

Data Linkage Guide

August 2017

Prepared by Greg Duncan Senior Research Fellow

Funded by Eastern Health Foundation

Contents

Overview..... 3

Definitions 4

Guiding Principles..... 4

Good Clinical Research Practice..... 4

Core considerations for all data linkages research 5

Figure 1: Process map for research planning for projects involving data and data linkage 6

Detailed guide to research planning for projects involving data and data linkage 7

 Clear research question (Use PICO format) with plan for analysis..... 7

 Identify databases or datasets of interest 7

Ethics..... 8

Managing research data..... 8

Contact customer services officer or similar for appropriate data linkages agency 8

Costs..... 8

Data linkage and transfer..... 9

Secure environment for data analysis..... 9

Time frame 10

 Figure 2: Summary of data linkage processes for Eastern Health researchers: Using local, state and cross-jurisdictional data (more detail provided in specific sections to follow) 11

Linking and analysing data from Eastern Health only..... 12

Researcher or survey collected data using REDCap 13

Figure 3. Process for using EH only data (including survey/researcher collected data via REDCap..... 14

Linking and analysing data from Eastern Health with other Victorian databases..... 15

Detailed step by step requirements for linking Eastern Health and Victorian state data..... 15

Further information for using VDL services for Victorian data linkage..... 17

Figure 4. Process for using EH and /or Victorian data (including survey/researcher collected data via REDCap 18

Linking and analysing data from Eastern Health with data sources from other jurisdictions in Australia..... 19

Datasets accessible..... 19

Data access 20

Linking cross jurisdictional data..... 21

Secure data transfer 21

Secure environment for data analysis..... 21

Figure 5. Process for using cross-jurisdictional data 22

Guide to Using Health Data for Research

Overview

The statutory routine collection of data across health services in Australia provides for a rich source of data that may give more effective measures of population outcomes when trialling interventions, evaluating risk and benefit of various services and identifying population groups at particular risk.

Eastern Health, in line with other health services is moving to the widespread use of Electronic Medical Record. Eastern Health's extended EMR will support Eastern Health research by providing standardised, internationally recognised datasets in real time.

To explore the benefits of using health data for research both internally and externally it might be useful to consider scratch that it is useful to consider some basic examples of how this may be of benefit. Provision of access to this data for researchers and clinicians allows for comparisons and evaluations to be made of current services, new interventions and resource use to facilitate improved service delivery, innovative strategies for patient care and outcomes assessment. For example, with the implementation of any service in a department review of the data available would allow for comparison of therapeutic economic or other outcomes to make before and after the intervention to evaluate effectiveness and benefit. In other circumstances, it may be that the dataset available provides researchers and clinicians with a more significant understanding of the demographics of the patient group improving the ability to implement new therapies or services by comparing the Eastern Health patient group with those on whom original research was done that informed and implementation process or guideline. When looking beyond the Eastern health environment and consider datasets available at both state and Commonwealth level more useful and profound comparisons can be made. This includes benchmarking with outcomes in similar health networks, effect of interventions in other settings and the potential to draw on these databases to provide large populations with which comparisons may be made.

There are many issues around the processes of using data that are being considered and addressed at organisational levels including data access rights, storage, transfer, linkage, costs, custodian responsibilities, consumer rights and many more. This guide has been informed by the current situation regarding these issues and will be amended as the landscape changes. What the guide provides is an overview for researchers on the core considerations and processes that currently exist for the use of a range of datasets in Australia for research purposes.

Definitions

There are several terms used in the discussion and processes of data use to help research context which may have variations in meanings in different settings. To avoid any ambiguity in this context, clear definitions are important to ensure all involved have the same understanding.

Data: digitally recorded, transferred, and stored values of variables that may be qualitative or quantitative in nature

Dataset: a collection of discrete items of related data that may be accessed individually, in combination or as a whole entity for reference, analysis, or review

Linkage: a technique for creating links within and between data sources so that information that is thought to relate to the same person, family, place or event can be connected for analysis.

Custodian: the person(s) or entity responsible for the safe custody, transport, and storage of a particular set or sets of data, and the implementation of rules around access and use.

Data security: this refers to the collective measures used to protect and secure a database or dataset from illegitimate use and malicious threats and attacks. It includes a range of processes, tools, methodologies, and strategies to ensure security with dataset environment

Guiding Principles

Before considering any research that involves data use, linkage and analysis, please familiarise yourself with the overarching principles that apply to this process provide by the National Health and Medical Research Council (NHMRC)

NHMRC Principles for accessing and using publicly funded data for health research

<https://www.nhmrc.gov.au/principles-accessing-and-using-publicly-funded-data-health-research>

In summary these principles state:

1. Maximise the use of publicly funded health and health related data for research
2. Data custodians should recognise their responsibilities and accountabilities when providing access to data from research
3. Researchers should recognise their responsibilities and accountabilities when using publicly funded health and health related datasets.

Good Clinical Research Practice

Courses and guidelines for researchers to enable a contemporary competency in research practice are available and recommended for all involved in research.

Staff who have a university appointment should check to see if the Research Office of your University offers courses directly.

