



NSW Government Standard for Data Quality Reporting

October 2015

v1.2

DOCUMENTATION AND ENDORSEMENT

Document history

Date	Version No.	Description	Author
August 2014	0.4	Final draft submitted to IM Community and ICT Leadership Group	Strategic Policy, OFS
September 2014	0.5	Final draft submitted to ICT Leadership Group	Strategic Policy, OFS
October 2014	1.0	Published	Strategic Policy, OFS
January 2015	1.1	Minor edits to reporting questionnaire and statement, star rating and data quality levels	Strategic Policy, OFS
October 2015	1.2	Amendments to reporting questionnaire and statement to reflect development of data quality reporting tool	Strategic Policy, DFSI

Approvals

Name	Date
Information Management Community of Expertise [v.1.0]	28 August 2014
NSW ICT Leadership Group [v.1.0]	5 September 2014
Information Management Community of Expertise [v.1.2]	11 September 2015

Document custodian

Name	Signature
NSW Department of Finance, Services & Innovation	

CONTENTS

1. PURPOSE	4
2. INTRODUCTION	4
2.1 Information Management Framework	4
2.2 About the Standard.....	4
3. IMPLEMENTATION.....	6
4. SCOPE.....	7
5. DEFINING DATA QUALITY.....	7
5.1. What is data quality?	7
5.2. Dimensions of data quality.....	8
6. EVALUATING RELEVANCE.....	10
7. REPORTING ON DATA QUALITY.....	11
7.1. Publishing data of variable quality.....	12
8. RELATED DOCUMENTS.....	13
9. REVIEW.....	13
10. CONTACT.....	14
APPENDIX A – REPORTING QUESTIONNAIRE.....	15
Improving the data quality rating.....	19
APPENDIX B – DATA QUALITY STATEMENT	20
APPENDIX C – GLOSSARY	24

1. PURPOSE

The purpose of this document is to establish common principles and protocols for reporting on data quality, so that agencies can create simple data quality statements and users can easily evaluate whether shared or published data is suitable for re-use.

2. INTRODUCTION

2.1 Information Management Framework

A key initiative of the [NSW Government ICT Strategy](#) is the development of an Information Management Framework to enhance government's use and management of data and information.

The Framework is a set of standards, policies, guidelines and procedures which are implemented either manually or, where possible, automated through technology. This will allow data and information to be managed in a secure, structured and consistent manner.

It ensures that data and information can be appropriately shared or re-used by agencies, individual public sector staff, the community or industry for better services, improved performance management and a more productive public sector.

This standard forms part of the Information Management Framework.

2.2 About the Standard

This standard outlines a method for describing and evaluating the characteristics of data. It does not prescribe a minimum level of data quality.

Data quality is determined by whether or not the data is suitable for its intended use. This is generally referred to as being "fit-for-purpose".

Agencies create or collect data and information to meet their operational and regulatory requirements. They will define their own acceptable levels of data quality according to these primary purposes.

Agencies are increasingly aware of the potential benefits of re-using their datasets for multiple purposes. They are also making their data available to a range of public and government users in support of the *Government Information Public Access Act*, the *NSW Government Open Data Policy* and the benefits of information sharing between agencies.

It is not possible for data producers, stewards or custodians to predict the needs of all users and the various ways in which their data might be used.

To make the most of open and shared data, users need a description of the characteristics of the data, to help them determine whether it will be fit for their specific purpose.

This standard provides a method for agencies to create simple data quality statements that describe their datasets for secondary users. These statements can be used by:

- staff within the agency, to evaluate data for internal re-use
- staff from other agencies, where data is available for sharing
- individuals, researchers and industry, where data is published.

The data quality statement aims to help a user understand how the dataset could be used and whether it can be compared with other, similar datasets.

This method can also be used by agencies when planning data collection exercises, or to evaluate datasets acquired from external sources.

ABS Data Quality Framework

The method for data quality reporting is based on the dimensions described in Australian Bureau of Statistics' *Data Quality Framework*, which is closely aligned with other jurisdictions including Canada, Europe and the USA. The reporting tool has been modelled on the National Statistical Service (NSS) Data Quality Tool.

