

Applying for Linked Data



Nadine Wiggins

Tasmanian Data
Linkage Unit

22 June 2016



UNIVERSITY of
TASMANIA

MENZIES 
Institute for Medical Research

What are the two most common questions asked of us?

1. How much will linked data cost?



2. How long will it take?



TDLU's Response

“Well.....It all depends!”

There are a number of factors that impact on cost and timeliness including;

- Complexity of the project (eg. multiple cohorts, use of controls, use of Commonwealth data)
- Size of the project (eg. number of datasets to be linked)
- Data availability
- Time for ethics and data custodian approvals
- Workload of the TDLU and the number of projects in the ‘queue’



Applying for Linked Data

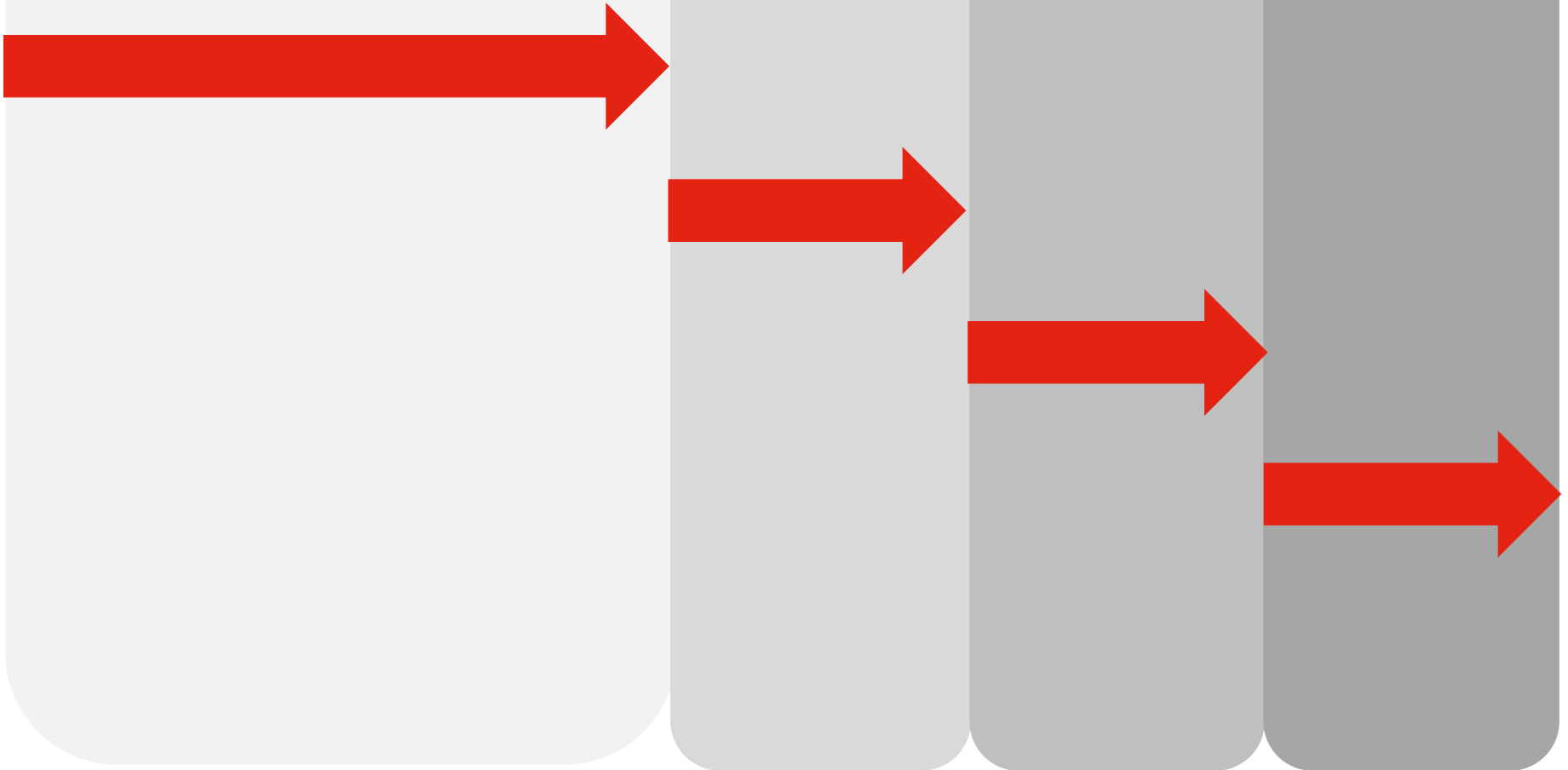
- In order to assist researchers plan for a data linkage project, the TDLU have developed a flowchart to step through the key tasks involved in applying for and receiving linked data from the TDLU

