Data Quality Guideline

Information Management Framework

Document Control

Applies to	All departments and Victoria Police	Authority	Information Management Group
Period	2018-2021	Advised by	CIO Leadership Group
Issue Date	March 2018	Document ID	IM-GUIDE-09
Review Date	March 2020	Version	1.0

Approval

This document was approved by the whole of Victorian Government Information Management Group under authority of CIO Leadership Group on 25/04/2018 and applies from the date of issue.

Version history

Version	Date	Comments
0.7	07/03/2018	First draft for review by stakeholders
1.0	26/03/2018	Incorporated feedback

Contents

Introduction

Overview

This guideline provides high-level advice to Victorian Government (government) departments and Victoria Police relating to assessing, maintaining and improving the quality of critical and shared data assets. This guideline will assist data owners and data custodians to understand and implement the IM-STD-07 Data Quality Standard (the standard) along with the supporting templates.

The guideline provides further advice and clarification on the dimensions of data quality, and implementation of tools and processes for the assessment and management of assets, and suggested data quality improvement activities. The guideline will assist in the formation of implementable action plans for monitoring, maintaining and improving data quality.

In this guideline, 'data asset' refers to a dataset containing structured data[1]. Unstructured data are not covered.

Rationale

Trusted high-quality data across the government enables confident decision-making, informs policy development, promotes data reuse, and supports service delivery. Providing a broad set of minimum requirements to be applied to all critical and shared data assets provides a solid foundation for a consistent approach to measure, communicate and improve quality of data across government.

Departments and agencies are becoming increasingly aware of the benefits of data sharing across the Victorian Public Sector as well as with other external organisations (e.g. research facilities) and the public. Sharing quality data:

▪ enables timely and informed decisions

▪ maximises accountability and transparency

▪ allows collaboration between government and the public

▪ helps to avoid the rework and duplication of additional data collection

▪ allows users to question data.

The government should share a consistent approach to data management, both at the enterprise level via an Enterprise Data Management Plan[2] and the data asset level via standards (including the Data Quality Standard). Implementing effective data quality practices will increase the value of data across an organisation both as a strategic and operational asset.

---

[1] 'Structured data' refers to data that can be organised and stored in fixed fields such as in a relational database record or spreadsheet. 'Unstructured data' does not conform neatly into a fixed field format. Examples of unstructured data include: data streams, social media data, documents, emails, videos, audio files, and images.

[2] As described by IM-STD-05 Enterprise Data Management Plan Standard
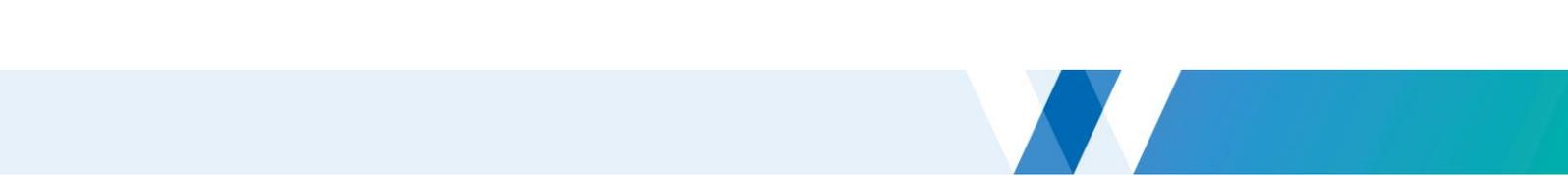
The standard has been developed to outline the minimum requirements that are needed for departments to maintain the quality of their data assets and to help them to comply with their statutory and administrative obligations.

The guideline has been developed to assist departments to assess, measure and describe the quality of their data.

Derivation, scope and glossary

Derivation

This guideline is derived from the WOVG Information Management Policy, and is guided by the Information Technology Strategy for the Victorian Government, 2016–2020 (IT strategy).

Scope

The following departments and agencies are formally in scope:

- Department of Economic Development, Jobs, Transport and Resources
- Department of Education and Training
- Department of Environment, Land, Water and Planning
- Department of Health and Human Services
- Department of Justice and Regulation
- Department of Premier and Cabinet
- Department of Treasury and Finance
- Victoria Police.

These are referred to collectively as 'departments' in this document.

While not required, this guideline may be adopted by agencies and partner organisations, if desired.

Audience

The *Data Quality Guideline* is intended for information managers and business data owners, but may also be applicable to information and IT practitioners.

Glossary

The glossary of terms and abbreviations used in this document are defined in the Information Management Glossary.

## Related documents, tools and references

- Data Integrity Manual (DTF)
- DataVic Access Policy
- Freedom of Information guidelines
- Information Privacy Principles (Guidelines) (Office of the Victorian Information Commissioner) (OVIC)

- [Public Record Office of Victoria (PROV) Standards Framework and Policies](#)
- [Victorian Protective Data Security Framework (OVIC)](#)
- [WOVG Information Management Governance Standard](#)
- [WOVG Information Management Framework](#)
- [WOVG Information Management Policy](#)

The below diagram indicates how the standard and associated documents relates to other WOVG information management material. The standard forms part of a wider enterprise level data management plan (green level), informing each department's strategies and plans (purple and blue levels). The associated Data Quality Guideline (this document) and templates operate at the data asset level (orange level). The red level is the overarching strategy for information management across the department.

