of scientific studies is frequently questioned, the *PFL* data offer a unique opportunity to test the reproducibility of, as well as extend, the original results, which will ultimately increase the scientific integrity and rigour of the *PFL* study.

References

American Economic Review (2017) *Special Edition: Replication in Microeconomics,* American Economic Review, Vol. 107 (5), pp. 27-64.

Campbell, F., Conti, G., Heckman, J.J., Moon, S.H., Pinto, R., Pungello, E., et al. (2014) *Early Childhood Investments Substantially Boost Adult Health,* Science, Vol. 343 (6178), pp. 1478-1485.

Doyle, O. (2017) *The First 2000 Days and Child Skills: Evidence from a Randomised Experiment of Home Visiting*, UCD Geary Institute for Public Policy Working Paper 2017/06.

Doyle, O., and *PFL* Evaluation Team (2016) *Assessing the Impact of Preparing for Life: Final Report*. Report to Preparing for Life Programme (Atlantic Philanthropies & Department of Children and Youth Affairs).

Doyle, O., Harmon, C., Heckman, J., Logue, C., and Moon. S. (2017) *Measuring Investment in Human Capital Formation: An Experimental Analysis of Early Life Outcomes,* Labour Economics, Vol. 45 (April), pp 40-58.

Doyle, O., Fitzpatrick, N., Rawdon, C., and Lovett J. (2015) *Early Intervention and Child Health: Evidence from a Dublin-based Trial,* Economics and Human Biology, Vol. 19 (December), pp. 224-245.

Doyle, O., Delaney, L., O'Farrelly, C., Fitzpatrick, N., and Daly, M. (2017) *Can Early Intervention Policies Improve Well-being? Evidence from a randomized controlled trial,* PLoS ONE Vol. 12 (1), e0169829.

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., Chang, S.M. and Grantham-McGregor, S. (2014) *Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica,* Science, Vol. 344 (6187), pp 998-1001.

Heckman, J.J, Moon, S.H., Pinto, R., Savelyev, P.A., and Yavitz, A. (2010) *Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program,* Quantitative Economics, Vol. 1 (2), pp. 1-46.

Heckman J.J, Pinto, R., and Savelyev, P.A. (2013) *Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes,* American Economic Review, Vol. 103 (6), pp. 2052–86.

Science. (2011) *Special Edition: Data Replication and Reproducibility.* Science, Vol. 334 (6060), pp. 1225-1233.

Author information

Dr. Orla Doyle is an Associate Professor in the UCD School of Economics and a Research Fellow at the UCD Geary Institute for Public Policy. The core focus of her research is a micro analysis of human behaviour. Her areas of expertise include the economics of human development, health economics, labour economics, political behaviour, early child development and education, and methods for evaluating policy interventions. She is the Director of the UCD Childhood and Human Development Research Centre. Dr. Doyle received her Ph.D. in economics from Trinity College Dublin and holds a B.A. in economics and social science (TCD).


fn
shift
caps
tab

Archiving the Preparing for Life data

Practical steps taken to archive the quantitative data

Lorraine Wong

Preparing for Life (*PFL*) is a prevention and early intervention programme that is operated by the Northside Partnership in Dublin. Since 2008, the programme has worked with families to help children achieve their full potential. Mothers, who were on an overall average of 23.4 weeks pregnant, were randomly assigned into the *PFL* programme (Doyle et al., 2010). Data were collected over time to measure developmental outcomes of children and households. The richness of the data allows many valuable analyses to be undertaken, utilising both cross-sectional and panel dimensions. Due to the small sample size and region-specific nature of the data, anonymising and statistical disclosure controls were carefully performed before the release of the data to the public data archives. The following technical note summarises the motivation, process, and risks of archiving the quantitative data from the *PFL* evaluation. As there are no international standardised protocols in data archiving, precautious steps were undertaken during *PFL* data curation process which may serve as a reference in archiving other longitudinal evaluation data.

Motivations for data archiving

The motivations for archiving and sharing data should have two goals. First, to protect the confidentiality of respondents (Elliot et al., 2016; University College Dublin (UCD) Library, 2017). No information on the identity of respondent or household should be revealed without lawful authority (Office of National Statistics (ONS), 2001; Irish Statue Book (ISB), 1988). Direct identifiers such as names, date of birth, geographic information[8], should be omitted as they reveal respondents' identity. Indirect identifiers, such as occupation, age, or wages, that can be obtained from local knowledge, should be processed carefully. Second, to release useful data where statistically valid conclusions can be drawn (Growing Up in Ireland (GUI), 2013).

Three practical procedures

In processing *PFL* data, anonymity of participants was ensured through three procedures: 1) Small cell adjustments were commonly performed on outcomes that constituted fewer than five observations, including zero observations, which can easily compromise confidentiality protection (ONS, 2001). These variables from *PFL* include sensitive information on drug use during pregnancy, domestic risks, multiple pregnancies[9], specialists' consultation, and postnatal depression. Due to the low number of reported observations, these variables were removed entirely from the archived dataset. 2) Extensive banding was applied to socioeconomic data such as occupation, level of education, and ethnic background, as respondents could be identified through cross tabulation (ONS, 2015). For example, instead of specific job titles, occupations of mothers, fathers, and grandparents, were grouped into broad categories following the Standard Occupational Classification 2010 (SOC2010). These categories are comparable to census data in the UK and Ireland (ONS, 2015; UKAN, 2013). In addition, rather than reveal the ethnic backgrounds of the minority in the sample, maternal ethnic group was broadly re-categorised as Irish and non-Irish. 3) Top and bottom coding was applied to information related to income, demographics, and the household. Where an individual or household output was an outlier, the statistical output was amalgamated into neighboring sample groups, such as age of first pregnancy that was "below 17 years", wages that were below or above a certain level, and family size that was "greater than seven" (GUI, 2013; ONS, 2001). While one can follow national or international guidelines on banding, small cell adjustments and top and bottom coding thresholds are rather data-driven.

Potential challenges

Since the thresholds constructed are data-specific, they can also constitute problems in the process of anonymisation. Depending on the utility of the data, the cut-offs being created should be useful for the purpose of socioeconomic analysis without compromising one's identity. For instance, the age range in the *PFL* dataset is between 16 and 38 years. This range would allow one to conduct analysis based on the standard categorisation of youth who are between 15 and 24 years. However, the outliers need to be removed as one may piece together several variables and identify the respondent from a small sample size. Another common problem in data processing is the issue of identifying missing observations or zero values. This requires understanding of the dataset, such as attrition due to social processes, or the existence of skip pattern in the survey. In the archived dataset, missing values are handled with caution as they can affect the final psychosocial scores.

Due to the nature of longitudinal data, extra risks in disclosing one's identity may include changes in demographic variables, such as change of marital status over the course of data collection period (UKAN, 2013). Thus, it is important to perform tabulations and cross tabulations to ensure the anonymity of individuals or households in cross-sectional and panel dimensions is maintained.

Conclusion

As a final note, a thorough and clear audit record documenting these procedures should be kept. For instance, Stata users can keep track of all relevant anonymisation activities, processes, and notes being performed on each wave of data in their .do files[10] (Stata Press, 2017). A clear audit record is useful in demonstrating the correct procedures and tracking mistakes.

In archiving the *PFL* longitudinal dataset, it made the evaluative data, that is rich in demographic content and contains measurement of children's development over time, available to the public. The above stated procedures have proven useful in preserving the identity of programme participants while making the evaluative results useful for researchers and policy makers. In sharing the experience of archiving the *PFL* quantitative dataset, this technical note will hopefully be deemed useful to the research community.

[8] The PFL original dataset contains detailed information such as the names of hospital, daycare, and location of interview. These are information that can compromise respondents' identity.

[9] Due to the small sample size and only a handful of respondents have more than five pregnancies, information such as maternal age beyond the fifth pregnancy is not disclosed in the archived dataset.

[10] A .do file is a text file that can be executed by Stata. Similar to a diary, the programmer can write all the commands along with their notes in this text file. It would then run by itself when one press the command 'do'.

