A guide to data development

National Data Development and Standards Unit

2007

Australian Institute of Health and Welfare
Canberra

AIHW cat. no. HWI 94

A complete list of the Institute's publications is available from the Institute's website <www.aihw.gov.au>.

ISBN-13: 978 1 74024 655 2

Suggested citation

Australian Institute of Health and Welfare (AIHW) 2007. A guide to data development. AIHW Cat. no. HWI 94. Canberra: AIHW.

Australian Institute of Health and Welfare

Any enquiries about or comments on this publication should be directed to:

Meera Rajendran
Australian Institute of Health and Welfare
GPO Box 570
Canberra ACT 2601

Phone: (02) 6244 1054

Email: datadevelopment@aihw.gov.au

Published by the Australian Institute of Health and Welfare

Printed by Union Offset

Contents

Acknowledgments

The Guide was written by Meera Rajendran of the National Data Development and Standards Unit (NDDSU) of the Australian Institute of Health and Welfare, who would like to acknowledge the contribution of the staff of the NDDSU, especially David Neilsen who documented the guidelines for formulating good data standards, and Shari Beaumont for formatting and proofreading the document.

The author would also like to thank the New South Wales Health Department, the Queensland Health Department and Standards Australia IT–14–2, all of whom contributed to the Guide.

Thanks are due to Trish Ryan, former Unit Head, NDDSU, whose idea it was in the first place to write this Guide, so that the Institute's wealth of knowledge and experience about data development could be made available, and contribute towards the development of good data collections.

Abbreviations

ABS	Australian Bureau of Statistics
AHMAC	Australian Health Ministers Advisory Council
AIHW	Australian Institute of Health and Welfare
ANIHI	Agreement on National Indigenous Housing Information
CSMAC	Community Services Ministers' Advisory Council
DEC	Data element concept
DSS	Data set specification
HACC	Housing and Community Care
HDSC	Health Data Standards Committee
HMAC	Housing Minister's Advisory Council
MDS	Minimum data set
METeOR	Metadata Online Registry
NCHI	National Committee for Housing Information
NCSDC	National Community Services Data Committee
NCSDD	National Community Services Data Dictionary
NCSIA	National Community Services Information Agreement
NCSIMG	National Community Services Information Management Group
NDDSU	National Data Development and Standards Unit
NEHTA	National e-Health Transition Authority
NHADD	National Housing Assistance Data Dictionary
NHDA	National Housing Data Agreement
NHDD	National Health Data Dictionary
NHIA	National Health Information Agreement
NHIMPC	National Health Information Management Principal Committee
NMDS	National minimum data set
SIMC	Statistical Information Management Committee

Summary

For more than a decade, the Australian Institute of Health and Welfare (AIHW) has been undertaking data development and assisting other organisations with related activities. In doing so, the AIHW has developed a body of expertise in data development. Much of this experience has been incorporated into this guide.

While data development activities are undertaken by many organisations, detailed information about the process of data development is not readily available. Data development requires a considerable investment of time, staff and other resources, and includes the development of data standards to support consistent data collection. The guide outlines the basics of data development and the steps to follow when developing data.

The use of electronic health records is increasing throughout the health care system and the feasibility of electronic continuous client records is being considered by the community care sector. The advent of electronic data capture and communication systems provides a unique opportunity to improve the quality and coverage of information in Australia and the potential to re-use and share data becomes even more relevant. In line with the ‘create once, use often’ objective of information management, it is important to look for opportunities to improve the information available for national reporting and analysis when undertaking electronic data collection initiatives. Otherwise, significant additional costs may be incurred as a result of the need to create parallel systems whose products may not necessarily be compatible, resulting in inefficiencies and risks to data quality. It is therefore important that data development be undertaken using principles that ensure data collected at the point of service delivery can be aggregated and used for secondary purposes.

The guide outlines sound data development practices that incorporate the necessary steps to produce high-quality, consistent data that meets the needs of all users of the data, and attracts the necessary level of agreement on content and authority behind the data specifications. The document promotes the role of metadata in describing data and the use of data standards to improve consistency in data recording where data consistency and comparability is required.

The document takes the reader from the basics of data development to the methodology behind good data development and the specification of data elements to describe and standardise data.

The principles and processes outlined are applicable to all subject areas where data are required, including community services, health (including clinical data collections) and housing assistance. The processes outlined are relevant for local data collections as well as state/territory or national data collections. The document provides information about governance processes for national data collections.

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.

2 Data development overview

This chapter provides an overview of data development and introduces the key components, such as data, information, data elements, metadata, data standards and their relationships. The importance of data standards to data development is explained and the relationship between terminology and data standards is discussed.