For those interested in online courses, there are a number of options including

NHMRC <http://australianclinicaltrials.gov.au/nhmrc-elearning-modules>

Core considerations for all data linkages research

There are three environments from which data may be sourced for Eastern Health researchers with specific processes to be followed for data access, linkage and analysis in each context. Specifically those three environments are:

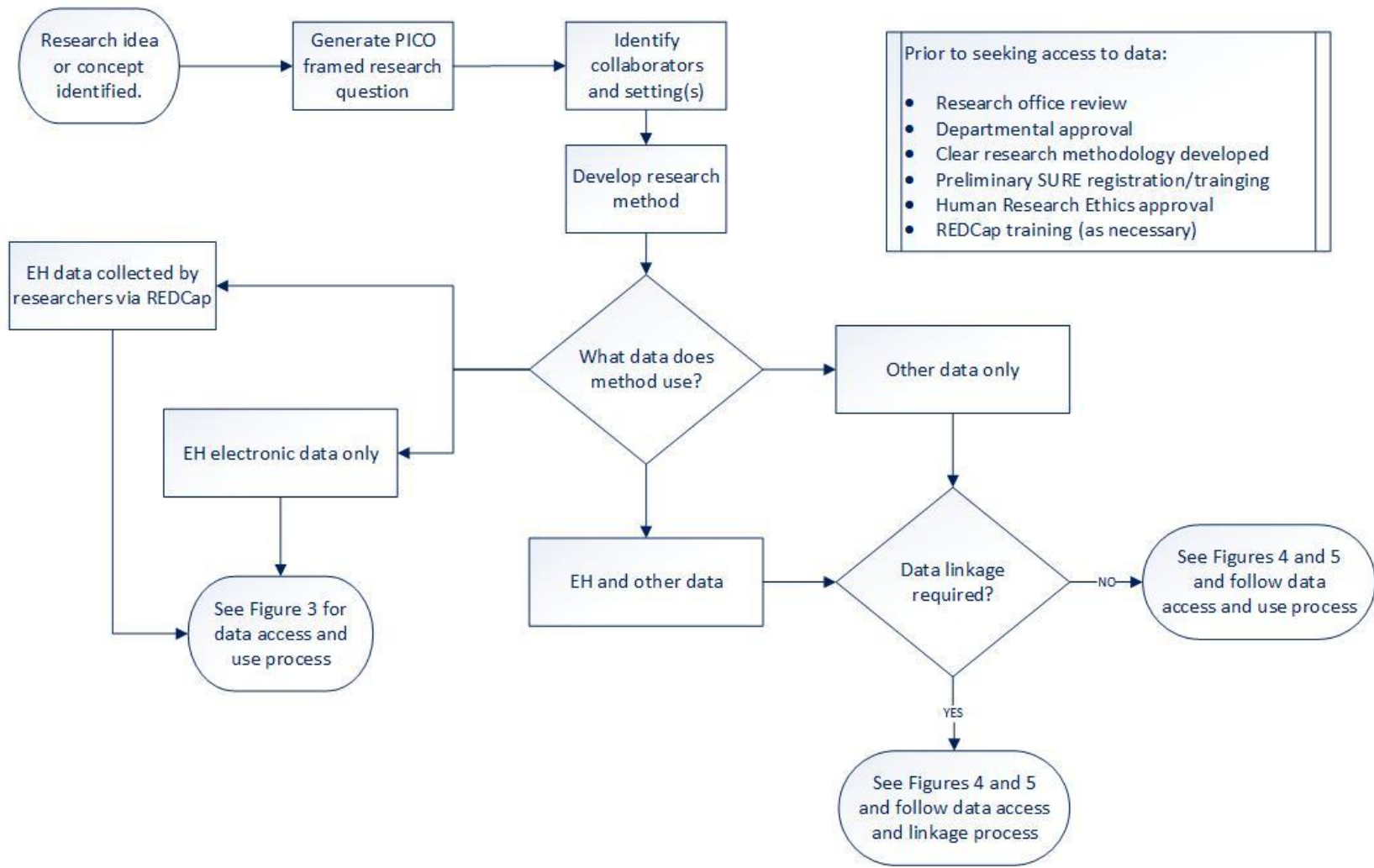
- Eastern Health data only (including researcher or survey collected data using REDCap)
- Linkage of Eastern Health data with or use of Victorian State databases (including researcher or survey collected data using REDCap)
- Linkage of Eastern Health data with or use of databases from other jurisdictions in Australia (Cross jurisdictional linkage)

Before exploring data access in these specific contexts there are several important activities to address that underpin any data linkage research. Specifically, these activities are detailed and comprehensive research design, implementation planning and ethics approval. These activities will overlap rather than be sequential. Researchers are advised to engage with the office and research of ethics and any external human research ethics body you may be involved with a project from an early stage of the research planning process.

Figure 1 provides a process map for the overall research planning for the use of data while the overall process for using and linking data in these three environments is described further on in this guideline, and summarised in a flowchart in Figure 2.

Figure 1: Process map for research planning for projects involving data and data linkage

Research planning process for EH researchers looking to use and/or link EH data and other data sources.



Detailed guide to research planning for projects involving data and data linkage

Clear research question (Use PICO format) with plan for analysis

The use of data sets in any environment is not a fishing trip to trawl for relationships between variables and then explore what that might mean. The NHMRC principles and those of good research ethics indicate there should be a clear question to be answered by the data being collected in whatever format that might be. To assist in this, it may be useful to frame your research or clinical question in the standard PICO format that has underpinned an evidence based approach to clinical practice for many years. Many research funders now require grant applications to frame a research question in this format so it is useful to be familiar with the process.