Figure 1 - Information management documents
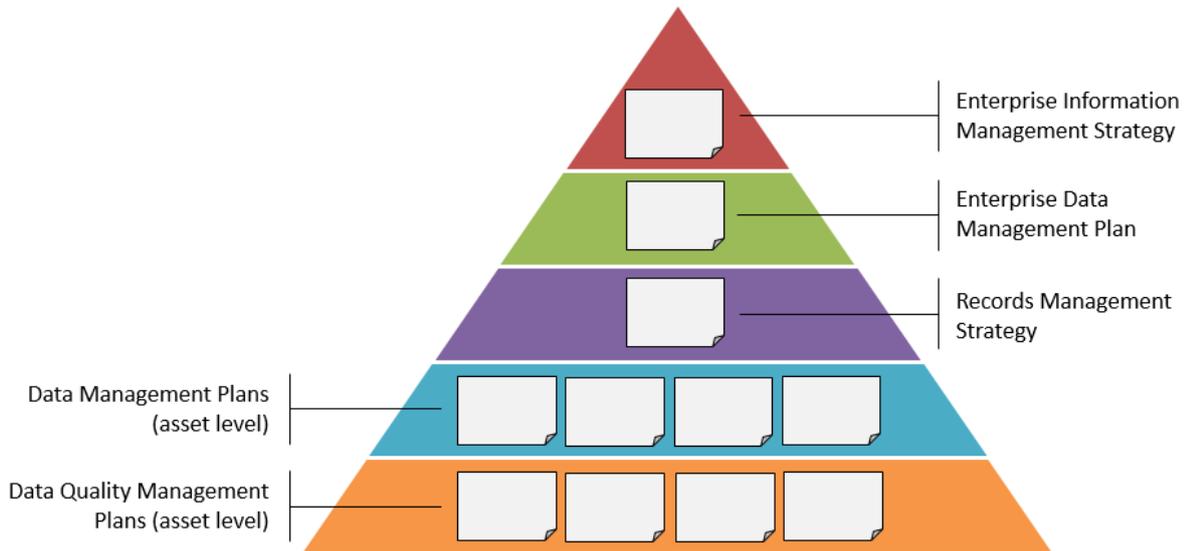
The guideline (this document) works with the standard and templates as per the following diagram. It provides guidance for implementing the standard, and using the templates provided, for each data asset as required.

Figure 2 - Data quality documents
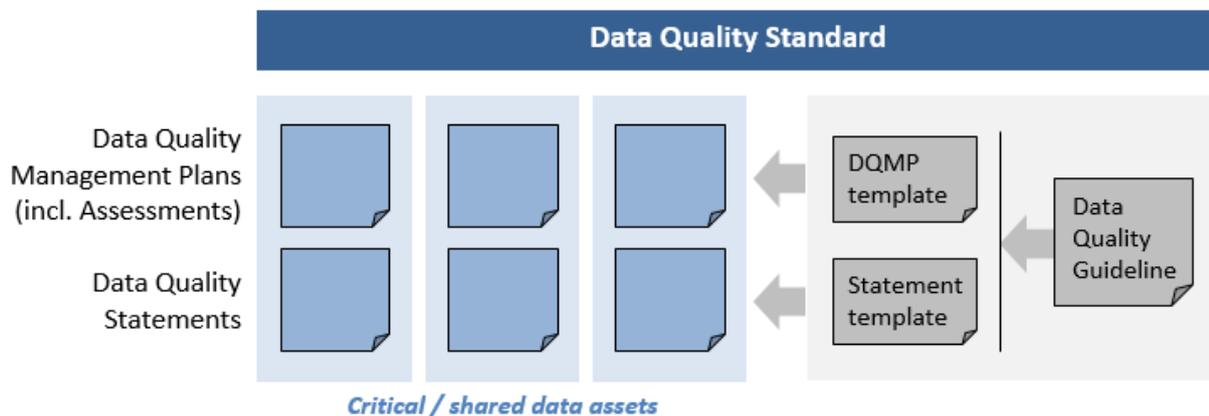
Guideline

The Data Quality Guideline (this document) and associated templates are intended to provide help in implementing the standard, by providing more information about the supplied templates and further definition of the data quality measures.

Specific parts of the standard addressed are:

1. Develop and regularly maintain a management plan (see the Data Quality Management Plan template) for each critical data asset (see the Information Management Governance Standard). Note: a management plan is not required where the data asset is closed (no longer being collected or created) or a one-off (collected or created once only).

2. Create a Data Quality Statement for all critical data assets.

3. A Data Quality Statement should also be created for data assets to be shared internally, with other departments or external partners, where the assets have been assessed as higher risk as a result of sharing. (Note: risk is to be determined by each department.)

> (!) Departments will be seen as adhering to this standard, where they have processes in place which equal or exceed the requirements outlined above. Use of the supplied templates is **optional**.

Data quality dimensions

The seven data quality dimensions specified below should be applied across Victorian Government to determine data quality (refer to Figure 3). These dimensions span all measurable aspects of quality for defining critical and/or shared data assets.

It is important to note that the cause of data quality issues within any of these dimensions may be related to people, process, technology or a combination of these. Therefore, plans for assessing and improving data quality should address all these areas when assessing each dimension.

Figure 3: Data Quality Dimensions

Each dimension is further divided into sub-sections (properties). These properties define certain characteristics to be considered in the evaluation of an individual dimension. Properties each define a singular, testable measure of quality and are the most granular level of the hierarchical model. Descriptions of the seven dimensions, along with examples of properties for each, can be found below. These can be used as appropriate to each data asset to populate the assessment tool found in the Data Quality Management Plan (management plan) template to measure and track data quality.

Analysis dimensions can assist departments in the development of both management plans and associated assessments as well as the development of Data Quality Statements. All dimensions should be applied consistently across government to allow for comparability of data quality processes, initiatives and progress.

Accuracy

*Is the data **accurate** and **valid**, and to what level?*

Accuracy refers to the degree in which the data correctly portrays the real world situation in which it was originally designed to measure. Data must be meaningful and useful to allow for correct and accurate interpretation and analysis. For data to be accurate, it must also be valid, meaning it must conform to a defined format whilst implementing and adhering to specific
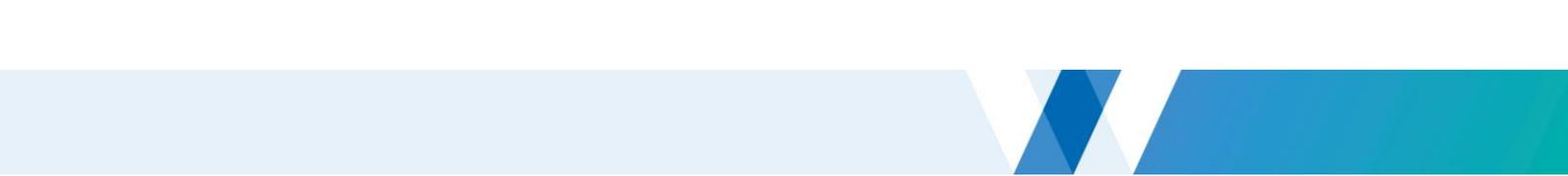
business rules, which may be recorded in a metadata repository (a system or application where information about the data (metadata) is stored and managed).