References

Doyle, O., McNamara, K., Cheevers, C., Finnegan, S., Logue, C., and McEntee, L. (2010) *Preparing for Life Early Childhood Intervention. Impact Evaluation Report 1: Recruitment and Baseline Characteristics.* UCD Geary Institute Discussion Paper Series; WP 10 50. [Online] Retrieved from: http://ideas.repec.org/p/ucd/wpaper/201050.html [assessed 9 September 2017].

Elliot, M., Mackey, E., O'Hara, K., and Tudor, C. (2016) *The Anonymisation Decision-Making Framework. Manchester*, The United Kingdom Anonymisation Network (UKAN) Publications. [Online] Retrieved from: http://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf [assessed 9 September 2017].

Growing Up in Ireland (GUI) (2013) *Infant Cohort: Design, Instrumentation and Procedures for the Infant Cohort at Wave One (9 months)*. Technical report number 2. [Online] Retrieved from: http://www.esri.ie/pubs/BKMNEXT252.pdf [assessed 9 September 2017].

Irish Statue Book (ISB) (1988) Data Protection Act. Number 25 of 1988. [Online] Retrieved from: http://www.irishstatutebook.ie/eli/1988/act/25/enacted/en/html [assessed 15 September 2017].

Office of Nation Statistics (ONS) (2001. Census 2001: Definitions. London, Office of National Statistics, General Register Office for Scotland, Northern Ireland Statistics and Research Agency. [Online] Retrieved from: https://census.ukdataservice.ac.uk/media/51185/2001_defs_intro.pdf [assessed 9 September 2017].

Office of National Statistics (ONS) (2015) 2011 Census UK Comparability. Office of National Statistics. [Online] Retrieved from: https://census.ukdataservice.ac.uk/media/450149/2011censusesukcomparabilityreport.pdf [assessed 9 September 2017].

Stata Press (2017) User's Guide. Chapter 16: Do-files. Stata Press. [Online] Retrieved from: [Online] Retrieved from: https://www.stata.com/bookstore/users-guide/ [assessed 15 September 2017].

UK Anonymisation Network (UKAN) (2013) European Union Statistics on Income and Living Conditions (EU-SILC). [Online] Retrieved from: http://ukanon.net/wp-content/uploads/2015/09/EUROSTAT-EU-SILC-DATA-Nov-2013-pdf.pdf [assessed 9 September 2017].

University College Dublin (UCD) Library (2017) *Research Data Management: Ethical Issues.* [Online] Retrieved from: http://libguides.ucd.ie/data/ethics [assessed 9 September 2017].

Author information

I am a PhD student at University College Dublin with interest in labour economics, applied econometrics, and international development. Stemming from a deep interest in human development and social justice, I have empirical and policy research experience on topics such as migration, poverty, education, and health.

With proven professional track record on related issues, I undertook a range of roles at the International Labour Organisation (ILO), International Organisation for Migration (IOM), United Nations Economic Commission for Europe (UNECE), and Eurofound.

At Preparing for Life, I provided support in data anonymisation and curation, to be deposited in the national data archives.

Challenges and lessons learned from archiving and anonymising quantitative research data for secondary analysis

Personal experience

Shane Leavy

In 2016 I was approached by the Children's Research Network to prepare for secondary analysis three quantitative projects that were part of the Incredible Years Ireland Study. This was a research project exploring the long-term outcomes of the Incredible Years Parent and Teacher Classroom Management programmes, undertaken by several of my current colleagues in Maynooth University's ENRICH project[11]. My task was to identify the appropriate data files, standardise formatting and naming conventions, and anonymise data, editing where necessary to protect the anonymity of participants and remove harmful information.

[11] EvaluatioN of wRaparound in Ireland for CHildren and families (ENRICH) is an ongoing five year, multi-component research programme designed to help promote child health and family wellbeing early in life, involving the evaluation of two wraparound-inspired models of care. For more information on ENRICH and similar studies, see the Centre for Mental Health & Community Research website at www.cmhcr.eu.

Identifying the data

An initial challenge was the identification of the relevant data files. No final, completed folder had been assigned with all SPSS data files; instead a number of data files were spread across several folders, often versions of the same file with very minor differences. Seeking the correct, final file was complicated further by the fact that the control group did not receive the second follow-up interview, which meant that there was no one merged file containing all the data for intervention and control at all three-time points. At the start of the project I was unaware of this and could not understand why the numbers appeared not to match. In retrospect, this was a simple mistake and it would be important to clarify the numbers expected in each wave of the survey before identifying the data files. This highlights the desirability of a data management plan.

Searching for the data files through a complicated list of folders and sub-folders containing a great many slightly differing data files was most time-consuming. Research data managers would do well to identify completed data files and place them in a clear unique folder to aid archival work in the future. Perhaps data managers could create an Excel file to serve as a map to the many folders and sub-folders involved in the project, describing and linking to important data.

A similar challenge lay in the compilation of all instruments (questionnaires, etc.) used in interviews. These documents were generally filed into a small number of folders, which was helpful, but there was still some scrambling to identify which documents applied to which project. The Incredible Years study featured three research projects, each of which had different questionnaires and scale measures; I was fortunate that a member of the Incredible Years research team was my colleague in ENRICH and her help in identifying relevant files was invaluable. Without the support of a member of the original research team, distinguishing between the slightly different questionnaires used in the three separate projects would have been difficult, which perhaps shows that the work to prepare anonymised research data for archiving can helpfully begin during the research project itself. At least, good housekeeping by researchers, leaving reference files in clearly-named folders, can greatly speed future work on archiving.

Naming conventions

Variables in the archived data were to be standardised to a simple formula: root/item number/suffix. For example, a Profile Questionnaire question 3b at baseline is rendered: PQ3bT0. The same question in the first follow-up survey is: PQ3bT1.

To individually change several hundred variables in each wave would take an enormous amount of time. Instead, I used some useful formulae in Excel to speed this up. Consider the original variable name for the Profile Questionnaire 3b at baseline: PQ_3b. This contains the "PQ" and the "3b" I want to include, but with an unnecessary underscore. I used the following formulae in Excel:

	A	B	C	D	E
1	PQ_3b	=LEFT(A1,2)	=MID(A1,4,1000)	T0	=B1&C1&D1
2	PQ_3c	=LEFT(A2,2)	=MID(A2,4,1000)	T0	=B2&C2&D2
3	PQ_3d	=LEFT(A3,2)	=MID(A3,4,1000)	T0	=B3&C3&D3

Table 1: Naming converntions table

The LEFT formula reproduces the leftmost characters of a cell. In this case, LEFT(A1,2) takes the two leftmost characters "PQ" from cell A1. The MID formula does the same thing, but starting a stated number of characters into the source cell, i.e. in this case MID(A1,4,1000) reproduces the 1,000 characters in the middle of cell A1, starting at character 4. I selected 1,000 characters as an arbitrary large number, simply to capture all the characters after the starting point. In column D I entered "T0", representing baseline. The final formula simply merges the other three: PQ and 3b and T0 becomes PQ3bT0. The table below shows how these cells look in Excel:

It is a simple task to apply such formulae to the entire list of variables. This kind of methodology allows fairly rapid standardisation of hundreds of variables.

Missing values

Many variables may have missing values either because an answer is not applicable, refused by the respondent, respondent answered "don't know", or for some other unknown reason. It can be important for analysis to distinguish between cells appropriately left blank because they were not applicable to that respondent and cells left blank because of the refusal of the respondent to reply or some other reason. I did not have access to original paper copies of the surveys, but usually it was clear if the missing value represented a valid "not applicable" response. Where such responses were already identified, a simple piece of SPSS syntax could replace blank cells with the number 96, representing undefined missing values.

	A	B	C	D	E
1	PQ_3b	PQ	3b	T0	PQ3bT0
2	PQ_3c	PQ	3c	T0	PQ3cT0
3	PQ_3d	PQ	3d	T0	PQ3dT0

Table 2: Naming converntions table

Anonymisation

By far the most onerous part of data archiving was the anonymising of variables, where risk assessment was based on two criteria: risk of identification and risk of harm. In particular, many variables featured open-ended textual questions, any of which could include personal details like names or private information on behavioural problems or illness, faithfully taken down from the respondent by the interviewer.