The *ABS Data Quality Framework* has been developed to support the collection, comparison, analysis and dissemination of data from many sources. Its design and language is best suited for managing surveys and statistical data.

NSW agencies create or collect a wide range of data types, including statistical, scientific, environmental, administrative and spatial data. The language and examples of the *ABS Data Quality Framework* have been modified only to the extent necessary to ensure this standard can be applied to many types of data and information.

Within the NSW Government the *ABS Data Quality Framework* has already been adapted for a variety of uses, particularly within the health and education clusters.

The *NSW Government Standard for Data Quality Reporting* will also be supported by:

- guidance to help agencies develop their own Data Quality Framework; and
- a template for a Data Quality Management Plan.

These documents are modelled closely on tools developed by the Queensland Government that have been adapted for use in several NSW Government agencies.

Benefits

The benefits of a standard approach to data quality reporting are:

- **enhanced accessibility:** enable users to understand and use data and information
- **enhanced development opportunities:** easier evaluation of data quality supports re-use and value adding by non-government sectors
- **reduction in administrative costs:** streamlines reporting, and reduces duplication through appropriate sharing and re-use of data and information
- **enhanced interoperability:** establish whether datasets can be brought together and integrated or compared
- **improved decision making:** increases the likelihood of existing data and information being found and re-used for planning, management, policy development and service delivery.

3. IMPLEMENTATION

NSW Government has adopted a business-driven approach to standards for data and information, based on the following five criteria:

Standards are:

- **aligned** – with the principles and outcomes of the IM Framework;
- **relevant** – to the specific business needs, objectives and operating environment of NSW Government agencies, so that their business value is evident;
- **proven** – established or endorsed by the industry sector or professional community;
- **aspirational** – supporting phased implementation or continuous improvement;
- **enterprise-wide** – support for developing people, processes and technology.

Standards are incorporated into the IM Framework where existing or potential business needs and opportunities are recognised. Agencies are expected to refer to the Framework, and use standards that are relevant to a specific project, process or corporate objective.

This standard will be implemented by agencies according to their operating environment.

A simple data quality statement should accompany future datasets published on data.nsw.gov.au or the Information Asset Register. Agencies may choose to maintain more detailed data quality statements depending on their business requirements.

Agency requirements for data management and quality vary according to data types and the purpose or processes the data support. Guidance and templates are being developed to help agencies develop a Data Quality Framework which meets their particular needs.

Further information about NSW Government approach to information management and standards can be found on the website of the *NSW Government ICT Strategy* at:

<http://finance.nsw.gov.au/ict/information-management-framework>

4. SCOPE

This standard is for use by all NSW Government Departments, Statutory Bodies and Shared Service Providers. In accordance with *Premier's Memorandum M1999-19 Applicability of Memoranda and Corporations*, it does not apply to State Owned Corporations (SOCs) although SOCs are encouraged to adopt this standard.

Other stakeholders are encouraged to adopt this standard, including:

- Public Trading Entities;
- Entities fully funded by the NSW Government; and
- Local Government Authorities.

5. DEFINING DATA QUALITY

5.1. What is data quality?

Quality describes the degree to which a set of inherent characteristics fulfils requirements¹.

Data quality is determined by whether or not the data is suitable for its intended use. This is generally referred to as being “fit-for-purpose”. Data is of sufficient quality if it fulfils its intended use (or re-use) in operations, decision making or planning.

Agencies create or collect data and information to meet their operational and regulatory requirements. They will define their own acceptable levels of data quality according to these *primary purposes*.

To make the most of open and shared data, public and government users (*secondary users*) will define what data quality means with reference to their specific aim or objectives.

Secondary users need to understand the characteristics of the data and how it was collected to help them decide how well it can meet their own needs or expectations.

¹ TC 176/SC (2005). *ISO 9000: Quality management systems - Fundamentals and vocabulary*

5.2. Dimensions of data quality

The Australian Bureau of Statistics' *Data Quality Framework* includes seven "dimensions" (or characteristics) of quality. Each dimension may impact the potential value of the data, depending on the intended use for the data.

