1. BACKGROUND

WA Health, like most organisations, is becoming increasingly dependent on information for decision-making. Ensuring that this information is of the highest possible quality is essential to both quality of care and effective decision-making.

Data quality refers to the accuracy, timeliness, comparability, usability and relevance of data within an organisation. These quality dimensions are defined as follows:

- **Accuracy** - refers to how well information contained within the system reflects reality.
- **Timeliness** – refers primarily to how current or up to date the data is.
- **Comparability** - refers to the extent to which data within the system is consistent over time and uses standard conventions, making them comparable to other related systems.
- **Usability** - reflects the ease with which the data can be understood and accessed.
- **Relevance** - reflects the degree to which the data meets the current and potential future needs of users.

These five dimensions can be viewed together in Figure 1. The hierarchical approach displays the ‘relevance’ dimension as the most important dimension. If data is not relevant, its value decreases substantially, even if the other four dimensions have been met. In order for the data to be relevant, it must meet both the current and future needs of the users.

The building blocks for this hierarchical model require that the data be accurate and use consistent conventions to ensure its comparability with like data. Once these are established, data must be as current as possible so that decisions are made with recent information. Also, to ensure its usability, the data must be accessible and easy to understand. These four dimensions of data quality, if achieved, allow the user to make decisions that are based upon accurate, comparable, timely and usable data.
2. SCOPE
This policy applies to all data collections, including those provided for by statute, held by or within WA Health. It includes collections of patient information, corporate, financial and workforce information where one or more of the following conditions are met:

- The data collection is used to meet business, operational or legislative requirements.
- The State of WA has a strategic need for the data.
- The data collection contains personal information.
- The data collection is used for reporting at a state level, national level or external to the health service where the data collection resides.
- The data collection is used across multiple health services.

The scope of this policy includes both paper based (including paper medical records) and electronic data.

For the purpose of this policy, a data collection includes both operational data collections and data repositories.

3. PURPOSE
The purpose of the Data Quality Policy is to ensure that data within WA Health is of high quality to support its intended uses. This is achieved through the allocation of accountability and responsibility for data quality to WA Health staff.

4. POLICY
Data quality is the responsibility of all WA Health staff. Staff create, capture and store information into systems on a daily basis as part of their duties.

To effectively manage data quality throughout WA Health the following roles and responsibilities have been allocated.
Data Stewards
Data Stewards are responsible for promoting and fostering a data quality culture throughout WA Health. They are also responsible for ensuring Data Quality Assessments are carried out for each data collection under their responsibility.

Data Custodians
Whilst data quality is the responsibility of all staff, it is the responsibility of the Data Custodian to ensure a coordinated and documented approach to data quality within the data collection(s) they are responsible for. This includes, but is not limited to:
- Ensuring and adhering to organisational data quality standards.
- Completing a Data Quality Assessment (attachment A) for endorsement by the Data Steward.
- Ensuring validity checks are completed for key data elements and invalid data is flagged and corrected.
- Ensuring validity checks are current.
- Ensuring current metadata documentation exists and is maintained.
- Ensuring that documentation of all data processing activities within the data collection is maintained.
- Ensuring technical specifications for the data collection are maintained by the Health Information Network or relevant Information Technology (IT) area.
- Ensuring that changes are made to data at source systems, not by the data recipient.
- Ensuring data entry quality control measures exist and are implemented.
- Ensuring standard data capture procedures exist and are followed by data entry staff.
- Ensuring that relevant staff are adequately trained in the data entry and use of data collections.
- Ensuring that data quality audits are carried out on a regular basis.

Data Entry Staff
Staff are responsible for ensuring the information they are entering into a system is accurate, current (timeliness), relevant and usable.

Responsibilities include, but are not limited to;
- Ensuring that the information collected is accurate and verified.
- Ensuring data is entered as close to real-time as possible.
- Ensuring required changes to data are reflected within the information system.

Supervisors/Line Managers
Supervisors and Managers are responsible for:
- Ensuring staff are aware of their responsibilities towards data quality.
- Supporting and encouraging a data quality culture amongst their team.
- Ensuring that data quality is monitored using audit and other operational reports and where anomalies are identified, they are raised with data entry staff and IT support.
- Ensuring data capture processes are reviewed regularly.
• Ensuring data processes are maintained to ensure data entry staff are consistent in their approach to the quality of captured data.
• Ensuring feedback is provided to data entry staff where discrepancies in data are identified.
• Releasing staff for training as required.
• Providing timely advice to data entry staff where changes to data capture are required.
• Ensuring data quality information/feedback from the Data Custodian is responded to appropriately.
• Maintaining agreed organisational data quality standards for collecting data.
• Advising the Data Custodian of any quality issues with the data.

Users
Data users are responsible for ensuring the information received is of the highest quality. They are required to advise the Data Custodian of any known or perceived data quality issues in the data sets or information products they receive.

Health Information Network
Health Information Network, in its role of leading the Information, Communications and Technology reform program are responsible for:
• implementing system changes to improve data quality; and
• assisting Data Custodians in completing the Data Quality Assessment.

5. National Aboriginal and Torres Strait Islander Health Data Principles
In October 2006 the Australian Health Ministers’ Advisory Council (AHMAC) endorsed the ‘National Aboriginal and Torres Strait Islander Health Data Principles’. These 11 principles set out a culturally respectful foundation for the collection, storage and use of Aboriginal health and health related information.

Of particular relevance for the Data Quality Policy is Principle 10. Principle 10 states that information collections require regular review and refinement in order to ensure ongoing relevance to service delivery and the potential for improved health outcomes.