APPLICATION
(allow at least 4 to 8 weeks)*

APPROVALS
(allow 8-10 weeks)*

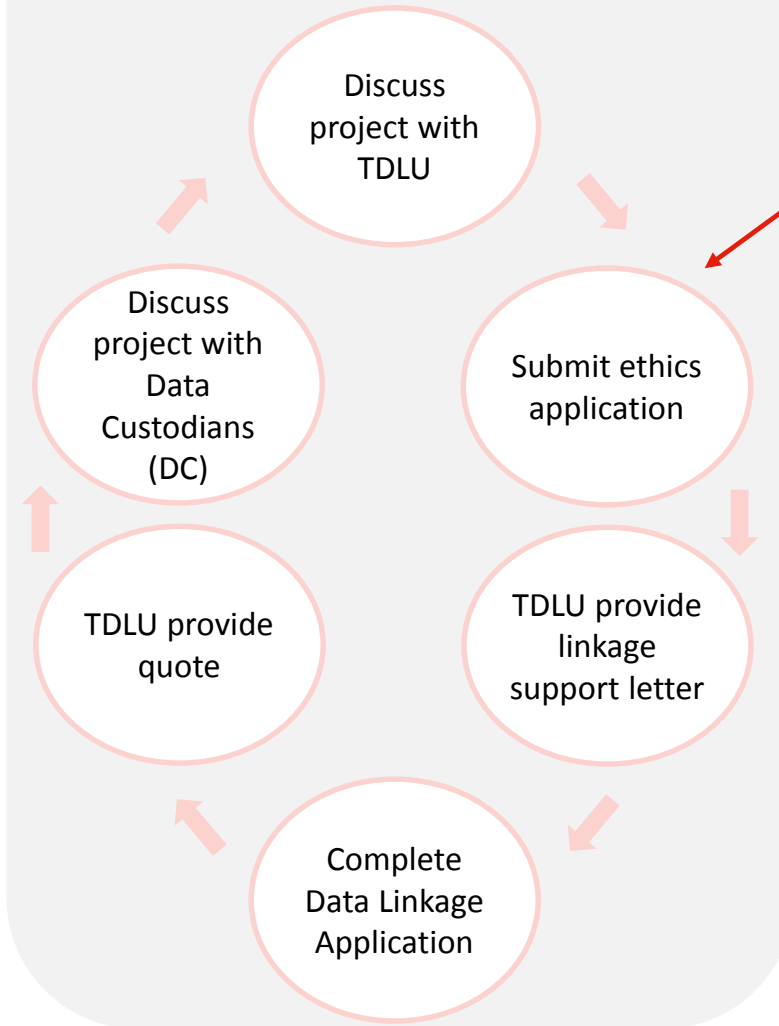
**DATA LINKAGE &
EXTRACTION**
(allow 8-12 weeks)*