This broad approach to the definition of data quality can be used to evaluate data in a variety of forms, for a wide range of possible uses.

These dimensions have been adapted to help NSW government agencies create simple data quality statements. A data quality statement can help a user understand how the dataset could be used and whether it can be compared with other, similar datasets.

Figure 1 illustrates the dimensions of the ABS *Data Quality Framework*.



Figure 1: The seven dimensions of the ABS *Data Quality Framework*²

² Australian Bureau of Statistics, National Statistical Service: Data Quality Online
<https://www.nss.gov.au/dataquality/aboutqualityframework.jsp>

Institutional environment

This dimension refers to organisational factors which may influence the effectiveness or credibility of the agency producing the data. It also considers the regulatory framework and governance arrangements that support data quality.

Information about institutional environment helps a user to decide if:

- data is produced and disseminated objectively or impartially
- data is collected for a regulatory or compliance reason
- the organisation has strong policies and procedures for data quality
- responsibilities and accountability for data quality are established.

Relevance

This dimension refers to how well the data meets the needs of the user. It considers whether the data addresses the issues of most importance to the user and their specific purpose. This is discussed in more detail in section 6 *Evaluating Relevance*.

Timeliness

This dimension refers to the currency of the information. It is closely related to the dimension of Relevance and is discussed in more detail in section 6 *Evaluating Relevance*.

Accuracy

This dimension refers to the degree to which the data correctly describe what they were designed to measure, monitor or report. When evaluating accuracy, it is helpful to consider the primary reason for the data collection.

Information about accuracy helps a user to decide:

- how well the data represent the user's area of interest
- the level of impact errors may have on their use of the data
- how useful and meaningful the data will be for interpretation or further analysis.

Coherence

This dimension refers to the internal consistency of the data over time. It also considers how the data aligns with other, related sources of information. The use of standard concepts, common classifications and data recording practices promotes coherence.

Information about coherence helps a user to decide:

- whether changes in the data can be analysed over time
- whether the data can be compared or combined with data from other sources.

Interpretability

This dimension refers to the ability to make sense of, or gain insight into the data. It considers how the data can be correctly used or analysed.

Interpretability is very important for secondary users of the data. Clear explanations about the concepts, classifications and sources used to create or collect the data will help users determine whether it can meet their specific objectives.

Information about interpretability helps a user to decide:

- how the data can be explained and understood
- which questions can be answered, or which inferences can be drawn from the data
- which questions or inferences cannot be understood from the data.

Accessibility

This dimension refers to the ease with which data can be searched, identified or retrieved, and the suitability of the format or medium through which the data is made available. The cost of retrieving or using the data also influences its accessibility.

Accessibility describes how the dataset can be found, used and shared.

6. EVALUATING RELEVANCE

As described in 5.2, relevance describes how well the data meets the needs of the user. Relevance is particular to each individual project. Understanding relevance enables an assessment of whether the data addresses the user's key areas of interest.

Timeliness is a separate dimension in the Data Quality Framework. However, it is very closely related to Relevance. For reporting, the two dimensions are covered together.

Timeliness refers to the time elapsed between the *reference period* (for which the data were collected) and the date when the data become available (i.e., the actual release date). Lengthy delays in timeliness can affect the currency or reliability of the data.

Users must carefully evaluate data according to their specific needs and environment.

Data custodians can help by providing as much information as possible about how and why the data were originally collected, the subjects covered or excluded, the meanings of technical terms and insight into the concepts being measured or represented.

The quality and relevance of a dataset should be evaluated by asking questions about aspects of each dimension, based on user needs.

For example, if credibility and trustworthiness of the data source are particularly important, then a careful examination of the Institutional Environment dimension will be especially important and this may have more weight in making an overall quality assessment.

Alternatively, if a key purpose is to compare and contrast data, then the Coherence dimension will be more significant.

7. REPORTING ON DATA QUALITY

The questionnaire (Appendix A) provides five points each for the dimensions of Institutional Environment, Accuracy, Coherence, Interpretability and Accessibility.