In a few cases it was appropriate to simply delete the variable, including for example questions about names of schools, teachers or phone numbers. It was decided, however, that most string variables were too valuable to omit entirely so there was no option but to read every single response and systematically recategorise them to conceal harmful or identifying data while preserving useful information. For example, a question may ask parents about the medication being used by their children. Below are fictional examples of responses:

- Asthma
- Allergy medication
- Inhaler
- She had breathing problems and was recently prescribed an inhaler by Doctor O'Malley. Better now
- ADHD
- Azelastine
- Methylphenidate

These diverse responses represent three basic categories: asthma medication ("asthma", "inhaler", and "She had breathing problems and was recently prescribed an inhaler by Doctor O'Malley. Better now."), allergy medication ("allergy medication" and "Azelastine") and ADHD medication ("ADHD" and "Methylphenidate"). The response describing the prescription of medication by a Doctor O'Malley shows how identifying information can be included in unexpected variables, illustrating the need to read every string variable and recategorise many into categorical variables.

In the fictional example above, some parents responded with exact names of medication like azelastine and methylphenidate, while others knew only the general illness or condition being treated. Since I am not knowledgeable about these forms of medication, I had to quickly browse the internet for corporate or scientific names to check their general purpose.

In some open questions participants gave several replies. There are a number of possible solutions to this. One could split the string variable into several categorical variables, or use SPSS's Multiple Response Sets, which similarly require the generation of categories from string data. I generally chose the former, generating up to three categorical variables from the one string variable. For example, supposing parents were asked to list any concerns about their child. Below I give a fictional example to show how these questions could be answered in the full string variable, and then how I categorised them into further variables.

In this example some respondents give three answers, some give two and some just one. I split these answers into three categorical variables, listing "not applicable" for the second and third variable where no answers were given. Note that the fourth example includes some potentially identifying information in the name of the sister Ellie, again illustrating the importance of categorising or editing these variables.

The Multiple Response Set command in SPSS follows a similar methodology. Categories are derived from the text variables and each category becomes a new binary (yes/no) variable.

This process of anonymising string variables was extremely time-consuming, involving large numbers of decisions on every variable. Individual responses could sometimes be ambiguous and could potentially fit in different categories. A response "shouts, screams, not good vocabulary" could feasibly fit in either the aggression/temper or speech development categories. While time-consuming and onerous, at least the data processor here

Original string variable	Category 1	Category 2	Category 3
'Hitting, biting, very aggressive. Nervous of strangers. Problems with sleeping.'	Aggression/ temper	Social skills	Sleeping
'Temper tantrums, out of control. Not speaking much, poor vocabulary.'	Aggression/ temper	Speech development	Not applicable
'Wakes often at night.'	Sleeping	Not applicable	Not applicable
'Very fussy eater. Fights sister Ellie often, hits other children.'	Eating	Aggression/ temper	Not applicable

Table 3

should try to be consistent, making all decisions on consistent criteria across the data.

Conclusion

I know from my own experience working with data on a research project that files tend to multiply and become difficult to organise. Working on one's own data, at least one has a shot at remembering past decisions in organising files and folders. This is much more complicated when coming fresh to other researchers' folders: ad hoc decisions made by data managers are invisible to those archiving old research projects.

Even the decisions made in anonymising data often required knowledge of the field. External data processors attempting to prepare old research projects are disadvantaged by their lack of local knowledge on the project. In the examples I gave above, I would not automatically know whether future analysts of the data would prefer to know brand names of medication or general areas of illness; this is a question answered by researchers or practitioners in the field.

All this suggests that a simple piece of organisation, undertaken by the data manager or relevant researchers towards the end of their project, could be very helpful for future archiving and analysis. An ideal situation might even involve the inclusion of data archiving into the timeline and budget of the project, allowing the original research team with their specialist knowledge to make the relevant decisions that both protect the anonymity of participants and preserve the most valuable data for future analysis. Indeed, such plans are now commonplace and required by major funders, such as the British ESRC and Horizon 2020, and it is likely that these will facilitate future archival projects and the productive secondary analysis of archived data.

Author information

Shane Leavy is the Data Manager and Research Assistant with the EvaluatioN of wRaparound in Ireland for CHildren and families (ENRICH) research project at Maynooth University. He also currently prepares and analyses work-related accident data with the Health and Safety Authority, coding every fatal work-related accident in Ireland since 1989. He has previously worked in data analysis and research with the Economic and Social Research Institute, including an internship with the Growing Up in Ireland project, and print journalism, writing for publications in Ireland and abroad.

Data management for archiving

Keeping the secondary data analyst in mind

Seamus Fleming & Liam O'Hare

This article touches on some basic processes relating to Data Management (DM) that aid straightforward Data Archiving (DA), and facilitate future secondary data analysis (SDA). Prevention and Early Intervention (PEI) projects can be logistically challenging and costly, with limited resources for exhaustive data analysis. Therefore, it is prudent to archive the data for subsequent researchers to analyse and extend the research beyond the scope of the original PEI project. It is becoming common practice for funding organisations (e.g. the Education Endowment Foundation, 2017) to request relevant provision, and planning for DA (Van den Eynden, Corti, Woollard, Bishop and Horton, 2011).

Researchers, who have both experience of conducting PEI projects and preparing data sets for archiving, write this short commentary. Examples of previously archived projects, from the Irish Social Science Data Archive (ISSDA, 2017), are included to highlight some pitfalls of the DM process and provide tips for more efficient DA.

When a data analyst works with a secondary dataset, relevant data information must be available. Variable names must be appropriate and in a logical convention, data labels should be relevant and consistent, and contain precise values. However, if this essential information is missing, the DA process becomes labor intensive and requires the researcher or archivist to rework the data file before archiving.

In the case of the Mate-Tricks dataset, from a randomised controlled trial evaluation of an afterschool programme (O'Hare et al. 2012), there were over 800 variables, therefore a logical and consistent variable naming convention was implemented with appropriate data labels and explicit data values. Variable names may be in a format that allows the person who created them to understand what they relate to (as they set up the data file), but they may not be explicit enough to inform a subsequent user to understand what they represent. Example variables in the Mate-Tricks dataset are CPT_TEI_1 through to CPT_TEI_75. The CPT refers to Child Post Test (time point of measurement), TEI denotes the Trait Emotional Intelligence questionnaire (the measure used), and 1 to 75 represents the item number on the measure. When a standard variable naming convention is not used, the variables will have to be renamed, which requires additional work. If data labels and values are missing, then the data archivist will also need to refer to the measures/questionnaire in order to modify the variable characteristics in the data file.

These basic DM procedures are implemented easily, and when done as a matter of course they allow for straightforward DA and efficient navigation of the variables for the secondary data analyst.

By investing a small amount of time, adhering to basic DM procedures mentioned above during the data preparation stage, the DA process becomes less time consuming, and more beneficial to an SDA. Consequently, the data will help extend research and can provide a direct credit to the researcher as a research output in its own right (Van den Eynden et al., 2011).

References

Education Endowment Foundation (2017). Available online at: https://educationendowmentfoundation.org.uk/our-work/resources-centre/submitting-your-data-to-the-fft-archive/

ISSDA (2017) Available online at: https://www.ucd.ie/issda/

O'Hare, L., Kerr, K., Biggart, A., & Connolly, P. (2012) *Evaluation of the Effectiveness of the Childhood Development Initiative's Mate-Tricks' Pro-social Behaviour After-school Programme.* CDI, Dublin. Dataset available online at: www.ucd.ie/issda/data/cdimate-tricks/

Van den Eynden, V., Corti, L., Woollard, M., Bishop, L.Horton, M. (2011) *Managing and Sharing Data.* Available online at www.data-archive.ac.uk/media/2894/managingsharing.pdf

Author information

Dr Seamus Fleming is a Research Assistant in the Centre for Evidence and Social Innovation at Queen's University Belfast. He is currently a member of the Children's Research Network for Ireland and Northern Ireland (CRNINI). His past research focussed on general psychopathology and the effect of adverse environmental experiences, such as traumatic events, on mental health outcomes. He is currently working on the data archiving of a series of past CRNINI-PEI projects to the ISSDA and is interested in research relating to prevention and early intervention in the area of child education.