REVIEW



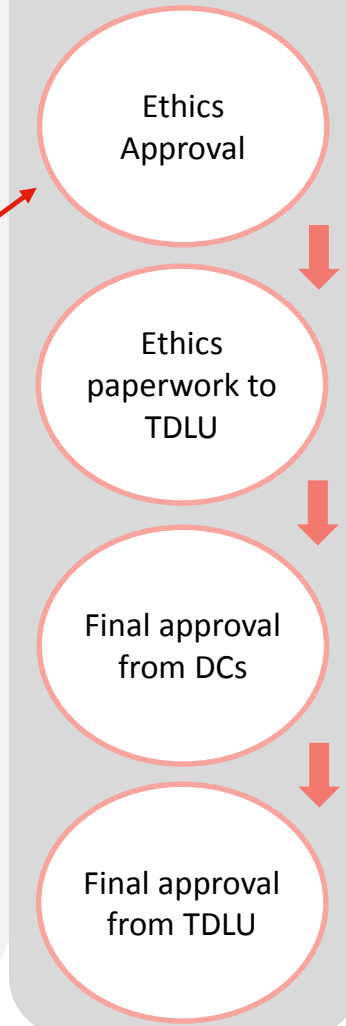
APPLICATION

(allow at least 4 to 8 weeks)*



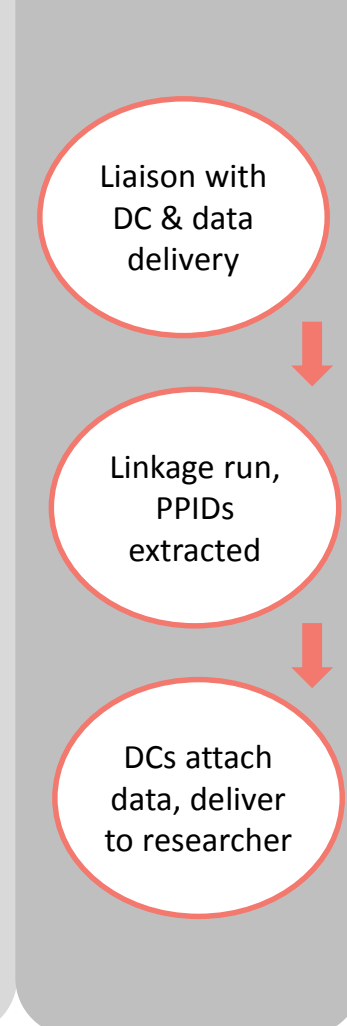
APPROVALS

(allow 8-10 weeks)*

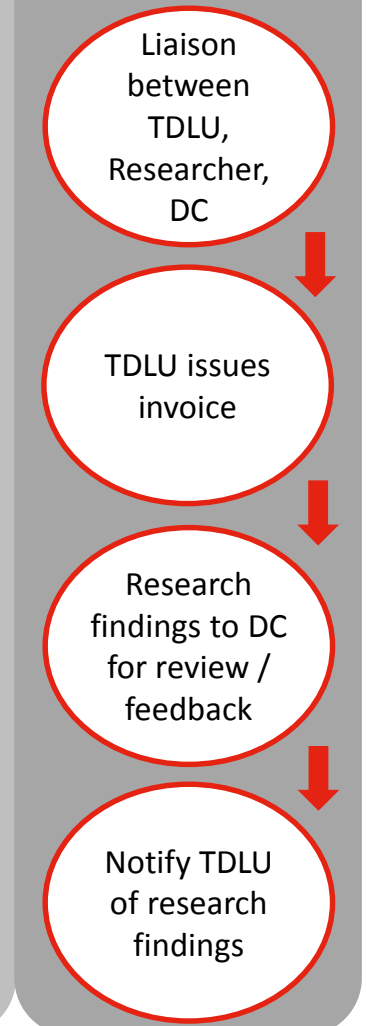


DATA LINKAGE & EXTRACTION

(allow 8-12 weeks)*

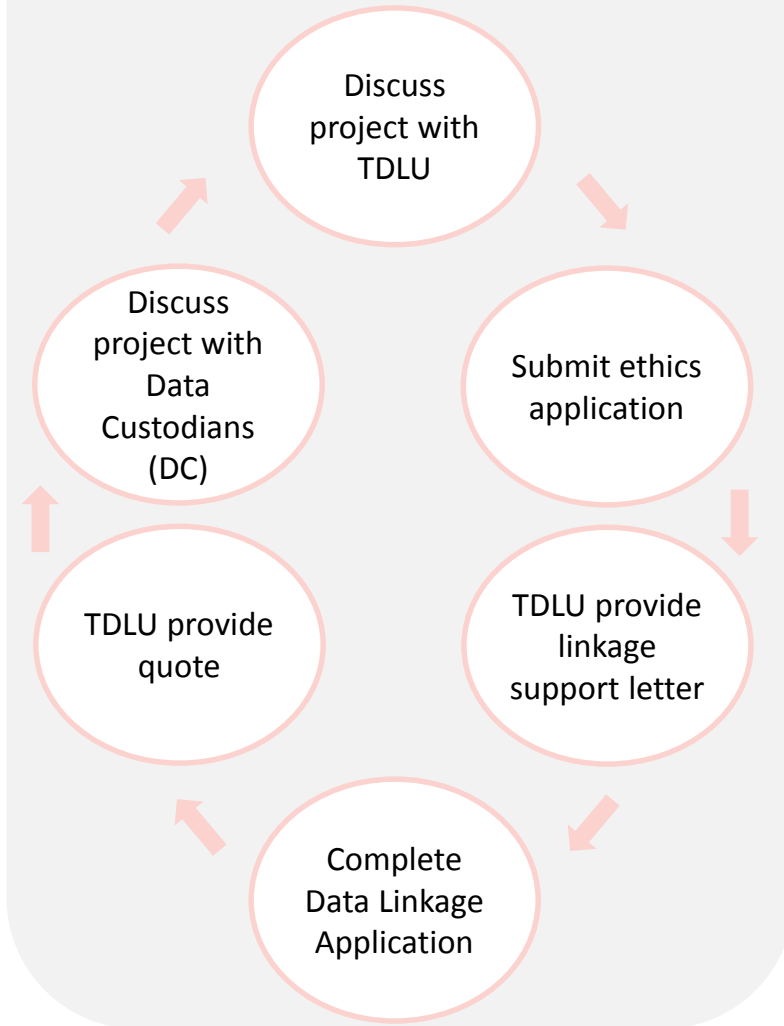


REVIEW



APPLICATION

(allow at least 4 to 8 weeks)*



Application Phase

- There is no one starting point
- It can be an iterative process
- We strongly suggest discussing your project with us in the first instance so that we can help streamline the process as much as possible
- Allow time to complete this process
- Plan your project well in advance, with clear aims and objectives

APPLICATION

(allow at least 4 to 8 weeks)*

Discuss
project with
TDLU

Discuss the Project with the TDLU

An initial discussion will typically cover:

- Project aims and objectives
- Datasets to be linked
- Availability of data
- Feasibility of project
- Project timeframes
- Ethics requirements
- Indication of costs
- Where to from here

APPLICATION

(allow at least 4 to 8 weeks)*

Discuss
project with
Data
Custodians
(DC)

Discuss the Project with Data Custodians

- Have data custodians already been approached about the project?
- Are they happy to participate?
- Does the TDLU have an existing data supply agreement in place?