Properties

Example properties for Accuracy include:

Data collection: Errors may occur at the source of collection, usually by providers of data or during data entry. They can occur at multiple stages of the collection process and may be caused by multiple factors including poor form design, providers having an unclear understanding of concepts and/or the form, and human error. Poor data collection may misrepresent findings and lead to incorrect conclusions, or cause gaps in the data asset.

Data collection errors can be identified when data is compared to other appropriate sources or through follow up investigations and may be prevented by implementing checks and validations for data entry. Feedback may be established between data owners and suppliers to enable rectification of consistent errors in collection. Note that it is important to consider the associated risk when correcting data collection errors, e.g. implications to sensitive or personal data, as certain data may need to remain 'uncorrected'. (See also Collection dimension.)

Data is commonly understood: An understanding of the use and meaning of the data by those collecting and using it reduces risk of errors in entry, and misinterpretation of the data. To increase understanding, the data should be consistent with documented formats and standards, and the intent of each data field defined, and the definitions and related details readily available (for example, in a data dictionary or glossary). Data that is well understood is more likely to correctly represent what it is intending to capture

Process: This relates to any process where data is collected or manipulated, such as changing format or moving data. This may be done by a person or automatically by a system. Business processes should be created and in use to monitor and improve quality. Processes should be documented and discoverable, and automated where appropriate to remove human error and reduce errors.

Incorrect processing of data can have an impact on data quality. Data process errors can be created by both people and systems. Examples can include incomplete matching of one set of data to another (e.g. codes or references), modifications and adjustments made to data to ensure confidentiality, and rounding errors involved with collection.

People: It is important that all staff that are involved with data collection and processing are aware of the importance of data quality, and what they can do to contribute to ensuring quality is maintained or improved. For example, if a user has to repeatedly correct a data asset manually, they should feed this information back to the data collector or provider to fix any issues in the original data asset or collection mechanism. Also relates to Collection dimension.

Validation: Validations are used to verify that data fits certain criteria and standards specified for that particular type of record. Validation rules should be documented and be accessible by all staff. Where possible, automated validations should be implemented in systems used for data collection and processing to reduce errors. For example, field or form level validation used at the point of collection will ensure that only allowable values may be entered. Rules should be reviewed to ensure correctness and currency throughout the life of the data asset.

Format: Format relates to the specific pattern or structure of data. Formats that are too general (such as large open text fields) decrease the validity of the data and increase the risk of errors. Structured data should be entered in a certain way, so that when it is collected it is easier to interpret and use. For example, an address is easier to process when entered as separate parts - street number, name, suburb, postcode - rather than a single text field. Data should be validated against format rules (e.g. certain data must not contain numbers or special characters), and where possible, based on recognised standards (e.g. address data standard).

Completeness

*How **complete** is the data? Are there known gaps?*

Completeness relates to the extent the data is complete, i.e. is it 100% complete (if 100% is required)? This dimension also reflects the ability to determine what data is missing, and whether omissions are acceptable (for example, optional data). Departments must determine and understand whether a data asset contains unacceptable gaps, as these may place limitations on the data leading to an increased reliance on assumptions and estimations or preclude the asset for use altogether. It is also useful to note the level of completeness, particularly if 100% is not required to meet the original purpose of the dataset. Also if the dataset is considered complete as at a particular point in time, e.g. beginning or end of month.

Properties

Example properties for Completeness include:

Process: As with the Accuracy dimension, processes must be in place to ensure that data entered is as complete as possible. Feedback and review processes should be implemented to ensure that any gaps or coverage issues are discovered and resolved. This may be done manually by a person or automated by a system. Business processes should be documented and implemented to check completeness where possible.

Gaps: This property relates to the ability to determine and document gaps in the data. Gaps may be identified due to a breakdown in system or processes, lack of available resourcing, or when data is collected for a defined time period. In some cases, 100% completeness may not be required to be fit for purpose, and this should also be noted. Known gaps should be documented so that users of the data can be aware and compensate if needed.

Representative

*Is the dataset **representative** of the conditions or scenario to which it refers?*

Representativeness relates to the relevance of data and the extent to which it meets the defined purpose which initiated its collection or creation. For data to be representative it must be indicative of the environment in which it was collected or created as well as reflective of the situation it is attempting to describe. Raw data may not always be representative, however it should be noted that this can also be achieved using analytical techniques such as weighting.

Properties

Example properties for Representativeness include:
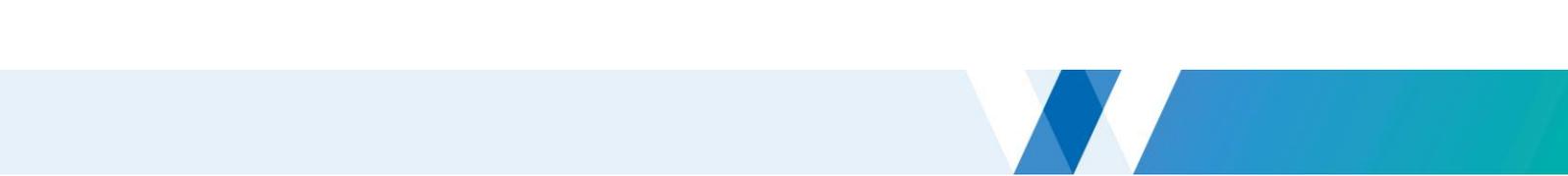
Coverage: Coverage relates to the proportion of a population or sample that have been incorrectly included or excluded. This can be discovered when comparing the actual results against expected results, response rates and totals. The coverage should be appropriate for the data in terms of population groups, geographic span and conditions, and the sample size and breadth adequate for the required use of the data. For example, a small sample of a large overall population may not yield a useful result.

