



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

# WA Health Data Linkage Services

## Minimum Data Requirements for Linkage

**Version 2.0**

March 2026

Classification: Official

[health.wa.gov.au](http://health.wa.gov.au)

# 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	May 2023	B Singh	Original Version
1.1	November 2023	B Singh	Reviewed the content of Table 1 within Section 2.1.  Updated the content of Table 2 within Section 2.2.
2.0	April 2026	R Miln	Clarified interpretation of mandatory data requirements  Strengthened definition of unique record identifiers  Expanded recommended data items  Enhanced guidance on data formats, event-based data, and delta updates  Updated document classification, template, and wording for clarity and consistency.

## 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	1
1.1. Background	1
1.2 Purpose	1
2. Minimum data requirements	2
2.1 Mandatory data items	2
2.2 Recommended additional valuable data items	2
2.3 Accepted and preferred data formats	3
2.4 Additional data requirements	4
2.4.1 NULL Values	4
2.4.2 Non-human and dummy records	4
2.4.3 Consistent formatting within data items	4
2.4.4 Metadata	4
2.4.5 Duplicate rows	4
3. Data Updates	5
3.1 Deltas	5
3.1.1 Documentation of data extraction processes	5
3.2 Data changes	5
4. Appendices	6
4.1 Data provider checklist	6

# 1. Introduction

## 1.1. Background

Data linkage is a method used to integrate information from different sources (e.g. different government agencies) thought to relate to the same person, place, family, or event to provide a complete picture of an individual's experience and service interaction over time.

Population-based linkage services in Western Australia (WA) are delivered through the internationally recognised dynamic and enduring Western Australian Data Linkage System (WADLS). The WADLS is a core function of the WA Department of Health (Department), with the infrastructure having supported significant improvements for health policy and care in WA as well as supporting research projects using linked data for analysis.

The data linkage process is complex and primarily utilises demographic data (or otherwise known as 'reasonably identifiable information' such as name, date of birth and address) to bring together records likely to belong to the same individual. To be feasible, data linkage requires a minimum set of demographic and unit level data. Linkage quality and efficiency increase with enhanced data quality, completeness, and consistency.

## 1.2 Purpose

The purpose of this document is to provide guidance to data providers on the minimum requirements of their data to enable the application of high quality and efficient data linkage practices, and to maintain the integrity of the WADLS.

## 2. Minimum data requirements

### 2.1 Mandatory data items

The following data items are typically mandatory for effective data linkage:

Table 1. Mandatory Data Items for Linkage

Data item	Data item description	Field type	Example
Unique and enduring record identifier (ID)	A system-generated primary key that maps back to specific records in the original system or collection. It cannot be used for any other record in the dataset, and it cannot be changed or re-used. At times, it might be comprised of multiple fields.	Alpha numeric	9F3C21ED
First name	Person's first name	Character	JOHN
Surname	Person's surname	Character	DOE
Date of birth	Person's date of birth	Character	01/03/1960
Address	Person's residential address at time of event	Character	12 SMITH STREET
Postcode	Person's residential postcode at time of event	Numeric	6000

Please note that if a significant portion of records have missing values (i.e. NULLs) in the mandatory fields above, the data *may* fall below the minimum requirements for linkage. Suitability of data for linkage is assessed on a case-by-case basis.

### 2.2 Recommended additional valuable data items

There are additional data items that are not compulsory for linkage but are very valuable as they can increase linkage quality, completeness, and efficiency. For example, IDs that deterministically connect one data collection to another, such as WA Unit Medical Record Numbers (UMRN), Medicare Numbers, or Elector Numbers. Table 2 details some examples of additional value adding data items.

Table 2. Supplementary Data Items for Linkage

Data item	Data item description	Field type	Example
Person ID	Person's ID within a dataset used to group all records belonging to one individual within the original system/collection.	Alpha numeric	
Phone number	Person's contact number	Numeric	
Email address	Person's email address	Character	

Unit medical record number	Person's WA unit medical record number	Alpha numeric	A1234567
Medicare number	Person's Medicare number	Numeric	
Medicare individual reference number	Person's Medicare reference number	Numeric	
Sex	Person's sex recorded at birth	Character	1, 2, 9 M, F, U
Gender	Person's gender	Character	Male, Female, Non-binary, Prefer not to answer
Date of event/date of service	Date the event or service occurred	Character	01/06/1995
Date of death	Date the person died	Character	09/11/2006
Middle name(s)	Person's middle name(s)	Character	EDWARD
Suburb	Person's residential suburb at time of event	Character	PERTH

## 2.3 Accepted and preferred data formats

The Department's Data Linkage Services branch prefers UTF-8 encoded tab-delimited or comma separated (csv) files, particularly for large datasets (over 50,000 records). Excel files are accepted, but each file must contain only one worksheet of data. CSV files must contain a single table of data. Any text containing commas, quotation marks, or special characters should be enclosed in quotation marks to ensure it remains within a single field or column.