APPLICATION

(allow at least 4 to 8 weeks)*

TDLU provide
quote

TDLU Provides Quote

The TDLU will provide a data linkage quote based on the following factors:

- Size of project
- Complexity of project
- Number of datasets for linkage
- Size of cohort dataset(s)
- Data quality issues
- Level of clerical review required
- Additional programming required

APPLICATION

(allow at least 4 to 8 weeks)*

Complete
Data Linkage
Application

Complete a Data Linkage Application

1. TDLU Application
or
 2. PHRN Online Application?
- State based linkage only
 - National / cross-jurisdictional linkage

The TDLU will work with researchers to develop a linkage protocol.

APPLICATION

(allow at least 4 to 8 weeks)*

TDLU provide
linkage
support letter

Data Linkage Support Letter

The TDLU will provide a letter of support which states that the project is technically feasible.

APPLICATION

(allow at least 4 to 8 weeks)*

Submit ethics
application

Submit an Ethics Application

Human Research Ethics Committee (HRECs) play a central role in the ethical oversight of research involving people.

HRECs review research proposals to ensure that they are ethically acceptable and in accordance with relevant standards and guidelines.

All projects requiring linked data provided by the TDLU must have HREC approval.

Ethics approval involves the submission of an application to an approved Ethics Committee; in Tasmania this body is jointly constituted between the UTAS and the DHHS.

APPLICATION

(allow at least 4 to 8 weeks)*

Discuss
project with
TDLU

Discuss
project with
Data
Custodians
(DC)

