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 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.

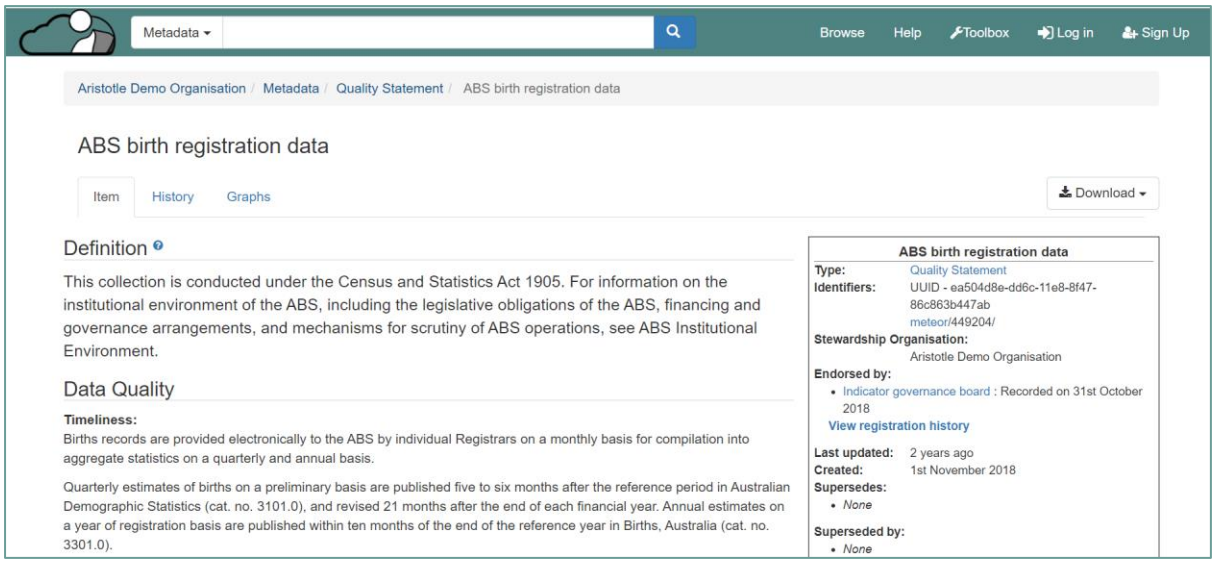
A **Quality Statement** is a brief summary of the key issues to be aware of when using the data. When an organisation assesses the quality of its data against these 7 key headings in a **Quality Statement** it increases levels of trust from the perspective of the stakeholders and users of the data.

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.

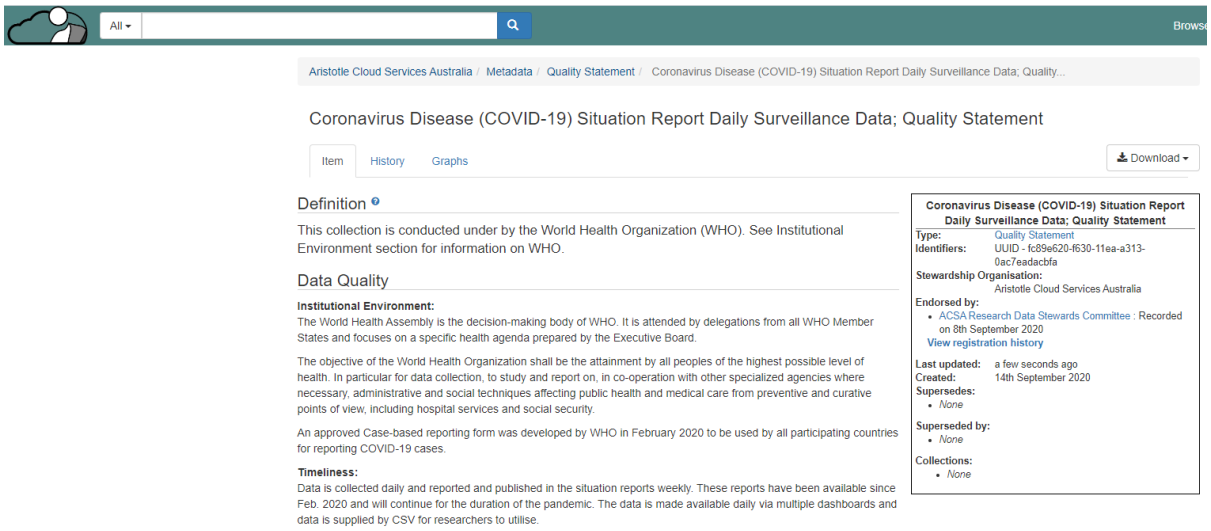
Examples of Quality Statements

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.



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



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