Dr Liam O'Hare is a Senior Research Fellow in the Centre for Evidence and Social Innovation at Queen's University Belfast. He has been a member of the CRNINI committee since 2014 and he is currently the Chair of the Committee. Liam has worked in children and young people's research for over 15 years. His main area of interest is the design, implementation and evaluation of prevention and early intervention and programmes. He has published widely and received numerous grants to carry out work in this area. He is also currently Principal Investigator on a research project to archive a series of prevention and early intervention data sets with the ISSDA.

You lost me at hello

A non-research perspective on sharing data

Marian Quinn

Introduction

I'm not into data really. Statistics don't do it for me. Facts and numbers just aren't my thing. And archiving is definitely not something I get excited about. The suggestion of archiving conjures up for me images of fusty libraries with tomes of dust covered ledgers; the Maesters in Game of Thrones, cloistered away for years on end, hoping to find the secret to youth in those leather- bound books or amidst the jackets of ancient manuscripts. I have however, through the journey I'm about to share with you, met quite a few people who really do get animated about archiving. And they are generally very nice, quite normal people! For them the mere whisper of a data cleansing plan, the whiff of anonymised quants, or a hint of cumulative comparative processes, has them chomping at the bit to get down and dirty with the data.

Whilst data may not light my fire, I am unequivocally passionate about sharing information. My friends tell me I don't know when to stop! I get frustrated when we reinvent wheels; I am irritated when we don't adequately learn lessons, and I find it inexcusable that we hold insights close to our chests for fear of diminishing their value, when actually, opening them up for others to view and scrutinise would increase the knowledge multiple times over. I may not be good on detail, but the concept of an effective archiving process does resonate with me as something which is fundamentally important.

Of course, when we started out on this journey ten years ago, we hadn't thought far enough ahead to know that we would want or need an archiving process; we just knew that we wanted to ensure that whatever we learned, whatever data we gathered, and whatever implications these had for policy and practice, would be disseminated thoughtfully, transparently, accessibly and honestly. Simple!

Background

The Childhood Development Initiative (CDI) was formally established in 2007, after a three year consultation and assessment process with those living and working in Tallaght West (TW). Funded through the Department of Children and Youth Affairs (DCYA), and The Atlantic Philanthropies (Atlantic), CDI set out to design, deliver and evaluate a suite of interventions aimed at improving outcomes for children and families. In our first year of operating, we began an EU wide tendering process to commission eight independent evaluations, including three randomised controlled trials and two quasi-experimental studies, with a total spend of just under €2 million committed to various academic institutions by the end of 2008. Since then, we have commissioned a number of further studies and spent €2.7 million, or almost 12.5% of our total budget on research related activities. This comprehensive and often complex process was enabled by leadership from our Board of Management, and an Expert Advisory Committee (EAC) overseeing and advising the research programme. In addition, Atlantic provided legal guidance in relation to copyright and intellectual property rights (IPR) which subsequently proved to be invaluable.

Atlantic required that all our contracts stated that IPR would be held by CDI, which would in turn grant the researcher:

> A perpetual, transferable, non-exclusive, royalty-free licence (carrying the right to grant sub-licences including to the Department of Children and Youth Affairs and The Atlantic Philanthropies (Ireland) Limited) to use for any purpose any such Intellectual Property Rights.

In essence, this clause ensured that CDI would own any intellectual property, including qualitative and quantitative data, produced through or arising from, research and evaluations funded by us. In 2008, it was extremely unusual for IPR to be held by anyone other than the researcher, and even more unusual for contracts to explicitly state IPR

ownership in favour of the commissioning agent. The implications of this were not immediately apparent to the various research teams with whom we worked, and it wasn't for a couple of years that we realised ourselves how important this clause was.

Unravelling the implications

Given the number of evaluations commissioned, and the fact that there were so many academic institutions and Principle Investigators involved, for the first three years or so, CDI brought the eight evaluation teams together on a quarterly and then half-yearly basis. These meetings proved to be extremely useful for the purposes of scheduling data collection, identifying synergies between approaches and maximising the utilisation of the information being collected. They were also informal networking opportunities and created a sense of camaraderie or shared purpose amongst many of the research teams.

It was at one of these meetings, in early 2009, that the implications of the contractual commitment first became apparent for many of the research teams. In discussing the eventual use of the data, and the potential dissemination plans, one academic suggested that this would require the consideration of their ethics committee, as the potential for archiving data had not been agreed as part of their ethics approval process. CDI noted that our ownership of the IPR determined that the dissemination process was, in effect, in our hands. The ensuing discussion highlighted a very significant difficulty for some of the evaluation teams because the process of agreeing, scrutinising and signing off on the contract with CDI had been undertaken by people who had no connection to or dialogue with those who oversaw the ethical approval process. Likewise, the ethics committees did not review contracts or consider their content in any way. There was an almost universal separation of roles in terms of the legal considerations and those relating to ethical approval. And so, we had a situation where individual evaluation teams had signed a contract with CDI which contradicted the terms of their internal ethical approval agreements.

Inevitably, unravelling these complications took some time, and to date they have not been universally resolved. Establishing a legal context in which CDI was able to drive a significant and comprehensive dissemination strategy required considerable time, effort and expertise, but we were fortunate: we had money! We bought in expertise to undertake the following tasks:

- Review the IPR and legal context of CDI's research and evaluation data; identify those which could potentially meet the required standards for submission to a national archive, and set out a plan for progressing the archiving of each available dataset;
- Write an Archiving Toolkit based on CDI's experience, to enable others to replicate the process (http://www.twcdi.ie/wp-content/uploads/2016/11/CDI-Sharing_Social_Research_web.pdf)
- Engage with and support each individual research team to archive relevant data.

In a similar, parallel process, we were invited to engage in a project with the Irish Qualitative Data Archive (IQDA), which entailed being the pilot site for archiving qualitative data, and supporting the documentation of the process and lessons learned (https://www.twcdi.ie/resources/publications/). Resulting from a connection with a member of our EAC, this project really cemented our commitment to engaging in and driving effective archiving of our own research and maximising the capacity of others to also do so.

Critical factors

Commissioning Dr Brid McGrath and Robin Hanen to support the archiving of internally held data, and CDI owned data held by independent evaluation teams was critical to progressing this initiative. Their expertise and detailed understanding of the sometimes very

technical tasks to be completed (such as the management of imputed data, or developing a consistent approach to metadata which could be consistently applied across all archives) were invaluable. However, their passion for the work, their ability to get excited by the complexity of numerous seemingly disparate datasets, and their ability to communicate the potential impact of effectively sharing this information was not only energising but transformative. I was converted!

Being surrounded by others who had the knowledge, skills and feelings to support a long, slow, sometimes very frustrating process was also vital. The required knowledge related to a thorough understanding of research and data, but also of how organisations work, and the kinds of dynamics which can enable collaboration or mitigate against it. The skills needed were those relating to being solution focused, and having the capacity to work around obstacles. The relevant feelings were to do with having the commitment, enthusiasm and motivation to not only do the right thing but also to do it right.

Finding the right people for a particular job can sometimes be a matter of luck, but the chances of doing so are significantly increased if you know what you're looking for. In this case, through our engagement with the IQDA, coupled with the expertise within our EAC, and drawing on our experience with the evaluation teams to date, we were able to draw up a fairly concise and precise tender document. This undoubtedly informed our selection of the archiving supports.

In addition to the above, for an organisation to commit to sharing its data, and having all its research products and findings shared and laid open, requires a particular ethos and mind-set. Not all organisations will readily agree to share the data, irrespective of the research findings or conclusions. This is not a process which can be readily adapted deepening on the stomach for it so up-front discussion about 'what if…' is needed. What if…the research doesn't find any change? What if…the findings reflect poorly on us? What if…the evaluation indicates that our work didn't do what it set out to do?

Dialogue and reflection throughout the organisation to consider these possibilities is therefore an essential component of an archiving strategy.

