



# Data and Metadata Association

21 July 2009

Narrisa Gilbert

[narrisa.gilbert@abs.gov.au](mailto:narrisa.gilbert@abs.gov.au)

Melissa Gare

[m.gare@abs.gov.au](mailto:m.gare@abs.gov.au)

# Questions to cover

- Level of detail? – collection, product, both?
- How do we write a quality statement when various data sources get aggregated into our data warehouse?
- COAG Quality Statements

# ABS Data Quality Framework!

- History of ABS DQF
- Brief discussion of 7 dimensions
- Examples of how to use it

# History of ABS DQF

- Adopted from two frameworks – Statistics Canada and Eurostat
- Originally six dimensions – Relevance, Timeliness, Accuracy, Coherence, Interpretability, Accessibility
- Additional dimension (2007): Institutional Environment

## Dimension 1: Institutional Environment

- The institutional and organisational factors that have a significant influence on the effectiveness and credibility of the agency producing statistics
- Questions to assess:
  - Which organisation(s) have supplied the data? What sort of organisation is this?
  - Is statistical confidentiality guaranteed, and if so, under what authority?
  - Impartiality / Independence
    - Advertised release dates in advance

## Dimension 2: Relevance

- How well the statistical release meets the needs of users in terms of the concepts(s) measured, and the populations(s) represented.
- Questions to assess:
  - About whom, or what, were the data collected?
  - How useful are these data at small levels of geography?
  - Does this data source provide all the relevant items or variables of interest? Does the population presented by the data match the data need?

## Dimension 3: Timeliness

- The delay between the reference period (to which the data pertains) and the date on which the data becomes available; and
- the delay between advertised date and the date at which the data become available (i.e., the actual release date).
- Questions to assess:
  - Are there likely to be subsequent surveys or data collection issues for this topic?
  - What is the gap between the reference period, the time when the data were collected, and the time when the data became available?

## Dimension 4: Accuracy

- Refers to the degree to which the data correctly describe the phenomena they were designed to measure.
- Questions to assess:
  - What controls or checks are in place to address data entry error, recognition error or editing error?
  - Are there particular questions which are hard to understand and to which respondents may provide an inaccurate response?
  - Have the data been adjusted in any way to account for non-response?

## Dimension 5: Coherence

- Refers to 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.
- Questions to assess:
  - To what extent can a user meaningfully compare several data items within this collection?
  - Have these data been confronted with other data sources, and are the messages consistent from all data sources?

## Dimension 6: Interpretability

- Refers to the availability of information to help provide insight into the data
- Questions to assess:
  - Are terms used in the data release which are confusing or ambiguous for a user?
  - Is there information available to help the user gauge the potential magnitude of error in the data?
  - To what extent can a user of the release or dataset find supporting information about the data to enable improved interpretation?

## Dimension 7: Accessibility

- Refers to the ease of access to data by users
- Questions to assess:
  - How easily can a user obtain this information? Is it publicly available?
  - What range of products are available, and what are their costs?
  - Are the data available in suitable formats?

# Uses of the ABS DQF

- A data user:
  - Defining a data need
  - Assessing a dataset in the context of a data need (includes idea of assessing a quality statement)
- A data producer:
  - Assess the quality of the data produced (Creation of a quality statement)
  - Design a statistical collection or product which is fit for purpose (not covered today)

# Example: A User's Perspective

- Data Need: Number of children at school in Australia.
- ABS DQF: What questions would YOU as a **user** want to ask to get clarification for your data need and to also determine what data source best fits your need?

# User: Institutional Environment

- Who (organisation) collects this information?
- Is it a private, public, not for profit etc. organisation?
- What authority/legislation was the data collected? E.g. is it mandatory?

# User: Relevance

- What age group(s) am I interested in?
- Am I interested in school years rather than age?
- What level of geography do I require? E.g. Australia, State level, Local level?
- What type of schools am I interested in? E.g. public, private, primary, secondary, university
- Am I interested in a sub-population E.g. Indigenous, Regional Data

# User: Timeliness

- What is the year/time frame am I after?
- When is the data collected?
- When do I need the data by?

# User: Accuracy

- What issues are associated with the particular method of collection?
- Is all the data collected in the same way? E.g. is there more than one form? Is it a mix of Survey and Administrative record?
- Has imputation been used on this data for missing values/non-answers?
- How was the data collected? E.g. Personal interview, report taken.

# User: Coherence

- Is this data comparable over time with itself?  
E.g. can I get a 'time series'?
- Are other data sources available for comparison with this source?
- Has there been a change to any definitions?
- Has there been a change to laws that may influence the starting school age / school retention age?
- What about State comparison?
- What about international comparison?

# User: Interpretability

- What standard classifications have been used (if any)?
- Is there supporting information to explain concepts, methodologies?

# User: Accessibility

- What format(s) are the data in?
- What isn't available? E.g. level of detail not obtainable?
- How much is it going to cost me?

# Clarification of need

- The above list of questions will help a User clarify what data they are after.
- The above list of questions will also help a User determine where there are data gaps i.e. data not available to meet all specifications/needs

# Uses of the ABS DQF

- A data user:
  - Defining a data need
  - Assessing a data set in the context of a data need (includes idea of assessing a quality statement)
- A data producer:
  - Assess the quality of the data produced (Creation of a quality statement)
  - Design a statistical collection or product which is fit for purpose (not covered today)

# Producer perspective

- What are the most important things to tell a User about your data?
- Use the ABS Data Quality Framework with the User in mind.

# Further work on DQF

- DQF will continually be revised (and has been already)
- Looking to develop additional support information to assist users
  - Templates of quality statements
  - Case studies
  - Support/resource material

# Quality Declarations

(very similar to Quality Statement idea)

- Designed for web
- Links on web page to other information
- Use a layered approach
- Top level: key quality issues by quality dimension
- Second level: optional; more detailed information
- Organise information by seven dimensions of quality

# Quality Declarations

- Content of ABS Quality Declarations not specified – 7 dimensions were mandated, but content within each dimension was left up to the discretion of the author areas.
- Quality Declarations for all statistical releases / products including:
  - Compendia publications
  - Confidentialised Unit Record Files (CURFs)
  - Publications / spreadsheets

# For more information...

- You can directly access the DQF on the ABS website [www.abs.gov.au](http://www.abs.gov.au)
  - Catalogue 1520.0 ABS Data Quality Framework (May 2009)
- Examples of ABS quality declarations
  - [National Accounts: 5206.0](#)
  - [Labour Force: 6202.0](#)
  - [Consumer Price Index: 6401.0](#)

# For more information...

- The Methodology and Data Management Division in ABS is happy to provide information and support
  - Narrisa Gilbert (02) 6252 5283 (Canberra) or at [narrisa.gilbert@abs.gov.au](mailto:narrisa.gilbert@abs.gov.au)
  - Melissa Gare (02) 6252 7147 (Canberra) or at [m.gare@abs.gov.au](mailto:m.gare@abs.gov.au)