2.1 What is data?

Data are representations of real world facts, concepts or instructions in a formalised manner suitable for communication, interpretation or processing by human beings or automatic means (Standards Australia 2005). Data relates to events, people, transactions and facts. For example, some of the data collected when a person buys products at a supermarket include:

- cash register identifier (id) (for example, 123)
- cashier identifier (id) (for example Z456)
- item description (for example, apple juice, jam, bread, coffee, milk)
- item identifier (id) (for example, X123)
- item unit price (for example, $1.20)
- quantity (for example, 2)
- total cost (for example, $10.30)
- date of service (for example, 26.10.2005)
- time of service (for example, 14:30)
- payment method (for example, cash, credit card, cheque).

2.2 What is information?

Information is data that are interpreted, organised and structured in such a way as to be meaningful to the person who receives it (Standards Australia 2005).

At the point of service delivery, data about items purchased by a customer in a supermarket are converted into information and provided to the customer in the form of a receipt. The same data would also be useful to the supermarket manager. For example, information in the form of a report showing total sales in the day and the best-selling products would help with inventory control. For this purpose, all supermarket transactions occurring in a day are captured, and then processed into information in the form of a management report.

2.2.1 Uses of information

All businesses (such as supermarkets or banks), services (such as community or health services) and individuals need information to support service initiatives and to develop business strategies. Some examples of how information is used are summarised below.

1. **Service delivery**—information about each event or transaction is needed to deliver effective services to clients. For example, information about purchase details is required to prepare receipts or invoices for clients. A doctor needs particular information to correctly prescribe medication for patients.
2. **Customer profiling**—information about clients (such as their habits and preferences) makes it possible to design and deliver better and more effective services. If agencies do not know the extent to which they are currently serving customers, they cannot determine what policy or procedural changes would be most effective in improving services.
3. **Planning**—up-to-date and comprehensive information about resources (such as cash, people, machinery, equipment and property) is required for planning for the effective management of resources. Information about the markets or climate in which organisations operate can also assist in service planning. At the planning stage, information is a key factor in decision making.
4. **Decision-making**—information is needed to make sound decisions. For example, how an organisation should distribute its resources, or confidently forecast its profits, requires reliable information to inform and support decision making. Doctors require high-quality information to assist with diagnosis and to select the best treatment option available to a patient.
5. **Policy making**—policies should be based on reliable and transparent information. The need for transparency in the way information is selected and used is a key factor when determining the legitimacy and acceptability of policy decisions, such as where, when and how something needs to happen.
6. **Measuring and controlling performance**—information is required to measure and evaluate the impact or effectiveness of a service. Information is needed to determine whether service provision is better or worse than expected, and to identify ways in which corrective action may be taken. Information that compares local performance with agreed benchmarks is used as the main way of measuring performance. For example, this can be done by collecting and analysing information on services delivery costs, waiting times, incidence of disease, and so on.

2.3 What is metadata?

Metadata is data about data that we need to help us understand and accurately interpret information. Metadata provides us with answers to questions such as ‘What does the data mean?’, ‘Are the data derived, and if so how?’ or ‘How are the data represented?’ Metadata is important when collecting, storing and using data.

2.3.1 Metadata is the key to data

Metadata is instrumental in transforming data into meaningful information. For example, metadata tells us that ‘19031905’ is the date of birth, and not the date of death or the date of service, and that the date is represented in the format DDMMYYYY, where DD is the date, MM the month and YYYY the year. There are many ways of representing a date of birth, such as YYYYMMDD, YYYYDDMM, and so on. As different organisations store data using different formats, the metadata helps us to understand and interpret data accurately.

Metadata removes the ambiguity about data and helps to avoid situations where individuals put their own interpretation on information. It helps to avoid situations where one section (or department) reports to the Minister that a service has declined by 10% and another reports (using the same data) a rise in the same service by 10%. Metadata makes data meaningful by creating a single version of the truth. It increases our confidence in the information used because metadata allows us to interpret data accurately and consistently.

The absence of metadata has been likened to a filing cabinet stuffed with papers, but without any folders or labels. The lack of proper directions, labels or signs makes it difficult to find anything and reduces confidence that what is found is what is required. Information without metadata can also be likened to a huge library of books not arranged in any particular order, with no catalogue or index to guide searching. It may be possible to find a book of interest by browsing through large numbers of books, but you can never be sure that what you have found is all that there is on the topic of interest.

2.3.2 Uses of metadata

In libraries, where metadata in the form of the library catalogue has been used for centuries, the primary function of metadata is resource discovery. In this setting, the term metadata refers to any data that aids in the identification, description and location of resources.

Various metadata communities have developed a number of standard metadata schemas and formats, each fulfilling a specific purpose and developed in response to various information management needs.

