![ACT Government | ACT Health]

Data Quality Policy and Procedure

Document number	AHDPD-03:2023
Effective date	10 August 2023
Review date	10 August 2024
Author branch	Data Analytics Branch
Endorsed by	Executive Board (Operational)
Audience	All staff and contractors
Version number	1.0

Contents

Policy statement

Data is a vital asset for organisations and when managed effectively and responsibly, it can be utilised to generate benefits for the public, such as the delivery of improved services or better health and welfare outcomes. ACT Health Directorate (ACTHD) and Canberra Health Services (CHS) use data to record clinical care episodes, inform operational or health system policy decision-making, support research and evaluation, enable performance monitoring and reporting, and to underpin funding justifications. Those who rely on this information must be able to trust that the foundational data they access are of the highest quality, and that any problems or limitations are fully disclosed.

Purpose

Where there is poor quality data that contains missing values, inaccuracies, is incomplete, truncated or out-of-date, it may not be fit-for-purpose. The unknowing utilisation of these data may adversely impact end-user decision-making; lead to inaccurate or misleading reporting, research or evaluation findings; or result in other erroneous clinical, operational or health-system outcomes that diminish data user or community trust.

The purpose of this policy is to support the communication of data quality issues to data users. It also aims to facilitate the collection and management of high quality, fit-for-purpose data to support ACTHD business and operational decisions.

Scope

This policy applies to all ACTHD staff, including permanent, temporary and casual employees and students, along with external contractors, consultants, and volunteers who use, report, share, release or interpret data. These personnel are required to comply with all obligations in relation to data management as directed in relevant legislation, policies and procedures.

As data are shared between ACTHD and CHS, it is imperative that data quality issues are addressed in a uniform manner across these organisations. As such this policy applies to all data collected or hosted by ACTHD or CHS, including data shared between these organisations. It applies to data acquired from external sources that are accessible through ACTHD IT infrastructure. The policy also applies to data held solely by ACTHD that relates to the conduct of ACTHD business or operations.

Data Quality Statements

A Data Quality Statement (DQS) (also known as a Data Quality Declaration) is a 'presentation of information about the quality of a statistical collection or product'.[1] It provides information on the quality and suitability of the collected data for its original intended use and potential re-use.

To help data users effectively and confidently utilise data and assess whether the data are fit for their intended purposes, all datasets, data outputs and reports should be supported by a DQS.

---

1 Australian Bureau of Statistics, ABS Data Quality Statement Checklist.

Datasets

In collaboration with data custodians, data stewards are responsible for ensuring, promoting and enabling the collection and maintenance of high-quality data for the datasets assigned to them. Data stewards are also responsible for preparing, updating and publishing DQSs on the central data catalogue. Each DQS should be time specific and reviewed and updated at each reference period.

> **The DQS for a dataset should be published on the central data catalogue.**

Data custodian and steward roles are described in the *Data Accountabilities Policy and Procedure.*

Outputs and reports

DQSs should be included in all outputs that are drawn from a dataset. For routinely published dashboards, publicly released outputs (such as to ACT Open Data), or other regular reporting, it is not necessary to include a DQS. Instead, where there are data quality issues that may affect a reader's interpretation of the report or dashboard, a footnote should be provided as close as possible to the output. For example, when reporting Aboriginal and Torres Strait Islander status, a footnote may be necessary to draw attention to the uncertainty around identification of Aboriginal and Torres Strait Islander people within administrative datasets.

> **DQSs should be included in all outputs that are drawn from a dataset.**

Seven dimensions of quality

Datasets should be assessed against the following **seven dimensions of quality**[2] as set out by the Australian Bureau of Statistics:

1. Institutional environment:[3]
   o Institutional and organisational factors which may have a significant influence on the effectiveness and credibility of the agency producing the statistics.
   o Enables an assessment of the context, which may influence the validity, reliability or appropriateness of the product.

2. Relevance:[4]
   o How well the statistical product or release meets the needs of users in terms of the concept(s) measured, and the population(s) represented.
   o Includes information that allows end-users to decide whether the dataset or data holding (refer to Glossary) addresses the issues of importance.
   o Enables an assessment of whether the product addresses the issues most important to end-users and to the broader Australian community
   o Includes any cautions of which end-users should be aware.

---

2 Australian Bureau of Statistics, ABS Data Quality Statement Checklist.
3 Australian Bureau of Statistics, Data Quality Statement, Institutional Environment 1520.0
4 Australian Bureau of Statistics Data Quality Framework, Relevance 1520.0

3. Timeliness:[5]
    o Refers to the delay between the reference period (to which the data pertain) and the date at which the data become available; and the delay between the advertised (published) date and the date at which the data become available (i.e., the actual release date).
    o Lengthy delays between the reference period and data availability, or between advertised and actual release dates can have implications for the currency or reliability of the data.