Relevance: The degree to which the data reflects the real-world situation meeting the needs of users. Data with high relevance efficiently tells you what you want to know without having to sort through irrelevant data and information. The data should accurately describe the situation or environment it is intending to measure or analyse. User satisfaction should be measured periodically to ensure the data is relevant to their needs.

Timeliness/Currency

*Is the **timeliness** and **currency** of the data appropriate?*

Timeliness refers to how quickly data can be made available when required, and the delay between the reference period (period to which data refers, such as a financial year) and the release of information. Factors that may impact this include collection method and processing. Data must be discoverable, available and accessible throughout all stages of the data asset lifecycle from creation to retirement, to be available for greater internal use, external use (external partners, other government departs and researchers) and the public. If delays occur during the provision of data, currency and reliability may be impacted.

Properties

Example properties for Timeliness/Currency include:

Availability: Data must be made available as soon as possible both internally to the department as well as externally to public entities (where applicable). Delays commonly occur between the release date of the data and the reference period it spans. Data must be available quickly and frequently enough to support information needs, to meet contract terms and reporting deadlines, and to influence service or management decisions. Data used to support findings in reports should be available when the report is released.

Currency: Based on the needs of the user, if data is no longer current, it may no longer be fit for its intended use. For example, data for activities occurring within a set period of time, or where data expires or becomes irrelevant after a set date or time. Where it references specific time periods, data should be current and released before it is superseded. This may occur with data assets that are collected on a one-off basis or infrequently. Metadata describing the age, currency and expiry dates of data and information should be provided with the data asset, including collection dates, collection periods and coverage.

Collection

*What was the **collection** method and was it consistent?*

Data collection methods must be appropriate depending on the type of data being collected. For example, to collect certain data, a survey may be a more appropriate method of collection

than using data entry software. Collection must be consistent, especially if the same data is collected multiple times, or is to be compared to other data assets. (This is also covered in the Consistency dimension; however this dimension deals specifically with how the data is collected.)

Properties

Example properties for Collection include:

Data is commonly understood: As for the Accuracy dimension, an understanding of the use and meaning of data by those collecting and using it reduces risk of data collection errors. Definitions of the data itself, and of the data collection and processing methods should be made available organisation-wide to provide a common understanding. This should include what data was collected, how the data was collected, who collected the data, how the data was processed (system or person) and details of any edits or manipulation. When this is well documented, collection over time is consistent independent of the person or system used to collect the data.

Appropriateness: Data must be collected using the most appropriate method possible for the data being recorded. Some possible methods of data collection include surveys, forms, data feeds, and reporting from service providers. Collection methods must be chosen depending on the level of data quality required for analysis, and the level of risk. For example, a survey may be appropriate to collect general information from people, whereas collecting entered data might be more appropriate for analytical use. It is important to consider the associated risk, e.g. implications to collecting sensitive or personal data. The method selected should be documented.

Duplication: Data should not be duplicated in other data assets. Where duplication exists, it should be identified and managed. Documentation of a glossary of terms and context help to avoid duplication and ensure data comparison is appropriate and relevant.

Consistency

*Is the data **consistent** with related datasets, agreed standards and formats*?

Consistency of data means that the data is collected, grouped, structured and stored in a consistent and standardised way. This requires standard concepts, definitions and classifications to be implemented across departments, and agreed upon as to their meanings and interpretation.

Data must also be consistent in the context of its use. For example, data may appear similar but have different meanings or uses in different departments. Duplication, or different meanings for similar data, may result in confusion or misinterpretation of data and render such data unsuitable for comparison with related assets. Also, it may be unclear if trends are due to a true effect or due to problems with inconsistent data collection.

Properties

Example properties for Consistency include:

Comparison with similar data assets: There must be the capacity to make meaningful comparisons across multiple data assets. This is achieved through common data definitions and standards. Common data definitions should be agreed and shared across the department, and any inconsistencies should be managed.

Consistency over time: This refers to the ability to track a data asset over time. Changes may be made regarding scope, definition and collection, however, the data must remain consistent so it can be compared with previous data assets. For example, collecting data annually to track changes over time. Documentation of changes and the frequency or timing of updates must occur to determine that comparisons made are appropriate.

Documentation: Changes to data and any related processes should be well documented and traceable. Data dictionaries specifying business rules, validations, specifications, and glossary terms should be available, regularly maintained and updated with any changes made to data. For example, the definition, naming conventions, or scope of data that is collected periodically may change over time. Documentation of related information should be recorded with the data, such as time, date, reason for change, and person who made the change, as relevant to the data asset. This allows for mapping of changes between versions for traceability, which is required for retrospective analysis. Providing a Data Quality Statement - a summary of known characteristics which may affect the quality of a data asset - informs and enables prospective users of the asset to make decisions about its use.

Data is commonly understood: As for other dimensions above, documentation of data definitions, collection, groupings, and terminology helps with understanding of data, to ensure that continuity and consistency can be maintained. Where applicable, synonyms for terms should be included (e.g. "client" and "customer" may mean the same in one context but differ in another). Technical definitions may also be included where this is important (e.g. consistency in the length of a data entry field or use of the same codes or reference data).

Fit for purpose

*Is the data **fit for the purpose** of its original or intended use?*

Data is considered fit for purpose when it is appropriate for its intended use. The purpose could include decision making, policy development, service delivery, reporting, and legislative or administrative requirements.

In this context, the purpose against which the data is measured is the original intended purpose. A future use of the data may not be known at the time the original data was collected.

An understanding of who potential users might be, and their expectations of data quality, must be balanced against the original business intent of the data asset. Potential users may be consulted in the development or plan phase of a data asset to ensure that data collected meets user expectations in terms of quality and relevance.