Some of these purposes and the metadata schemas used include:

- bibliographic and archival description—*ISBD, MARC and AACR-2*
- record keeping—*AS/ ISO 15489-1:2002; Australian standard for records management (Standards Australia 2002))*
- resource discovery—*Dublin Core and AGLS metadata*
- Geo-spatial description—*AS/NZS ISO19115:2005, Geographic information – metadata*
- learning object metadata for technology supported learning environments—*IEEE 1484.12*
- digital search and retrieval interoperability—*ANSI/NISO Z39.50-1995, Information Retrieval (Z39.50): Application service definition and protocol specification (ISO 1995)*
- common warehousing interchange specification — OMG Common Warehouse Metadata Interchange (CWMI) Specification
- electronic text encoding—*Text encoding initiative guidelines for electronic text encoding and interchange*
- standardisation of semantics and representation of data—*ISO/IEC 11179*.

The type of metadata described in the guide is about the meaning (semantics) of data and the standardisation of data. The definition of metadata adopted by the guide is the one provided in the second edition of ISO/IEC 11179-1:2004, which is:

> 'Metadata is data that defines and describes data.'

The purpose of metadata in this context is to describe and represent data that helps users understand the meaning and content of data.

While metadata is commonly used as a mechanism for describing or representing data, Australia was one of the first countries to use metadata, and ISO/IEC 11179 in particular, as a tool for describing national data standards in order to facilitate the exchange and sharing of information in a meaningful way.

2.4 What is ISO/IEC 11179?

ISO/IEC 11179 addresses the semantics of data, the representation of data and the registration of the descriptions of that data. It is through these descriptions that an accurate understanding of the semantics and a useful depiction of the data are found. ISO/IEC 11179 provides a standardised metadata format to describe and represent data to make it easier to understand the meaning and content of data.

ISO/IEC 11179 provides the essential components to promote:

- identification and standardised description of the components of data
- common understanding of data across organisations and between organisations
- re-use and standardisation of data and its components over time, space and applications
- management of the components of data.

2.4.1 Parts of ISO/IEC 11179

ISO/IEC 11179 is a six part standard. A short description of each part follows:

- Part 1—*Framework*—establishes the relationships between the parts and gives guidance on their usage as a whole. It contains an overview of the standard and describes the basic concepts. It introduces and discusses fundamental ideas of data elements, value domains, data element concepts, conceptual domains and classification schemes essential to the understanding of this set of standards (ISO/IEC 11179-1, 2004).
- Part 2—*Classification*—describes how to manage a classification scheme in a metadata registry (ISO/IEC 11179-2, 2000).
- Part 3—*Registry metamodel and basic attributes*—specifies metadata for data elements, data element concepts, value domains, conceptual domains, classification schemes and other components to be registered. Part 3 also provides the basic conceptual model, including the basic attributes and relationships, for a metadata registry. The registry metamodel is expressed in the Unified Modelling Language (UML) (ISO/IEC 11179-3, 2003).
- Part 4—*Formulation of data definitions*—provides rules and guidelines for forming quality definitions for data elements and their components. It provides guidance on how to develop unambiguous data definitions. A precise, well-formed definition is one of the most critical requirements for shared understanding of data; well-formed definitions are imperative for the exchange of information. Only if every user has a common and exact understanding of the data can it be exchanged without creating problems (ISO/IEC 11179-4, 2004).
- Part 5—*Naming and identification principles*—describes how to form conventions for naming data elements and their components (ISO/IEC 11179-5, 2005).
- Part 6—*Registration*—specifies the roles and requirements for the registration process in an ISO/IEC 11179 metadata registry. It provides guidance on these procedures. It provides instruction on how a registration applicant may register metadata with a

registration authority. Maintenance of metadata already registered is also specified in this document. The tasks and roles of the registration authority, data steward, registrar and submitting organisation are described (ISO/IEC 11179-6, 2005).

2.5 What is a data element?

A data element is used to standardise the representation of data. The ISO/IEC 11179 metadata format for data elements provide the structure that enables one to describe, in a standardised way, what we need, or want to know, about a piece of data. Each data element represents a basic unit of identifiable and definable data of interest. It is a unit of data for which the definition, identification, representation and administration are specified by means of a set of fields or attributes (see table 1).

Table 1: Attributes of a data element

Attribute category	Attribute
Identifying attributes include:	identifier, version, name, synonymous name, registration authority
Definitional attributes include:	definition
Representational attributes include:	permissible values, representation class, data type, format
Administrative attributes include:	registration status, submitting organisation, steward

More detailed information about data elements and its components are provided in Chapters 4 and 5.

An example of a data element is provided below.

