

The Development of the Public Health Research Data Management System

Deborah van Gaans¹, Katina D'Onise¹, Tony Cardone², and Robyn McDermott³

¹ School of Population Health, University of South Australia, South Australia, Australia

² Business Intelligence and Planning, University of South Australia, South Australia, Australia

³ Medical and Veterinary Sciences, James Cook University, Queensland, Australia

Abstract

The design and development of the Public Health Research Data Management System highlights how it is possible to construct an information system, which allows greater access to well, preserved public health research data to enable it to be reused and shared. The Public Health Research Data Management System (PHRDMS) manages clinical, health service, community and survey research data within a secure web environment. The conceptual model underpinning the PHRDMS is based on three main entities: participant, community and health service. The PHRDMS was designed to provide data management to allow for data sharing and reuse. The system has been designed to enable rigorous research and ensure that: data that are unmanaged be managed, data that are disconnected be connected, data that are invisible be findable, data that are single use be reusable, within a structured collection. The PHRDMS is currently used by researchers to answer a broad range of policy relevant questions, including monitoring incidence of renal disease, cardiovascular disease, diabetes and mental health problems in different risk groups.

Keywords: Public Health; Modelling; Database Management Systems; Secondary Use

1 Introduction

Epidemiological and health related statistical information provide the evidence base for health care and policy, by providing accurate and reliable data including the health of minority and vulnerable populations [1]. However, in public health research, data management is the poor cousin of analysis, as it is often undervalued and underfunded [2]. Without accurate data there is little capacity to monitor changes in health status, to evaluate access to services and the response of services to needs, or to quantify the resources expended on health services and programs [1].

Managing the life cycle of scientific data presents many challenges including deciding responsibilities, funding, resource allocation, what data should be kept and for how long [3]. Research data is a valuable asset and while data management is a necessary part of

good research it is not always undertaken well by the researcher. Ackerman and Osborne (2005)[4] highlight the importance of an integrated system for managing health research data to ensure the smooth transfer of data from the Hospital's patient record database to the research database, and finally to statistical software for analysis.

In a system that emphasizes competition rather than collaboration among researchers, data sets resulting from multimillion dollar investments from tax payers sit idle inside locked computers, only available to a small number of researchers despite their containing the seeds that would allow for the exploration of a vast number of important research questions that could change the healthcare landscape [5]. There are indications that public and foundation funders of public health research wish to strengthen data sharing policies, shepherding epidemiologists down the road already travelled by ge-

neticians [2]. Secondary research refers to the use of research data to study a problem that was not the focus of the original data collection [6]. This secondary analysis may involve the combination of one data set with another, address new questions or use new analytical methods for evaluation [6]. The benefits of data sharing are many and include:

- Allowing, the same data to be used to answer new questions that may be relevant far beyond the original study [2].
- Accelerating investigations already under way and taking advantage of past investments in science [3].
- Obtaining a statistically meaningful number of cases quicker than studies in a single centre, so the applied research results can be used quicker as well and particularly for rare diseases a critical mass of cases can be obtained in sufficient quality that no single institution could obtain [7].
- Generates opportunities for additional publications through collaboration, and may increase the citation rate of primary publications [8].
- Once investments in infrastructure have been made, recycling and combining data provide access to maximum knowledge for minimal additional cost [2].
- Sharing data increases the visibility and relevance of research output [8].
- Being able to extend the study dataset through linking to other data sources has the potential to enable the important research questions for the study to be better answered, with the added benefit of generally reducing the burden on respondents [9].

To enable reuse, data must be well preserved. Community standards for data description and exchange are crucial as these facilitate data reuse by making it easier to import, export, compare, combine and understand data [3]. As Pisani (2010) [2] states improved documentation will lead to data being combined more easily across time, locations and sources.

The development of public health information systems requires an understanding of the principles, practices, structures and settings in which these systems operate [10]. Issues of conflicting data standards, the need for interoperable tools for exchanging and sharing data and the need for innovative solutions to address

integrated disease surveillance, among many other issues, are driving forces to formalize design strategies in public health information [10]. Details regarding the specific design and features of such databases are not readily available in the literature and yet, this type of practical information would be valuable for clinicians and researchers who wish to design database systems tailored to their particular requirements [4].