TDLU provide
quote

Complete
Data Linkage
Application

Submit ethics
application

TDLU provide
linkage
support letter

APPROVALS

(allow 8-10 weeks)*

Ethics
Approval

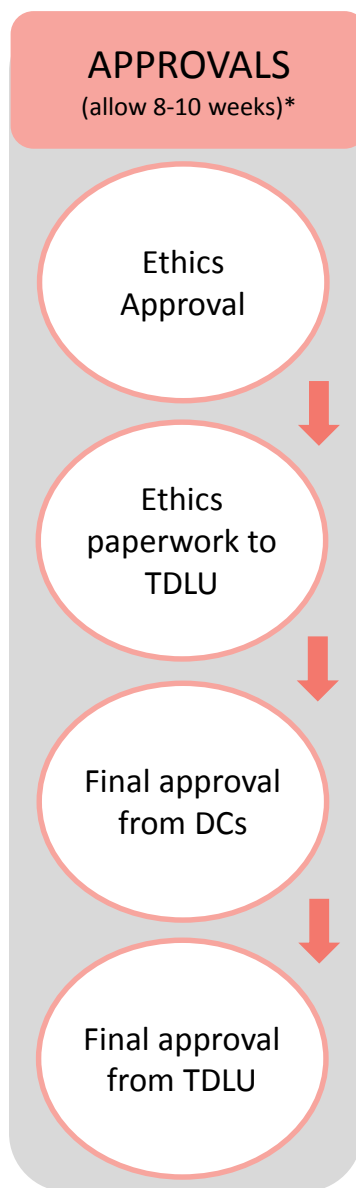
Ethics
paperwork to
TDLU

Final approval
from DCs

Final approval
from TDLU

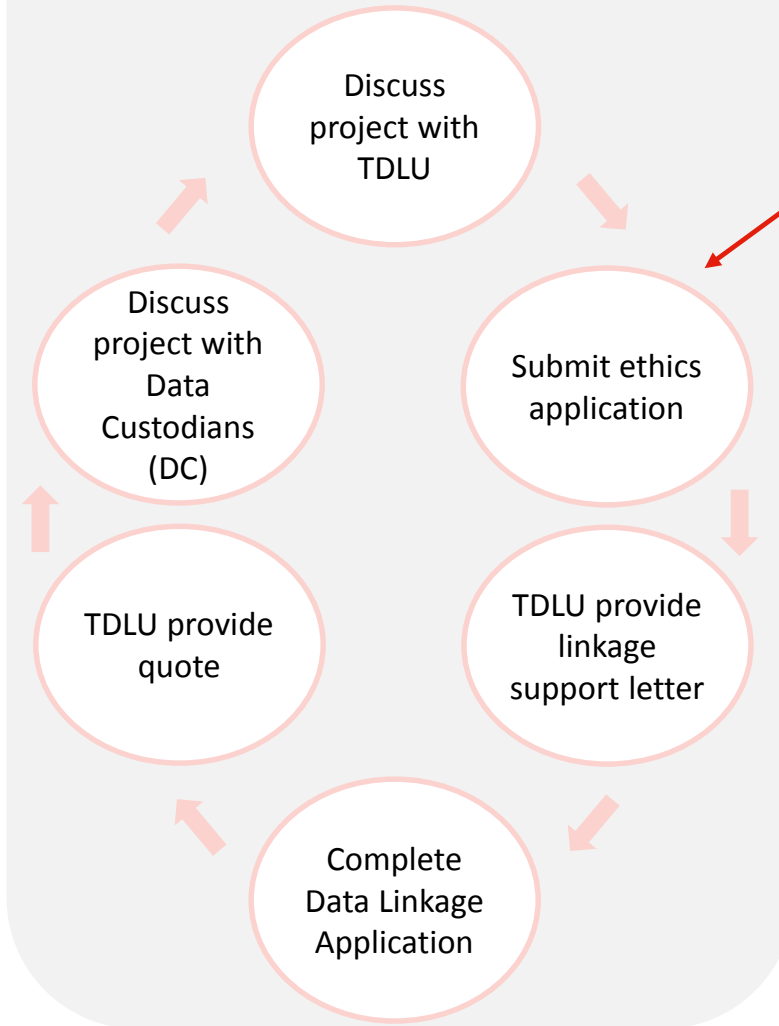
Approvals

- Ethics approval letter and further HREC amendments and approvals to be forwarded to TDLU
- TDLU seek written approval from data custodians
- Timeframe for receiving data custodian approvals varies
- After ethics and data custodian approvals the TDLU will review all paperwork and provide final approval



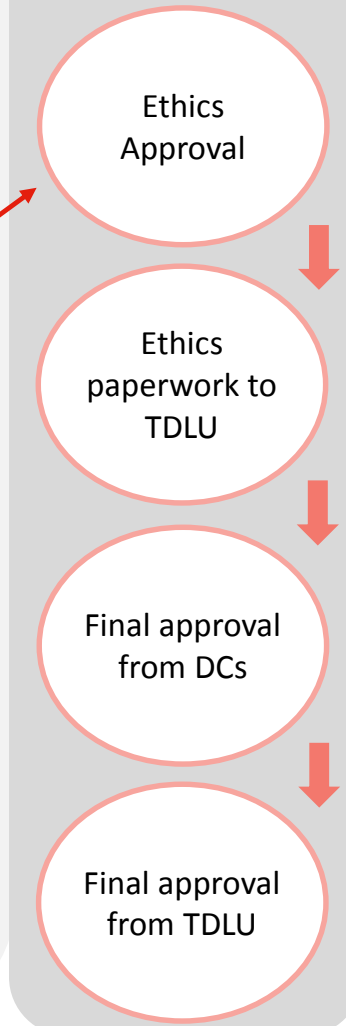
APPLICATION

(allow at least 4 to 8 weeks)*



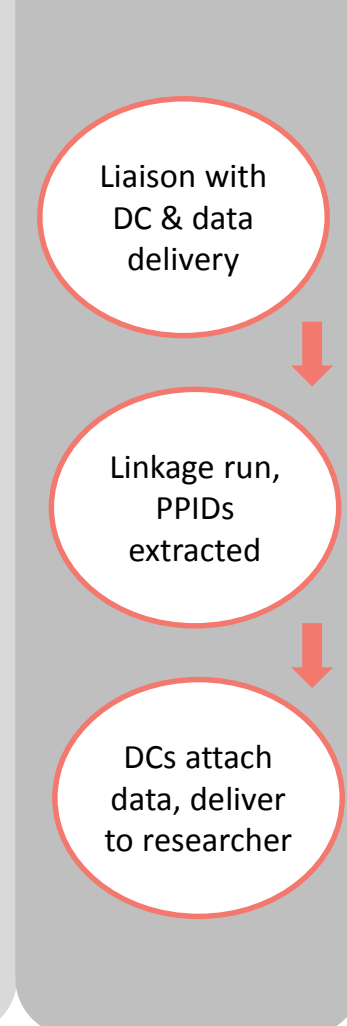
APPROVALS

(allow 8-10 weeks)*



DATA LINKAGE & EXTRACTION

(allow 8-12 weeks)*



**DATA LINKAGE &
EXTRACTION**
(allow 8-12 weeks)*

Liaison with Data Custodians

- Arrange secure delivery of data from custodians
- Timeframes for linkage key delivery arranged
- Seek advice on data quality issues

Liaison with
DC & data
delivery

Linkage Conducted and PPIDs extracted

This process is another presentation in itself!

