

## Supporting the timely, easy and cost effective access to high-quality linked data via the Custodian-Controlled Data Repository

Schneider, M<sup>1</sup>, Radbone, C<sup>1</sup>, Stanley, A<sup>1</sup>, Woollacott, A<sup>1</sup>, Churches, T<sup>2</sup>, Farrow, J<sup>3,1</sup>, Basso, P<sup>4</sup>, Hardin, T<sup>5</sup>, McNamara, M<sup>6</sup>, and Harrison, J<sup>7</sup>

<sup>1</sup>SA NT DataLink

<sup>2</sup>Ingham Institute for Applied Medical Research and South Western Sydney Clinical School, UNSW Sydney

<sup>3</sup>Farrow Norris

<sup>4</sup>Ballarat Health Services

<sup>5</sup>SA Health, Government of South Australia

<sup>6</sup>Sax Institute

<sup>7</sup>Flinders University

### Introduction

While linkage units perform population data linkage with high efficiency, other parts of the workflow from custodians to researchers remain largely outside the control of linkage operators. Most importantly, resource constraints at data custodians often limit quality control (QC) efforts and lead to delays in the data delivery to researchers.

### Objectives and Approach

to overcome these challenges, we have created the Data Integration Unit (DIU) to undertake content data management and delivery in conjunction with the custodians who remain the principal data curators. Data is managed in the Custodian-Controlled Data Repository (CDDR) a highly secure virtual repository for data storage, analysis and access, established and operated by the DIU. Stringent controls for user access and data flows ensure that data is provided safely by custodians. Overseen by the custodians, DIU staff undertake QC activities, and integrate and deliver multiple datasets for approved linkage projects.

### Results

Long-term data storage in the CDDR decreases data custodian workloads by reducing the frequency of content data provision to periodic updates. Feedback loops built into the QC process allow custodians to improve their datasets by learning from data issues identified by the DIU.

Extensive QC undertaken by DIU staff on individual datasets and data validation across multiple datasets held in the CDDR ensure that the quality of data provided to researchers is improved. Moreover, DIU staff dedicated to data integration provide faster content data delivery to re-

searchers. Lastly, the CDDR reduces the number of custodians researchers need to liaise with for data provision.

### Conclusion/Implications

Operational since February 2018, the DIU has delivered content data for several linkage projects, based on key datasets stored in the CDDR. The incorporation of additional datasets is currently negotiated.

Recognising recent developments in secure analytics infrastructure, the further evolution of the CDDR towards a cloud-based model is anticipated.