2 Methods

The conceptualization of the Public Health Data Management System (PHRDMS) occurred through a series of consultative meetings between public health researchers, information technology business intelligence specialists and data managers. The Public Health Data Management System (PHRDMS) stores data, metadata and documents that are generated throughout the life-cycle of research projects. The PHRDMS provides a structure to allow research data to be maintained in accordance with a large number of laws, regulations and conventions, and was designed specifically to meet the standards of: University of South Australia, (2012) UniSA Framework for the Responsible Conduct of Research [11], James Cook University, (2012) Code of Conduct [12], and National Health and Medical Research Council, (2007) Australian Code for the Responsible Conduct of Research [13]. The guidelines were synthesised into the following core set, that have underpinned the development of the PHRDMS:

- Researchers should retain research data and primary materials for sufficient time to allow reference to them by other researchers and interested parties. For published research data, this may be for as long as interest and discussion persist following publication.
- When considering how long research data and primary materials are to be retained, the researcher must take account of professional standards, legal requirements and contractual arrangements.
- Research data should be made available for use by other researchers unless this is prevented by ethical, privacy or confidentiality matters.
- Research data should be retained for at least the minimum period specified in the institutional policy.
- The institutional policy on the secure and safe disposal of primary materials and research data must be followed (note that for patient records these are to be kept indefinitely).

- Researchers must manage research data and primary materials in accordance with the policy of the institution.
- Sufficient materials and data are retained to justify the outcomes of the research and to defend them if they are challenged. That security and confidentiality of the data is undertaken and maintained.
- Keep clear and accurate records of the research methods and data sources, including any approvals granted, during and after the research process.
- Ensure that research data and primary materials are kept in safe and secure storage provided, even when not in current use.
- Provide the same level of care and protection to primary research records, such as laboratory notebooks, as to the analysed research data.
- Retain research data, including electronic data, in a durable, indexed and retrievable form.
- Maintain a catalogue of research data in an accessible form.
- Manage research data and primary materials according to ethical protocols and relevant legislation.
- Maintain confidentiality of research data and primary materials. Researchers given access to confidential information must maintain that confidentiality.
- Primary materials and confidential research data must be kept in secure storage. Confidential information must only be used in ways agreed with those who provided it. Particular care must be exercised when confidential data are made available for discussion.
- Ability to deidentify / reidentify participants if necessary.
- Ability to link data from other sources.
- Ability to create reports for: individual participants, communities, health services, projects.
- Allow for version control of project documents and derived datasets.
- Data fits with International/ national standards where possible.
- Temporal view of data.
- Logging of data extracts.
- Formal process of data upload and extraction.
- Metadata development, cleaning, maintenance.
- Developing and implementing protocols regarding storage, retrieval, security and integrity of the data to be used by key stakeholders.

3 Results

All of the data, metadata and documents that form part of any public health research project are captured within the PHRDMS. As can be seen in Figure 1 this includes ethics agreements, reports, questionnaires, methods, approvals, publications, data dictionary, and study protocols.

A copy of the plain language statement for each research project, as required by ethical standards of research, is held within the PHRDM System as a .pdf file. Participant consent agreements are stored as a .pdf file for each project participant within the PHRDM System. Through the security structure of the PHRDM System research participants are deidentified however the system also has the capability to make data reidentifiable (to system administrator roles only) so that reports can be sent to individual participants, participants can be contacted for further involvement in research projects, and also for data linkage purposes.

The participant's consent agreements often have a series of statements relating to particular data / information and the participant can choose to consent to the individual statements. These statements often reflect being contacted for further research projects, having the participants data forwarded on to their primary health care clinic etc. These statements are captured with the participants consent within the PHRDM System, so that the agreement between the participant and the project can be maintained during data extraction and reporting.

The PHRDMS was constructed by the Information Strategy and Technology Services Unit within the University of South Australia through consultation with population health researchers. During the design phase of the PHRDMS, specific researcher requirements were identified, these included:

- Ensure data is accessible to who need it: including remote regions, different universities.
- Easily used by researchers as it fits with their business process Eg. Data entry forms look like the questionnaire.