The Data Quality Reporting Tool is an online application, designed to guide agencies through the reporting questionnaire and generate a standard data quality statement.

By selecting each point which applies to the data, the tool will generate standard responses in the data quality statement (Appendix B). It is also possible to add brief explanatory notes or key technical information which may be helpful to the user.

The simple data quality statement will help a user:

- understand how and why the data was collected
- clarify what the data represents
- review or assess the data against their own requirements.

The statement helps a user understand how a particular dataset could be used and whether it can be compared with other, similar datasets. It provides a description of the characteristics of the data to help the user decide whether the data will be fit for their specific purpose.

With its standard format, the statement also helps a user to compare the quality elements from one dataset to another to determine which one can best meet their needs.

The Data Quality Reporting Tool can produce a standard data quality statement in editable MS Word format or as a publishable PDF document. It can also be output as an XML file.

Business requirements for data quality can vary according to the purpose or process the data supports, the technical aspects of different types of data collection and the needs of primary or secondary users.

Agencies will implement data quality management at different levels. Data and information of greater value or higher risk may require more a detailed data quality statement.

7.1. Publishing data of variable quality

The data quality statement is designed to enable a user to determine whether the data is fit for their specific secondary purpose.

A data quality statement does not indicate the potential value of the data. It draws the attention of a secondary user to particular data quality dimensions, characteristics of the data and how it was collected. These are factors the secondary user should consider in their evaluation of fitness for purpose.

As “purpose” will vary widely among users, each user can have a different view of the quality of one dataset or data source.

Data collection exercises can encounter unexpected issues that lead to gaps or inconsistencies. The data collected may not fully represent the factors it was intended to measure, monitor or report. These may impact the types of conclusions that can be drawn from the data.

Even if the data collection did not meet the objectives of the primary user, the data may hold value for a secondary user. The data quality statement will assist the secondary user to understand any limitations in the data and determine whether it is fit for their purpose.

Example

A survey is carried out every year, from 2010 until 2014. It provides a standard set of questions and seeks responses from 150 agencies across the public sector. In 2012, one agency (Agency XYZ) does not submit a response. The survey team make an adjustment for this by substituting the 2011 data of Agency XYZ in the 2012 survey data. They publish the data along with a data quality statement which contains the relevant caveats.

Secondary User A: wants to make a high level analysis of general trends across the public sector. eg Has whole of government spending increased or decreased in the 5-year period? The user considers the 2010 and 2011 responses from Agency XYZ and notes they only account for a small percentage of total government spending in a given year. The user makes an informed decision that the data quality is acceptable for their purposes, and adds a footnote about Agency XYZ in their conclusions.

Secondary User B: wants to make a specific analysis of spending by Agency XYZ in 2012. The user makes an informed decision that this data source is not of suitable quality for their purposes.

Secondary User C: wants to make a comparative analysis of several agencies over the 5-year period. The user makes an informed decision to exclude Agency XYZ from the study, but finds the data quality acceptable for a review of three other agencies.

The data quality statement encourages secondary users to consider some of the following questions to assess data quality according to their specific needs and environment. The list is not exhaustive. Users are encouraged to generate their own questions according to their needs.

- What was the primary purpose or aim for collecting the data?
- How well does the coverage (and exclusions) match the secondary user's needs?
- How useful are these data at small levels of geography?
- Does this data source provide all the relevant items or variables of interest to the secondary user?
- Does the population presented by the data match the secondary user's needs?
- To what extent does the method of data collection seem appropriate for the information being gathered by the secondary user?
- Have standard classifications (eg industry or occupation classifications) been used in the collection of the data? If not, why? Does this affect the ability of data from different sources to be compared or brought together by the secondary user?
- Is there a time difference between the secondary user's reference period, and the reference period of the data?
- What is the gap of time between the reference period (when the data were collected) and the release date of the data?
- Will there be subsequent surveys or data collection exercises for this topic?
- Are there likely to be updates or revisions to the data after official release?