Age

Identifying attributes

Metadata item type:	Data Element
Technical name:	Person—age, total years N[NN]
METeOR identifier:	303794
Registration status:	NHIG, Standard 08/02/2006 NCSIMG, Standard 29/04/2006 NHDAMG, Standard 10/02/2006

Definitional attributes

Definition:	The age of the person in (completed) years at a specific point in time.
Context:	Age is a core data element in a wide range of social, labour and demographic statistics. It is used in the analyses of service utilisation by age group and can be used as an assistance eligibility criterion.
Data element concept:	Person-age

Representational attributes

Representation class:	Total
Data type:	Number
Format:	N[NN}
Maximum character length:	3

Supplementary codes:	999 Unknown/not stated
Unit of measure:	Year
Guide for use:	Age in single years (if aged under one year, record as zero). If age (or date of birth) is unknown or not stated, and cannot be estimated, use Code 999.
Collection methods:	Although collection of date of birth allows more precise calculation of age, this may not be feasible in some data collections, and alternative questions are: Age last birthday? What was .... age last birthday? What is .... age in complete years?
Administrative attributes	
Submitting organisation:	National Public Health Information Working Group
Origin:	Australian Bureau of Statistics, Standards for Social, Labour and Demographic Variables. Reference through: <http://www.abs.gov.au/Ausstats/abs@.nsf/StatsLibrary>
Implementation in Data Set Specifications:	Computer Assisted Telephone Interview demographic module DSS NHIG, Standard 04/05/2005 **Information specific to this data set** In CATI surveys, age refers to completed age of respondent on day of interview. If collecting age in single years is not possible, age may be collected as a range. Refer to the data element Person—age range, code NN.

2.6 What is a metadata registry?

Metadata can be stored in a data dictionary or a metadata registry (also referred to as a metadata repository). The aim of the metadata registry (or data dictionary) is to make metadata available to users. Data becomes more useful when metadata is readily available. While a data dictionary can be either paper or electronic based, a metadata registry refers to an electronic database that is used to store and manage metadata. Metadata registries are often implemented by organisations or national bodies, rather than by individual programs, and as such they are an effective mechanism for providing a perspective of metadata at an organisational or national level.

Since the metadata registry stores metadata in a database, the registry provides mechanisms for querying, identifying and retrieving the metadata stored. In addition, it has the potential to produce reports on the stored metadata and use of that metadata. For example, 'Which data collections contain date of birth information?' or 'How many users queried the metadata repository yesterday?'

ISO/IEC 11179 metadata registries support the functionality of registration, whereby metadata items are assigned a unique identifier and registration status. Registration is both a process and a goal. The registration process allows for metadata life-cycle management, whereby it is possible to track the development of a uniquely identified metadata item from the time of development to when it becomes a standard. The registration status specifies the state of a metadata item in the metadata register, such as 'candidate', 'standard' or 'retired'.

The registration status categories address improvement and progression towards levels of perfection in the quality of the metadata of the item.

Functional operating procedures are needed for people who develop, operate, and/or maintain a metadata registry. ISO/IEC 11179-6 (2005) specifies participation of certain roles, such as registration authority, registrar, submitting organisation and steward. The registration authority has one or more registrars responsible for facilitating the registration of metadata items and making those items widely accessible and available to the community. The registrar may be viewed as the contact for the registration authority. Submitting organisations submit metadata items for metadata registers. A submitter is a contact for a submitting organisation for a metadata item. Stewards are authoritative sources for the attributes of metadata items.

2.7 What is data development?

Data development is undertaken in order to improve the quality, relevance, consistency and availability of information. The drivers for data development arose from the need for better information, whether it is clinical, administrative, statistical or any other type of information.

Some specific reasons for data development could include:

- a need for more complete information about clients and/or services
- to find out about the needs or the unmet needs of clients
- to measure services, targets and outcomes
- a need for information about the population being served
- a need for information about a new population to be serviced
- to meet reporting requirements of government departments, boards or other regulatory bodies
- to compare information at different levels (for example, local, state, national, international)
- to compare similar information (for example, about services or clients) collected in different places (for example, in different states or territories)
- to inform agency or service planning
- to develop a management information system for agencies or services
- to exchange data seamlessly between systems, organisations or jurisdictions by making clear and unambiguous information available (for example, data set development for use in electronic health records enable seamless delivery of care to patients).

Data development begins with the articulation of a business need for data. It includes identifying what data are needed to support business requirements, and determining if the data already exists. Where gaps in the data exist, these are prioritised and consideration is given to how the data can be collected in practical terms (see Figure 1).

Data development results in the building of a data set (also referred to as a data collection) for a specific purpose regardless of the how these data are collected.

Good data development practices result in the production of data standards that support standardised and consistent data collection and usage. Good data development also emphasises the need for consultation and authoritative approval of the data set and supporting data standards.

It is important to note that data development is not about software or system interoperability and it does not specify systems, schemas or application program interfaces. In fact, good data development is system independent That is, data are well defined and standardised such that data can be compared regardless of the system or application that captures them.