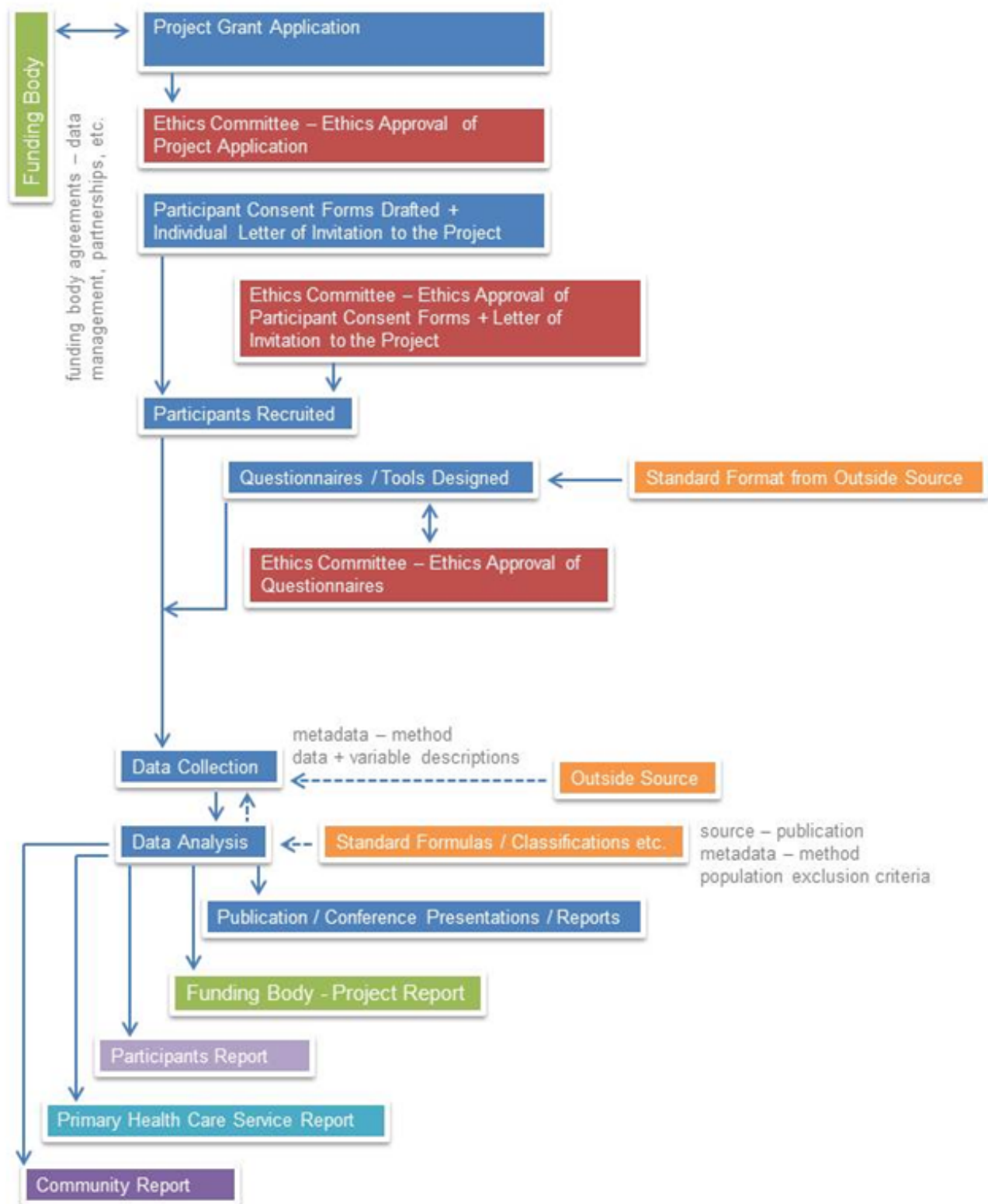


Figure 1: Data, Metadata and Documents that are Captured within the Public Health Research Data Management System

The research projects often have agreements with Communities, Primary Health Care Clinics, hospitals, data custodians etc. Copies of these agreements are held as .pdfs within the PHRDM System for each research project.