Fitness for purpose can be subjective and therefore difficult to measure, and is a key component of data quality across all dimensions. Therefore, no specific properties are suggested for this dimension. A statement may be made against the overall data asset's fitness for purpose.

Roles and responsibilities

Good information and data governance is essential for maintaining and improving the quality of data assets. Governance structures provide accountability by defining responsibilities for the management of data assets within a department.

Various roles within a governance structure have different responsibilities. All should work to contributed to a consolidated and consistent approach to data quality within an organisation. It is important to note that tasks and responsibilities held by these roles may be delegated.

This guideline specifies responsibilities for data quality processes only, for a full description of all roles and responsibilities, please refer to _IM-STD-03 Information Management Governance Standard_.

Owner

Under the Financial Management Act, the department head or agency chief executive officer (or equivalent role) has ultimate accountability for the department's asset and risk management.

In practice, the accountable officer (owner) may delegate responsibility of information assets to a delegated owner who in turn delegates to an information custodian[3].

Delegated Owner

The delegated owner is an executive level officer responsible for specific data assets, ensuring they are accurate, current, protected, and accessible and shared where possible. The owner provides high-level direction and advice on activities and decision making to ensure data remains in line with business processes and objectives and remains fit for purpose. Owners typically are not directly involved in day-to-day management and maintenance of an asset.

The owner should ensure data quality is improved upon and maintained at a stated level of quality. Owners should encourage the development of management plans, data quality statements, and commit to supporting the implementation of improvement processes and activities.

Although the delegated owner may choose to delegate tasks to another executive level or custodian, overall accountability still remains with them.

Custodian

The custodian is appointed by the owner and is responsible for the day-to-day management of the data asset. Custodians should have strong business knowledge of the asset and are typically a subject matter expert in their chosen area. Custodians guide how information is managed in their area ensuring legislative and administrative compliance. Custodians are

---

[3] This information has been taken directly from IM-GUIDE-06 Information Management Governance Guideline
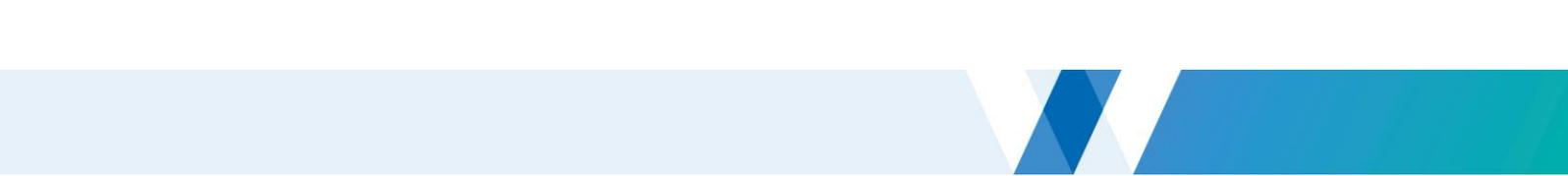
responsible for ensuring data asset quality is in line with user and business needs and fit for purpose.

Data quality responsibilities include:

- assisting and supporting the development of management plans, assessments and development of data quality statements
- ensuring deviations from current plans and processes are documented
- confirming the status of data quality statements for all critical and shared data assets in the department's information asset register
- ensuring processes meet internal data quality standards, guidelines and policies.

Administrator

The Administrator is the role with detailed knowledge of the data with a hands-on approach to data management. Administrators are responsible for day-to-day processes, which ensure data assets remain of high quality and meet legislative and statutory obligations set out by the custodian.

Data quality responsibilities include:

- development and execution of management plans
- development of data quality statements for all critical and shared data assets
- conducting data quality assessments using the data quality assessment tool.

Data Quality Statement

What is a data quality statement?

Data quality statements (statements) summarise known characteristics which may affect the quality of a data asset, to inform and enable prospective users of the asset to make decisions about its use.

Statements should highlight both strengths and weaknesses of an asset to assist users in determining whether an asset is fit for their intended purpose. They also form part of a disclaimer around the use of a data asset. For example, a statement outlining the quality of data relating to air quality (gathered for a specific purpose) helps to protect the provider of that data asset if the data quality is found to be insufficient for providing information to asthma sufferers about air pollution levels.

Statements must be recorded for all critical data assets held by the government, and for shared data assets where the risk has been assessed as higher as a result of sharing, and these statements should be discoverable to users to assist in interpretation of data use.

Data custodians must ensure completion of data quality statements for all critical and selected shared data assets for which they have responsibility and record the status in the department's
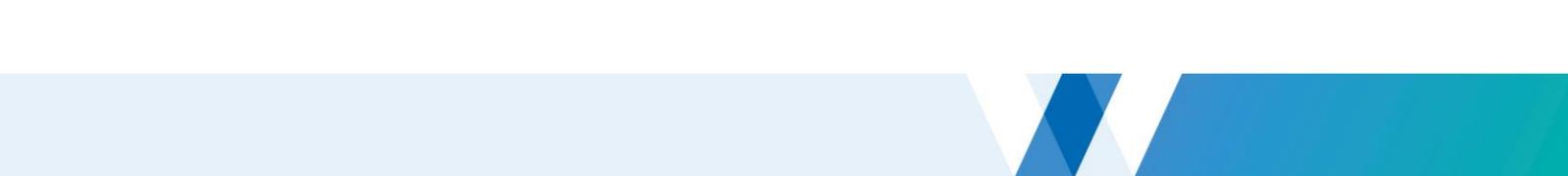
information asset register. Departments can define the quality of a data asset within their information asset register or their data management platform, as long as a data quality statement can be produced when required. Although the Administrator will be responsible for completion of tasks, compliance responsibility remains with the Custodian.

How to complete a data quality statement

A data quality statement template has been developed to assist departments to provide information about the quality of data. Use of this template is **optional**. This template will allow consistent capturing of data quality and related information across data assets and over time.

The data quality dimensions are listed in the template with various properties for consideration.