8. RELATED DOCUMENTS

This standard should be read in conjunction with the following related documents:

- *NSW Data and Information Custodianship Policy* which addresses the responsibilities of agencies with respect to specific datasets or information assets
- *NSW Government Open Data Policy* which aims to simplify the public release of appropriate government data for use by industry and the community

9. REVIEW

This standard will be maintained by the ICT Leadership Group. The standard is subject to review at least every two years, or as appropriate.

10. CONTACT

For assistance with any aspect of this document, please contact:

Principal Policy Officer, Information
Office of Finance & Services (Strategic Policy)
Level 23, McKell Building
2-24 Rawson Place
SYDNEY NSW 2000
(02) 9372 8291

APPENDIX A – REPORTING QUESTIONNAIRE

IDENTIFY

1. Name of agency publishing the data quality statement.
2. Custodian agency of the dataset or data source.
 - 2a. If applicable, explain why agency publishing is different from custodian agency.
3. Name of contact person or position at custodian agency.
4. Email of contact person or position at custodian agency.
5. Phone number of contact person or position at custodian agency.
6. Name of the dataset or data source.
7. Brief description of the dataset or data source.
8. To which government function does the data relate? (choose one)
 - Business & industrial development
 - Education & training
 - Finance
 - Health
 - Labour
 - Law & order
 - Recreation & culture
 - Conservation & environment
 - Emergency management
 - Government administration
 - Infrastructure & communications
 - Land & resource management
 - Primary industries
 - Social & community services

INSTITUTIONAL ENVIRONMENT

- Is the agency publishing this data the recognised data custodian?
- Is the data collected and managed according to a data quality framework?
- Are data governance roles and responsibilities clearly assigned for this dataset or data source?
- Is the data collection mandated or required by a law, regulation or agreement?
- Has the custodian any commercial interest or conflict of interest in the data?

Additional information:

- Where can the user find out more about the data quality framework?
- Where can the user find out more about the data governance responsibilities?
- Name of law, regulation, or agreement to collect data.
- Declaration of commercial or other interest in the data.

ACCURACY

- Has the data been subject to quality assurance processes? (For example: Checking for errors at each stage of data collection and processing, or verifying data entry and making corrections if necessary.)
- Did the data collection meet the objectives of the primary user? ie Does the data correctly represent what it was designed to measure, monitor or report?
- Are there any known gaps in the data? (For example: non-responses, missing records, data not collected.)
- Have there been any adjustments, changes or other factors that could impact the validity of the data? (For example: weighting, rounding, de-identification of data, changes or flaws in data collection or verification methods.)
- What is the revision policy, if errors are identified in the data after publication?

Additional information:

- Where can the user find out more about the quality assurance processes?
- What was this data designed to measure, monitor or report?
- Briefly describe gaps in the data and enter appropriate caveats or explain where the user can find more detailed information.
- Briefly explain which data items were adjusted, why and how. Describe any changes or other factors that could impact the validity of the data and enter appropriate caveats or explain where the user can find more detailed information.
- Briefly describe the revision policy. Explain how often revisions are produced or disseminated.

COHERENCE

- Have standard definitions, common concepts, classifications and data recording practices been used?
- Can elements within the data be meaningfully compared?
- Is this data generally consistent with similar or related data sources?
- Have there been any significant changes in the way data items are defined, classified or counted over time?
- Is this data consistent with previous releases? Have there been any changes in methodology or external impacts since the last data release.

Additional information:

- Briefly describe the types of standards that have been used and where the user can find out more.
- OR describe any non-standard or uncommon practices, why this approach was taken, likely impacts.

- Describe anything that might affect the ability to compare multiple items within this dataset. For example: changes or differences in the way data is collected, classified or processed by different parties, or at different times.
- List other known data sources which report similar information. How does this data compare?
- OR list any known data sources with which this data cannot be meaningfully compared or combined, and why. For example: differences in scope or definitions.
- When was the data in this series first collected or produced?
- Describe any changes that could cause a break in the consistency of data items over time.
- Describe any data collection exercises that may be planned for the future.
- Describe any changes or events which have impacted the data since the last release.