6. Management of Personal Information – Good Practice and Opportunities for Improvement – Western Australian Ombudsman
On 28th March 2011 the Ombudsman published a report, “The management of personal information – good practice and opportunities for improvement”, following an investigation of selected State Government agencies that deal with large amounts of personal information.

Good practice principles were developed to assist Government Agencies in managing personal information they collect and hold. These principles are based on national and state legislative requirements, agency-specific legislation and internationally accepted good practice.
Of specific relevance is Principle 3 – Accuracy: Personal information collected, used and disclosed is accurate and current.

7. IMPLEMENTATION
All WA Health staff must take reasonable steps to ensure information created, collected, stored, accessed and used within their working environment is accurate, timely, comparable, useable and relevant.

Data Custodians are required to complete a Data Quality Assessment Plan (Attachment A) on a yearly basis and address any issues which are identified from this assessment. The Data Quality Assessment Plan must be submitted to the Data Steward for endorsement. The Data Quality Assessment Plan may be conducted at longer intervals if endorsed by the Data Steward.

8. DEFINITIONS
A Data Collection is a systematic gathering of data for a particular purpose from various sources, including manual entry into an application system, questionnaires, interviews, observation, existing records and electronic devices. This includes both operational data collections and data repositories.

Data Custodians are responsible for the day-to-day management of data from a business perspective. The Data Custodian aims to improve the accuracy, usability and accessibility of data with the data collection.

A Data Repository includes data that is collected from various sources, including operational data collections for the primary purpose of monitoring, evaluation, reporting and research. Examples of data repositories include data held within the Hospital Morbidity Data Collection, Finance Data Warehouse and the Emergency Department Data Collection (EDDC).

Data Stewards have delegated responsibility for setting the overall strategic direction of the specific data collection to ensure the collection is developed, maintained and utilised in accordance with the strategic goals of WA Health. Data Stewards are also responsible for authorising the access, use and disclosure of data from the data collection for clearly defined purposes that comply with WA Health’s statutory obligations.

Information Management is the discipline that directs and supports effective and efficient management of information in an organisation. Holistic, effective management of information requires mobilisation of three enterprise capabilities: people; processes (policies and procedures); and technology.

Metadata is structured information that describes, explains, locates or otherwise makes it easier to retrieve, use or manage an information resource. The purpose of metadata is to add value to the data it describes.
An **Operational Data Collection** includes data that is collected as part of the day-to-day activities of an area for the primary purpose of tracking and managing the operational aspects of the area. The operational data collection is typically a transaction-based system which contains detailed data elements to represent the activities of the area. Examples of operational data collections include data held within Patient Administration Systems, TRIM, Financial Systems and Human Resource Management Systems.

9. **ASSOCIATED POLICIES**

*Data Stewardship and Custodianship Policy OD 0321/11*
*Information Access and Disclosure Policy OD 360/12*

10. **RELEVANT LEGISLATION**

*Health Act 1911*
*Health Services (conciliation and review) Act 1995*
*Hospital and Health Services Act 1927*
*Human Reproductive Technology Act 1991*
*Mental Health Act 1996 (WA)*
*Freedom of Information Act 1992*
*Financial Management Act 2006*
*Public Sector Management Act 1994*
*State Records Act 2000*

11. **SUPPORTING DOCUMENTS**

Data Quality Policy, ACT Health


Australian Institute of Health and Welfare (AIHW), Quality and availability of Indigenous statistics, National Aboriginal and Torres Strait Islander Health Data Principles

Management of Personal Information – Good Practice and Opportunities for Improvement – Western Australian Ombudsman
ATTACHMENT A – DATA QUALITY ASSESSMENT PLAN

1. INTRODUCTION ........................................................................................................ 8
2. ASSESSMENT CRITERIA .......................................................................................... 10
   2.1 Accuracy Dimension .......................................................................................... 10
       2.1.1 Coverage .................................................................................................. 10
       2.1.2 Capture and Collection ......................................................................... 11
       2.1.3 Missing Records ..................................................................................... 13
       2.1.4 Missing Values ....................................................................................... 14
       2.1.5 Data Entry Errors .................................................................................. 14
       2.1.6 Edits ....................................................................................................... 15
       2.1.7 Data Processing .................................................................................... 15
   2.2 Timeliness Dimension ....................................................................................... 16
       2.2.1 Data Currency at the Time of Release ..................................................... 17
       2.2.2 Documentation Currency ........................................................................ 17
   2.3 Comparability Dimension ................................................................................ 18
       2.3.1 Data Dictionary Standards ..................................................................... 18
       2.3.2 Standardisation ....................................................................................... 18
       2.3.3 Linkage ................................................................................................... 18
       2.3.4 Equivalency ............................................................................................ 19
       2.3.5 Historical Comparability ........................................................................ 20
   2.4 Usability Dimension ......................................................................................... 20
       2.4.1 Accessibility ............................................................................................ 21
       2.4.2 Documentation ....................................................................................... 21
       2.4.3 Interpretability ......................................................................................... 22
   2.5 Relevance Dimension ....................................................................................... 23
       2.5.1 Adaptability ............................................................................................. 23
       2.5.2 Value ...................................................................................................... 24
1. INTRODUCTION

This Assessment Plan will identify strengths in the data as well as issues that need addressing. An action plan will be required that identifies how and when issues identified will be addressed.

The Assessment Plan examines five dimensions of data quality – accuracy, timeliness, comparability, usability and relevance. Each dimension is divided into related characteristics, and each characteristic is further made up of several assessment criteria. These criteria are questions about aspects of data quality. Once answered, it is this information that will inform the Data Custodian about the presence of data quality issues.