Lessons learned

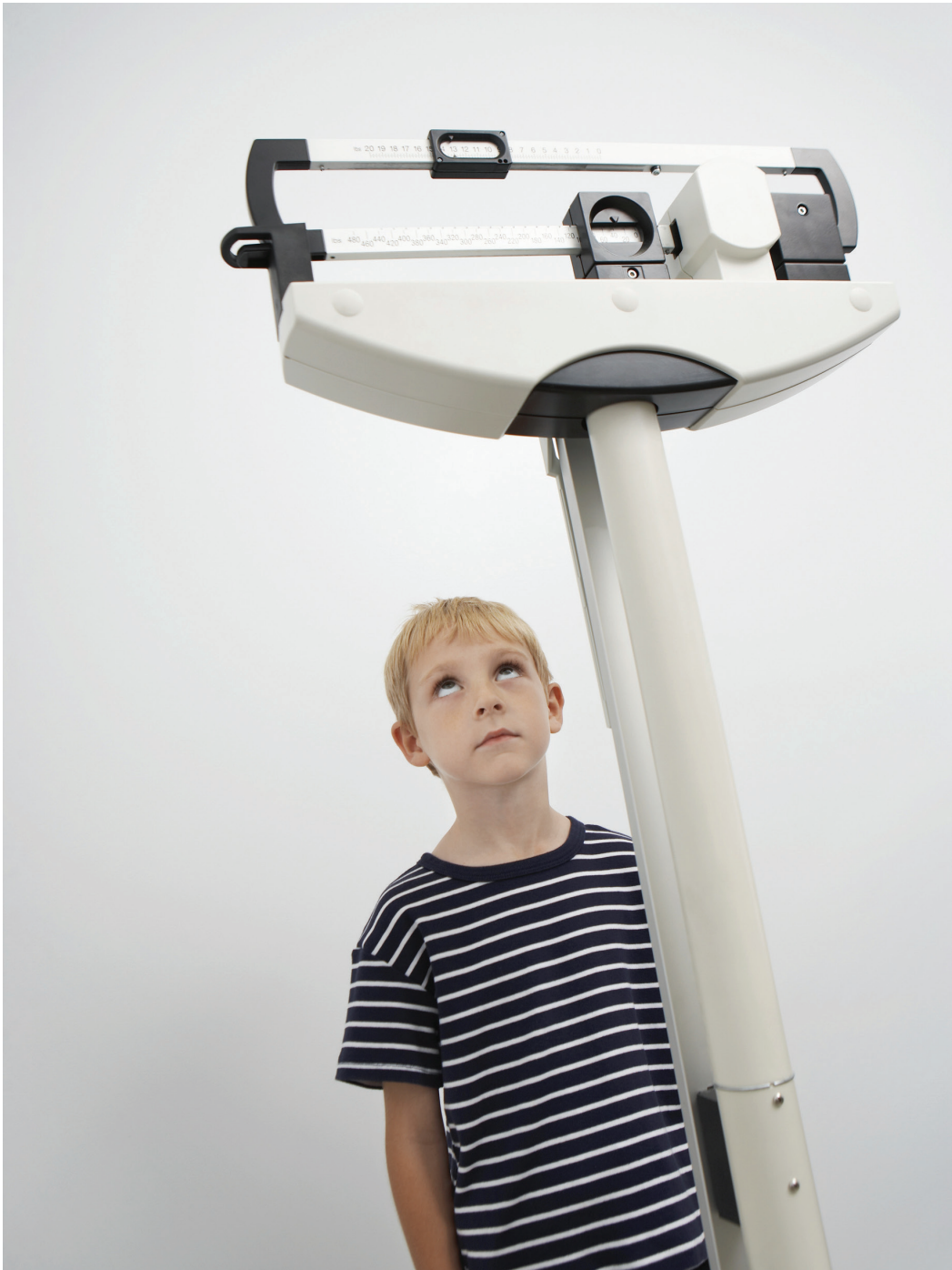
The central mechanism to enable effective archiving is to build in the possibilities for this from the outset. When the study is complete, the numbers crunched and conclusions drawn, it may be agreed that there is little of value to be archived; that the anonymisation process is too complex to undertake, or would produce relatively meaningless information; the findings may be indeterminate or dull, and the available effort to disseminate them reduced accordingly. None of this can be predicted, so start with the assumption that you will want to archive as much as possible. This will shape your ethical approval and consent processes; it may even inform your methodology in terms of using standardised surveys or the balance of qualitative vs quantitative data.

Clarity of ownership is also vital, and this should be explicit in any service level agreement or contract, and ideally be made clear from the outset of the procurement process.

Finally, as with pretty much any development, surround yourself with the best! Source people with relevant expertise, and bear in mind that often this doesn't require funding. Most people are delighted to share their insights, especially if this will support a process aimed at knowledge transfer.

Author information

Marian Quinn is Chief Executive of the Childhood Development Initiative (CDI), an organisation co-funded by Government and philanthropy, to design, deliver and evaluate a range of services aimed at improving outcomes for children and families. Marian previously worked in the Department of Justice where she had responsibility for children and families in the asylum process, and the Health Services Executive as Director of Children's Services. She has also worked in the voluntary sector with early school leavers and young people at risk. She wrote and managed a national crime prevention initiative, and has been published widely in relation to youth crime. She is a qualified Life Coach and has a Masters in Adult Education. She is a member of the Board of Management for the new National Childrens' Hospital and also the Board of the Airfield Trust, and is currently Chair of the Prevention and Early Intervention Network. Marian is co-author of Click Click, the number one best seller telling the true story of the three Kavanagh sisters and their journey to overcome child sex abuse.



Ethical challenges within a Healthy Schools Project and lessons learned

Catherine Comiskey

Background to the Healthy Schools evaluation

The Tallaght West Child Development Initiative (TWCDI) Healthy Schools (HS) programme sought to improve children's overall health outcomes and increase their access to primary care services. The programme was developed as a result of previous research conducted by TWCDI (2004) which identified the health needs of children living in Tallaght West. The HS programme was a manualised initiative which was based upon seven primary outcomes. These were that: (1) children demonstrate age-appropriate physical development; (2) children have access to basic health care; (3) children are aware of basic safety, fitness and health care needs; (4) children are physically fit; (5) children eat healthily; (6) children feel good about themselves and; (7) parents' involvement in their child's health.

The principal objective of this study was to evaluate the implementation and outcome of the Healthy Schools programme. The evaluation was a longitudinal comparative study which followed the children and all key stakeholders from intervention and comparison schools throughout the implementation of the Healthy Schools Programme. The evaluation was divided into two components: (1) an examination of the health outcomes for children, and (2) a process evaluation of the programme.

The sample frame consisted of children attending junior infant class to fifth class in five intervention schools and two comparison schools. All schools were DEIS Band 1 schools. The intervention schools self-selected in liaison with Tallaght West Childhood Development Initiative (TWCDI) prior to the commencement of the evaluation study.

In each intervention school the principal was asked to complete an interview to identify their understanding and views of the implementation of the Healthy Schools programme. Interviews were also carried out with the two Healthy Schools Coordinators, the Director of Public Health Nursing as well as two members of staff from CDI to examine the rollout of the HS programme.

All participating children (from Junior to Fifth class) had their BMI measured by a qualified nurse and member of the research team during school time. Self-report questionnaires (Kidscreen 27, Health Related Behaviour Questionnaire (HRBQ) and the Childhood Depression Inventory (CDI)) were completed by children or their parent. Outcomes for all participating were measured at baseline, 12 months and 24 months follow-ups. Details of baseline findings are available from Comiskey, C. M. O'Sullivan, K., Quirke, M., Wynne, C., Hollywood, E and McGilloway, S. (2012).

Ethical and contractual challenges and solutions

The study was primarily carried out by Trinity College Dublin (TCD), however the National University of Ireland, Maynooth was subcontracted for part of the research. The signing of the main contract between TCD and TWCDI took a considerable amount of time as it was important that all matters of existing and future intellectual property were agreed. The final contract was a service agreement as opposed to a research contract and was agreed by solicitors for both parties following a face-to-face meeting. All parties were satisfied and contracts were signed after a period of approximately six months. NUI Maynooth then received an identical sub-contract from TCD and all parties were legally contracted and protected.