One of the added features of the PHRDMS is that it also maintains an audit trail and history of all data modifications. The audit trail begins from the time the data has been entered into the system and all modifications to the data are recorded in audit tables which are maintained as part of the system. The audit tables are a log of the change that has been made to the data, at what time and by who. Data that has been manually entered into the database can be corrected through the data entry screens and bulk uploaded data will be backed out of the PHRDM system and then reloaded.

All surveys/ questionnaires that are used within a research project undergo an ethics approval process before they are administered. Sometimes a single survey/questionnaire will undergo a number of revisions. All versions of the surveys/questionnaires that have been used within the research project are maintained within the PHRDM System.

The PHRDMS stores demographic, vaccinations, diagnosed chronic conditions, medications, lifestyle measures, pathology results, mental health, management plans, allied health and specialist referrals, gestational data and child data. The system allows the system administrator to add clinical variables as needed by the research project, as well as surveys.

3.1 Design

The PHRDMS is a very flexible user friendly system. The data model that underlies the PHRDMS is based on three distinct entities and the relationships between them (Figure 2). This data model allows users to customise their view of the database to the variables that they are collecting for their own research project. Users can therefore add new variables to the participant, community, or Primary Health Care Centre entity. The system also allows new questions and answers from questionnaires to be added. The PHRDMS does not store derived variables, only raw data, which allows the users to classify the data according to individual researcher requirements.

The PHRDM System produces a number of standard research reports and individual clinical variables can be extracted into an excel spreadsheet. The system also allows codes to be assigned to data so that it can be used directly within Stata [13] once it has been extracted from the system. The system also produces a log report to capture the history of changes made to the data within

the system due to data corrections. All data extracts are recorded within the PHRDMS to maintain a history of what data was extracted by who, at what time, for what purpose.

The PHRDM system allows for data linkage to external data sets. Data from external data custodians is able to be linked to the participant, community, primary health care centre, or the participant's pathology result. The data is initially held in a staging area while it is reviewed against the current set of data variable rules. Any external data that does not match the existing data variable rules is able to be reviewed by the system administrator and either be corrected (in the case of a data error) or rejected from the data upload. The data upload and cleansing process is captured within the PHRDMS to maintain an activity log for administration purposes.

Due to the nature of Public Health research many of the projects contained within the PHRDMS collect the same clinical variables and often administer the same questionnaires. The PHRDMS maintains projects separately but with many of the research staff working across numerous projects it is possible for data to be viewed as a total collection (Figure 3), allowing for variables from a number of projects to be reused to answer new research questions.

3.2 Access

Initial access to the database is provided through the Australian Access Federation, which will allow researchers into the database, who belong to institutions that are registered with the Australian Access Federation. Therefore allowing researchers access to the system anywhere they are able to get access to the internet. Researchers are then able to be granted access to project data for which they have signed project confidentiality agreements. Access to the project data is then governed by the role that is assigned by the system administrator. The PHRDMS manages data access through the following roles: System Administrator, Researcher, and Data Entry. Functionality within the database is applied to each role with all roles other than system administrator being applied to a specific project.

4 Conclusion

The Public Health Data Management System stores and manages a large cohort of Indigenous adults and children, both "well" and who already have a chronic condition on study enrolment. The dataset will grow due to recruitment of participants over time and increase in scope as new datasets are linked to the cohort. The information generated from the system will be used for