1. **Review and assess the data asset:** To provide a data quality statement, the data asset should first be assessed against all data quality dimensions and properties using the template provided or an equivalent tool.

2. **Create the data quality statement:** Using the statement template provided, create a data quality statement for the data asset.

3. **Enter details for each dimension:** Enter details of the assessment findings as a brief statement per dimension to summarise the data quality and any specific details about that dimension.

4. **Provide summary information:** Summary details about the data asset should be entered in the table provided, as per the instruction text.

Data quality statements should then be approved and endorsed by the Data Owner and reviewed and/or updated whenever a data quality assessment is conducted, and the assessed dimensions change.

Data Quality Management Plan

A Data Quality Management Plan (management plan) is required to be created for all critical data assets, and for shared data assets that have been assessed with a higher level of risk as a result of sharing.

The plan assists departments to develop, measure and maintain data quality by setting actions and tracking assessment results over time. It provides a mechanism to summarise results, and to collect references to more detailed test plans and other documentation. It also enables consistent reporting of data quality issues and tracking of the progress of improvement initiatives across an organisation.

Plans should be established at the data asset level, in accordance with the Enterprise Data Management Plan. Plans should be realistically implementable in terms of timeframe, resource allocation, current state data quality management processes and current capability levels of staff.

This plan/template is **not mandatory** nor is it intended to replace any more detailed plans or assessment tools already in place. The intention is for the assessment and planning step to take place. The plan template is referenced here as a guide.

How to complete a data quality management plan

A management plan should be regularly maintained for all critical data assets. Note: a management plan is not required where the data asset is closed (no longer being collected or created) or a one-off (data only collected or created once).

The basic steps for completing a management plan are as follows.

Figure 4 - Data quality management process

1. **Define the properties:** The first step is to define the properties within each dimension and enter each into the Assessment template. Note that plans can also highlight areas of strength as well as issues to be addressed.

2. **Perform the assessment:** An initial assessment is done to provide a baseline of the level of quality pertaining to each dimension, after which subsequent reviews can be performed at intervals (six months is suggested, but this will depend on factors within each department and the risk associated with the data assets being assessed). Each review is intended to show where improvements have been made, in particular as a result of improvement activities since the previous review. For each review, enter the date of the review, and in that column enter a rating against each property. Refer to the 'Data quality assessment' section below for more information regarding this process and the ratings.

3. **Improvement plan:** Identify areas for improvement. Define a target state that is achievable within the resource levels and timeframe of the plan, and measures of success. Consultation may need to occur with users of the data, additional subject matter experts and information management professionals.

4. **Actions:** Based on the identified areas for improvement, provide details of actions that will address each issue and increase data quality. Note that some actions may fix multiple issues. For example, a data cleansing activity may address the accuracy of data as well as gaps. In such cases, an already listed action can be referenced for subsequent issues. Refer to 'Data quality improvement activities' below for further details about actions.

5. **Review periodically:** The data asset should be reviewed and reassessed at regular intervals to ensure that quality is maintained and to measure improvements. Action plans should be reviewed and/or updated at least annually by the Data Custodian.

6. **More detailed information:** The plan contains only a summary level of information. Details of more complex actions or repeatable test plans that can be executed for consistent measurement can be referenced by the plan.

The following items are currently defined in the template (note: use of this template is **not mandatory**):

- Date of assessment
- Assessed dimensions
- Assessed properties
- Review result (assessment scale rating)
- Area(s) for improvement
- Objective describing how the action will improve/resolve the issue
- Planned actions (improvement strategies/activities designed to address data quality issues).

Departments requiring more detailed plans may also document the following items:

- Start date of the action plan
- Name and position of the Data Owner of the data asset
- Name and position of the Data Custodian of the data asset
- Overall quality improvement objectives (separate to activities)
- People/positions of responsibility for implementation
- Timeframe for completion of the plan
- Measurement technique of improvements
- Next review date
- References to other documents and more detailed test plans.

Data quality assessment

The management plan contains an Assessment tool (the tool) that can be used to assess the current environment of data quality. Data quality issues should first be identified before targeted activities and initiatives can be implemented. The tool is initially used to develop a baseline quality assessment before further tracking ongoing progress. Issues identified in this assessment will be used to inform improvement activities and initiatives recorded in the asset's data quality improvement action plan.

The tool incorporates all seven data quality dimensions to be applied consistently across all critical and shared data assets. Dimensions are sub-divided into properties that describe practices and/or that inform data quality. An overall rating is given to each dimension at the completion of the assessment.

Properties are first defined by planning the assessment before they can be rated and analysed. Results from the assessment can then be used in the formation of an improvement action plan.
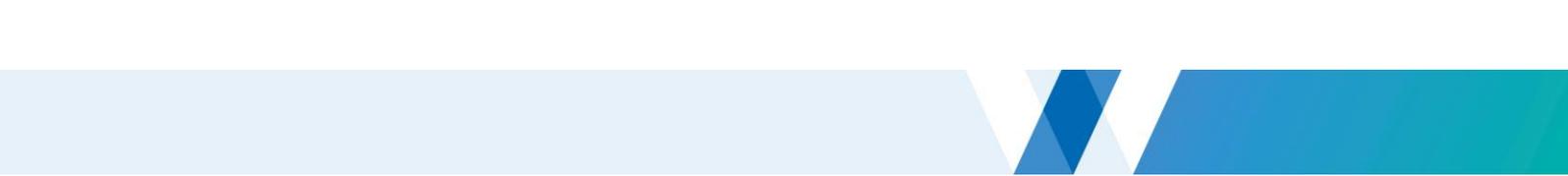
Users should aim to complete all properties as part of the assessment. If properties are not applicable to the data asset being evaluated, they may be removed from the tool, replaced with more relevant items, or left blank. Further properties can be added by following the instructions in the Introduction sheet.

The template merely provides a starting point to ensure adequate comparability of ratings across data assets, and is **not mandatory**. Any tool used for this purpose should be standardised across an organisation, limiting changes per asset assessment.

What is the purpose of the assessment?

