

## HINTS & TIPS Defining A Data Need

This document provides a list of questions to be considered when thinking about obtaining data. The list is not exhaustive or prescriptive, merely a suggestion of issues to consider in order to specify your data need more completely using the seven dimensions of the ABS Data Quality Framework. This template can be used to help you make a judgement about the overall fitness for purpose of a particular data item, or collection of data items, in the context of a data need.

ABS Data Quality Framework Dimension	Questions to consider
<p><b>Institutional Environment</b> refers to the institutional and organisational factors which may have a significant influence on the effectiveness and credibility of the agency producing statistics.</p>	<ul style="list-style-type: none"> <li>• From what type of organisation is it acceptable to obtain data? e.g. Private, public, non-profit, Professional leader/expert in topic etc.</li> <li>• Do we care how the data has been collected? e.g. Police report (so individual does not report on themselves), self assessment, personal interview, independent reviewer etc.</li> </ul>
<p><b>Relevance</b> refers to how well the statistical product or release meets the needs of users in terms of the concept(s) measured, and the population(s) represented.</p>	<ul style="list-style-type: none"> <li>• What population are we interested in?               <ul style="list-style-type: none"> <li>◦ If the data we are thinking of using has some exclusions from this population (e.g. the population may exclude people in hospitals or nursing homes, or prisons, or a particular age group, or people who don't have an Australian Business Number, or people who don't earn over a particular amount) is it still okay for our purposes?</li> </ul> </li> <li>• What geographic level of detail do we require?</li> <li>• What reference period are we interested in the data for?</li> <li>• What are the main data items / outputs we require?</li> <li>• What sort of analysis do we wish to conduct on the data?               <ul style="list-style-type: none"> <li>◦ Do we require group/range or individual/continuous variables?</li> <li>◦ Do we need grouped/range information for income (e.g. \$0 - \$5000), or actual income responses (e.g. \$1230)?</li> </ul> </li> </ul>



ABS Data Quality Framework Dimension	Questions to consider
<p><b>Timeliness</b></p> <p>refers to the delay between the reference period (to which the data pertains) and the date on which the data becomes available; and the delay between the advertised date and the date at which the data become available (i.e., the actual release date).</p>	<ul style="list-style-type: none"> <li>• When do we need the data?</li> <li>• Will we want to get a time series of this information?</li> <li>• What is an acceptable delay between the reference period and the availability of the data for our purposes?</li> </ul>
<p><b>Accuracy</b></p> <p>refers to the degree to which the data correctly describes the phenomenon they were designed to measure.</p>	<ul style="list-style-type: none"> <li>• Do we have an acceptable level of quality that is required for the data we require? e.g. What is the relative standard error on an estimate that is acceptable?</li> <li>• Do we want to include data that has been imputed*? <ul style="list-style-type: none"> <li>◦ If so, what is the maximum percent that can be imputed for any given data item?</li> </ul> </li> <li>• Do we want to include data that has been edit adjusted** (e.g. changing responses by fixing typo, misunderstanding questions)?</li> <li>• Do we want original data or seasonally adjusted or trend data?</li> <li>• How big does the data source need to be? e.g. A census, a specified minimum sample/collection size etc.</li> <li>• What is our minimum response rate required for a collection?</li> <li>• Do we want raw data*** or weighted data?</li> <li>• What is our minimum item non-response rate for a question?</li> <li>• Are we interested in levels or the movement of levels? i.e. Do we care about the number being 10 or is it of more interest that the number has changed by a movement of +1 since last month/quarter/year etcetera?</li> </ul>
<p><b>Coherence</b></p> <p>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.</p>	<ul style="list-style-type: none"> <li>• Do we want to combine several sources of data? <ul style="list-style-type: none"> <li>◦ If so, are the different sources comparable? For example in terms of definitions, scope, collection method etc?</li> </ul> </li> <li>• Do we need data that conforms to particular standards? e.g. International, Australian, health etc.</li> <li>• Do we need to create a time series of information? <ul style="list-style-type: none"> <li>◦ If so, is the data over time consistent for our purposes?</li> </ul> </li> </ul>



ABS Data Quality Framework Dimension	Questions to consider
<b>Coherence</b> <i>Continued</i>	e.g. Data items the same, underlying methodology for collecting the data the same (e.g. question wording, sampling method, collection method etc.), real world effects / impacts on the data etc?
<b>Interpretability</b> refers to the availability of information to help provide insight into the data.	<ul style="list-style-type: none"> <li>• Do we want a copy of the questions asked?</li> <li>• Do we require a list of definitions for the concepts measured?</li> <li>• Do we need other explanatory material in order to decide on whether to use the data or not?</li> </ul>
<b>Accessibility</b> refers to the ease of access to data by users, including the ease with which the existence of information can be ascertained, as well as the suitability of the form or medium through which the information can be accessed.	<ul style="list-style-type: none"> <li>• What level of detail do we need for this data? e.g. Local government area by sex by employment status</li> <li>• What is our budget for purchasing data?</li> <li>• What format(s) do we need the data in? e.g. Spreadsheet, csv file etc.</li> <li>• Do we need a contact name in case we need to ask for more information about the data?</li> <li>• Do we require a different breakdown to that which is published? e.g. We need single year ages, not the 5 year age groups published.</li> <li>• If we have a special data request when do we need the data by?</li> <li>• Would we consider obtaining average/mean data at finer levels if this alleviated possible disclosure issues?</li> </ul>

**\*Imputed:** Assigning a value for data that is missing.

**\*\*Edit Adjusted:** Adjusting existing data to correct obvious errors (e.g. typos).

**\*\*\*Raw data** is data that has not been adjusted in any way. For example if there were 20 fully responding people to a survey, the Raw Data count of the number of respondents would be 20.

For more information go to <http://www.nss.gov.au/DataQuality> or email [inquiries@nss.gov.au](mailto:inquiries@nss.gov.au).