Essentially the TDLU will do the following tasks;

- Clean new dataset
- Import new dataset to linkage engine
- Run a de-duplication process
- Link the dataset with existing MLM datasets
- Conduct clerical review
- Conduct QA
- Extract linkage keys and provide to data custodians
- Prepare QA report and provide to researcher

DATA LINKAGE &
EXTRACTION
(allow 8-12 weeks)*

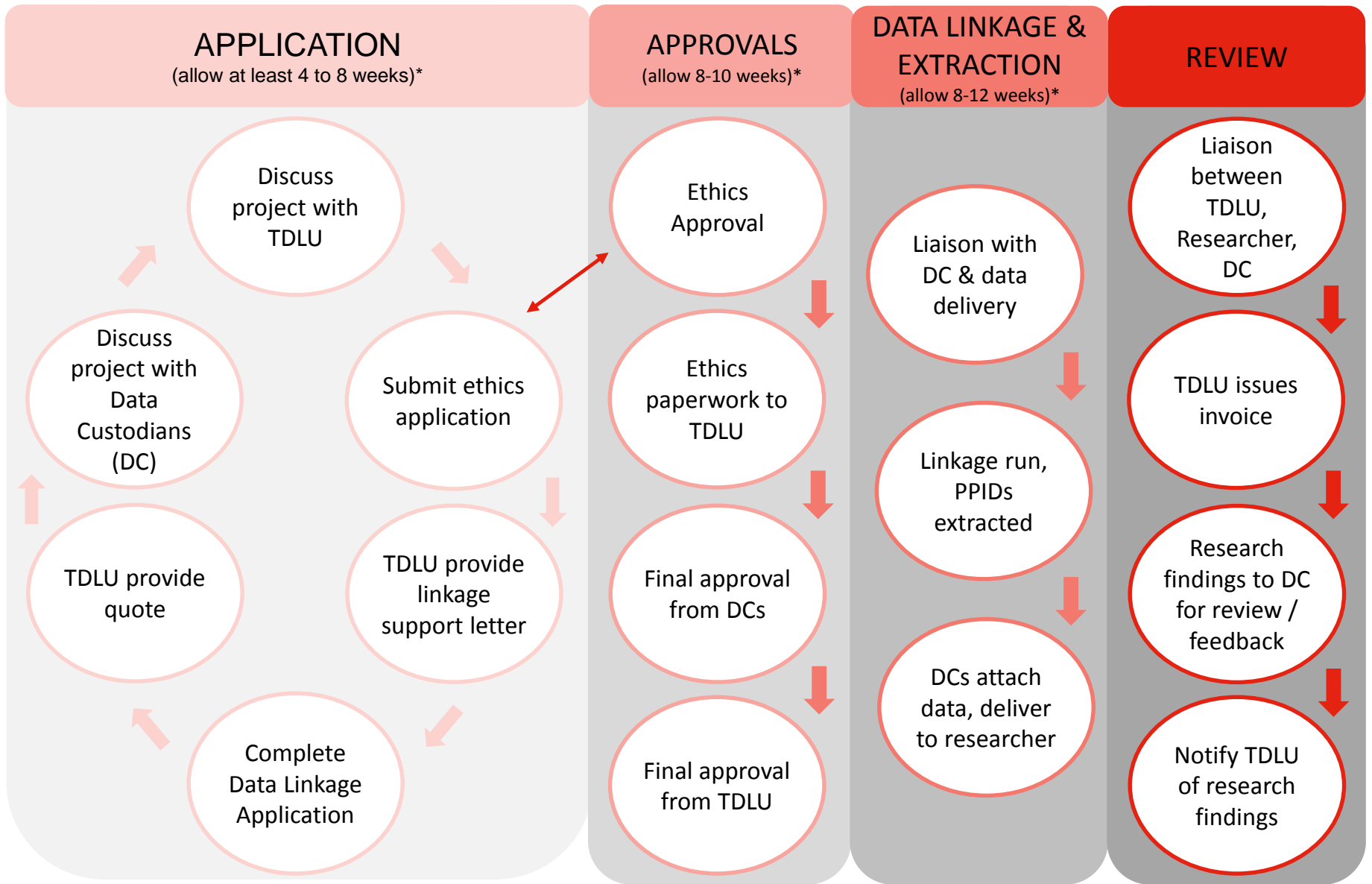
Linkage run,
PPIDs
extracted

Data Custodians Provide Data to Researcher

- The data custodian(s) receive a linkage key file containing a PPID and their source identifier
- The data custodian attaches the relevant content/clinical data as per the approvals provided and strips off the source identifier leaving a PPID and content/clinical data.
- The data is transferred securely to the researcher or to the SURE
- The data provided will not enable researchers to re-identify individuals

