

How to record data quality in the Aristotle Metadata Registry

User story: I have some data sets, some of which are of good quality and others that are of average quality. I need to link this information to the data sets to inform users about whether the data is fit for purpose. These data sets should be shared but not before an assessment has been made available about its quality.

KEY WORDS: Quality Statement, Data Quality Check, Data Quality Management

What metadata object should I use: Quality Statements

The type of item that can be used in this instance is a **Quality Statement** in the Aristotle Metadata Registry. An assessment of the quality of data in a data set can be made using a template with headings that permit data quality to be measured.

- A **Quality Statement** is a summary of the key issues to be aware of when using the data. It covers seven categories to be taken into consideration, Institutional Environment, Relevance, Timeliness, Accuracy, Coherence, Interpretability and Accessibility.

One way to understand a **Quality Statement** in an Aristotle Metadata Registry is by using the analogy of swimming pool water quality testing.

The quality of pool water can be variable across different pools. The water may look clear, but it may not be safe to swim in. The keys to maintaining water quality in a swimming pool rely on filtration, chlorination, pH level, water hardness and more. A set of criteria and tests will assist the pool user in understanding quality of the water for swimming. The results of the water quality test list each of the criteria and what falls within the range for safe use. The swimming pool water quality can be understood through the results of the quality tests.

Most importantly, the quality of the water can be checked before anyone swims in the pool.

Within the Aristotle Metadata Registry, a **Quality Statement** is like the pool water quality testing kit for data holdings. It includes all the headings or keys that should exist for good data quality and the results of the tests measured against each of the headings. So, a determination can be made about whether the data quality is safe to use before the analyst uses it or what conditions the analyst needs to consider prior to using the data.

A **Quality Statement** is a brief summary of the key issues to be aware of when using the data. When an organisation assesses the quality of its data against these 7 key headings in a **Quality Statement** it increases levels of trust from the perspective of the stakeholders and users of the data.

Based on Australian Bureau of Statistics [ABS Data Quality Statement Checklist](#) it consists of these dimensions of data quality;

- **Institutional Environment** - This is a description of the origin of the data collection and what arrangements or governance is in place. This information gives context about the reliability and validity of the Information Asset.
- **Timeliness** - How long ago was the information collected? If the data is old, is it part of a time series where changes over time can be viewed. Is there a lag between when the data was received and when it was published?
- **Accessibility** - Is the information easily obtained? Is it available through a web site or is it obtainable by request? If permission is required to access the data is it clear how to obtain these permissions.
- **Interpretability** - Is there an abundance of information available about this subject. Is there sufficient metadata to support the data including concepts and classifications?
- **Relevance** - An assessment of whether the collection was fit for purpose. What was the target population and whether the collection is representative of that population? Was the response to the collection of the data sufficient to provide reliable information?
- **Accuracy** - Whether the data describes what was to be measured correctly. The types and number of errors in the data should be considered when rating this category. If it is known that there were errors in the response or sample or other errors in accuracy of the data, it should be noted.
- **Coherence** - Does this collection adhere to known standards about the subject, is the methodology known and understood widely.

Examples of Quality Statements

The **Quality Statement** clearly shows the definition, registration history and related items. It can be used to assess the quality of a routinely collected data set of a large population of people such as the ABS birth registration data.

The screenshot shows a web interface for the Aristotle Demo Organisation. The breadcrumb trail is: Aristotle Demo Organisation / Metadata / Quality Statement / ABS birth registration data. The page title is "ABS birth registration data". There are tabs for "Item", "History", and "Graphs", and a "Download" button. The main content is divided into two columns. The left column contains a "Definition" section with a paragraph about the Census and Statistics Act 1905, and a "Data Quality" section with a "Timeliness" subsection. The right column contains a metadata box titled "ABS birth registration data" with fields for Type, Identifiers, Stewardship Organisation, Endorsed by, Last updated, Created, Supersedes, and Superseded by.

Definition

This collection is conducted under the Census and Statistics Act 1905. For information on the institutional environment of the ABS, including the legislative obligations of the ABS, financing and governance arrangements, and mechanisms for scrutiny of ABS operations, see ABS Institutional Environment.