For those not familiar with this process, PICO is an acronym for the four key elements of a specific and focussed clinical or research question. The acronym stands for

- P** – Population/ problem/ patient – what are the important aspects a population of interest?
- I** – Intervention – what is the intervention/therapy/service of interest?
- C** – Comparison – what would I compare it to – current practice/gold standard/placebo etc?
- O** – Outcome – what is the measurable outcome of interest?

For examples and more information on this process there are a number of excellent freely available resources including:

1. The Centre for Evidence Based Medicine at Oxford University <http://www.cebm.net/asking-focused-questions/>
2. A tutorial from Oxford University at http://learntech.physiol.ox.ac.uk/cochrane_tutorial/cochlibd0e84.php
3. University of Canberra PICO resource <http://canberra.libguides.com/content.php?pid=591487&sid=5015147>

There are many other guides available as well a very useful text that goes into this in more detail is: *Users Guide to the Medical Literature. Essentials of evidence-based clinical practice* (3rd Edition). G Guyatt, D Rennie, 2015, McGraw-Hill Professional.

Identify databases or datasets of interest

Once a research question has been framed the most likely or appropriate source of the data to enable comparisons to be made and evaluate the intervention of interest must be considered. If looking beyond Eastern Health there are links for both Victorian and other jurisdictional databases that may be accessible. These databases/datasets are linked in the relevant sections below. In the process of formulating a PICO question it may be useful to consider the likely sources of data that

will assist in answering that question. It is worth noting that access to datasets may require significant time due to governance issues with each respective data custodian.

Appendix 2 contains a spreadsheet not only of the various databases that may be used for research but where appropriate also identify the specific person or position within contact should be made initially.

Ethics

Any use of data collected on humans will require ethical approval before access, linkage and analysis may occur. The Eastern Health Office of Research and Ethics is leading this process in creating a clear and structured framework for data use for research at Eastern Health. Any researchers or clinicians considering research on this context should contact the Office of Research and Ethics for a discussion about the process and any specific implications for ethics.

Once data linkage has occurred by an accredited external agency the dataset is de-identified however it is the researcher's responsibility to manage data security.

Managing research data

A detailed guide the process of managing research data in context of the nature of the research been discussed is provided in the suite of resources from the Office of Research and Ethics. A summary of the key aspects of this process is discussed with each aspect of data linkage.

Researchers should ensure detailed familiarity with managing research data before commencing research planning, design or delivery.

Contact customer services officer or similar for appropriate data linkages agency

In any circumstance in which a researcher wishes to use and link to any data base it is essential to make contact with the relevant data custodian or linkage agency to discuss the process. This discussion will be informed by the focussed research question (PICO) and guidance from the Office of Research and Ethics as well as the departmental manager or other relevant research supervisor. These discussions will give a sense of whether the research itself is feasible given the data available, the time frames for access and linkage procedures, any further requirements in terms of ethics approval, how security issues must be addressed and, of course, any costs associated with the process.

Costs

There are three aspects (beyond the costs of conducting the research generally) of using and linking data that may involve a cost. Depending on the dataset sought there may or may not be costs involved with this process. This needs to be carefully and strategically factored into the research

budget, again indicating the essential nature of early discussion with the key agencies to get an indication of potential costs.

Firstly, access to the data and having a dataset linked at this stage only occurs a cost in the cross-jurisdictional environment. Internal Eastern Health Data and linking to Victorian state data does not have a cost at this stage but this may change over time.

Secondly, secure data transfer may incur costs as a specific system is used (SUFEX – see details below under “Data linkage and transfer”) for uploading Eastern Health data then downloading of a de-identified dataset for analysis.

Finally, to ensure data security a secure environment is necessary in which data analysis may take place. The Office of Research and Ethics is exploring server options but at this stage it is recommended (and some agencies require) the use of, the Secure Unified Research Environment (SURE – see link below) for which there are costs involved. These costs may include training for using SURE as well as the use of the secure environment for analysis and may be only on a project by project basis.

Data linkage and transfer

Data security protecting patient and organisational confidentiality is a critical issues in this type of research. Some data agencies use courier delivery for data sets to reduce the risk of inappropriate access during electronic transmission. Significant improvements in data encryption techniques and processes now provide for electronic data transfer systems to be used. For Victorian data, researchers are required to upload data using SUFEX (see below), the secure data transfer platform, and similarly for cross jurisdictional data. It is essential that researchers are familiar with the process and requirements (organisational and technical) for use of SUFEX.

Secure data transfer (SUFEX) <http://www.phrn.org.au/for-researchers/services-for-researchers/sufex/>

Secure environment for data analysis

As mentioned previously, to ensure data security a secure environment is necessary in which data analysis may take place. The Office of Research and Ethics is exploring server options but at this stage it is recommended (and some agencies require) the use of, the Secure Unified Research Environment (SURE). If using Cross-jurisdictional data the SURE process is likely to be required by one of the two accredited linking agencies unless there is approval of a definite secure environment in an organisational setting. At this stage, this is very uncommon. If a research project requires the use of cross jurisdictional data, the process of registering and engaging with SURE as early as

possible during a study (or beforehand if time allows) as training is required and there are other approval processes that need to be addressed. The relevant information for using SURE can be found at:

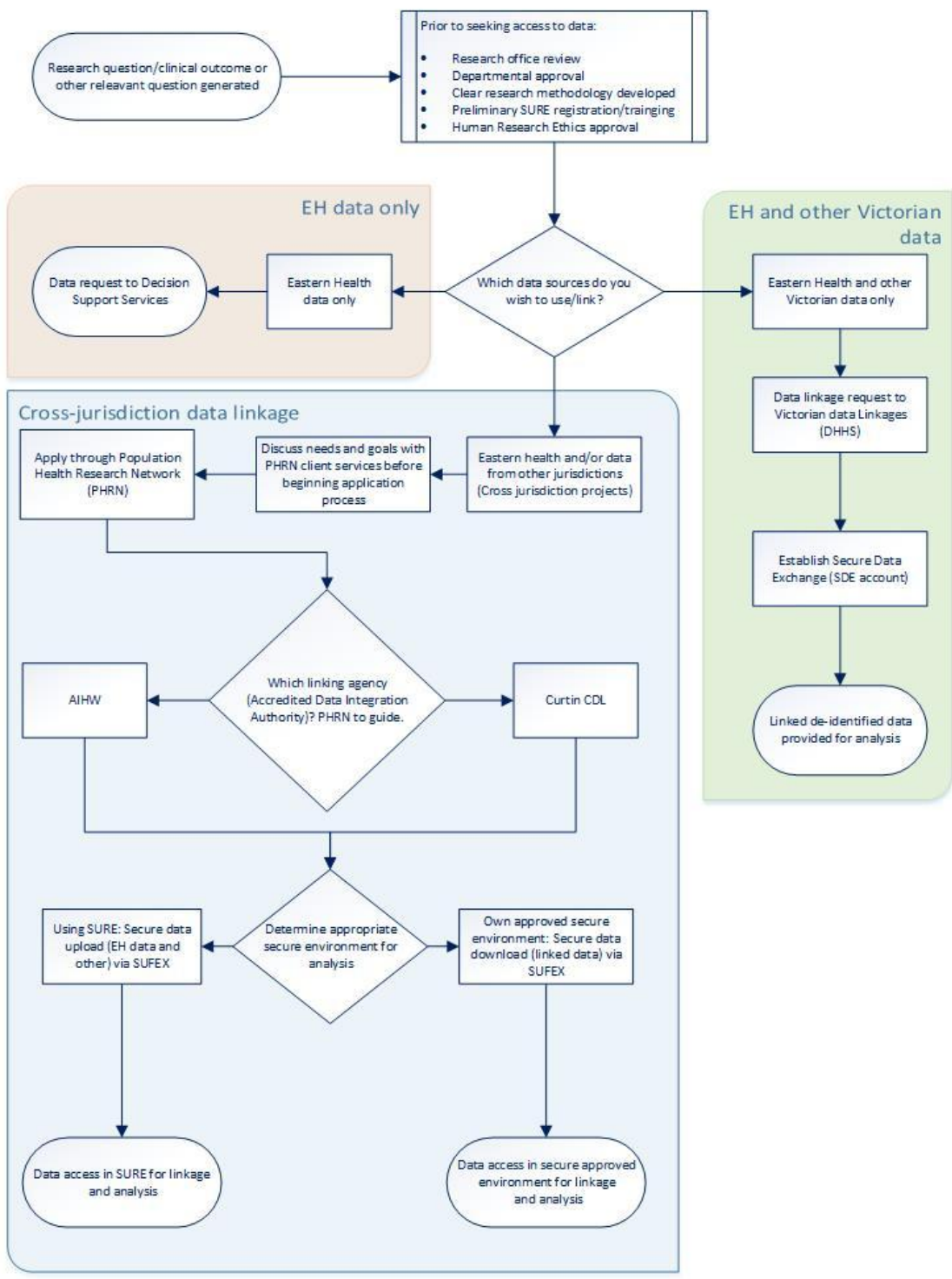
The Secure Unified Research Environment (SURE) <http://www.saxinstitute.org.au/for-researchers/collaborate-with-us/> and <http://www.phrn.org.au/for-researchers/services-for-researchers/secure-unified-research-environment-sure/>

Time frame

when developing a detailed plan for research is important to consider not only how long it will take to access, clean and analyse the data but also how quickly (or not) the data may be made available. This varies from jurisdiction to jurisdiction and dataset to dataset. Some are available almost immediately while others may have a lag time of many months and even years. The nature of the research will influence the time frame – if the research just requires the use of retrospective data, then the issues won't be too significant, but if the plan is to explore the impact of an intervention and use prospective data there may be significant time gaps between the intervention and data availability. Information on the time related issues will be a key discussion point with data custodian and data linkage agencies involved. Again, we reiterate the need to lots of early engagement to facilitate effective research planning.

Figure 2: Summary of data linkage processes for Eastern Health researchers: Using local, state and cross-jurisdictional data (more detail provided in specific sections to follow)

**Summary data linkage process for Eastern Health researchers:
Using Eastern Health, Victorian and cross-jurisdictional data.**



Linking and analysing data from Eastern Health only

The most straightforward use of data for research will be for projects which only use internal Eastern Health electronic data and/or researcher or survey collected data using REDCap. As only Eastern Health, and possibly universities with which Eastern Health researchers are associated, are involved, the ethics process is more streamlined and support from the Office of Research and Ethics and data custodian at Eastern health, is easier. Figure 3 provides a summary flowchart for this internal EH only process.