Each criterion is rated (numerically) as met (3), partially met (2), not met (1), not assessed (0) or not applicable (n/a). The response to each of the questions below will guide the creation of an action plan for improvements.

Data Custodians are required to complete the Assessment Plan on a yearly basis and address any issues which are identified from this assessment. The Data Quality Assessment Plan may be conducted at longer intervals if endorsed by the Data Steward.

It is not expected that all criteria be assessed. When determining which criteria are to be assessed, this must be based on the importance, complexity and sensitivity of the data collection.

Some criteria may not be applicable due to the lack of data quality functions within the data collection application.

<table>
<thead>
<tr>
<th>Dimensions/ Characteristics</th>
<th>#</th>
<th>Criteria</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td>1</td>
<td>The scope of data is explicitly stated in all releases of the data.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Known sources of under or over-coverage have been documented.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Data has been validated by comparison with external and independent sources.</td>
<td></td>
</tr>
<tr>
<td>Capture and Collection</td>
<td>4</td>
<td>Data capture requirements minimise burden on staff.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Data submission requirements minimise burden on data providers.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Data elements collected are of sufficient breadth and depth for their purpose.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Practices exist that encourage cooperation for data submission.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Practices exist that give support to data providers.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Standard data capture and data submission procedures exist.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Data entry quality control measures exist and are implemented.</td>
<td></td>
</tr>
<tr>
<td>Missing records</td>
<td>11</td>
<td>The magnitude of missing data is minimal.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>The numbers of records are monitored to detect unusual values.</td>
<td></td>
</tr>
<tr>
<td>Dimensions/Characteristics</td>
<td>#</td>
<td>Criteria</td>
<td>Assessment</td>
</tr>
<tr>
<td>----------------------------</td>
<td>---</td>
<td>---------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td><strong>Accuracy cont...</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing values (within records)</td>
<td>13</td>
<td>Missing values for core data items.</td>
<td></td>
</tr>
<tr>
<td>Data Entry Errors</td>
<td>14</td>
<td>The level of errors as a result of data entry.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>The degree of consistency in the data.</td>
<td></td>
</tr>
<tr>
<td>Edits</td>
<td>16</td>
<td>Validity checks are done for each data element and invalid data is flagged.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>Edit checks are current and logical.</td>
<td></td>
</tr>
<tr>
<td>Data processing</td>
<td>18</td>
<td>Documentation for all data processing activities are maintained.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>Technical specifications are maintained.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>System changes have been tested.</td>
<td></td>
</tr>
<tr>
<td><strong>Timeliness</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Currency at the Time of Release</td>
<td>21</td>
<td>The difference between the actual date of data release and the end of the reference period is reasonably brief.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>Data processing activities are regularly reviewed to improve timeliness.</td>
<td></td>
</tr>
<tr>
<td>Documentation Currency</td>
<td>23</td>
<td>Data quality documentation was available at the time of data or report release.</td>
<td></td>
</tr>
<tr>
<td><strong>Comparability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Dictionary Standards</td>
<td>24</td>
<td>All data elements are evaluated against the corporate metadata repository.</td>
<td></td>
</tr>
<tr>
<td>Standardisation</td>
<td>25</td>
<td>Data is collected at the finest level of detail practical.</td>
<td></td>
</tr>
<tr>
<td>Linkage</td>
<td>26</td>
<td>Geographical data is collected using a standard geographical classification system.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>Identifiers are used to differentiate facilities or organisations uniquely for historical linkage.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>Identifiers are used to differentiate persons uniquely for historical linkage.</td>
<td></td>
</tr>
<tr>
<td>Equivalency</td>
<td>29</td>
<td>Data mapping and conversions are logical and maintained.</td>
<td></td>
</tr>
<tr>
<td>Historical Comparability</td>
<td>30</td>
<td>Documentation on historical changes exists and is easily accessible.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>Trend analysis is used to examine changes in core data elements over time.</td>
<td></td>
</tr>
<tr>
<td><strong>Usability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accessibility</td>
<td>32</td>
<td>A final data set is made available at specified intervals.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>Key documents, reports and data sets are defined and published.</td>
<td></td>
</tr>
<tr>
<td>Documentation</td>
<td>34</td>
<td>Current data quality documentation for users exists.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>Current metadata documentation exists.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>A caveat accompanies any preliminary release.</td>
<td></td>
</tr>
<tr>
<td>Interpretability</td>
<td>37</td>
<td>A mechanism is in place whereby key users can provide feedback to, and receive notice from, the Data Custodian.</td>
<td></td>
</tr>
<tr>
<td>Dimensions/Characteristics</td>
<td>#</td>
<td>Criteria</td>
<td>Assessment</td>
</tr>
<tr>
<td>----------------------------</td>
<td>---</td>
<td>---------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Relevance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptability</td>
<td>38</td>
<td>Mechanisms are in place to keep stakeholders informed of developments within the data collection.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>The data collection has been developed so that future modifications can be made easily.</td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td>40</td>
<td>The mandate of the data collection fills a WA Health information gap.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>The level of usage of data is monitored.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>User satisfaction is periodically assessed.</td>
<td></td>
</tr>
</tbody>
</table>

2. ASSESSMENT CRITERIA

2.1 Accuracy Dimension

Accuracy refers to how well information captured within an information system reflects the reality.

2.1.1 Coverage

Coverage deals with the scope of the data collected and whether the collection has the required coverage.