The assessment is a tool for departments to assess and record the current level of data quality, specifically:

- establish a baseline assessment for current data quality practices
- track progress over time
- highlight areas needing improvement
- inform future plans and improvement initiatives.

Who should complete the assessment?

Data Custodians should ensure completion of the assessment and support the Data Administrator who will undertake the work. Roles may need to consult with various data users to ensure data is fit for purpose as well as information management professionals for further guidance.

What are the benefits of the assessment?

By completing this assessment, Data Custodians will be able to:

- easily identify strengths and weakness
- prioritise areas for improvement
- effectively manage operational risk arising from poor data quality
- effectively manage resource allocation relating to data quality improvement initiatives.

What does the assessment involve?

The assessment requires departments to self-assess current data quality practices across the seven dimensions of data quality. Each dimension is further divided into sub-sections (properties). These properties define certain characteristics to be considered in the evaluation of an individual dimension, and are the most granular level by which quality is defined in this template, with each defining a singular, testable measure of quality. Figure 2 below shows the hierarchy of dimensions, properties and descriptions in the template.

Figure 5: Data quality dimension levels
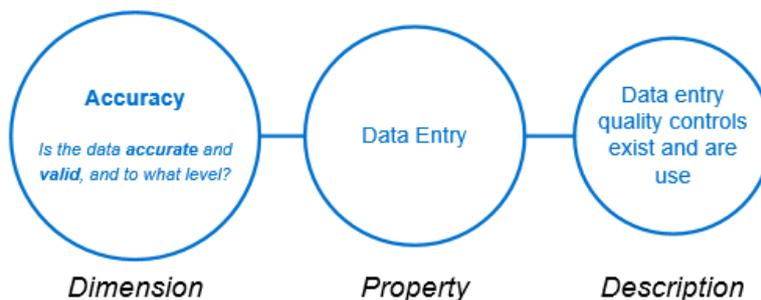
When should the assessment be undertaken?

It is recommended that the assessment be completed annually. Departments may wish to conduct assessments within shorter timeframes if desired. Completing the assessment regularly allows for departments to compare and track progress of all critical and shared data assets.

How to complete the assessment

Custodians and Administrators should first define all measurable aspects of data quality before conducting the assessment. The assessment will be conducted using the Assessment sheet in the management plan template.

Assessors should work through the Assessment sheet of the tool evaluating each property on a rating scale of 1-3 (see Figure 3 below). Departments can define the specifics for each level per property if desired – i.e. what constitutes a rating of 1, 2 or 3 for a particular property.

Figure 6: Data Quality Assessment scale

Once all properties are individually scored, totals are consolidated to provide a single average score for each dimension in the Results sheet.

These scores can be weighted if desired. A dimension that is considered more critical can be weighted with higher importance (which actually lowers the score so that improvement of this area is prioritised), and less important areas given lower weightings. Leaving the default settings under the 'Weighting' column will result in no weighting being applied.

Next to the 'raw' score, a numerical 'weighted' value between zero and three will appear for each dimension, colour coded according to the assessment scale to create a heat map within the table. The weighted scores are also shown in both a summary bar chart for all dimensions, and individual dimension line graphs to show progress over time.

How to interpret the results

Once the assessment has been performed and the Results sheet has been populated for each dimension, the results can then be studied. Definitions for the assessment scale ratings can be seen below in Table 1.

Table 1: Data Quality Assessment Tool Assessment Scale definitions

Rating	Meaning
1	**Low (needs attention)**: A rating of one highlights areas of data quality that require attention. This means that the property is applicable to the data asset however, the organisation has not met the requirement. These dimensions should be targeted for improvement in the Data Quality Management Plan management plan as a priority.
2	**Medium (adequate):** A rating of 2 identified areas that have been assessed as adequate by the assessment. This means the requirement stated by the property has been partially met however, there is still room for improvement. These areas should still be targeted for improvement in the management plan.
3	**High (excellent):** Dimensions with a rating of 3 identify areas that have good data quality. This means the organisation has met the requirements of the assessed property and dimensions well. Although this is the highest score in the assessment, organisations should aim to maintain this level of quality by incorporating maintenance activities into management plans.

Based on the results of the assessment, an Improvement Action Plan can be developed. Areas that have been identified as insufficient should be directly targeted by this plan. Results will give departments an indication of the effort required to improve data quality organisation wide. It is up to each department to determine if similar or related datasets can be grouped for both assessment and treatment (e.g. those with similar quality issues).

This assessment should be conducted regularly. A trend analysis of assessments performed using the heat map within the tool to track progress and identify stagnant areas over time.

Data quality improvement activities

Quality improvement activities can be used to monitor, analyse and improve the quality of data assets. Data quality issues should first be identified before they can be rectified.

While the tangible benefits of these activities are not always immediately evident, these activities are important to undertake to improve the value of the data asset. By improving data quality, the data assets become more usable both internally and if shared, by others, thereby increasing in value. Reducing rework on data assets that are low quality by correcting issues or automating processing will reduce resource costs and allow staff to focus on using the data

asset rather than spending time correcting it manually before it can be used. Risk is also reduced by having higher quality data to enable decision making.

When planning improvement activities, it is important to be realistic. If baseline data quality is extremely poor, foundational activities (such as those discussed below) may need to be planned and completed prior to the implementation of more complex activities. If specific issues have been identified relating to an individual data asset, specific activities targeting that data asset will need to be developed.

Foundational data quality improvement activities

Foundational activities establish a core basis which can be built upon by additional initiatives. Examples of foundational improvement activities can include:

▪ **Data dictionaries:** Data dictionaries are a reference of standardised concepts including data definitions, business rules, validations and allowable formats for data which should be applied. Implementation of data dictionaries creates a common understanding of data items which can be applied consistently by data suppliers.

▪ **Implementation of existing whole of Victorian Government (WOVG) standards:** Existing standards and guidelines issued by Enterprise Solutions, Department of Premier and Cabinet provide high-level guidance on the management of government information and data. As a minimum they are to be implemented by all government departments and Victoria Police. Together the standards provide a consolidated approach to information and data management across government, whilst enhancing capabilities of individual departments.