The first step is the formulation research plan that includes a clear and detailed methodology. It is also important that all researchers are detailed within the research description as well. To ensure this process meets the standards required, make sure you refer back to the earlier discussion about formulating a research proposal. Significant work up front in the development of your research proposal will streamline the processes for ethical approval, decision support engagement, access to data, and a data validation process. In your research proposal you will need to identify the dataset or datasets from which you wish to source data. This includes emergency department data, inpatient record data, pharmacy data, discharge data, and others. If you wish you may engage with Eastern Health Decision Support Services (DSS) at this stage to clearly define the reports that were required from these datasets and begin to describe a process of validation for the data. At this stage this is only a guidance function from Eastern Health DSS and does not include, or guarantee, data access.

Once your research proposal is complete, the next step is submission to the office of research and ethics for ethical approval. Consistent with the preliminary discussions in this document not only to the research methods need to be described clearly but also the issues around data security, confidentiality, who will have access, and other logistical considerations should be addressed at this point. The more detailed your ethics submission is, the more streamlined the process of ethical approval will be.

Once you've received ethical approval you may then engage with EH DSS, as they require ethical approval before taking any action. Not only will ethical approval of your research proposal indicate which data you have permission to access, but also data access will be restricted to researchers named in this context in the research proposal. The process of generating reports will be guided by Eastern Health decision support which will include recommendations of which system to use. For example reports from the live system may not be the same as the backup or nonproduction system this backup system scratch that. This backup system may be more reliable than the live system and should be considered if only using retrospective data analysis. Once you have the guidance, support, and approval of Eastern Health decision support you will then have a framework for accessing the

data of interest for your research proposal. While access to much of the data within the Eastern Health system is straightforward, especially if using a single dataset, and not restricted the issues of security and confidentiality need to be considered and the challenges of linking different data sets need to be addressed.

Prior to any data analysis, validation of the data that has been obtained needs to be completed. There may be gaps in the data, fragmentation, and other limitations, so a clear process of validation needs to be developed and described to ensure validity of any results generated. Furthermore, data audit for accuracy of linkage may be required and factored into the data validation and cleaning process.

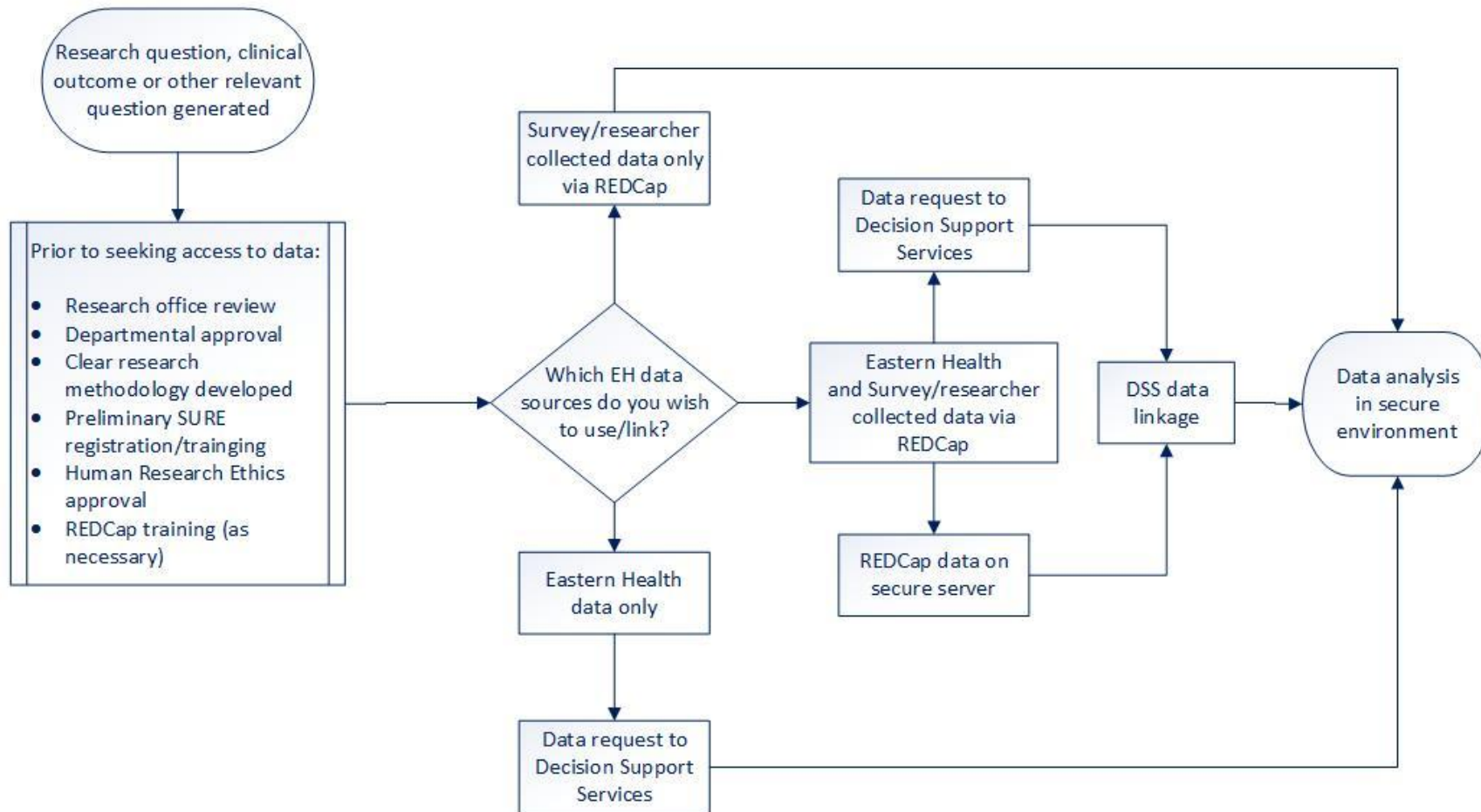
Issues of data security and the environment in which data analysis will be conducted should be explored in a research proposal for ethics approval and decision support endorsement. If there is uncertainty, or perceived requirement for data access outside this and health network, for example researchers from associated universities, then a clear strategy and reasoning should be provided to ensure the integrity and security of the data. If it is perceived that data may need to be analysed outside of the Eastern health network then access to secure data transfer and analysis environments as described in the first section of this document (for example the SURE and the SUFEX resources).