Prior to the initiation of any research or data collection in the study locations, the study, its design, instruments, processes, methodology and all letters of introduction, information leaflets, information posters for participating schools and consent forms received ethical approval from the Faculty of Health Sciences, Trinity College Dublin. The ethics committee of the Faculty is a legally constituted committee which reviews applications from the four constituent Schools of the Faculty. These include the Schools of Nursing

and Midwifery, Medicine, Dentistry and Pharmacy and Pharmaceutical Science. As applications for ethical approval to the Faculty often involve vulnerable patient groups and new treatments the Faculty has strict legal guidelines to which it must adhere.

The Healthy Schools project received full ethical approval and the research began and was completed on schedule and within budget. Details of the main outcome results are available within Comiskey, C.M., O'Sullivan, K., Quirke, M.B., Wynne, C., Kelly, P. and McGilloway, S. (2012). Details of the process and implementation evaluation are available within Comiskey, C. M. O'Sullivan, K., Quirke, M., Wynne, C., Hollywood, E and McGilloway, S. (2015). Finally results on body mass index and health related quality of life are presented by Hollywood E., Comiskey, C.M., Snel, A., O'Sullivan, K., Quirke, M., Wynne, C. (2013); Wynne C, Comiskey C, Hollywood E, Quirke MB, O'Sullivan K, McGilloway S. (2014) and Wynne C, Comiskey C, McGilloway S. (2015).

The ethical challenges which we faced began after the study was completed and the contract successfully executed. As over 600 children were recruited and measured at baseline, 12 months and 24 months and many parents, teachers and service providers interviewed, the study team found that it had a wealth of additional detailed data that could be analysed at a more detailed level and could provide additional evidence of the health and wellbeing of the children. For example, we had data on bullying and while rates were reported on within the reports and outcome papers no detailed analysis of the bullying data was undertaken. The study team wished to open up the database for sharing with other academic professionals. When the study team returned to the Faculty ethics committee for advice on sharing the anonymised quantitative data on a national data repository we were advised that this posed an ethical dilemma as the original signed consent from parents, teachers and key stakeholders and verbal assent from children did not include notice that the data would be archived. We were advised that we may need retrospective consent as we had not planned for archiving. Given we had over 600 children who had passed through the study and many of whom had since moved on to secondary schools this was not a feasible option and data archiving could not be permitted.

To overcome this challenge and to allow additional experts to work on the anonymised data the research team expanded and additional visiting researchers were invited to join the team and work on the database within the secured internal Trinity College database. While this did mean additional resources had to be expanded by Trinity College in the form of setting up and approving visiting academics with College identity cards and these visiting academics had to travel to Trinity College to access the data, the process worked and additional analyses were conducted. A fine example of this is the additional analysis of the bullying data, which can be found within Hyland, J., Cummins, P and Comiskey, C.M. (2017).

The internal team is planning further expansion with EU collaborators who have an interest in Healthy Schools and urban disadvantage. To conclude, while we were not in a position to archive the data as we would have wished we were able to compromise and ensure that the data was accessed and used to its best advantage to ensure that the evidence continues to be mined and disseminated for the good of the children, the schools and the families that participated. The key lesson learned by the research team was to ensure that all future large research studies include within the consent form, consent to archive the anonymised data for further data mining by additional bona fide researchers after the study has ended.

References

Comiskey, C. M. O'Sullivan, K., Quirke, M., Wynne, C., Hollywood, E and McGilloway, S. (2012) *Baseline results of the first healthy schools evaluation among a community of young, Irish, urban, disadvantaged children and a comparison of outcomes with international norms.* Journal of School Health Volume 82, Issue 11, pages 508–513

Comiskey, C.M., O'Sullivan, K., Quirke, M.B., Wynne, C., Kelly, P. and McGilloway, S. (2012) *Evaluation of the Effectiveness of the Childhood Development Initiative's Healthy Schools Programme*. Dublin: Childhood Development Initiative (CDI). ISBN 978-0-9570232-4-6 available at http://www.twcdi.ie/wp-content/uploads/2016/11/CDI-HSP_Report_12.11_web.pdf (last accessed 9th October 2017)

Comiskey, C. M. O'Sullivan, K., Quirke, M., Wynne, C., Hollywood, E and McGilloway, S. (2015) *An analysis of the first implementation and impact of the World Health Organisation Health Promoting School Model within urban disadvantaged schools in Ireland.* Vulnerable Children and Youth Studies DOI:10.1080/17450128.2015.1080394

Hollywood E., Comiskey, C.M., Snel, A., O'Sullivan, K., Quirke, M., Wynne, C. (2013) *Measuring Body Mass Index for a cohort of 500 children attending disadvantage schools.* Journal of Advanced Nursing (JAN) Volume 69, Issue 4, pages 851–861

Hyland, J., Cummins, P and Comiskey, C.M. (2017) *Victimisation in urban disadvantaged primary schools: Associations with health-related quality of life, depression and social support* (in press) Child and Adolescent Mental Health

TWCDI (2004) How Are Our Kids? (2004) available at http://www.twcdi.ie/wp-content/uploads/2016/11/2004_How_Are_Our_Kids_Children_and_Families_in_Tallaght_West_Co__Dublin1.pdf (last accessed 9th October 2017)

Wynne C, Comiskey C, McGilloway S. (2015) *The role of body mass index, weight change desires and depressive symptoms in the health-related quality of life of children living in urban disadvantage: testing mediation models.* Psychology and Health DOI:10.1080/08870446.2015.1082560

Wynne C, Comiskey C, Hollywood E, Quirke MB, O'Sullivan K, McGilloway S. (2014) *The relationship between body mass index and health-related quality of life in urban disadvantaged children.* Qual Life Res. 2014; 23:1895-905. DOI:10. 1007/s11136-014-0634-7

Author information

Professor Comiskey is the former Head of the School of Nursing and Midwifery, Trinity College Dublin (2014-2017), founder and inaugural Director of the Trinity Centre for Practice and Healthcare Innovation (2012-2014) and former Director of Research (2008-2012). Comiskey holds a B.A.(Mod) degree in Mathematics and Philosophy from Trinity College, Dublin University and M.Sc. and Ph.D. degrees in biomathematics, biostatistics and epidemiology. In 2007 she was appointed by Minister of Education and Science to serve on the board of The Irish Research Council. From 2011-2014 she served as the Inaugural Chairperson of the Children's Research Network of Ireland and Northern Ireland.

Archiving research data

A case study

Nóirín Hayes

With the rise in attention to evidence-based policy making (Sutcliffe and Court, 2005), the growth of digital technologies (National Academy of Science, 2009) and the realisation of the power of 'big data' (Lynch, 2008) the importance of archiving research data has become a topic of interest internationally. Indeed, many funding organisations, including the Irish Research Council, require applicants for grant awards to show how their data will be made available to other researchers. This has created the need for researchers to take account of procedures for archiving their data from both the ethical and research design perspective.

Where the issue of archiving research data has not been a central consideration at the commencement of a research study it can be difficult to retrospectively adapt it. Such was the case with a number of evaluation studies commissioned in the mid-2000s by the Childhood Development Initiative [CDI]. Included was an evaluation led by the Dublin Institute of Technology [DIT] designed to evaluate the CDI Early Years Programme - a two-year early childhood intervention for children aged between 2 years and 6 months and 4 years. The evaluation comprised both a quantitative assessment of the programme and a qualitative assessment of the implementation process and, as such, had a rich data set with the potential for archiving (Hayes, Siraj-Blatchford and Keegan, 2013).

However, at the time of study design no request for, nor commitment to, archiving the research data was discussed. This was not unusual with community-based evaluations at the time but, as a consequence, no contingency for archiving had been included when developing information and consent literature for participating children, families and settings. To discuss how, if at all, the data from the evaluation study could be archived CDI and the research team met on a number of occasions.

The research team had a number of specific concerns as the study was, at this stage, already well underway. These concerns centred around three key issues, (i) the selection of what data, if any, could be archived, (ii) the ethical issues that archiving raised and (iii) the resources [time and funds] necessary to anonymise the data. Each issue is addressed briefly below.