4. Accuracy:[6]
    o The degree to which the data correctly describe the phenomenon they were designed to measure.
    o Relates to how well the data portray reality, which has clear implications for how useful and meaningful the data will be for interpretation or further analysis. Statistical outputs for analysis are generally not the primary reason for the collection of administrative data.

5. Coherence:[7]
    o The internal consistency of a statistical collection, product or release, as well as its comparability with other sources of information, within a broad analytical framework and over time.
    o The use of standard concepts, classifications and target populations promotes coherence, as does the use of common methodology (across surveys for example).
    o Indicates factors that may affect comparability over time.

6. Interpretability:[8]
    o Information that helps to provide insight into the data, and assist with interpretation - variables, metadata availability, including concepts, classifications, and measures of accuracy.

7. Accessibility:[9]
    o The ease of access to the data; the ease with which the existence of information can be ascertained; suitability of the form or medium through which information can be accessed; and access costs.

Assessments are to be conducted using the Data Quality Statement template at **Appendix A**.

Using the information provided in a DQS, the data user should consider whether any reported issues may affect the confident use of those data for their intended purpose. Data that are fit for one purpose may not be fit for another.

Data that are fit for one purpose may not be fit for another.

---

5 Australian Bureau of Statistics Data Quality Framework Timeliness 1520.0
6 Australian Bureau of Statistics Data Quality Framework, Accuracy 1520.0
7 Australian Bureau of Statistics Data Quality Framework, Coherence 1520.0
8 Australian Bureau of Statistics Data Quality Framework, Interpretability 1520.0
9 Australian Bureau of Statistics Data Quality Framework, Accessibility 1520.0

Where data quality issues are not understood (for example, there is no metadata or DQS), or where known quality issues cannot be rectified, it is recommended that the data not be disclosed (made available to others). This is important to prevent the use of data in decision-making, research, reporting or other activities where the data user cannot confidently or adequately assess whether the data are fit for their intended purposes.

> **Where data quality issues are not understood or cannot be rectified, it is recommended that the data not be used or disclosed.**

The data custodian, in consultation with the data steward, determines whether the extent of issues or lack of data quality documentation should prevent the disclosure of those data.

> **The data custodian determines whether the extent of issues or lack of data quality documentation should prevent the disclosure of those data.**

Improving data quality

The quality of data can be assured or improved through data validation and quality assurance checks.

Data validation

Data validation is 'the confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled'.[10] This helps to assure the data user about the accuracy and quality of data through the application of rules to check the correctness and meaningfulness of data. Business areas, in consultation with the data steward, will develop business rules to validate data.

Data stewards are responsible for ensuring that data have been validated and anomalies rectified. Where errors require changes to data in source systems, these changes must be approved in writing by the relevant data steward, with changes logged in the IT system and the enterprise records management system (Objective) to enable auditing. The process for resolving validation errors should also be documented.

Data quality assurance

Even where data validation processes have ensured the quality of data within a dataset or data holding, errors may occur during data extraction or manipulation. As such, all data must have an appropriate level of review, prior to its use or disclosure.

Where data are to be released for research purposes or *ad hoc* requests, the relevant business area will undertake a data quality assurance process which may include a quality assurance checklist or multiple levels of review and approval. The outcome of this activity is to be included in the advice to

---

10 International Standard ISO 8000-8 Data Quality Part 8.

the data custodian when seeking approval for the disclosure. The assurance process is to be documented in the enterprise records management system (Objective).

Data revisions

All staff have a responsibility to bring data quality issues to the attention of data stewards or their managers so they may be rectified as quickly as possible. Data stewards will work with data users and relevant data custodians to address quality problems to help data be ready and suitable for use by ACTHD staff, or for sharing with approved external data requestors.

The active communication of data quality issues, including unresolvable problems, assists in promoting a data culture at ACTHD where data issues are transparently documented and addressed to maximise the utility and value of datasets and data holdings.

> **The active communication of data quality issues assists in promoting a data culture at ACTHD where data issues are transparently documented and addressed to maximise the utility and value of data holdings.**

Roles and Responsibilities

Responsibilities for all staff are detailed in **Table 1.** Staff members may have more than one role at a time and their roles may change for different datasets.

Table 1: Key data quality roles and responsibilities

Position	Responsibility
All staff	Ensure that data are: • collected and recorded promptly and accurately • collected in accordance with standards and documented procedures • appropriately updated to reflect changes in accordance with standards and documented procedures • corrected in a timely manner based on documented evidence in accordance with standards and documented procedures • maintained legally (*Information Privacy Act 2014*, *Territory Records Act 2002*, *Health Records (Privacy and Access) Act 1997*, *Public Health Act 1997* and the *Public Service Management Act 1994*) • maintained in accordance with any contractual obligations (e.g. custodian agreements, contract of employment, service contract) • maintained ethically (professional codes of practice) • proactively managed to improve data and datasets, by alerting data stewards to data issues.
Executives	• lead and promote the development of a data quality culture • determine the records required to demonstrate compliance with this policy for areas within their responsibility.