Researcher or survey collected data using REDCap

REDCap stands for Research Electronic Data Capture and is a software tool designed for easy use by researchers and clinicians. Developed at Vanderbilt University, this software is now being used by thousands of organisations across the world, with Eastern Health now a partner in the consortium. This means that EH researchers can use this very flexible data collection tool across a range of platforms (tablets, phones, computers and the web) to enhance their research and optimise access to data in a secure environment. For detail about REDCap access and use, please see the REDCap guide form the Office of Research and Ethics.

Figure 3. Process for using EH only data (including survey/researcher collected data via REDCap)

Summary data linkage process for Eastern Health researchers:
Using Eastern Health data only (including survey/ researcher collected data via REDCap).



Linking and analysing data from Eastern Health with other Victorian databases

Victorian Data Linkages (VDL) is a State government agency under the Department of Health, with funding from Commonwealth and State sources to develop new data linkage capacity in Victoria. They are the managing agency for accessing and linking data that is routinely collected in Victoria. VDL is also a state node of the Population Health Research Network (PHRN) and will be involved in any cross jurisdictional work in which Victorian data is to be included along with that of other jurisdictions.

Before pursuing this option, familiarise yourself with the VDL in terms of structure, services and expectations – all found at <https://www2.health.vic.gov.au/about/reporting-planning-data/victorian-data-linkages>.

Core databases that are accessible for data linkage by VDL are

- The Victorian Admitted Episodes Dataset (VAED)
- The Victorian Emergency Minimum Dataset (VEMD)
- Victorian Births, Deaths and Marriages
- Other state datasets to be determined

VDL provide a very structured framework for engaging with their services. This step by step description may seem complex but it is both comprehensive and essential and must guide any work in this area. To ensure all steps are completed and assist in the timeliness of this process a checklist to support state-based data linkages applications is available in the suite of resources provided by the Office of Research and Ethics.

Detailed step by step requirements for linking Eastern Health and Victorian state data

Prior to commencing this formal process, please ensure you have done the background work to prepare for the process and had informal discussion with VDL customer services officer and the EH Office of Research and Ethics to ensure preparedness.

Step 1: The researcher completes and submits a Data Request form to VDL's inbox vdl@health.vic.gov.au. The form can be downloaded from VDL's website: <http://www.health.vic.gov.au/vdl/researchers.htm>

Step 2: VDL reviews the Data Request form, provides feedback and advises researchers of any data availability or data quality limitations as well as identifies further information needed.

Note: Further linkage feasibility analysis is undertaken, technical issues are discussed within VDL team. If VDL is in doubt of data linkage feasibility, a decision is made whether to proceed or not.

Step 3: Data Custodian agree to provide the dataset 'in principle'.

Step 4: The Researcher proceeds with submitting an application to a recognised Human Research Ethics Committee (HREC) and sends Ethics application approval to VDL.

Step 5: VDL presents the completed Data Request form to the Data Custodian for approval.

Step 6: Once the Data Custodian approves of the project, VDL and the Data Custodian will work together to prepare a data extraction plan prior to the extraction process commencing. The nature of this process depends on the complexity of the project and the variables available. Some projects can require multiple iterations to determine cohort selection.

Step 7: The data recipient organisation and the Researcher sign DH Deed of Acknowledgment and Confidentiality/Condition of Release (COR).

Note: A dataset that is not readily available in the department will require endorsement from the DH Secretary to acquire the dataset into DH for a Departmental purpose before linkage can be undertaken. This will add significant delay to the acquisition process of the dataset.

Note: If DH Secretary endorses the collection of external datasets into the department, a Memo of Understanding (MoU) may need to be negotiated between the relevant DH and Data Custodians outlining Data Custodian obligations, legal limitations on disclosure of the dataset.

Step 8: If data is requested from the Victorian Death Index the Researcher submits an application to the Registrar for Births, Deaths and Marriages (BDM) (if required VDL can liaise with BDM on behalf of the researcher).

Step 9: Prior to sending and receiving the dataset, the Data Custodian and Researcher submit an eBusiness application requesting access to the department's Secure Data Exchange (SDE) portal. The application is assessed and authorisation is enabled accordingly.

Note: Once all completed supporting documents have been received, reviewed and approved, preparing the linked data requires 4 to 6 weeks.

Note: Researchers interested in receiving new data or using previously released data for a new project should follow instructions to submitting a request.

Step 10: The Data Custodian provides VDL with Linkage variables (linkage variables for a particular project will have been specified during the application process and listed in the Data Request application form from the Researcher).

Step 11: VDL will conduct the data linkage using the provided linkage variables and create the linkage map consisting of the project specific Linkage IDs¹ along with the Record IDs² to provide to the Data Custodian and extract data records from the department datasets belonging to those linked Record IDs to provide to the Researcher.