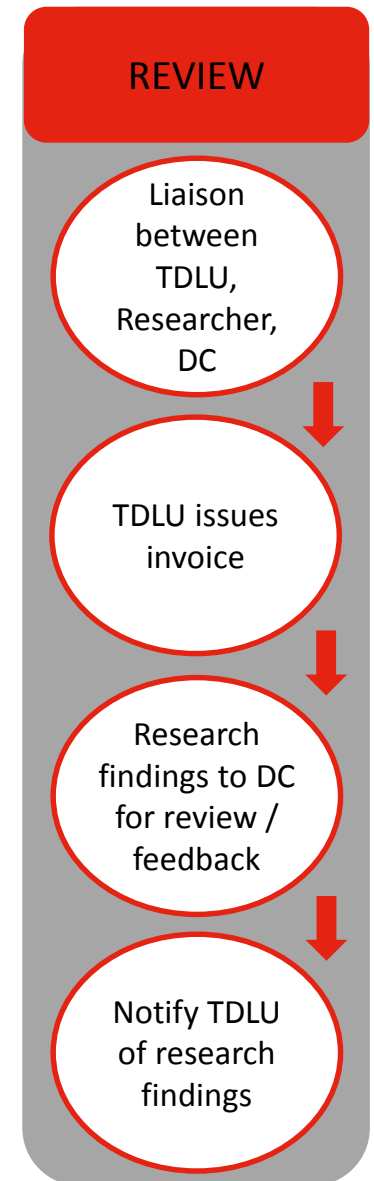
DATA LINKAGE &
EXTRACTION
(allow 8-12 weeks)*