INTERPRETABILITY

- Is a data dictionary available to explain the meaning of data elements, their origin, format and relationships?
- Is information available about the primary data sources and methods of data collection? (For example: instruments, forms, instructions.)
- Is information available to help users evaluate the accuracy of the data and any level of error?
- Is information available to explain concepts, help users correctly interpret the data and understand how it can be used?
- Is information available to explain ambiguous or technical terms used in the data?

Additional information:

- Where can the user find out more about the data dictionary?
- Where can the user find out more about the primary data sources and methods of data collection?
- Where can the user find information to evaluate the accuracy of the data and any level of error?
- Where can the user find information about concepts and how to understand or interpret the data?
- Where can the user find information to explain ambiguous or technical terms used in the data?

ACCESSIBILITY

- Is the dataset or data source available online with an open licence?
- Is the dataset or data source available in a machine-processable, structured format?
- Is the dataset or data source available in a non-proprietary format?
- Is the dataset or data source described using open standards and persistent identifiers?
- Is the dataset or data source linked to other data, to provide context?

Additional information:

- Enter the appropriate licence type and provide any additional information about restrictions on use of or access to the data.
- List the different digital file types or formats in which the data is available.
- Provide any other information about the use of identifiers and links to other data.

INFORMATION TO HELP USERS EVALUATE RELEVANCE:

Scope and coverage

- About whom, or what, was the data collected? (target audience, population, event)?
- Were any individuals, groups or occurrences excluded from the data collection?
- Who or what were excluded? Does this have any impact or cause any bias?
- Which jurisdiction applies to the data?

Geographic detail

- Is the data collected from across the whole State or only from particular Localities (Suburbs)? If only from particular Localities, please list which ones apply.
- For which other levels of geography are the data available? (ie postcode, Local Government Area, Local Health District, road segment, address, XY geocode, travel zone, other – please explain)
- How are the data represented or apportioned at lower levels of geography?

Outputs

- In what form(s) are the data available?
 - Primary (original raw numbers)
 - Transactional (structured content, generated in a business process)
 - Authored (unstructured content, documents, multimedia or applications)
 - Indexes
 - Other
 - Analytical (structured content, aggregated or derived information)
 - Published (unstructured content, assembled into a form suitable for wide dissemination)
 - Estimates

Other cautions

- What does the data not represent or cover?
- Any other issue or caution that affects the use or interpretation of the data?

Reference period

- What is the period for which the data were obtained? (eg 2012-13 financial year, 2013 academic year, week ending Sat 27 Sep 2013)
- Were there any exceptions to the collection/observation period (eg delays in receipt of data, changes to recording processes)

Timing

- When did the data become available?
- Are there likely to be updates or revisions to the data after its release?

Frequency of production

- How often is the data collected or expected to be collected?
- Are there other, less frequent data sources that contain more detailed data that can be used in other reporting years when available?

Improving the data quality rating

Institutional environment:

- Put a governance framework in place and assign responsibilities.

Accuracy:

- Plan your data collection exercise.
- Attach relevant caveats to help users understand what the data represents and enable them decide if it could be used for their secondary purpose.
- Define and publish a revision policy explaining how updates are made.

Coherence:

- Make use of agreed data standards and apply them consistently over time.
- Identify similar or related data sources. Explain whether they are comparable or not, and why.

Interpretability:

- Make procedural and technical documentation available to help secondary users make sense of the data.

Accessibility:

- Publish data with an open license in the most re-useable formats possible.

APPENDIX B – DATA QUALITY STATEMENT

The standard data quality statement helps a user understand how a particular dataset could be used and whether it can be compared with other, similar datasets. It provides a description of the characteristics of the data to help the user decide whether the data will be fit for their specific purpose.

The reporting questionnaire asks five key questions for each of these data quality dimensions:

- Institutional Environment
- Accuracy
- Coherence
- Interpretability
- Accessibility

For each question: “yes” = 1 point; “no” = 0 points

Other questions describe additional information that can help users interpret the data.

The number of points determines the Quality Level for each dimension (high, medium, low).

Dimensions with four or five points receive a star.