**Criterion 1:** The scope of data is explicitly stated in all releases of data.

It is very important for users of data to know who or what is being examined. Hence, it is important that the scope of data together with details of any issues or anomalies be explicitly stated in all releases of data. A release includes any report, data extract or output from the data collection.

For example, does the scope of data being released include both public and private patients? This inclusion or exclusion needs to be outlined.

This criterion is met if the scope of data together with details of any issues or anomalies is stated in all key releases of data.

**Criterion 2:** Known sources of under or over-coverage have been documented.

The known sources of under or over-coverage of data within the data collection should be documented on a regular basis. Over-coverage occurs when a data collection contains additional records which are not considered to be within the scope of the data collection or duplicate records exist within the data collection. Under-coverage occurs when there are records missing from the data collection, resulting in under-reporting.

If the under- or over-coverage can be corrected, the data should be corrected before results are published and the source of the coverage error documented for internal use. If the data cannot be corrected, the known sources of under- or over-coverage should be mentioned in all data quality documentation.

This criterion is met if the known sources of coverage errors are documented internally or externally as required.
Criterion 3: Data has been validated by comparison with external and independent sources.

In order to detect errors within the data, it may be necessary to compare the data to sources external to and independent to the data collection. Although data validation and integrity checks should be done on a regular basis, it is necessary to compare the contents of the data to an external source that is independent of the data collection to ensure that the data collection is valid and up-to-date. In conducting any comparisons with external data, the credibility of the external source should be noted and, where possible, multiple sources used.

Data sources are generally considered independent if they are derived from different sources that are not related to the data collection in question. In many cases, finding an external source of information that is completely independent is not possible. If an independent data source cannot be found with the level of detail required, a comparison at an aggregate level can be done to detect errors.

This criterion is met if data is compared with external sources on a regular basis to determine the presence of errors within the data collection.

2.1.1 Capture and Collection

Capture and collection refers to the data entry practices at the point of collection.

Criterion 4: Data capture requirements minimise burden on staff.

This criterion assesses whether measures are used to ensure that the effort required by staff is minimised without compromising data quality. This is otherwise known as minimising the response burden. There are many ways this can be done: electronic capture and submission, reasonable submission schedules, exclusion of unnecessary data elements and reducing duplication in data entry.

This criterion is met if there are adequate measures in place to minimise data capture requirements.

Criterion 5: Data submission requirements minimise burden on data providers.

This criterion assesses whether measures are used to capture the data, ensuring that the initial source of the data has to do as little unnecessary work as possible. This is otherwise also known as minimising the response burden. There are many ways this can be done: electronic interfaces from source systems, reasonable submission timeframes and exclusion of unnecessary data elements.

This criterion is met if there are adequate measures in place to minimise data submission requirements from data providers.

Criterion 6: Data elements collected are of sufficient breadth and depth for their purpose.

This criterion examines if the capture and collection of data within the data collection is sufficient in managing the requirements of the organisation. The data elements are considered to be of sufficient breadth and depth if the data collection is able to respond effectively to strategic, tactical and operational needs of the organisation.

This criterion is met if the data elements collected in the data collection satisfy the strategic, tactical and operational needs of the organisation.
Criterion 7: Practices exist that encourage cooperation for data submission.
Practices that encourage cooperation are important regardless of whether the submission of data is voluntary or mandated. Where data providers are not obligated to submit data, a collaborative effort will be necessary to receive quality data.

Practices to encourage the data provider to submit data can be as simple as stressing the importance of participation, the assurance of confidentiality or the acknowledgement of the receipt of data in a letter or email thanking the data provider for the information.

This criterion is met if there are any practices in place that encourage cooperation.

Criterion 8: Practices exist that give support to data providers.
Providing support to data providers is essential to ensure that data is submitted promptly and correctly. This support can be done before data capture and during data capture.

Prior to data capture, education sessions and training can be administered, as well as having persons in place to promptly respond to emails and phone calls from the data providers. During data capture, emails and phone questions should still be answered promptly. In addition, technical and coding support should be made available to data providers by providing access to supporting documentation, coding guidelines and relevant manuals.

This criterion is met if there are practices in place before and during data capture by which support is given to the data providers.

Criterion 9: Standard data capture and data submission procedures exist.
Standard data capture and data submission procedures make the data collection process consistent, easier and may also result in improved timeliness and reduced errors. The availability of documented policies and procedures which identify data capture and data submission procedures will ensure that data collection is done as consistently as possible across sites.

A data collection in which some data providers submit annual data on paper and others submit monthly data electronically may not have standard procedures. It is however, noted that these practices exist for some data collections due to technical or business reasons.

In the same regard, data elements that are collected across sites should be the same, as well as having similarities among mandatory and optional data elements. This will facilitate meaningful comparisons across sites.

This criterion is met if the data capture and data submission procedures used in a data collection are standardised and followed.

Criterion 10: Data entry quality control measures exist and are implemented.
Data entry quality control greatly increases the quality of data within a data collection. Quality control measures include data entry edit checks, visual verification of the data, exception reporting and quality reporting.
Data entry edit checks can greatly increase the quality of data as they allow verification or corrections whilst the original data is present. Data entry edit checks generally consist of validity edit checks by data element (eg, checking to see if the gender is reported as male or female or if the telephone number is the appropriate length). In addition, consistency edit checks verify the relationship between data elements (eg, verifying a clinical intervention that can only be performed on females).

Visual verification consists of having a second person examine the original data and the captured data for any differences. It provides a level of confidence that key pieces of information are not missed during data entry.

Exception and quality reporting should occur post data entry. This reporting is designed to flag potential errors in the data. It is important that outputs from exception and quality reports are reviewed by managers and errors rectified accordingly.