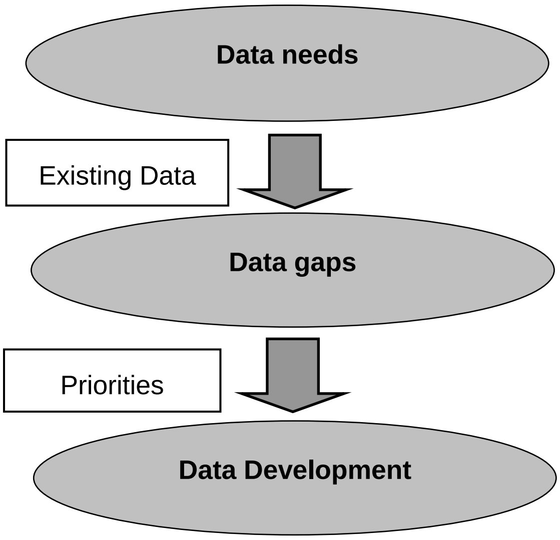

Figure 1: Initial steps in data development

2.7.1 Data development for primary and secondary purposes

The need for data development arises at all levels, from senior decision makers at the national and international levels to the state and territory, local government and individual levels. Data are the prerequisites for information that is required to support all levels and types of business. For example, data recorded on a patient's medical record, such as blood pressure, glucose level and heart rate, assists a clinician to diagnose and treat an individual patient. The same data, including diagnosis and treatment data, when obtained from large groups of patient records, provides a researcher with the information required for outcomes analysis. Parties who fund health services require the data for billing and reimbursement purposes and health service administrators require data to assess the use of (and need for) resources. In many organisations, data required for different purposes is developed and collected independently, often with much duplication and possible errors. Good data development practices should be mindful of opportunities to re-use available data, where appropriate, and ensure that data collected for one purpose can be made available for other tasks, so that additional data should be collected only where this is not yet available. It is important that data required for secondary purposes (for example, the generation of statistics, policy making and planning) is obtained as a by-product of service delivery (or the care process). Similarly, data developed for mainly primary purposes should also be used to compare and inform about the effectiveness of service delivery, and measure performance. This would:

- reduce the administrative burden of collecting data
- speed up the generation of reliable and accurate information
- improve service delivery as a result of feedback received.

2.8 What is a data set?

A data set is a set of data that is collected for a specific purpose. There are many ways in which data can be collected—for example, as part of service delivery, one-off surveys, interviews, observations, and so on. In order to ensure that the meaning of data in the data set is clearly understood and data can be consistently collected and used, data are defined using metadata. Examples of data sets include:

- Home and Community Care (HACC) data set
- (Health care) client identification data set
- (Service) provider identification data set
- Vital statistics (births, deaths, marriages and divorces) data sets
- Juvenile justice data set
- Supported accommodation assistance program (SAAP) data set
- Children services data set
- Diabetes (clinical) data set.

2.8.1 Data set specifications

A data set specification (DSS) contains the set of data items to be collected and details the standardised output of data that has been agreed upon by stakeholders. A crucial feature of a DSS is that there does not need to be any obligation to collect or report it. That is, the collecting or reporting of a DSS can be mandatory or optional. An example of a DSS is the National Diabetes Data Set Specification. It consists of a set of data items developed by the National Diabetes Data Working Group and is intended to guide providers involved in the treatment of patients diagnosed with diabetes about what data should be collected and how they should be recorded in a standardised way. The specification provides agreed definitions for data items related to the data set, and aims to ensure standardised methodology for data collection.

A minimum data set (MDS) is the name given to a selective core set of data that have been identified by users and stakeholders as the minimum for collection for a specific purpose. Examples of existing minimum data sets include the NSW Trauma Minimum Data Set and the Home and Community Care (HACC) Minimum Data Set. A minimum data set does not preclude the collection of additional data to meet individual agency or local needs.

In Australia, national minimum data sets (NMDS) are mandated national data collections for all states and territories. An NMDS is dependent upon national agreement to collect and supply uniform core data towards a national collection.

Arguably, the most important aspect of a NMDS is the agreement between all relevant parties (see Chapter 6). Without agreement, a NMDS does not exist. A NMDS agreement includes data standards specified using data elements, as well as the scope for the application of those data elements.

An example of an NMDS is the Admitted Patient Care NMDS. It specifies what information needs to be collected nationally about episodes of care for admitted patients in all public and private hospitals in Australia.

2.9 Data development and metadata

During data development, a number of characteristics, or attributes, of the data are defined in order to help understanding, interpretation and use of data. These characteristics of data are known as metadata. For example, the phrase ‘living arrangement‘ on its own is probably not meaningful, or at least has the potential to be interpreted differently by different people. It could refer to the living arrangement of children, such as whether they live with both parents, whether they live with siblings or whether they live with grandparents. It could also refer to the home in which a person lives, such as ‘own home’, ‘in a family home’ or ‘in a boarding house’. On the other hand, it might refer to whether a person lives alone or with others.