(i) **Selection of data** – While the archiving of quantitative data is well established and straightforward the situation is more complex with qualitative data. Qualitative data does not lend itself so readily to archiving and re-use and the challenges can be structural, contextual and ethical (Fink, 2000). There is, for instance, the issue of copyright and ownership of data created in a relational context between the researcher and the researched (Parry and Mauthner, 2004). No discussion of copyright had occurred with respondents in advance of data collection. In addition, given the personal and detailed nature of the qualitative data it was felt that the respondents could be easily recognised. The extensive modifications necessary for it to be effectively anonymised would significantly compromise the usefulness of the data.

It was agreed that only the quantitative data could be considered for archiving.

(ii) **Ethical issues** – The ethics of retrospectively preparing the quantitative data for archiving was a consideration. A review of the information sheets, consent forms and ethical approval statements confirmed that it would be necessary to seek additional ethical approval from the DIT ethics committee. To this end the team requested and received

(a) ethical approval to pass anonymised quantitative data on children, families and childcare settings to Tallaght West Child Development Initiative once the evaluation ends

(b) ethical approval for Tallaght West Child Development Initiative to archive that anonymised data for the purposes of future research.

(iii) **Resources** – The original research proposal did not include funding for

archival preparation. However, following the conclusion of the study the DIT research team facilitated CDI through the Children's Research Network in preparing the quantitative data from the Early Years evaluation study for archiving. There is restricted access to this data through the Irish Social Science Data Archive.

References:

Fink, A.S. (2000) *The Role of the Researcher in the Qualitative Research Process: Potential Barriers to Archiving Qualitative Data.* Forum: Qualitative Social Research, Vol1, No 3, Art. 4

Hayes, N., Siraj-Blatchford, Keegan, S. (2013) *Evaluation of the Early Years Programme of the Childhood Development Initiative.* (Dublin: Child Development Initiative)Lynch, C. (2008) Big Data: How do your data grow? Nature 455, 28-29 doi:10.1038/455028a

National Academy of Sciences (2009) *Ensuring the integrity, accessibility and stewardship of research data in the digital age.* Washington DC: National Academy of the Sciences

Parry, O. and Mauthner, N.S. (2004) *Whose Data Are They Anyway? Practical, Legal and Ethical Issues in Archiving Qualitative Research Data.* Sociology Vol. 38(1): 139-152

Sutcliffe, S. and Court, J. (2005) *Evidence-Based Policymaking: What is it? How does it work? What relevance for developing countries?* London: Overseas Development Institute. Available online at https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/3683.pdf

Author information

Nóirín Hayes is Visiting Professor at the School of Education, Trinity College Dublin. Working within a bio-ecological framework and a children's rights lens she researches in early childhood education with a particular interest in curriculum and pedagogy. She has authored many books, reports and research articles and has served on a number of government working groups, commissions and advisory groups. Her most recent books include *'Early Years Practice: Getting it Right from the Start'* (2013) and, with co-authors O'Toole and Halpenny, *'Introducing Bronfenbrenner: A Guide for Practitioners and Students in Early Years Education'* (2017).

Secondary data analysis with young people

Some ethical and methodological considerations from practice

Leonor Rodriguez

Introduction

Carrying out secondary data analysis poses various ethical and methodological dilemmas for researchers including issues with informed consent, so why do it?

The reality of the current times is that data archives exist now and researchers need to be aware of the changes and adaptations that need to be effected to the design, methodologies and implementation of research to be able to respond and adapt efficiently to this new reality. Data archives are an innovative way of disseminating data and providing material than can be effectively used for research and teaching (Bishop, 2009; Bishop, 2012) by utilising information technologies (Parry and Mauthner, 2004).

Research has identified benefits and advantages of carrying out secondary data analysis specifically. One of these benefits is the expectation of research data to be transparent and reproducible as open data can increase evaluation and reproduction of research (Roche et al., 2014) as well as provide a platform for auditing the quality and reliability of research data and findings.

Research is often publicly funded and, therefore, secondary data analyses can be a way of aggregating value to the original investment, increasing the cost-efficiency of funded research (Camfield and Palmer-Jones, 2013; Roche et al., 2014). Another benefit of secondary analyses is that it prolongs the use of data over time and participants will often engage in research not only to contribute to a particular project, but also to the broader body of knowledge (Bishop, 2013). Researchers have a duty to benefit society through their work by improving policy and practice; secondary data analysis can be another means to achieve this goal with reduced risk to participants (Bishop, 2009).

This secondary data analysis was carried out with young people involved in the Big Brother Big Sister mentoring programme, which have been described as a 'vulnerable' population which may experience poor social skills, low self-esteem and/or economic disadvantage (Dolan, Brady, O'Regan, Brumovska, Canavan and Forkan, 2010). This posed further challenges to the justification of carrying out such an analysis despite the criticisms and challenges that exist towards re-using data.

The original research study evaluated the benefits of mentoring relationships between an adult and a young person by focusing on the role of social supports, emotional well-being, education, risk behaviour, relationships and outcomes of matching a young person and an adult (Dolan, Brady, O'Regan, Brumovska, Canavan and Forkan, 2010).

This secondary analysis was focused on exploring one aspect of mentoring relationships that had not been explored previously: the role of empathy in mentoring relationships. It was, therefore, considered that the context of both research questions was similar enough for the data to be valuable and useful in providing further understanding of what the primary data had already achieved about mentoring relationships and their impact on young people. The secondary analysis would build on the first analysis to provide further evidence of how to improve the mentoring programme, maximise the benefits for young people and expand on the body of knowledge to inform policy and improve practice in the field.

Consent and informed consent

The first major issue that this secondary data analysis had is that participants were not asked to consent for their data to be archived or included in further research. This was a limitation that had to be carefully dealt with. Original participants gave consent to a research study that in essence provided further understanding of mentoring relationships, supports, benefits and outcomes. The secondary data analysis is expanding further on this knowledge by introducing a new variable: empathy. It was considered that the purpose of

the primary analysis also covered the purpose of the secondary analysis and was, therefore, deemed viable. According to Bishop (2013) some secondary researchers have argued that they are entitled to use data in new ways as long as confidentiality and integrity of the data are not breached at any time. Issues with consent were approached with the selection of secondary data analysis instead of archiving the data. Secondary data analysis allowed original researchers to be involved and 'supervise' the type of analysis that was carried out with the data.

Secondary researchers are aware that ideally consent should have been sought from participants at the time the data was collected. However, the next issue is determining if consent at the time of primary data collection is really 'informed' as researchers may not foresee specifically what the archived data or secondary data analyses in the future will be before the research is designed (Bishop, 2012). Bishop (2014) argues that researchers need to inform participants of the benefits and risks of taking part in research, but this does not mean that researchers are in a position to decide 'what is best' for participants. Therefore, consent should be sought and explanations of potential uses of the data should also be specified to help participants make their own informed, if limited, decisions.

Another issue of informed consent may be less explored in the literature: researcher consent. Consent from researchers needs to be sought as well as participants because if data is constructed mutually, both parties contributed to the production of data and, therefore, share ownership. More importantly, Camfield and Palmer-Jones (2013) would also argue that researchers may reveal and report personal information that may have contributed to build rapport. This information may also be sensitive; therefore, informed consent needs to be sought from participants and original data collectors and analysts, particularly in the case of secondary data analysis, as archived data is usually covered by data licenses agreed with the repository.

The 'voice' of young people

One important aspect of carrying out primary and/or secondary research is considering what impact the research findings will have on participants. Researchers have a specific interest in a field which is important to trigger their motivation and commitment to carry out the research in the first place; however, research needs to have an ethical and responsible approach to safeguard the well-being of young people in this case both during, and subsequent, to the research. Secondary data analysis can be used to develop insight into hard to reach or vulnerable populations by reducing the level of potential participant distress (Irwin, 2013). Participants should be exempt from unnecessary intrusion, if primary data that can answer a research question already exists, collecting more unnecessary data can be an intrusion (Bishop, 2009).