This criterion is met if suitable and robust data entry quality control measures exist within the data collection.

2.1.2 Missing Records
Missing records can cause misleading information to flow through to all aspects of a patient’s care, future planning, program development and medical research.

Criterion 11: The magnitude of missing data is minimal.
This criterion examines if there is data which should be on the data collection but has not been recorded on the system. This can occur for several reasons:-
a) Information is received but not entered on the data collection;
b) Information is not received or reported; or
c) The data collection does not have the required functionality to capture the details.

Investigations into the cause of missing data and the solution to resolve the issue may involve the data provider(s), Health Information Network (for the technical aspects) and others, as required, to enable the capture of the missing data.

This criterion is met if the level of missing data is known, clearly documented and measures are put in place to ensure that all information is entered.

Criterion 12: The number of records are monitored to detect unusual values.
Similar to the previous criterion, this criterion examines if there are changes to the number of records over a period of time which may indicate problems with the data. For example, if the number of records received in a month suddenly increases from the average of 50 records to 200, this may indicate a problem which needs to be analysed further. The increase may not necessarily indicate a problem with the data, as there are other possible reasons for this (eg, late submission of data, new reporting protocols, advertising campaigns, etc).

This criterion is met if the records within the data collection are monitored over time to detect unusual trends.
2.1.3 Missing Values

Each field requires a value and missing values can impact on the information obtained from the data collection.

**Criterion 13:** Missing values for core data items.

Core data items are data elements within the data collection which are routinely used to inform key decision making. Data items such as unique patient identifiers, clinical diagnosis and procedures are considered core data items that are used as part of clinical decision making, analysis and reporting.

Missing values for core data items will impact on the ability to adequately manage functions of the organisation and also on management reporting. It is important to carefully distinguish between data items left blank (as it is not applicable) versus missing values (e.g., patient’s surname).

Where possible, Enterprise data collections should have the fields which are used for operational, legislative, State or National reporting made mandatory to ensure all data is captured.

This criterion is met if the level of missing values for core data items is minimal and action is being taken to prevent this from occurring in the future.

2.1.4 Data Entry Errors

Errors can occur for many different reasons, but ways to minimise the number of errors include: data entry edit checks, data transfer via electronic means (no manual transfer), training and onsite audits.

**Criterion 14:** The level of errors as a result of data entry.

It is often difficult to assess the level of errors in data entry without conducting a site audit. In the absence of audits being undertaken, this criterion will examine the level of data entry errors based on values entered in related fields.

Examples of data entry errors include, the ‘date of discharge’ is before the ‘admission date’, or ‘date of birth’ for a patient is after the ‘current date’.

This criterion is met if the level of data entry errors is minimal and action is being taken to prevent this from occurring in the future.

**Criterion 15:** The degree of consistency in the data.

Consistency is a concern for all data collections, as consistency for each data element in a data collection is highly dependent on the opinion or interpretation of clinical coders, nurses, registrars etc.

Differences in work practices and different interpretations of business rules, legislation and policy may also result in inconsistencies in the data. The level of inconsistency is best determined by conducting an audit on the actual data. Essentially, the core elements of a piece of information should be identical if the same information is captured by a different person each time.

This criterion is met if the level of data consistency is high across sites.
2.1.4 Edits

Editing is the application of checks to identify whether units or data elements are missing or invalid. Such identification will point to data items which are potentially in error.

Criterion 16: Validity checks are done for each data element and invalid data is flagged.

Validity checks ensure that the proper data entry format is used and that the value entered is appropriate for that field. Validity checks can consist of comparing the entered value to a list of acceptable values (eg, list of suburbs and postcodes, list of diagnosis codes) or simply ensuring that the entered value is in a correct format (eg, making sure dates are entered as dd/mm/yyyy).

It is important to realise that a valid response does not necessarily mean that the response is correct; it just means that the response conforms to a set of expected values. For example, if surgery is reported to have occurred on February 27, this is a valid date, but it may or may not be the correct date. However, if the surgery was reported to have occurred on February 31, the reported date is known to be incorrect since this date is not valid.

Invalid data in a database will quickly raise questions about the quality of the data, and as invalid data is nearly impossible to justify, it is very important that invalid data be identified and corrected. Depending on the nature of the database, invalid data may be excluded, sent back to the data provider for correction or simply flagged as invalid and dealt with separately.

This criterion is met if all collected data elements are checked for validity and any invalid data is flagged and subsequently corrected.

Criterion 17: Edit checks are current and logical.

This criterion examines the edit checks performed on core data items within the data collection to ensure the checks are current and logical. Edit rules are logical if they make sense with regard to the data that is collected.

For the technical side of implementing or checking the currency and logic of edit checks, the assistance of Health Information Network may be required. For the main Patient Administration Systems this should already be in place but monitoring to ensure the currency and logic of edit checks should be done at regular intervals.

This criterion is met if edit checks are current and logical and applied consistently across sites.

2.1.5 Data Processing

Data processing is the sequence of steps carried out when extracting, loading, editing, mapping data, etc.

Criterion 18: Documentation for all data processing activities are maintained.

Movement of key staff, due to resignations, transfers, leave, etc, can result in a loss of knowledge about the processing steps, which can in turn result in errors in the data. The documentation for all data processing activities should ideally be in one location.
The processing steps need to be sufficiently documented so that anyone new to the area can use the documentation to follow and implement the processes.

This criterion is met if all processes that are run are adequately documented and made available to key staff within the organisation.