During data development, it is metadata in the form of a definition, value domain, data type, field length, and so on, that describes what data are collected and how it is represented. Metadata provides a standardised way in which to describe data, for example ‘living arrangement’, within a specified context.

EXAMPLE—‘Living Arrangement’

Definition:	Whether a person usually resides alone or with others	
Data type:	Numeric	
Maximum Size:	1	
Data Domain:	1	Lives alone
	2	Lives with others
	9	Not stated / Inadequately described
Guide for use:	This item does not seek to describe the quality of the arrangements, but merely the fact of the arrangement. It is recognised that this item may change on a number of occasions during the course of an episode of care.	

2.10 Data development and data standards

With the increasing use of electronic data capture systems, data has the potential to be collected once and used many times. It is therefore important that the data collected and recorded in electronic records is accurate and clearly understood by anyone that needs to use it. While this was true with manual systems, the potential to re-use and share data becomes much more significant when computers are used to represent and present information.

Metadata provides the underlying definitions and representation that supports collection, reporting and use of data within a specific context. Data standards describe the agreed meaning and acceptable representation of data for use within a defined context. While metadata helps us interpret data and information, it is the use of agreed data standards that enable data from different sources, organizations or systems to be exchanged and compared in a meaningful way.

Soon after 11 September 2001, a Director of the United States Counter Intelligence Agency was interviewed on television. He was asked, ‘with all of the money that is spent by the U.S. government on intelligence, why was it that no one was aware that a terrorist attack was imminent?’ The answer was that the Central Intelligence Agency (CIA), the Federal Bureau of Investigation (FBI) and the National Security Agency (NSA) all managed their data independently of each other. Without sharing of data it was impossible to identify and prevent problems across the organisations (Seiner 2001).

Another report claimed that two of the terrorists involved were on the FBI's watch list and, even though they were travelling using their real names, they were not flagged by the airlines. Again, because there was no sharing of information between the FBI and the airlines, the airlines were not aware of the security threat (Silverston 2001).

Besides the motivation or incentive to share data and technology to ensure software interoperability, data standards are needed in order to share data from different sources. A significant component of data development is the development of data standards to ensure comparability and consistency of data collected by different systems and organisations, and of information generated from these collections. Data standards describe the agreed meaning and acceptable representation of data for use within a defined context. The need for consistency of meaning and representation of data is vital to facilitate information sharing among primary users of data, and ultimately secondary users of data.

While metadata helps us to interpret data and information, it is the use of agreed standards to define the meaning and representation of data that will ultimately enable the meaningful sharing of data from a variety of sources.

2.10.1 Why standardise data?

To ensure consistency and comparability of content and definition

To increase efficiency, data must be optimally used or shared either within an organization, or with external agencies. . If data are to be shared, then clear understanding of the data are needed for accurate data collection and interpretation. When data are shared all users of the data must be able to clearly understand the meaning regardless of how the data are collected or stored. Use of common data standards ensure consistency and comparability of shared data.

To avoid duplication and diversity

Consistent use of data allows safer communication of that data and makes systems easier to use. Data standards are required to prevent diversity (for example, different definitions of emergency department waiting times) or duplication of the same data that can lead to misinterpretation. Data standards limit the ways information is collected (for example, by use of recommended questions) interpreted and exchanged between different groups, allowing for uniformity and synergy between multiple development efforts.

To ensure reduction in cost of data development

Standardising data reduces organisational costs and time by avoiding duplication and simplifying implementation. Data can be collected once and used for multiple purposes as required, thereby reducing costs in terms of time and money.

2.11 What are national data standards?

National data standards enable nationally comparable and consistent information to be produced, regardless of how the data are collected by the individual state/territory or organisation. A national data standard specifies the nationally agreed meaning and output representation of data. When data are collected according to the national data standard, data can be consistently compared across different jurisdictions, settings and sectors. National data standards also make data collection activities more efficient by reducing the duplication of effort in the field.

In Australia, national health, community services and housing data standards are contained in the National Health Data Dictionary (NHDD), the National Community Services Data Dictionary (NCSDD) and the National Housing Assistance Data Dictionary (NHADD) respectively. Some national data standards are common or integrated across these three sectors thus promoting the comparability of national health, community services and housing information.

2.12 Terminology

Individual components of a data element, such as permissible values or terms within the definition, can be mapped or linked to a recognised terminology standard to ensure clarity of meaning.