Points	Quality Level	Star / No Star
0	LOW	No Star
1	LOW	No Star
2	LOW	No Star
3	MEDIUM	No Star
4	MEDIUM	Star
5	HIGH	Star

Data Quality Reporting Tool

Visit Data.NSW to access the Data Quality Reporting Tool. It is an online application designed to guide agencies through the reporting questionnaire. The tool can generate a standard data quality statement as an editable MS Word document, a publishable PDF document or as an XML file.

The Data Quality Reporting Tool will gather all the metadata required to register a dataset on Data.NSW or the Information Asset Register (IAR).

Contact DataNSW@finance.nsw.gov.au for more information about the Data Quality Reporting Tool.

INSTITUTIONAL ENVIRONMENT

✓ Agency publishing this data is the recognised data custodian.	✗ The agency publishing this data is not the data custodian.
✓ Data is collected and managed according to a data quality framework.	✗ Data is not collected or managed according to a data quality framework.
✓ Data governance roles and responsibilities are clearly assigned for this dataset or data source.	✗ Data governance roles and responsibilities are not assigned for this dataset or data source.
✓ Data collection is mandated or required by a law, regulation or agreement.	✗ Data collection is not mandated or required by a law, regulation or agreement.
✓ Custodian has no commercial interest or conflict of interest in the data – or, has declared any interest in the data.	✗ No information available regarding whether the Custodian has any commercial interest or conflict of interest in the data.

ACCURACY

✓ This data has been subject to quality assurance processes. ie Checking for errors at each stage of data collection and processing, or verifying data entry and making corrections if necessary.	✗ This data has not been subject to any quality assurance processes. ie Checking for errors at each stage of data collection and processing, or verifying data entry and making corrections if necessary.
✓ The data collection met the objectives of the primary user. The data correctly represents what it was designed to measure, monitor or report. Refer to Scope and Coverage in the last section of this report for more information.	✗ The data collection did not meet the objectives of the primary user. The data may not fully represent what it was designed to measure, monitor or report. Refer to Scope and Coverage in the last section of this report for more information.
✓ There are no known gaps in the data. (For example: non-responses, missing records, data not collected.) OR ✓ Gaps are identified in caveats attached to the dataset or data source.	✗ There is no information available about gaps in the data. (For example: non-responses, missing records, data not collected.)
✓ There have been no adjustments, changes or other factors that could impact the validity of the data. (For example: weighting, rounding, de-identification of data; changes or flaws in data collection or verification methods.) OR ✓ Adjustments are identified in caveats attached to the dataset or data source	✗ There is no information available about adjustments, changes or other factors that could impact the validity of the data. (For example: weighting, rounding, de-identification of data; changes or flaws in data collection or verification methods.)
✓ Revision policy: If errors are identified, data is revised and the revision is publicised	✗ There is no revision policy

COHERENCE

✓ Standard definitions, common concepts, classifications and data recording practices have been used.	✗ Standard definitions, common concepts, classifications and data recording practices have not been used.
✓ Elements within the data can be meaningfully compared.	✗ Elements within the data cannot be meaningfully compared.
✓ This data is generally consistent with similar or related data sources.	✗ No similar or related data sources have been identified.
✓ This dataset is a single collection. It is not impacted by changes in methodology or external events over time. OR ✓ This data is part of a time series. There have not been any significant changes in the way data items are defined, classified or counted since the start of the series.	✗ This data is part of a time series. There have been significant changes in the way data items are defined, classified or counted since the start of the series.
✓ This data is part of a time series. It is consistent with previous releases. There have been no changes in methodology or external impacts since the last data release.	✗ This data may not be consistent with previous releases. There have been changes in methodology or external impacts since the last release in this series.

INTERPRETABILITY

✓ A data dictionary is available to explain the meaning of data elements, their origin, format and relationships.	✗ There is no data dictionary available to explain the meaning of data elements, their origin, format and relationships.
✓ Information is available about the primary data sources and methods of data collection. (For example: instruments, forms, instructions.)	✗ There is no explanatory information available about the primary data sources and methods of data collection.
✓ Information is available to help users evaluate the accuracy of the data and any level of error.	✗ There is no further information available to help users evaluate the accuracy of the data and any level of error.
✓ Information is available to explain concepts, help users correctly interpret the data and understand how it can be used.	✗ There is no further information available to explain concepts, help users correctly interpret the data and understand how it can be used.
✓ Information is available to explain ambiguous or technical terms used in the data.	✗ There is no further information available to explain ambiguous or technical terms used in the data.