**Criterion 19:** Technical specifications are maintained.
The way systems and applications are developed can affect data quality. It is therefore important that the technical specifications be implemented as intended and fully tested. It is equally important to have these technical specifications maintained in the form of system documentation.

The reasons for documentation are simple: documentation allows easy validation of the systems, programs and applications, and if changes must be made, the documentation makes it easier to implement changes. Good documentation should be accessible and easily understood by someone new to the project. Furthermore, such documentation should be regularly reviewed, and any updates to the technical specifications should also be reflected in the documentation.

Assistance from the Health Information Network may be required in developing and maintaining technical specifications documentation.

This criterion is met if technical documentation exists, is kept up-to-date and made available to key staff within WA Health.

**Criterion 20:** System changes have been tested.
Although revisions are often necessary to accommodate modifications to a data collection, changes to programs can have unexpected consequences. Changes to systems will require extensive unit, system and user acceptance testing, whether it is just a small change to the data element name or an addition or deletion of a data element.

The modified programs should be checked to ensure that expected results are achieved. It is also prudent to verify the downstream effects of the changes (ie, on the reporting platform). As an example, changing the format of a data element from numeric to character can affect programs that treat the data element as numeric. Unit, system and user-acceptance testing should be performed to prevent unexpected results in a production environment.

Assistance from the Health Information Network may be required in performing modifications to systems or programs.

This criterion is met if the data collection and downstream systems are tested when changes are made.

**2.2 Timeliness Dimension**
Timeliness refers primarily to how up-to-date the data is at the time of release. Timeliness is closely associated with relevance, in that if the data is not timely, the data may no longer be relevant for the needs of users.
2.2.1 Data Currency at the Time of Release

Data currency is measured by taking the difference between the date of release and the last date to which the data relates. The duration should be short enough so the data remains relevant for the decision making process.

**Criterion 21:** The difference between the actual date of data release and the end of the reference period is reasonably brief.

The date of release is defined as the official date when data becomes available to users. The reference period refers to the period of time that the data actually spans or to which it relates. The start of the reference period is the first date to which the data relates, and the end of the reference period is the last date to which the data relates. Data currency at the time of release refers to the time lag between the release of the data and the reference period.

Data collections will have different standards relating to what is considered ‘reasonably brief’. As a general rule for annual data releases, a 6 to 9 month period between the end of the reference period and the release date is desirable, with up to 12 months being acceptable.

This criterion is met if the difference between the date of release and the end of the reference period for the data is reasonably brief.

**Criterion 22:** Data processing activities are regularly reviewed to improve timeliness.

The programs or systems that are used to prepare and analyse the data should be reviewed in an ongoing manner to ensure that they are as efficient as possible to produce timely reports.

For example, multiple programs may be combined to reduce the amount of manual input and time required for data management, analysis and report creation. Existing methods should be reviewed in light of new technologies that may be more efficient and may result in more accurate data at the same time.

This criterion is met if the data processing activities for the data collection are reviewed at least once a year.

2.2.2 Documentation Currency

The purpose of data quality documentation is to inform users of the major limitations associated with the data, so that they can decide whether the data is fit for their intended use. This type of information is also necessary for the correct interpretation of results based on the data. It is therefore crucial that data quality documentation be made available, along with any major data release or report.

**Criterion 23:** Data quality documentation was available at the time of data or report release.

It is important that data quality documentation be made available once users can internally or externally access the data or the reports summarising the data.

Data quality documentation needs to include information such as the data source, scope (inclusions/exclusions), data accuracy, data comparability, known issues with the data, reporting definitions used and contact details for further information.
This criterion is met if data quality documentation was available at the time of data or report releases.

2.3 Comparability Dimension
Comparability is defined as the extent to which the data collection is consistent over time and uses standard conventions, making it similar to other relevant data collections within the Department.

2.3.1 Data Dictionary Standards
The Data Dictionary is a repository containing information on data such as the meanings of the data elements, relationships to other data, usage and formats.

**Criterion 24:** All data elements are evaluated against the corporate metadata repository.

The goal is to have all databases use the same definitions for common data elements, thereby eliminating confusion amongst data users. It is important that all data elements in an existing database be reviewed periodically against the corporate data dictionary. The evaluation will determine if the standards used for the data elements in a database are in agreement with those in the corporate data dictionary. Those that do not agree with the data dictionary standard should be noted and considered for future modification.

This criterion is currently not applicable as there is no corporate metadata repository available.

2.3.2 Standardisation
Standardisation pertains to the uniformity of data within the collection to enable comparability with other data collections. Standardisation of the data benefits not only the ease with which the data can be used but also the accuracy of the data.

**Criterion 25:** Data is collected at the finest level of detail practical.

A fine level of detail in data elements is important, because it allows flexibility to conform to different standards. It is important to note that the level of detail depends on the users of the data collection and, in some cases, the data provider.

For example, for acute care stays, it is usually sufficient for length of stay to be measured in days. In contrast, wait time in emergency should be measured in minutes. Another example relates to the collection of data by using a person’s birth date rather than capturing age groups, even if the information is reported in certain age group categories. A simple regrouping into different age ranges is possible when age has been reported in years.

In summary, fine detail may not always be required for common uses, but it may be necessary in order to create new groupings.

This criterion is met if all core data elements are collected with the necessary detail required for operational, tactical and strategic management.

2.3.3 Linkage
Linkage refers to the process of joining records from two or more data collections by the use of one or more common linking data elements. Linkage of data from
different sources provides a more complete picture. In addition, data quality can be improved with linkage, as duplicates can be identified and removed. Keep in mind that the Department’s privacy and confidentiality guidelines must be adhered to when linking data between data collections.