Table 1 continued: Key data quality roles and responsibilities

Position	Responsibility
Managers	• provide the necessary resources for staff and data stewards to manage data quality • encourage a data quality component at the commencement of initiatives and projects • provide direction and prioritise quality improvement of datasets • ensure staff members are appropriately trained to fulfil their data quality responsibilities.
System trainers	• promote data quality and embed best practice into training • include training materials that support best practice data quality.
Data custodians	• work with data stewards to address data quality issues.
Data stewards	• ensure, promote and enable the collection and maintenance of high-quality data for data assigned to them • prepare, update and publish DQSs on a central register • ensure that data have been validated and anomalies rectified.
ICT personnel	• address system-related data quality issues with the appropriate stakeholders as they are identified • address system issues in a timely manner and ensure that stakeholders are notified of issues that may impact data quality • identify ways to improve source data systems.
System administrators	• address data quality issues with the appropriate stakeholders as they are identified • ensure system documentation and metadata meet established agency criteria • work with the appropriate business units to ensure metadata correctly reflects system fields and data elements.
Analysts, data users	• have the appropriate skills to use and analyse data • access and understand DQSs prior to utilising data.
Data entry staff (such as clinicians, and administrative staff)	• understand the purpose for which the data were collected (primary purpose) • ensure data are captured and validated at the point of source • respond to any data quality issues identified at the source • ensure data are correct, current and complete at the time of data entry • ensure data are entered as close to real-time as possible • report data quality issues as they are identified to the appropriate manager or data steward • escalate data quality issues that impact on immediate health and welfare (e.g. patient safety) to the manager as a priority • participate in data quality education activities • address feedback from management regarding data quality • participate in the resolution of issues as appropriate.

Evaluation

Outcome Measures	Method	Responsibility
Data Quality Statements are developed for all datasets	The central data catalogue will be periodically reviewed.	Chief Health Data Officer
Explanatory footnotes will be appended to all dashboards or reports where there are data quality issues.	Business area quality assurance.	Relevant business areas

Records Management

All records are to be maintained in accordance with the *ACT Health Directorate Records Management Policy* and *Territory Records Act 2002*.

Version Control

Version	Date	Comments
0.1	August 2022	Initial draft
0.2	October 2022	Added Data Quality Statement and Checklist
0.3	November 2022	DAB, Office of the Chief Health Officer feedback
0.4	January 2023	Data Working Group (CPHB, ACTHD)
0.5	March 2023	ACTHD Legal Policy review
0.6	April 2023	DG feedback
1.0	June 2023	Final revision

Glossary

Term	Definition
Data holding	An aggregation of datasets in a specific repository.
Dataset	A cohesive set of data with measurable value that is designed to address a specific set of business needs.

Appendix A: Data Quality Statement template

This template has been adapted from [DATA VIC](#) and the [ABS Data Quality Statement Checklist.](#)

Using this template, complete a DQS for all releases that draw on a dataset. A DQS for each dataset should be published on the central data catalogue.

Data Quality Statement

Data Quality Statement	
Dataset name	
Date statement completed	
Date review required (within 2 years)	

Domain	Summary
Institutional environment	
Factors that influence the credibility of the organisation collecting the data; their authority to collect	
Institution overview	Insert institution statement here
Legislative, other authority to collect	
Data collection organisations	
Data compiling organisations	
Error detection and correction frequency	
Additional information	
Relevance	
The degree by which the data meets the needs of the user in terms of the population represented.	
Topic of data	
Original collection purpose	
Level of geography	
Key data items	
Numerator source	
Denominator source	
What is not represented	
Standard classifications	
Additional information	

Data Quality Statement

Data Quality Statement	
Dataset name	
Domain	**Summary**
Timeliness	
Refers to the time between the availability of the data and the event it describes.	
Collection frequency	
When data became available	
Reference period	
Planned updates or revisions	
Other relevant data sources	
Additional information	
Accuracy	
The degree to which the data correctly describe the condition it was designed to measure	
Method of collection	
Adjustments	
Sample size	
Dataset size	
Standard errors	
Under or over counts	
Potential for bias	
Processing corrections	
Non-response/reporting rates	
Missing sub-populations	
Additional information	

Data Quality Statement

Data Quality Statement	
Dataset name	
Domain	**Summary**
Coherence	
Refers to the internal consistency of the data; consistency in methods and standards compared with other information sources.	
Consistency over time, impact	
Consistency over jurisdictions, nationally	
Numerator and denominator comparability, impact	
Time series consistency	
Changes to underlying data	
Impact real-world events	
Similar data sources	
Comparability other data sources	
Additional information	
Interpretability	
Refers to the availability of information, such as metadata or guides, to assist the user understand the data	
Context considerations	
Supporting information	
Clarification of terms	
Additional information	
Accessibility	
Refers to the ease of access to the data, including the medium of access; findability (existence) of the data	
Unpublished data availability	Available Not available: [add reason]
Data request contact details	
Additional information	