Step 12: The Data Custodian uses the project specific Linkage IDs along with the Record IDs provided by VDL to extract the approved research variables, removing all personal identifiers from their dataset.

Step 13: The Researcher is thus provided with de-identified content data of each record and its corresponding project Linkage ID by VDL via DH Secure Data Exchange (SDE) application.

Step 14: Using the project Linkage ID, the researcher can determine which records from different datasets belong to the same individual without having access to the personal information in order to create a merged dataset for their analysis and address research questions.

Step 15: Data Custodian review publication in accordance with the application.

Step 16: Data is returned to VDL on completion of the project to be destroyed.

Step 17: Researcher provides VDL with a copy of research/ manuscript documents for departmental review before publication.

Further information for using VDL services for Victorian data linkage

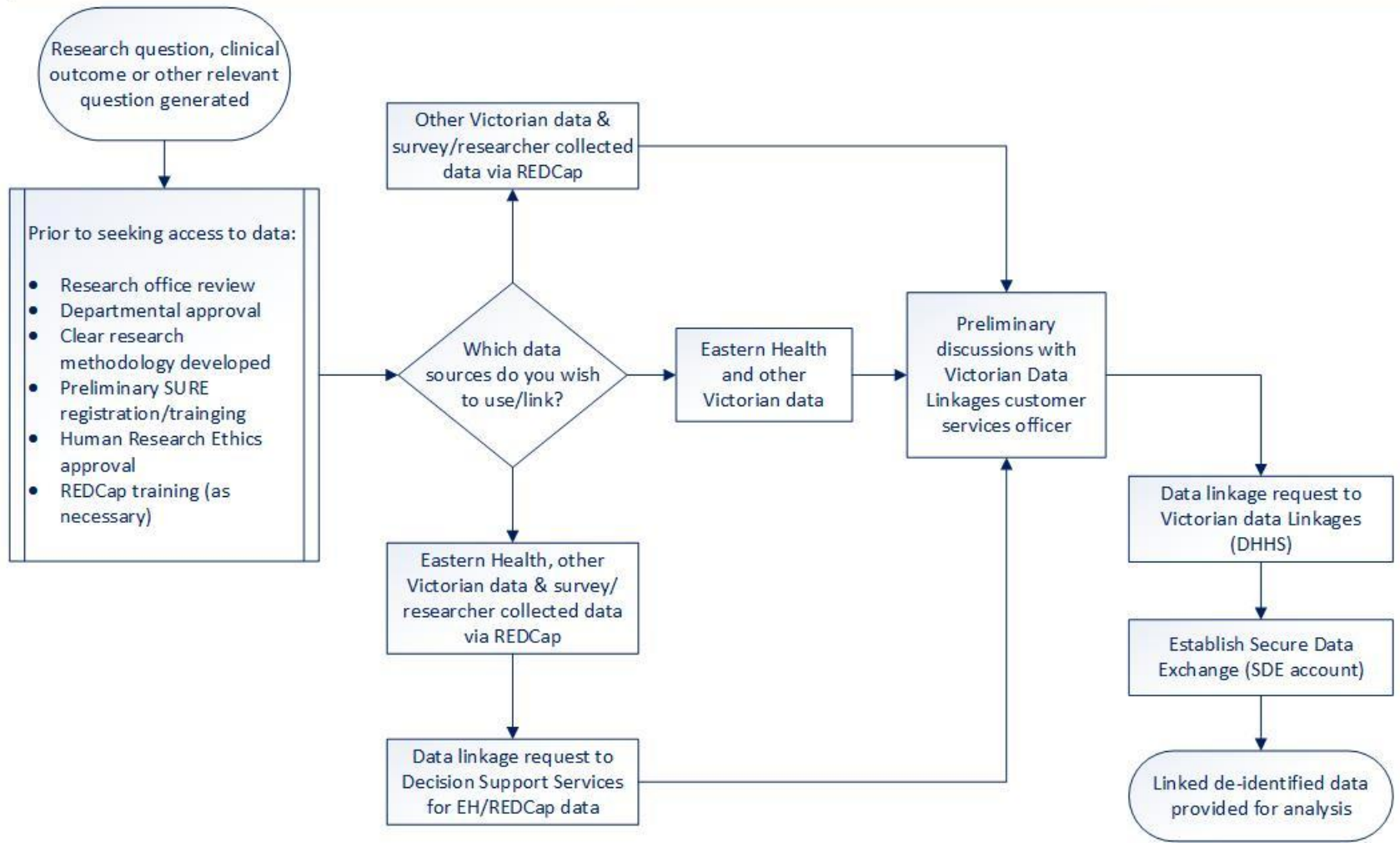
VDL for Researchers <https://www2.health.vic.gov.au/about/reporting-planning-data/victorian-data-linkages/for-researchers>

Instructions for Registering SDE Account

<file:///C:/Users/gregd/Downloads/User%20guide%20for%20SDE%20download%20and%20upload%20Rights%20-%20PDF.pdf>

Figure 4. Process for using EH and /or Victorian data (including survey/researcher collected data via REDCap)

Summary data linkage process for Eastern Health researchers:
 Using Eastern Health (including survey/ researcher collected data via REDCap) and/or other Victorian data.