People working in different specialties or subject areas often use different terms to mean the same thing. For example, the term ‘agency’ is often used within the community services sector to mean an organisation or establishment that provides a service, such as a child protection agency. In the health sector, the preferred term is ‘service’, even where referring to the organisation (or establishment) providing the service, such as the blood transfusion service. Another example is the use of the word ‘separation’ that in a statistical context means the end of an episode of care, and in a gynaecological context refers to the detachment of the placenta from the uterus.

Unique identification of concepts, along with the context in which the different terms are used, is important to ensure that clear and consistent meaning is maintained. Unless a term is properly defined, doubt may arise about whether the same term used in different context refers to the same thing. Terminology standards enable interoperability of meaning. For example, a terminology standard would specify that the concept ‘myocardial infarction’ (id: 87654), is often referred to by terms such as ‘heart attack’, ‘cardiac infarction’ or ‘MI’. This provides added clarification of meaning through the set of terms linked to it.

When developing data, terminologies can be used to provide the list of permissible values in the value domain. Terminologies provide sufficient granularity of terms required from a data collection perspective. For purpose of data analysis and reporting, a standardised way of aggregating the terms or concepts to suitably stable reporting categories or classifications is necessary.

2.13 Data development and data quality

While the quality of data may be affected by factors, good data development practices can contribute to the quality of data by minimising the chance of personal interpretation of the data, and by ensuring that data are consistent and comparable over time. When accompanied by data standards, the data to be collected is clearly understood by those collecting the data and by those using it for analytical purposes.

Good data development also ensures that data to be collected is relevant to primary users of the data (such as service providers) and therefore is more likely to be accurate. Where data to be collected is not relevant, or not perceived to be useful to primary users, there may not be the same commitment to ensuring data accuracy.

Good data development practices ensure that the limitations of the data are acknowledged and recorded. For example, measures of suicide identified from administrative records do not necessarily give a true measure of the extent of the problem in society. These records simply represent a system's response to suicide or an ability to identify the problem. It will exclude any cases that are not brought to the attention of, or are not substantiated by, the relevant authorities. Measures can also be sensitive to variation in practice over time—for example, a similar case of child abuse may be substantiated in one period of time, but not in another, making comparisons over time problematic.

It is important to note that there are potential data quality risks associated with allowing system users to create and maintain their own version of standard reference tables (or code sets). The development and promotion of corporate guidelines that assist business areas requiring extensions to the corporate reference tables for local purposes can help to minimise this risk. Having a single source for obtaining corporate reference tables and by 'locking' corporate reference tables within information systems can prevent local modifications.

3 Data development principles and methodology

The first part of this chapter outlines the main principles that underpin good data development. These principles are applicable regardless of the purpose of data development, such as statistical purposes or electronic records, and are not listed in any order of importance.

The second part of the chapter describes a sequential, but overlapping, approach to data development. It outlines the steps involved in data development from the time a need for data development is identified to the endorsement of a fully developed data set.

3.1 Principles of good data development

3.1.1 Creating data standards is part of data development

The quality of data, including its consistency and comparability, is enhanced when data standards are available to support the collection and use of a data set. The development of data standards is not something that is done at the end of the data development process. It is very much a part of the data development process.

Data developers must develop operational procedures to ensure that data standards become a key component of the data development process, and to ensure their ongoing relevance and maintenance.

There is a cost associated with creating data standards, but the cost of not creating data standards is likely to be even higher. This includes the loss of information that occurs due to staff changes, data redundancy, data conflicts, liability, misapplications and decisions based upon poorly documented data. These costs should be factored into the data development budget.

3.1.2 National and international standards should be used wherever available and applicable

When data are being developed, it is important to ensure that the specifications of the data are consistent with national and international data standards. This avoids duplication of effort and the development of conflicting data standards.

Data sets should be based on a single set of agreed definitions and data standards, such as the NHDD, so data developers can 'pick and mix' from existing agreed definitions, ensuring a high degree of consistency and reducing data development time and cost.

Local needs sometimes require more specificity that national or international data standards can provide. Where this is the case, it is possible to specify more granular data standards as long as they are mappable to national or international data standards. Tips and examples on what to do when local requirements demand more detailed or granular data are provided later in this chapter in the section entitled ‘What to do when a modification to an existing data element is required’ and in the following chapter in the section entitled ‘Value Domain’.

3.1.3 Be clear about the purpose of the data collection

Before deciding on what data to collect and develop, it is important to be clear about the purpose of the data collection including the important policy or planning questions, or service provision needs for which data collection is required, and how having the data will help deliver more efficient and effective services. It can be tempting to decide on what data to include in the collection before the purpose of the data collection is fully defined.

3.1.4 Data included must be required to meet the objectives of the data collection

Once the purpose of the data collection has been defined, the data required to meet the objectives of the collection can be identified and developed. Data developed and included in the data collection must meet the collection’s objectives as the cost of developing, collecting, compiling, validating and reporting data can be expensive.