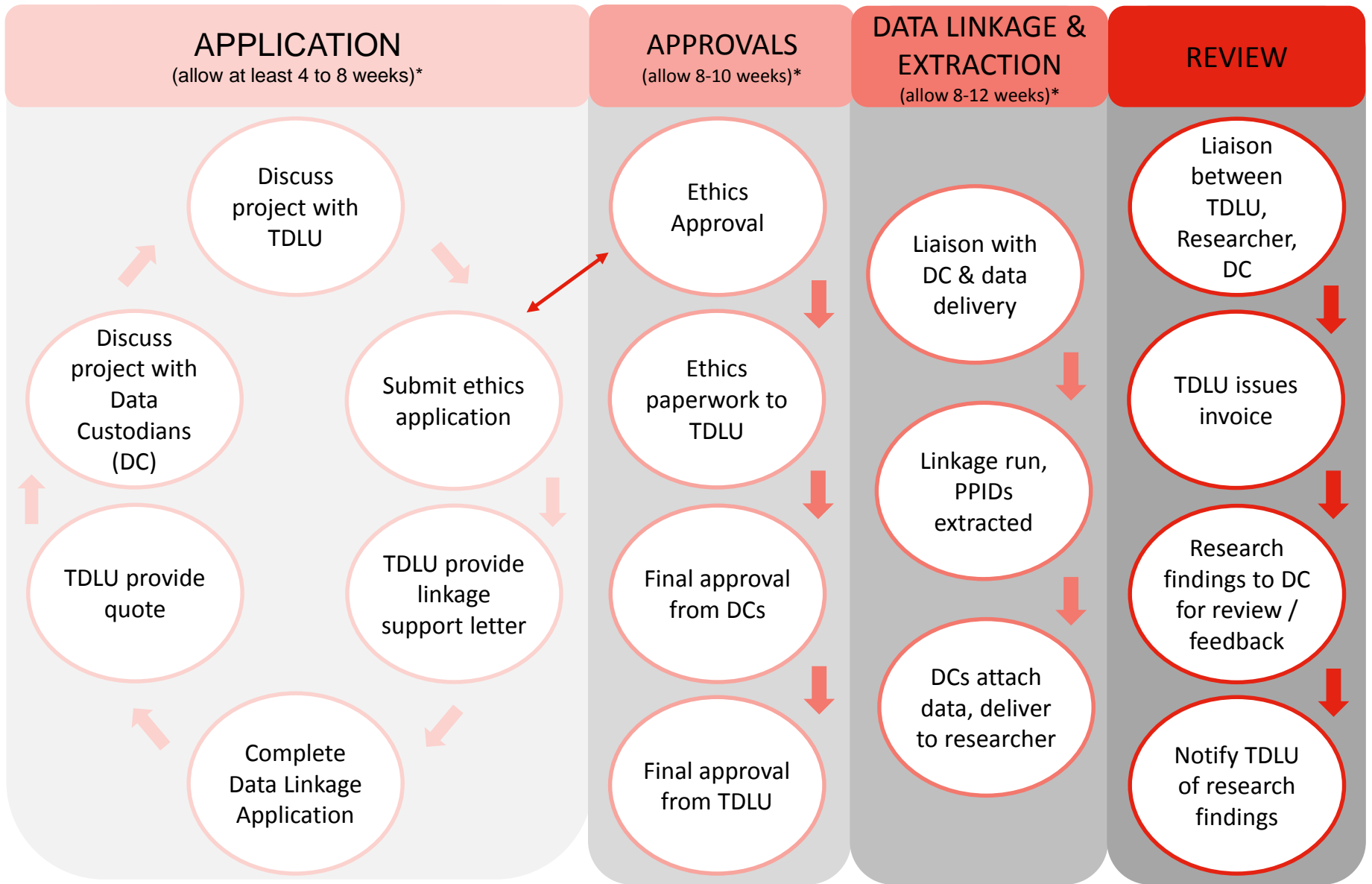
DCs attach
data, deliver
to researcher



Review Phase

- During the review phase there may be ongoing discussions between the TDLU, researcher and data custodians relating to data extraction, data quality, metadata etc.
- Invoice issued to researcher
- Some data custodians may also charge for data extraction
- The researcher is obliged to notify data custodians of their research findings for review and/or feedback in accordance with their data release approval conditions
- The researcher is to also notify the TDLU of publications, media and conference presentations relating to the linked data supplied.





In Summary

- Talk to us
- Allow plenty of time
- Plan your project and complete the necessary paperwork and approval processes well ahead of required data delivery
- Often the actual data linkage process is quick and relatively straightforward, it is the application and approval processes that takes time
- There is a 'queue' and within the actual linkage system itself only one dataset can be linked at a time
- We will always try our best to accommodate researcher requirements and timeframes, however short timeframes are not always possible

APPLICATION

(allow at least 4 to 8 weeks)*

APPROVALS

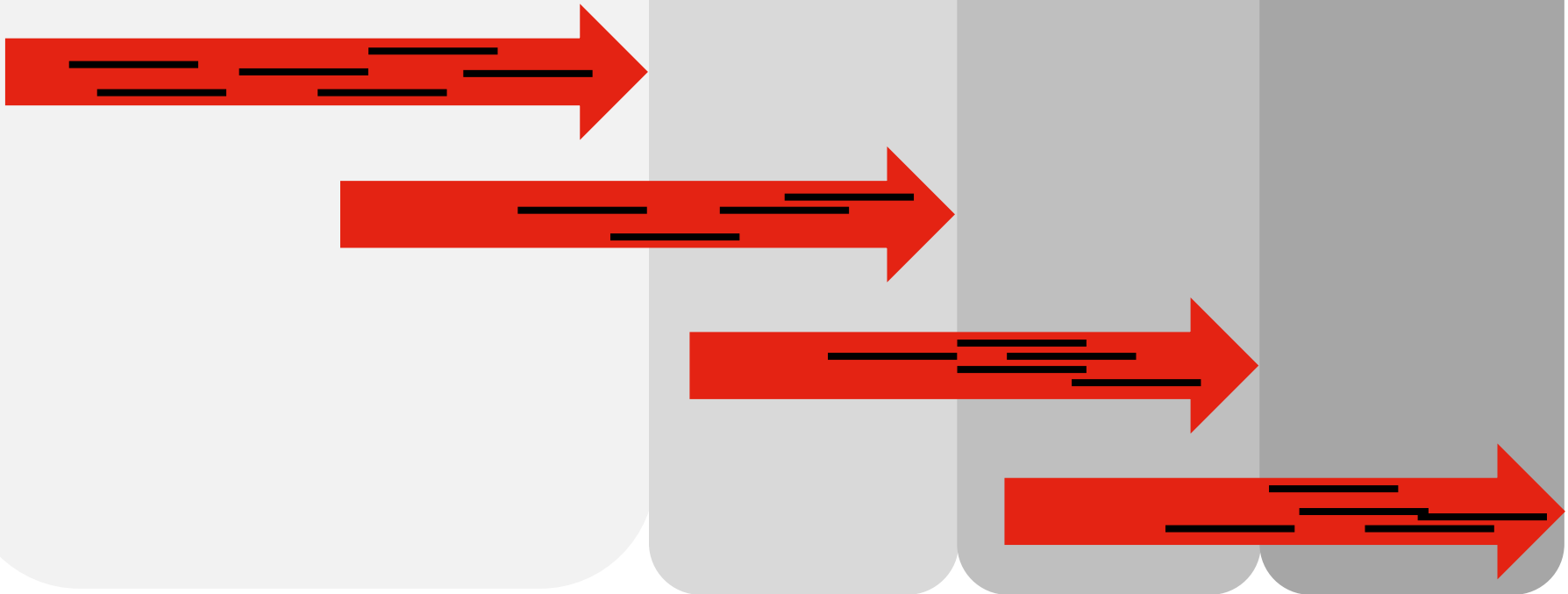
(allow 8-10 weeks)*

DATA LINKAGE & EXTRACTION

(allow 8-12 weeks)*

REVIEW

At any point in time, the TDLU is processing multiple applications that fall within different parts of this flowchart.



Further Information

- Information and contact details for the Tasmanian Data Linkage Unit can be found on the Menzies Institute for Medical Research website
- <http://www.menzies.utas.edu.au/research/research-centres/data-linkage-unit>
- Please contact us if you would like further information about data linkage

Questions???