Linking and analysing data from Eastern Health with data sources from other jurisdictions in Australia

As would be expected, the most complex research using available datasets occurs when linkage is required between datasets from other jurisdictions in Australia with Eastern Health data. These may be datasets from specific states that reflect the statutory required data collection in those states, which are often useful for comparison as it may allow the identification of individual health networks and hospitals that may be similar in terms of size, population served or other key characteristics. There are also datasets of routinely collected data by the Commonwealth. As each jurisdiction has legal and regulatory authority and responsibility for the data and the people the data represent, the application and approvals process is necessarily more complicated. This complexity has been recognised by the Commonwealth and strategies have been devised to facilitate access to the range of datasets available through centralised agency and process. The Population Health Research Network (PHRN) represents a program backed by the Australian government as part of the National collaborative research infrastructure strategy and supported by all states and territories. The PHRN is designed as Australia's first national data linkage network. For researchers considering any project involving what is described as cross jurisdictional data and essential starting point is familiarity with the PHRN, its activities, services and resources. The PHRN was established as an initiative of the Australian Government as part of the National Collaborative Research Infrastructure Strategy (NCRIS). The network represents a group of collaborators at state and federal levels including the Australian Institute of Health and Welfare (AIHW) whose Data Integration Unit plays a key role in the secure linkage process. More information about the Population Health Research Network can be found at <http://www.phrn.org.au/>. The process for using data in the context is summarised in Figure 5.

The PHRN is the starting point for any cross jurisdictional data linkage research but depending on the nature and setting from which data is obtained other agencies may be involved. The PHRN gives guidance as to which agencies will need to be involved and provide links and contacts for all relevant agencies. As with the processes for using Victorian data, there are also guidelines on issues around secure data transfer and a secure environment for data analysis.

Datasets accessible

The PHRN list the datasets available in each jurisdiction, describing the content and process for that data been provided, identifies data custodians and any relevant regulatory issues relating to that dataset and importantly identifies the year of most recent update. The list of core datasets and years

of available data can be found at http://www.phrn.org.au/media/81210/ganttchart_june-2017.pdf .

The PHRN also lists the nature and detail of data available including

- A description of the data collection
- The date range for which data is available for linkage
- Which jurisdiction the data represents
- The data linkage unit responsible for linking the data collection
- Contact details for the data collection
- Links to variable lists for each data collection
- Links to any metadata available (including mode and method of collection, changes over time) where available
- Links to any validation studies of the data available

When reviewing these datasets, significant variation will be observed with the lag time between data generation and availability varying from months to years. This is an important consideration in the research planning process and one of the reasons why engagement with data custodians and linkage agencies very early in the planning process is critical to ensure a realistic timeframe is provided for any research project.

Data access

Once an initial research plan has been developed and key databases of interest identified, as with other forms of research, engagement with the office for research and ethics is a primary starting point. This engagement will include guidance on the process for gaining access to specific cross jurisdictional datasets and provide support and advice to facilitate the process and reduce the risk of refusal of access to the data. Guidance will also be provided by the PHRN on which of the two Commonwealth government accredited data integration authorities that should be contacted for the research project of interest. This linkage process is described in more detail further on. An overview of this process known as Cross Jurisdictional Linkage of Data (PHRN) may be found at

<http://www.phrn.org.au/for-researchers/data-access/cross-jurisdictional-linkage-of-data/>

In the process of seeking data access, clear and specific reasons must be provided for access to any particular dataset and the variables of interest within that dataset. Researchers will be limited to data that is directly relevant to the clearly described research project (with a focus on a PICO question), and that data may only be used for that specific project at that point in time. One of the

core responsibilities of the PHRN is the maintenance of data integrity and protection of an individual's confidentiality. Therefore, access to datasets is not provided for researchers to go on a "fishing expedition" and see what may be trawled from the data.

Linking cross jurisdictional data

The PHRN will give advice on which Commonwealth Government Accredited Data Integration Authority to engage with – either the AIHW Data Integration Services Centre Unit (<http://www.aihw.gov.au/organisation-detail/?id=10737420129>) or Curtin University (PHRN) Centre for Data Linkage (<http://www.curtin.edu.au/research/cphr/centre-for-data-linkage.cfm>).

Depending on the advice of the Data Integration Authority, the nature of the data and the resources and infrastructure available to the researcher

Secure data transfer

For the Data Integration Authority to link datasets they must be provided via a secure transfer mechanism – to be determined by the Data Integration Authority. It is likely that the SUFEX platform, described in the introduction, will be used for this process and for secure return of linked, de-identified data to researchers for analysis. It is essential that researchers are familiar with the process and requirements (organisational and technical) for use of SUFEX.

Secure data transfer (SUFEX) <http://www.phrn.org.au/for-researchers/services-for-researchers/sufex/>

Secure environment for data analysis

As mentioned previously, to ensure data security a secure environment is necessary in which data analysis may take place. When using Cross-jurisdictional data the SURE process is likely to be required by one of the two accredited linking agencies unless there is approval of a definite secure environment in an organisational setting. At this stage, this is very uncommon. If a research project requires the use of cross jurisdictional data, the process of registering and engaging with SURE as early as possible during a study (or beforehand if time allows) as training is required and there are other approval processes that need to be addressed. The relevant information for using SURE can be found at:

The Secure Unified Research Environment (SURE) <http://www.saxinstitute.org.au/for-researchers/collaborate-with-us/> and <http://www.phrn.org.au/for-researchers/services-for-researchers/secure-unified-research-environment-sure/>

Figure 5. Process for using cross-jurisdictional data

Summary data linkage process for Eastern Health researchers:
Using Eastern Health and/or other jurisdictions.
(Cross cross-jurisdictional data linkage)