Exploratory secondary data analyses can provide insights to the perspectives and views of young people. Secondary researchers may be able to answer research questions by carrying out secondary data analysis which means further primary research may not be required. This saves time and funding resources while making an effective contribution to the body of knowledge. If the secondary analysis is not sufficient to fully answer a specific research question, this can still inform future research that has to be carried out, but it will be a more targeted and effective way so that the research question can be fully answered. This would be an ethical approach to research as participants do not need to take part in unnecessary research processes. In terms of investment, funding can be targeted at specific needs and gaps in the research that existing databases definitely cannot provide.

Young people in this secondary data analysis were not explicitly asked about the topic of interest. One of the concerns of the secondary researchers was that this could lead to a limited understanding of the topic, as it was not directly approached it would have to be inferred.

Secondary data analysis requires a level of interpretation of the data which may or may not achieve the depth and accurate representation of what a young person might have said if asked about the topic directly. Secondary data analysis, as well as or moreover primary data, needs to have a rigorous, transparent and replicable analytical process to support the findings. Research findings need to be clearly supported by evidence identified in the primary data to ensure that the voices of young people are accurately captured and understood, avoiding bias from the researchers' interpretation.

One of the ethical considerations in this secondary data analysis was to include young people in the dissemination of the findings of the secondary analysis. Access to the original cohort of young people was not possible and, also, they would have out-grown this developmental stage. It was decided to include a youth advisory group consisting of young people currently involved in mentoring relationships to inform the dissemination of the findings and be the 'voice' of young people in issues that matter to them from their own perspective and approved by them. This would also provide an opportunity to audit and validate the findings and ensure that the current needs of young people are captured to inform policy and the original services targeted will continue to be improved by the data that was originally commissioned for the evaluation of their services. The findings will be adapted to the current service users to ensure that it is relevant and used for the 'public good' (Bishop, 2009).

Losing the context

One of the biggest concerns regarding secondary data analysis is the importance of context and rapport in the generation of qualitative research. Qualitative data can be defined as a mutual construction between researchers and participants, which is not possible in secondary data analysis (Irwin, 2013; Parry and Mauthner, 2004). Bishop (2012) argued that secondary data needs to include extensive and detailed descriptions of the context of the primary data. However, this does not mean that the original context can be or should be reproduced in secondary data analyses. Since the current tendency of research is towards archiving, researchers need to start accurately systematising in detail the context of their primary data collection.

Another issue related with methodological approaches and context is the appropriateness of specific methodologies to undertake secondary data analysis. According to Irwin (2013) only certain methodologies are appropriate to carry out secondary data analysis. In ethnography, for example, the researcher is involved in the setting to such an extent that data becomes a product and possession of the researcher, limiting the possibilities of analysis by external researchers. Semi-structured interviews can produce data that is more independent of the primary researcher, suggesting that the assumptions and data generation is more evident and transparent (Irwin, 2013). In this secondary data analysis, primary data was obtained through semi-structured interviews and, therefore, considered suitable for secondary data analysis.

Different views have emerged describing the removal of distance and emotional detachment from the data which can have benefits for the analysis process (Camfield, Palmer-Jones, 2013). Secondary data analysis may also benefit from a wider contextual data, more resources, and more complex theoretical and methodological approaches (Camfield, Palmer-Jones, 2013). Overall, researchers may develop a better understanding of their topic and acquire new methodological and analytical skills over time. This can contribute to improving their own research and provide a depth of understanding that was not possible at the time of the original data collection and analysis.

In 2007 Moore introduced the concept of 'recontextualisation' which emphasizes that all primary and secondary researchers engage in contextualisation. All researchers, independent of involvement in the original data context or

not, need to support the claims they make from the data, competing claims exist independently of primary or secondary data (Camfield and Palmer-Jones, 2013; Bishop, 2014). According to Bishop (2014) context is built on the research question, suggesting that the original context may not be relevant at all with the introduction of a new line of inquiry. Therefore, more than 'losing' the context, secondary data analysis needs to build a new context that is suitable for the existing data and proposed methodology.

Lack 'familiarity' with the data

Another consideration for secondary researchers was limited knowledge of the data at the time of seeking funding, which was crucial to be able to have access to the primary data. At the time of securing funding, researchers may or may not be familiar enough with the data and this can be an issue as the proposed methodology may not be suitable once the data is obtained. This issue may become more relevant if, and when, researchers do not have access to data archived unless they have secured funding. One of the important criteria in applications for secondary data analysis is the innovation aspect of the proposed methodology. However, secondary researchers need to ensure that they can deliver the study as designed and that the 'promised' contribution to knowledge is achieved.

One of the possibilities to mitigate the lack of familiarity with the context and the depth of the data is facilitating and encouraging communication between primary data collectors and data re-users (Roche et al.,2014). Dialogue with primary researchers can help secondary researchers access the original context where primary data was generated (Irwin, 2013) and acquire an understanding of the data available even before they have access to it.

In the case of this study, the main researcher had access to the original data collectors, data analyst and principal researchers which provided an opportunity for deeper understanding and clarification of how the data was obtained, analysed and reported.

Conclusions

This paper has provided some insight into the ethical, technical and methodological challenges faced by researchers carrying out secondary data analysis with young people. Researchers are aware that this is not an extensive exploration of the issues, but rather an invitation for further reflection and analyses of the implications of carrying out this type of research using an example from practice.

Secondary data analysis with young people is a cost-efficient and effective way to contribute to the body of knowledge, improve policy and practice while reducing the level of intrusion and distress in vulnerable and hard-to-reach populations such as young people engaged in mentoring relationships. This paper provides some reflection on informed decisions and considerations that were taken to safeguard the well-being and integrity of research participants when carrying out a secondary data analysis with young people.

Secondary researchers agree that ideally participants should have been asked for consent regarding future uses of their data at the time of the original data collection, particularly when considering archiving data and when contact and communication with primary investigators is limited or impossible.

References

Bishop L. (2009) E*thical sharing and reuse of qualitative data.* Australian Journal of Social Issues, Vol.44 (3), pp.255-275.

Bishop L. (2012) *Using archived qualitative data for teaching: Practical and ethical considerations,* International Journal of Social Research Methodology, Vol. 15 (4), pp.341-350.

Bishop L. (2013) *The value of moral theory for addressing ethical questions when reusing qualitative data,* Methodological Innovations Online, Vol.8 (2), pp.36-51.

Bishop L. (2014) Re-using qualitative data: *A little evidence, on-going issues and modest reflections*, Studia Socjologiczne, Vol.3 (214), pp.167-176.

Camfield L. and Palmer-Jones R. (2013) I*mproving the quality of development research: What could archiving qualitative data for reanalysis and revisiting research sites contribute?,* Progress in Development Studies, Vol.13 (4), pp.323-338.

Dolan P., Brady B., O'Regan C., Brumovska T., Canavan J. and Forkan C. (2010) Bi*g Brother Big Sisters (BBBS) of Ireland: Evaluation Study. Report: Randomised Control Trial and Implementation Report*. Galway: UNESCO Child and Family Research Centre.

Irwin S. (2013) *Qualitative secondary data analysis: Ethics, epistemology and context.* Progress in Development Studies, Vol.13 (4), pp.295-306.

Parry O. and Mauthner N. (2004) *Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data.* Sociology, Vol.38 (1), pp.139-152.

Roche D., Lanfear R., Binning S., Haff T., Schwanz L., Cain K. et al. (2014) *Troubleshooting public data archiving: Suggestions to increase participation,* PLOS Biology, Vol.12 (1), pp.1-5.

Author information

Leonor is a Postdoctoral Researcher at the UNESCO Child and Family Research Centre. She has experience in health and clinical psychology working with families, children and young people that experience chronic illness and completed a Maters in Clinical and Health Psychology in her native Costa Rica. Leonor completed her PhD in the School of Psychology, NUI Galway entitled: *Understanding adolescent adjustment to maternal cancer: A study of personal experiences and psychological factors that promote adjustment*.

Children's Research Network for Ireland and Northern Ireland

Supporting the research community in Ireland and Northern Ireland to better understand and improve the lives of children and young people

childrensresearchnetwork.org

Supported by the Department of Children and Youth Affairs and The Atlantic Philanthropies