Relationships between Project, Participant, and Primary Health Care Service

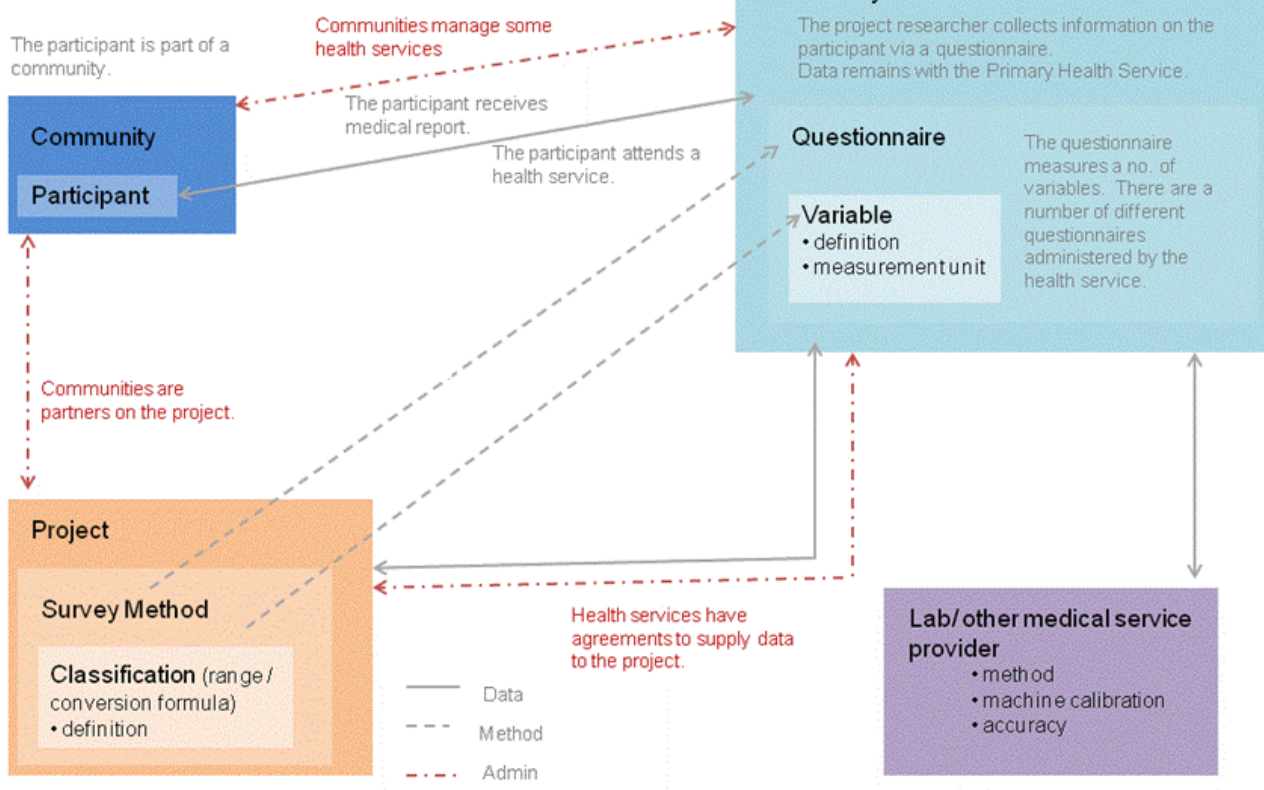


Figure 2: The Entities within the Public Health Research Data Management System

the immediate research aims of the Centre of Research Excellence in Prevention of Chronic Conditions and will be able to be used by researchers into the future to answer a much broader range of policy relevant questions, including monitoring incidence of renal, cardiovascular disease, diabetes and mental health problems in different risk groups. This cohort will include these participants at baseline, but also be able to identify incidence of disease in those free of problems at recruitment.

The design and development of the Public Health Research Data Management System highlights how it is possible to construct an information system which allows greater access to well preserved public health research data to enable it to be reused and shared. While the development of the PHRDMS has been based on Australian guidelines, the conceptual model underpinning the PHRDMS which is based on the three main entities: participant, community and health service could be used internationally.

Acknowledgements

The research reported in this paper is a project of the Australian Primary Health Care Research Institute, which is supported by a grant from the Commonwealth

of Australia as represented by the Department of Health. The information and opinions contained in it do not necessarily reflect the views or policy of the Australian Primary Health Care Research Institute or the Australian Government Department of Health.

Conflicts of Interest

None declared.