**Criterion 26:** Geographical data is collected using a standard geographical classification system.

In order to analyse data from a geographical perspective, it will be necessary to capture data on facilities (ie, service providers, health services), persons (ie, patients, carers or staff) and organisations (ie, Joondalup Health Campus, Health Review Board) using a standard geographical classification system.

This criterion is met if the entities about which data are collected (facilities, persons, organisations etc) are identifiable by either postal codes or a relevant geographical classification.

**Criterion 27:** Identifiers are used to differentiate facilities or organisations uniquely for historical linkage.

In order to analyse and report on data at a facility level (ie, service providers, health services), identification of each facility using a unique identifier will be required.

This criterion is met if a unique code acceptable for historical linkage purposes exists for each facility and is available on the data collection.

**Criterion 28:** Identifiers are used to differentiate persons uniquely for historical linkage.

The purpose of this criterion is to ensure that a suitable identifier is present that accurately distinguishes between persons in the data collection. In order to do this, the identifier must be unique, be consistent over time and have the capacity to accommodate future individuals.

In order to facilitate record linkage, the data element must be consistent across databases within WA Health as well. For clinical systems, this will most likely be the unique patient master index (PMI).

This criterion is met if a unique person identifier is available in the data collection that could be used to link to corresponding records across different time periods.

**2.3.4 Equivalency**

Equivalency refers to how well data can be mapped over time, especially when classification systems are used (for example, the International Classification of Diseases (ICD)). An example of equivalency includes the mapping required between data coded using ICD 9 and ICD 10.

**Criterion 29:** Data mapping and conversions are logical and maintained.

Data mapping and conversions need to be logical and maintained, especially when lookup/reference table data changes occur. For example, if the name of a hospital is changed in the reference list of hospitals, the handling of existing hospital names (with the original name) within the patient event table needs to be considered. How well this is handled and the level of documentation that exists to support any data mapping and/or conversions is assessed within this criterion.
This criterion is met if data mapping and conversions are adequately handled and documented within the data collection.

### 2.3.5 Historical Comparability

Historical comparability refers to the consistency of data concepts and methods over time, which in turn allows valid comparisons at different points in time. Many things can make the comparison of data over time difficult. Database enhancements that will improve a database for the future can sometimes inhibit historical comparability. In those situations, this limitation should be noted and documented.

**Criterion 30:** Documentation on historical changes to the data collection exists and is easily accessible.

This criterion assesses whether documentation on historical changes to the data collection exists and is maintained. This includes changes to screens, data elements, database structures and reports. Note that a set of manuals, each of which describes the current year changes, is not an acceptable form of historical documentation, as it becomes too difficult to track changes. In addition, storing comments within a computer program is also not an acceptable form when it is the sole source of historical documentation.

This criterion is met if documentation on changes to the data collection is consolidated in a central location and is readily accessible by key staff within the organisation.

**Criterion 31:** Trend analysis is used to examine changes in core data elements over time.

Trend analysis includes comparisons of counts or proportions over time to assist in investigating potential quality issues in the data. Graphing data is often particularly helpful for investigating temporal changes. One of the primary rationales for longitudinal analysis is to detect any potential problems in the data as a result of changes in concepts or methodologies. Note that no change across years may also be an indication of a problem if the data is expected to naturally trend upward or downward due to policies implemented or changes in legislation.

This criterion is met if trend analysis has been performed for core data elements within the data collection.

### 2.4 Usability Dimension

Usability reflects the ease with which data from a data collection may be understood and accessed. If data from a data collection is difficult to use, the data collection can be rendered worthless no matter how accurate, timely, comparable or relevant it may be.

Several factors contribute to the usability of a data collection. In general, the greater the number of limitations or exceptions associated with the data, the more difficult the data will be to interpret.

Efforts made to improve the standardisation of data benefit not only the ease with which the data can be used, but also the accuracy of the data. Inconsistent methods may also complicate interpretation. The benefits derived from the
introduction of new methods (for example, data element name or definition changes) should therefore be weighed against any loss in interpretability. Simply put, the fewer the limitations and changes, the easier the data will be to interpret.

2.4.1 Accessibility
Accessibility deals with the ease with which data can be accessed from a data collection. Data that users do not know about, cannot locate or cannot bring into their own working environment for operational purposes will not be of use to them.

Criterion 32: A final data set is made available at specified intervals.
The data that is used for analysis and the creation of key reports (eg, Annual Report) should be extracted from the data collection and stored in a secure location for future reference (for a set period). It is often necessary to refer back to previous sets of data in order to run further analysis. Having one version of the data set used in the creation of key reports will ensure that results based on any new analysis will be consistent with the previously released results.

This criterion is met if the data set from the data collection is extracted, stored and made available to key stakeholders at specified intervals.

Criterion 33: Key documents, reports and data sets are defined and published.
Key documents, reports and data sets from the data collection should be made available in a secure area (eg, intranet portal) which is accessible to authorised users of the data collection. The secure area can be used as a central library for all the information products that are available from the data collection. Through this area, users can upload and download documents, track version history of documents, view standard reports, access standard data sets, etc. The availability of a secure area will enable users to self-serve and facilitate a greater use of data and resources from the data collection.

This criterion is met if key information resources are defined and published in a central area which is accessible to authorised users of the data.

2.4.2 Documentation
Documentation is necessary for appropriate interpretation and utilisation of the data from the data collection. Documentation includes background information of the data collection, data sources and methodology, concepts and data elements, data accuracy and data comparability.