For linkage purposes, 'event-based' data is generally preferred. This refers to data where an individual's information appears across multiple records, each representing a separate event. Event-based data is particularly valuable for linkage purposes, as it can capture changes over time such as updates to a person's name or address, which improves linkage accuracy. Person-based data, where each individual has a single consolidated record, may not be as valuable for linkage as it does not reflect historical changes.

Multi-component fields such as name and address are best split into their subcomponents and provided as separate fields. For example:

- First name;
- Middle name;
- Surname;
- Street address;
- Suburb;
- Postcode.

If geocoding is also required, data must be provided as 'event based'.

## 2.4 Additional data requirements

### 2.4.1 NULL Values

Whilst the Department standardises data prior to linkage as best as possible, text equivalents of 'NULL' and 'N/A' can be time consuming to standardise. As such, data providers must identify and provide the Department with information on all the variants used.

### 2.4.2 Non-human and dummy records

Non-human and dummy records should also be excluded from the data file provided for linkage, noting that some datasets include animals, vehicles, equipment, etc. Examples of dummy data includes 'DUPLICATE', 'TEST', 'DO NOT USE', 'REFER TO'.

### 2.4.3 Consistent formatting within data items

Formatting within data items must be consistent to ensure data is handled correctly for linkage. For example, all date values provided must be in the same format

### 2.4.4 Metadata

Metadata is very useful when determining how to use fields for linkage, particularly when the data are understood by only a select group of people with specialised knowledge. Data providers must ensure they provide the Department with data dictionaries and/or code lists that correspond to the data provided for linkage.

### 2.4.5 Duplicate rows

As the data provided for linkage is usually only a subset of the full dataset, it is possible that when the data is extracted from the source system it will include duplicate rows (i.e. rows that contain exactly the same information). Data providers must ensure they remove all duplicate rows from the data file provided for linkage.

## 3. Data Updates

### 3.1 Deltas

If ongoing data updates will be provided to the Department, consider providing updates in the form of deltas. In some situations, it may be more efficient for the data provider to only provide records that are new, modified, or deleted since their previous provision.

#### 3.1.1 Documentation of data extraction processes

To ensure data updates are provided to the Department in a consistent format and timely manner, data providers must maintain sufficient documentation of the extraction process including any scripts that were used, information on the data that is extracted, and information on the frequency of data provision.

### 3.2 Data changes

Data changes such as the addition of new fields, moving the data to a new storage system, or reallocation of record IDs will impact processing requirements for data updates. In these instances, it is the responsibility of the data provider to format the data to look as similar as possible to the previous version (e.g. same format, same field order, same naming conventions, with newly added fields appended to the end of each record). Where this is not possible, the data provider must advise the Department in a timely manner to allow for any development work required.

## 4. Appendices

### 4.1 Data provider checklist

<input type="checkbox"/>	All records have a unique and enduring record ID
<input type="checkbox"/>	Data includes all of the following data items, if available: first name, middle name(s), surname, date of birth, address, suburb, postcode
<input type="checkbox"/>	Number of delimiters or fields are consistent
<input type="checkbox"/>	Data does not include duplicate rows
<input type="checkbox"/>	Data does not include instances where a record ID is assigned to more than one record
<input type="checkbox"/>	Single records do not wrap onto two lines
<input type="checkbox"/>	Data is formatted consistently within fields
<input type="checkbox"/>	Data extraction process has been documented (particularly if data updates are to be provided in the future)
<input type="checkbox"/>	Where the data is an update, it matches the exact same format as what was previously provided to the DLS, unless otherwise agreed
<input type="checkbox"/>	Dummy and non-human data have been excluded from the extract

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

© Department of Health 2026

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.