Data Quality

Timeliness:
Births records are provided electronically to the ABS by individual Registrars on a monthly basis for compilation into aggregate statistics on a quarterly and annual basis.

Quarterly estimates of births on a preliminary basis are published five to six months after the reference period in Australian Demographic Statistics (cat. no. 3101.0), and revised 21 months after the end of each financial year. Annual estimates on a year of registration basis are published within ten months of the end of the reference year in Births, Australia (cat. no. 3301.0).

ABS birth registration data

Type: Quality Statement

Identifiers: UUID - ea504d8e-dd6c-11e8-8f47-86e863b447ab
meteor/449204/

Stewardship Organisation: Aristotle Demo Organisation

Endorsed by:
• Indicator governance board : Recorded on 31st October 2018
[View registration history](#)

Last updated: 2 years ago
Created: 1st November 2018

Supersedes:
• None

Superseded by:
• None

View the ABS birth registration Quality Statement online at:

<https://aristotle.cloud/item/603578/qualitystatement/abs-birth-registration-data>

A Quality Statement may also be an assessment of the quality of the data collected for a Data Collection such as the Coronavirus Disease (COVID-19) Situation Report Daily Surveillance Data

The screenshot shows a web interface for Aristotle Cloud Services Australia. The breadcrumb trail is: Aristotle Cloud Services Australia / Metadata / Quality Statement / Coronavirus Disease (COVID-19) Situation Report Daily Surveillance Data; Quality Statement. The page title is "Coronavirus Disease (COVID-19) Situation Report Daily Surveillance Data; Quality Statement". There are tabs for "Item", "History", and "Graphs", and a "Download" button. The main content is divided into two columns. The left column contains a "Definition" section with a paragraph about the World Health Organization (WHO), and a "Data Quality" section with an "Institutional Environment" subsection. The right column contains a metadata box titled "Coronavirus Disease (COVID-19) Situation Report Daily Surveillance Data; Quality Statement" with fields for Type, Identifiers, Stewardship Organisation, Endorsed by, Last updated, Created, Supersedes, and Superseded by.

Definition

This collection is conducted under by the World Health Organization (WHO). See Institutional Environment section for information on WHO.

Data Quality

Institutional Environment:
The World Health Assembly is the decision-making body of WHO. It is attended by delegations from all WHO Member States and focuses on a specific health agenda prepared by the Executive Board.

The objective of the World Health Organization shall be the attainment by all peoples of the highest possible level of health. In particular for data collection, to study and report on, in co-operation with other specialized agencies where necessary, administrative and social techniques affecting public health and medical care from preventive and curative points of view, including hospital services and social security.

An approved Case-based reporting form was developed by WHO in February 2020 to be used by all participating countries for reporting COVID-19 cases.

Timeliness:
Data is collected daily and reported and published in the situation reports weekly. These reports have been available since Feb. 2020 and will continue for the duration of the pandemic. The data is made available daily via multiple dashboards and data is supplied by CSV for researchers to utilise.

Coronavirus Disease (COVID-19) Situation Report Daily Surveillance Data; Quality Statement

Type: Quality Statement

Identifiers: UUID - fc89e620-1630-11ea-a313-0a27eadeedfa

Stewardship Organisation: Aristotle Cloud Services Australia

Endorsed by:
• ACSA Research Data Stewards Committee : Recorded on 8th September 2020
[View registration history](#)

Last updated: a few seconds ago
Created: 14th September 2020

Supersedes:
• None

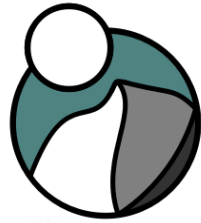
Superseded by:
• None

Collections:
• None

View the Coronavirus Disease (COVID-19) Situation Report Daily Surveillance Data Quality Statement online at:

<https://aristotle.cloud/item/605068/qualitystatement/coronavirus-disease-covid-19-situation-report-dail>

For more information on Aristotle.Cloud or the Aristotle Metadata Registry, visit us online at www.aristotlemetadata.com, contact the Aristotle team at hello@aristotlemetadata.com or visit our community forums at <https://community.aristotlemetadata.com>



ARISTOTLE
metadata registry