Criterion 34: Current data quality documentation for users exists.
The purpose of data quality documentation for users is to give both internal and external users of the data sufficient information so they can decide if the quality of the data is appropriate for their intended use. Contact information should also be provided with any release so that users can access additional information on the limitations.

A stand-alone data quality document for users which is updated on a regular basis should be made available for users of the data. This criterion is met if up-to-date data quality documentation for users exists each time data is released or extracted.
**Criterion 35:** Current metadata documentation exists.
To facilitate the interpretation and proper use of data, metadata documentation should be made readily available to staff working with the data. Whilst data quality documentation for users provides some background information and outlines the major data limitations, metadata documentation will provide users with an understanding of the data at its most granular level along with its known limitations.

This criterion is met if metadata documentation for the data collection exists and is kept up-to-date. This documentation should be reviewed regularly and updated as required.

**Criterion 36:** A caveat accompanies any preliminary release.
In an effort to improve timeliness, some data collections may provide preliminary releases of data or results. A preliminary release may include a release of data that is designed to help certify or validate data. For example, prior to publication, health indicator counts might be sent to Area Health Services for verification and validation. A preliminary release may also be defined as a release of possibly incomplete data for the purpose of improved timeliness. For example, data that is collected on an annual basis might be released six months prior to year-end so that business managers can gain an early indication of the level of incompleteness in the data.

For all preliminary releases, a caveat must be provided that advises that the data may not be complete and that it is subject to revision.

This criterion is met if a caveat accompanies any preliminary release of data.

### 2.4.3 Interpretability

Interpretability refers to the ease with which the user understands the data. Design features and underlying data quality limitations associated with the data will largely determine its interpretability. If standard concepts and classifications are used, the data will be easier to understand and use. Having the record layouts accompanying the various data extracts will also aid in the interpretability of the data.

**Criterion 37:** A mechanism is in place whereby key users can provide feedback to, and receive notice from the Data Custodian.

Contact information (name, phone number and email address) should be included with releases of data so that users (internal or external) can provide feedback on any major data quality limitations as they come to light. Major users should be encouraged to use the contact information to provide feedback on any limitations they may discover or concerns that they may have with the data.

Similarly, there should be a mechanism that allows contact with major users, so they can be notified of the existence of any limitations that are discovered after the release. Information on actions taken in light of the limitations and the effect of the errors should also be made available.

This criterion is met if contact information is included with any major data or report release and feedback is solicited.
2.5 Relevance Dimension
Relevance reflects the degree to which a data collection meets the current and potential needs of users. The purpose of the relevance dimension is to assess how well a data collection can adapt to change and whether the collection is perceived to be valuable.

2.5.1 Adaptability
The adaptability of a data collection relates to whether it is well positioned and flexible enough to address the current and future information needs of its main users. In order to remain relevant, a data collection may have to adapt in an ongoing manner to emerging issues in the field.

If existing or developing issues are known and tracked, then future information needs may be anticipated. It is important to remain proactive and not reactive to main user needs. Once something is anticipated, future information needs can be factored into the design of the data collection. Although it is impossible to predict the future needs of users with complete accuracy, one can try to design data collections that allow for change. It is also important that a data collection not be overly flexible to the wants and needs of users. The impact of changes should be fully evaluated before being incorporated.

Criterion 38: Mechanisms are in place to keep stakeholders informed of developments within the data collection.
Maintaining a relationship with key stakeholders will ensure that staff are kept abreast of the current and emerging developments within the data collection. Stakeholder liaison can occur under different forms (eg, mailing lists, bulletins, User Groups, Steering Committees, etc).

This criterion is met if liaison mechanisms are in place to help stakeholders stay abreast of developments in the field.

Criterion 39: The data collection has been developed so that future modifications can be made easily.
In order for the data collection to remain relevant, ongoing enhancements may be necessary. In order to address emerging issues or changes in legislation, existing data elements might need to be redefined or new data elements might need to be added. For example, new date and time data elements may be added to collect ambulance wait times in an emergency department database. A data collection should also be able to incorporate new clinical standards as they arise (eg, ICD 10).

Changes to the data collection may also be required to remediate data quality issues. For example, if negative lengths of stays are detected, new edits may be added.

Whilst the ability to enhance the data collection is important, the benefits of any changes should be weighed against the potential loss in data comparability or interpretability.
This criterion is met if the data collection has demonstrated the ability to adapt to an important emerging issue, changes in legislation or a major data quality limitation.

### 2.5.2 Value

The value of a data collection may be defined by its contribution to WA Health and its ability to fill an information need or requirement.

**Criterion 40:** The mandate of the data collection fills a WA Health information gap.

The mandate of the data collection should be periodically assessed to ensure that it continues to fill an information gap within WA Health. An assessment should also be conducted against other data collections within WA Health to ensure that information collected is not duplicated in other collections.

This criterion is met if the mandate of the data collection fills a WA Health information gap.

**Criterion 41:** The level of usage of data from the data collection is monitored.

The value of a data collection is related to the extent that the data is used. Evidence of usage may include high-profile users of the data, published reports, news items, citations, staff-authored papers, conferences and other forums. As well, the level of usage can be monitored by keeping track of the number and type of data requests received.

This criterion is met if the level of usage of data is monitored.

**Criterion 42:** User satisfaction is periodically assessed.

It is important to assess whether the data collection is satisfying the needs of users and to apply the results from the assessment as part of a review of the data collection. User satisfaction covers stakeholders, both from an operational and reporting perspective.

Stakeholder satisfaction may be used as direct evidence of the perceived value, accuracy, timeliness, comparability and usability of the data within the data collection.

This criterion is met if user satisfaction assessments are conducted regularly.