References

1. Thompson SC, Woods JA and Katzenellenbogen JM, (2012) The quality of indigenous identification in administrative health data in Australia: insights from studies using data linkage. *BMC Medical Informatics and Decision Making*, 12:133 doi:1186/1472-6947-12-133.
2. Pisani E, and AbouZahr C, (2010) Sharing Health Data: Good Intentions are Not Enough. *Bull World Health Organ* 2010;88:466–467, doi:10.2471/BLT.10.076943

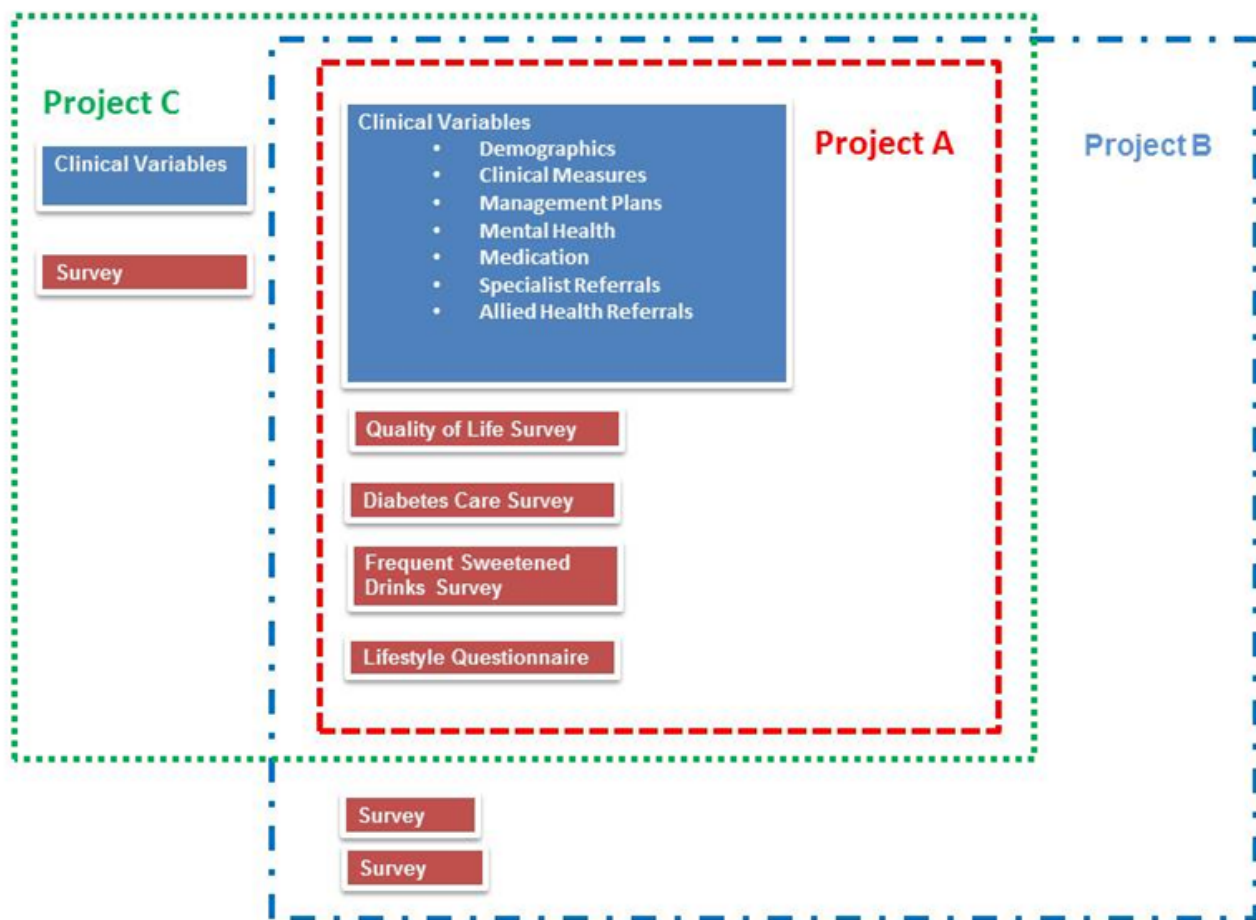


Figure 3: The Relationship Between Projects within the Public Health Research Data Management System

