



Government of **Western Australia**  
Department of **Health**

# **WA Health Data Linkage Services**

## **Limitations and Suitable Use of Linked Data**

**Version 1.0**

January 2024



## Version Control and Approval

This document should be considered a 'live document' and will be reviewed regularly and updated as required to:

- Reflect changes to policy and/or procedures
- Incorporate stakeholder feedback
- Determine effectiveness, relevance, and currency

Review and update of this document is coordinated by the Data and Information Systems unit within the Information and System Performance Directorate (ISPD).

| Version | Published date | Approved by                 | Amendment(s) |
|---------|----------------|-----------------------------|--------------|
| 1.0     | 8 January 2024 | Pavi Krishnan,<br>A/Manager |              |
|         |                |                             |              |

## Contact

Enquiries relating to this guide may be directed to:

Title:                      Manager, Data Linkage Services  
Directorate:              Information and System Performance  
Email:                     [DataLinkageStrategy@health.wa.gov.au](mailto:DataLinkageStrategy@health.wa.gov.au)



## Table of Contents

|   |   |
|---|---|
| 1. Introduction                             | 3 |
| 1.1 Background                              | 3 |
| 1.2 Purpose                                 | 3 |
| 2. What are the limitations to linked data? | 4 |
| 3. How should linked data be used?          | 4 |
| 4. Further information                      | 4 |



## 1. Introduction

### 1.1 Background

Data Linkage is a technique for establishing unique identifiers (data linkage keys) which are used to connect information from different data sources thought to relate to the same person, family, place or event via data integration processes. Information is created when a person encounters certain services, for example, when they visit an emergency department, stay in a hospital or register the birth of their child.

The Data Linkage Services branch at the Department of Health uses probabilistic linkage techniques to link data in WA. As datasets from various sources do not share a universal ID, records must be linked probabilistically to support data integration processes. The data linkage process is complex and primarily utilises a range of demographic data variables (or otherwise known as 'reasonably identifiable information' such as name, date of birth and address) to bring together service records that likely belong to the same individual, place, or event, generally utilising a software-driven probabilistic matching process.

The WA Data Linkage System (WADLS), managed and operated by the Information and System Performance Directorate within the Department of Health is used to create, store, update and extract links between datasets. Please visit the [Data Linkage Services website](#) for more information on the complex data linkage process.

### 1.2 Purpose

The purpose of this document is to articulate the limitations to and the appropriate use of linked data.



## 2. What are the limitations to linked data?

While data linkage techniques allow records to be linked where it would not otherwise be possible, it is important to consider the limitations caused by the probabilistic nature of linkage.

- Data linkage quality is dependent on the quality of the individual datasets, and each dataset varies greatly in quality. Data quality issues may be due to:
  - Errors in data collection (e.g., typographical errors, miscommunication due to information relayed over the phone)
  - Assumptions made about certain data items (e.g., address is usually presumed to be a permanent residential address, yet individuals may provide a temporary address).
- Due to the probabilistic nature of linkage, there is a potential for instances of 'false positives' where records are erroneously linked, as well as the possibility of 'false negatives' where records that should be linked are missed. Data linkage is an iterative process and links are regularly created, modified, and deleted as new information becomes available.
- Each dataset captures only a subset of the WA population, based on the purpose and circumstances of the collection. For example, a given data collection may only include records from public, not private health providers.

## 3. How should linked data be used?

The Data Linkage Services branch employs a variety of strategies to promote high quality linkage. However, 'perfect' data quality and linkage cannot be guaranteed. Therefore, **linked data is not recommended for use at the level of individual records**; that is, linked data should generally not be used to seek, interpret or act upon information about any one individual, unless deemed appropriate by the Data Steward or Data Custodian of the linked information.

Rather, linked data is suitable for *summarising and analysing a population or cohort*, such as to identify trends over time or to assess the level of engagement with a new service.

## 4. Further information

For more information, please contact [DataEngineeringDataL@health.wa.gov.au](mailto:DataEngineeringDataL@health.wa.gov.au).



Government of **Western Australia**  
Department of **Health**

**This document can be made available in alternative formats  
on request for a person with disability.**

© Department of Health 2024

Copyright to this material is vested in the State of Western Australia unless otherwise indicated. Apart from any fair dealing for the purposes of private study, research, criticism or review, as permitted under the provisions of the *Copyright Act 1968*, no part may be reproduced or re-used for any purposes whatsoever without written permission of the State of Western Australia.