1 Introduction

Data are fundamental components of information. If information is to be shared and used for decision making, then data must be available, reliable, accurate and collected to common agreed data standards.

Data development is the process of building a data set for a specific purpose. The process includes identifying what data are required and how feasible it is to obtain the data. Data development includes developing or adopting data standards in consultation with stakeholders to ensure uniform data collection and reporting, and obtaining authoritative approval for the data set.

1.1 Purpose

The guide helps to focus on relevant issues that must be addressed in order to achieve a successful outcome from a data development project and subsequent data collection. The guide does not aim to answer every question or solve every problem that may be encountered when undertaking a data development initiative. Instead it aims to provide the foundations for good data development work.

The guide provides:

- a methodology for data development, from the time an idea is conceived to a fully developed data set specification and accompanying data standards
- a specification for data standards using a metadata format that follows the concepts and principles outlined in ISO/IEC 11179-3(2003). ISO/IEC 11179 provides a standardised metadata format to describe and represent the meaning and content of data.

The primary purpose of the guide can be summarised as:

1. to provide guidance on the principles and methodology of data development that incorporate all the necessary steps to produce a high-quality data set with accompanying data standards, which meets the needs of users of the data and attracts the necessary level of agreement on content and authority behind the data specifications.

The secondary purposes of the guide can be summarised as:

1. promoting the use of metadata to describe data collections
2. promoting the use of the ISO/IEC 11179 metadata format to describe data standards, where data consistency and comparability is required.

The principles and methodology included in this guide are universally applicable. It can be adopted by those interested in developing data for use at the national, state/territory or local levels. Then principles and methodology are also applicable across all subject areas, such as health, community services and housing.

1.2 Audience

The guide provides a starting point for anyone who undertakes data development with the intention of collecting or using the data, and for those who want to know what is involved in the process of data development. These may include:

1. data developers who are seeking knowledge of the process, principles and methodology of data development to help avoid the problems and pitfalls associated with data development. Data developers may be anyone designing a data collection, including developing questions and data items to use, irrespective of how the data are to be collected, (survey, administrative collection, interview, and so on).

2. users of data sets who may be interested in understanding data (and metadata), including how it is developed and structured. Users of data sets would include service providers (who are often also responsible for data collection) as well as people using the data for secondary purposes, such as administration, research or statistical analysis.

3. system developers, including people involved in developing database specifications, data exchange mechanisms and validating programs. System development is expensive and time consuming. This makes it even more important that data from these systems are based on sound data development principles that meet agreed standards and are endorsed by authoritative bodies.

1.3 Stakeholders

While they may not themselves undertake data development, stakeholders have an interest in not only ensuring that data are made available, but also that the data are relevant, consistent, comparable and accurate. Stakeholders depend on the outcomes of good data development and may use the data for a variety of purposes, ranging from service provision, decision making, statistical reporting, service planning and policy making.

The key stakeholders of the guide include:

1. the Australian public—who rely on the outcomes of good data development where data are used to help make informed decisions about individual health and community care.

2. policy makers and service planners—including the Australian Government, state/territory governments and local governments—all of whom are interested in better data and information for a variety of purposes, including better service provision, enhanced decision making, informed policy making and service planning.

3. service providers—who require information about their clients in order to be able to provide better services.

4. funding bodies—which rely on information to make decisions about how to distribute funds.

5. information management and data standards committees—who have a role in the identification of priority areas for data development and are responsible for the governance of data and data standards.

6. educators and researchers—who need to understand and interpret data used as the basis for academic preparation and research. It also includes those who are directly involved in data analysis and use of data in order to identify areas for further research.

7. reporting organisations—such as the Australian Institute of Health and Welfare (AIHW), the Australian Bureau of Statistics (ABS) and others responsible for producing statistics and related information on specific programs, such as Home and Community Care (HACC), injury surveillance and supported accommodated assistances packages (SAAP).

8. Standards Australia—who are involved in standards development and provide infrastructure necessary for ensuring development of Australian standards.

9. other organisations—which need better data and are therefore interested in improvements in the accuracy, comparability, and consistency of data.

1.4 Structure of the guide

The guide is structured into seven chapters, each dealing with an aspect of data development or its supporting elements. Not all chapters in the guide may be relevant to all readers. For example, the more technical chapters, such as chapters 4 and 5, may be relevant to data developers rather than those wanting an overview of data development.

Chapter 2 provides an overview of data development. It explains what data development is, why data are developed, its relationship to metadata and data standards, the purpose of metadata registries and the need for national data standards to enable nationally comparable data to be produced.

Chapter 3 discusses the principles of data development and data development methodology. It aims to provide readers with an understanding of the basic principles and practice of data development. It identifies risks associated with data development and strategies that can help to minimise these risks.

Chapter 4 is a guide to developing metadata, regardless of whether the metadata is to be used to describe data or as a tool for developing data standards.

Chapter 5 provides an example of an implementation of a metadata registry based on the ISO/IEC 11179 standard. Australia's Metadata Online Register (METeOR) is an electronic source of national data standards across the health, community services and housing assistance sectors in Australia.

Chapter 6 provides an overview of the national agreements and governance structures that underpin national data development initiatives.

Chapter 7 contains the contact details of organisations and committees whose assistance can be sought in relation to aspects of data development.