3. Lynch C, (2008) How Do Your Data Grow? *Nature* Vol 455 4 September 2008, pp.28-29.
4. Ackerman IN and Osborne RH, (2005), Integrating Data to Facilitate Clinical Research: A Case Study. *Informatics in Primary Care*, 13 pp. 263-70.
5. Carvalho EC, Batilana AP, Simkins J, Martins H, Shah J, Rajgor D, Shah A, Rockart S, Pietrobon R, (2010) Application Description and Policy Model in Collaborative Environment for Sharing of Information on Epidemiological and Clinical Research Data Sets. *PLoS ONE* 5(2): e9314. doi:10.1371/journal.pone.0009314
6. Law M, (2005) Reduce, Reuse, Recycle: Issues in the Secondary Use of Research Data. *IASSIST Quarterly* (Spring), pp. 5-10.
7. Elger BS, Iavindrasana J, Iacono LL, Muller H, Roduit N, Summers P, Wright J (2010) Strategies for Health Data Exchange for Secondary, Cross-institutional Clinical Research. *Computer Methods and Programs in Biomedicine*. 99 pp.230-251.
8. Piwowar HA, Becich MJ, Bilofsky H, Crowley RS, caBIG Data Sharing and Intellectual Capital Workspace (2008) Towards a data sharing culture: Recommendations for leadership from academic health centers. *PLoS Med* 5(9): e183. doi:10.1371/journal.pmed.0050183
9. Soloff C, Sanson A, Wake M, Harrison L (2007) Enhancing Longitudinal Studies by Linkage to National Databases: Growing Up in Australia, the Longitudinal Study of Australian Children. *Int. J. Social Research Methodology*, Vol. 10, No. 5 December 2007, pp. 349-363. DOI: 10.1080/13645570701677060
10. Reeder B, Hills RA, Demiris G, Revere D, Pina J (2011) Reusable Design: A Proposed Approach to Public Health Informatics System Design. *BMC Public Health* 11:116
11. University of South Australia, (2012) UniSA Framework for the Responsible Conduct of Research, accessed: 26th November 2012, <http://w3.unisa.edu.au/RES/ethics/integrity/default.asp>
12. James Cook University, (2012) Code of Conduct, accessed: 26th November 2012, http://www.jcu.edu.au/policy/governance/conduct/JCUDEV_007161.html
13. National Health and Medical Research Council, (2007) Australian Code for the Responsible Conduct of Research, Australian Government. ISBN 1864964383 StataCorp. 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP.

Correspondence

Dr Deborah van Gaans (Corresponding Author)

Manager: Research Data

Centre for Research Excellence in the Prevention of Chronic Conditions in Rural and Remote Populations
School of Population Health, University of South Australia

Level 8, South Australian Health & Medical Research Institute (SAMHRI)

North Terrace, Adelaide, 5001

Tel: +618 830 22908

deborah.vangaans@unisa.edu.au

Research Associate

Dept. of Geography, Environment and Population,

The University of Adelaide,

North Terrace, Adelaide, South Australia, 5005

Dr Katina D'Onise

Senior Research Fellow

Centre for Research Excellence in the Prevention of Chronic Conditions in Rural and Remote Populations
School of Population Health, University of South Australia

Level 8, South Australian Health & Medical Research Institute (SAMHRI)

North Terrace, Adelaide, 5001

Tel: +618 830 21221

katina.d'onise@unisa.edu.au

Mr. Tony Cardone

Business Intelligence Specialist

Chancellery,

Business Intelligence and Planning, University of South Australia

City West Campus

North Terrace, Adelaide, 5001

Tel: +618 830 27286

tony.cardone@unisa.edu.au

Prof Robyn McDermott

Professor of Public Health Medicine

College of Public Health, Medical and Veterinary Sciences

James Cook University, PO Box 6811, Cairns QLD
4870 AUSTRALIA
Tel (07) 4232 1575
robyn.mcdermott@jcu.edu.au