ACCESSIBILITY

✓ This dataset or data source is available online with an open licence.	✗ This dataset or data source is subject to limiting or restrictive access conditions.
✓ This dataset or data source is available in a machine-processable, structured format.	✗ This dataset or data source is not available in a machine-processable, structured format.
✓ This dataset or data source is available in a non-proprietary format.	✗ This dataset or data source is not available in a non-proprietary format.
✓ This dataset or data source is described using open standards and persistent identifiers.	✗ This dataset or data source is not described using open standards and persistent identifiers.
✓ This dataset or data source is linked to other data, to provide context.	✗ This dataset or data source is not linked to other data.

Data Disclaimer

NSW Government is committed to producing data that is accurate, complete and useful. Notwithstanding its commitment to data quality, NSW Government gives no warranty as to the fitness of this data for a particular purpose. While every effort is made to ensure data quality, the data is provided “as is”. The burden for fitness of the data relies completely with the user. NSW Government shall not be held liable for improper or incorrect use of the data.

APPENDIX C – GLOSSARY

TERM	DEFINITION
Accessibility	A dimension of quality relating to the ease with which data or information can be retrieved, used and understood.
Accuracy	A dimension of quality relating to the degree to which the data or information correctly describes that which it was designed to measure.
Administrative data	Data or information created or collected through: <ul style="list-style-type: none"> • recording transactions or events • monitoring a process or service • querying another data or information source; or • regulatory activities and audit actions.
Coherence	A dimension of quality relating to the degree to which data or information can be compared with itself and other information over time.
Comparable	(1) Information that can be compared, is similar, is worthy of comparison. (2) Comparability refers to the extent to which differences between statistics for different places or times can be attributed to real differences between the things being measured.
Custodian	The agency, body or position designated with the custody of a specified dataset or information asset. Responsible for: <ul style="list-style-type: none"> • the development, management, care and maintenance of a specified dataset or information asset; • ensuring that all legal, regulatory and policy requirements are met in relation to the management of the specified dataset or information asset; and • determining the conditions for appropriate use, sharing and distribution of the specified dataset or information asset.
Data Quality	<p><i>Quality</i> describes the degree to which a set of inherent characteristics fulfils requirements.</p> <p><i>Data Quality</i> is generally accepted as "fitness for purpose".</p> <p>The ABS <i>Data Quality Framework</i> is comprised of seven dimensions (or characteristics) of quality: Institutional Environment; Relevance; Timeliness; Accuracy; Coherence; Interpretability; and Accessibility.</p> <p>Data quality is evaluated in terms of how well the characteristics of the data meet the needs or objectives of a user.</p>

TERM	DEFINITION
Institutional environment	A dimension of quality relating to the institutional and organisational factors which may have a significant influence on the effectiveness and credibility of the agency producing the data or information.
Interpretability	A dimension of quality relating to the degree to which data or information can be understood, explained and used.
Quality assurance checks	A system or series of activities for ensuring the maintenance of proper standards especially periodic interrogation and sampling of the product.
Relevance	A dimension of quality relating to how well data or information meets the needs of the user in terms of the concept(s) measured and the population(s) presented.
Time Series	A record of activity where data is measured at regular intervals over a period of time (eg a monthly unemployment rate). Time series assist understanding of the current situation, enabling the most recent data observations to be placed in a meaningful historical perspective.
Timeliness	<p>A dimension of quality relating to:</p> <ul style="list-style-type: none"> • the time taken between the occurrence of the characteristics/events being measured and the release of the data or information output; and • whether the data or information output is sufficiently up-to-date for the user's purpose.
User	End consumer of a data or information resource; those who use data or information for reference, or as input to solve problems and/or make decisions.