▪ **Establishment of an information asset register:** Significant and critical data assets should be identified and registered in the department's information asset register. This facilitates discovery, accessibility, protection, and the management of assets throughout their lifecycles. Information related to the quality and completeness of data assets may be recorded in the register with links to documented underlying metadata. By establishing a register and recording data quality attributes, assets with a low level of quality are more discoverable. (For further use of information asset registers, see Complex data quality improvement activities below.)

▪ **Defining a governance model:** Under the *Information Management Governance Standard*[4] departments are required to document a custodianship model, defining responsibilities for the management of data. Implementing a defined governance framework provides accountability for data quality analysis and improvement. The accountable person should champion and support initiatives highlighting the importance of data and increasing its quality across the department.

▪ **Awareness training:** Staff may undergo awareness training relating to the importance of data quality. High-level educational material can be developed that is applicable organisation wide. Training scope should be broad, as specific issues are addressed in more targeted training (see Specific activities below).

Topics may include:

– why data quality is important

– what are the uses of data in your organisation

---

[4] IM-STD-03 Information Management Governance Standard

– implications of poor data quality.

▪ **Feedback:** Feedback should be received from data users as they can provide information on whether data is fit for purpose, useful, accurate and timely available.

Specific activities

Specific activities target exact issues or problems which have an effect on data quality of a specific data asset. Activities can address issues relating to people, process, technology or all of the above.

People

▪ **Capability training:** Staff may be required to undergo training that is specific to data management capabilities of a specific data asset or type of data. Training may be in the form of online learning, classroom workshops or mentoring.

Training may include:

– relevant legislation and standards relating to the data

– available resources to assist in understanding the data, e.g. data dictionaries

– business processes relating to management of the data asset e.g. data collection, analysis or interpretation.

▪ **Data capability statements or definitions:** Data capability statements or definitions defining competencies and skills required for the management of a specific asset or type of data. Development and application of statements for complex data may prevent issues relating to accuracy of collection, data processing and analysis as only staff that are assessed to have the appropriate capabilities are responsible for data management.

Process

▪ **Identify processes:** Data Custodians and Administrators should work with staff and suppliers to document current business processes from data source through to data use. This can assist in identifying possible changes, combating inefficiencies or quality issues.

▪ **Analyse sequence of events:** This involves mapping a series of events to identify possible inefficiencies such as data duplication and processing bottlenecks. Areas of inefficiency can then be prioritised for improvement.

Technology

Although WOVG standards do not require or mandate implementation of specific data management systems, departments should ensure that current technology/ies:

▪ review current business systems to determine whether they are fit for purpose in terms of data management capability

▪ automate validation processes where possible to increase quality of data

▪ implement data management standards where possible

▪ ensure that software has appropriate data entry controls, such as field validations, length and type restrictions, and reference lists to help with quality of data input.

Specialist resources may be required to undertake activities aimed at technology improvements.

Complex data quality improvement activities

Complex data quality improvement activities should be implemented at organisations that have a strong existing data quality and data management capability, and should be undertaken by those with specialist expertise in the area. Examples of these activities are below.

Data profiling

Data profiling is the process of uncovering defects and anomalies in the data. It allows organisations to gain a complete insight into the quality of their data, as well as identify issues with people, processes and technology that contribute to data quality, and measure the magnitude of the problem. This process analyses data for aspects such as:

- Correctness
- Completeness
- Uniqueness
- Consistency

Data profiling can assist organisations in gaining an insight into possible operational risks. Results can be utilised for tactical and strategic planning, for both short-term data rectification and long-term tactics to improve overall quality.

Data defects should first be identified through data profiling to gain an understanding of the data before they can be rectified through data cleansing.

Data profiling methods:

- Tools and software are available from the market to assist in this process, but not currently centrally available within government. This can be a costly exercise. Specialised expertise is required to ensure that value is maximised.

Data cleansing

Data cleansing is the process of data rectification after records containing 'dirty data' have been identified by data profiling. One example is correction of data that doesn't meet format or standard requirements, such as an address or date in the wrong format.

An important aspect to correcting and cleansing data is to ensure that the issue isn't repeated through data entry errors or practices, and that any data once corrected is also fixed at the source i.e. in the original database. If data is not corrected at the original entry point or source, there is a high risk that it will need to be corrected again when that data is derived or used. When dealing with agencies or third-party sources for data collection that are beyond the departments' control, it is important to raise awareness by reporting such errors or collection problems to avoid rework. Where this is not possible, processing should be put in place to rectify known issues upon receipt of collected data. This will likely need specialised expertise.

Data cleansing methods

- Data cleansing is often a time-consuming labour-intensive process. However, tools and software from the market are available to assist, as are specialised resources.

Information Asset Register

An information asset register is used to describe all significant information holdings held by a department or agency[5]. Registers can be used to efficiently identify critical and shared information and data assets, and can be accessible within the entire department. Information asset registers may incorporate a number of attributes including legislative requirements, governance roles, information lifecycle, security and access requirements, as well as information about the quality of the data asset.

The Sample Information Asset Register (IAR) Template[6] produced by OVIC details minimum fields that an organisation needs to register for each information asset, for the Victorian Protective Data Security Framework, core and supplementary attributes for inclusion.

The presence or absence of a data quality statement should be recorded in the department's information asset register. This is a defined statement about the quality of the data and purposes for which it can be used. The location of the statement should also be recorded in the register, typically in the form of a file path name or URL.

All government departments are required to keep an individual information asset register as well as contribute to a WOVG register once developed. Attributes describing data quality can be recorded in the WOVG information asset register to inform other departments, promoting reuse.

Further information

For further information regarding this standard, please contact Enterprise Solutions, Department of Premier and Cabinet, at: enterprisesolutions@dpc.vic.gov.au.

---

[5] An Information Asset Register standard is currently under development

[6] Sample Information Asset Register (IAR) Template