3.1.5 Create once, use often

The availability of existing data sources should be explored and used where possible.

An important principle of data development is that data needed to support secondary (or downstream) information purposes (such as reporting, policy, governance or decision support) should be derivable from primary data (point-of-service delivery data). Otherwise, data needed for downstream requirements would have to be developed and collected separately, resulting in significant additional costs because of the need to establish parallel data collection systems to support existing and new data flows, whose products may not necessarily be fully compatible. Similarly, data developed and collected for mainly statistical purposes should be used to provide feedback to improve and enhance primary service delivery.

3.1.6 Acknowledge the limitations of data

While it is important to be mindful of the opportunity to re-use existing data, it is also important to be aware of any limitations of the data and to ensure that these are acknowledged. For example, measures such as child abuse can be sensitive to variations to practice over time, such that a similar case may be substantiated over one period of time but not in another, making comparisons over time problematic.

It is also important to note that data that are suitable for some purposes may have limited use for other purposes. For example, while data about ‘presenting problem’ and ‘reason for encounter’ may be useful measures of hospital emergency department services for treatment of patients and for management of department resources, they may not be adequate to inform about the incidence of domestic violence or child abuse.

3.1.7 Data development may be incremental

Data development should support incremental development, such that the scope of the data set is expanded over time.

It may not be possible to develop all data required for a data set at the same time. Some data may be more readily agreed upon and easily collected, while other data may be more problematic and require more time to develop. For example, for a problem gambling support services data collection, data developers may find it easier to develop demographic data about clients seeking help in relation to their gambling problems, but more difficult to develop and agree on data that can be used to measure outcomes of problem gambling educational programs or assistance services. Depending on the timelines and resources available, it may be better to stage the data development such that the scope of the data collection is expanded over time.

3.1.8 Data development is system independent

Data development must ensure that data in the data set is well defined and standardised so that it can be compared independent of the organisation, system or tool that captures the data.

Data development must not be limited by the capability of any particular system.

3.1.9 Data development should be mindful of privacy concerns

Data development processes must take into account security policies and privacy issues, including ensuring compliance with the information privacy principles. Data sets should avoid the inclusion of data that may be regarded as private or confidential in nature. Otherwise data provided may not be reliable or accurate because respondents may be reluctant to provide the data.

3.1.10 Data development should minimise collector/recording burden

Good data development should ensure compatibility of data collection and reporting requirements to avoid situations where the same data has to be collected, counted or reported differently for different programs. This will reduce the reporting burden on service providers and help to reduce cost..

3.1.11 Data development should reflect not drive practice

Data must be able to be collected as a by-product of service delivery or administrative practice. The data must be relevant and meaningful to those collecting the data and be of benefit to service providers. Data development must take into account the business needs, feasibility of data collection, appropriateness of the data, as well as scientific evidence and recommendations (guidelines) of subject matter experts. Where possible, data development should be based on data that service providers already want or need to collect about clients and service provision. Most importantly, data development should ensure that data collectors are not constrained or forced to operate in ways outside usual practice. For example, data about the country of birth of a client is usually not relevant for service delivery and should not be part of routine data set collections. Of more relevance to service delivery is data about whether interpreter services are required, client's preferred language, ethnicity and religion. Data about country of birth, which may be useful for statistical purposes, can be collected as part of a 'one-off' survey of clients.

3.2 The data development process

Data development is a methodological process, based on an understanding of the information to be derived using the data. It includes modelling data needs and clarifying the relationships between data. Key data concepts are identified and standardised using data elements. Data development results in the production of a set of data standards to ensure consistent collection and use of the data set.

3.2.1 Business context and information need identification

It is important to be clear about the purpose of the data development project and the benefits it has to offer, otherwise the effectiveness of the resulting data set may be greatly reduced. It is therefore necessary to obtain an understanding of the business context within which the information is needed, and the high priority policy issues, questions and/or service needs before proceeding with data development. For example, where the objective is to ensure equitable health care to non-English-speaking populations, an important policy question could be: 'Do we provide an equitable level of health services to people from non-English speaking backgrounds?'

Needs analysis results in documentation of the business requirements for data (or a problem statement), the target population (for example, healthcare professionals) and service environment, a description of the problem, gaps between expected and ideal outcomes, and the relative priorities of the business requirements. This helps to guide further data development, including identification of what data are required.

While the need for data development may relate to a specific problem or purpose, the expected benefits at all levels (for example, policy, program development, performance management and service provision levels) should be acknowledged and clearly stated.

3.2.2 Feasibility analysis

Developing a data set can be costly. Therefore, it is important to undertake feasibility analysis in order to provide an indication of the scope of the data development and the resources required to support the project. Feasibility analysis includes the identification of stakeholders and their relevant level of interest, the data needed to support information requirements, analysis of what data are currently collected, and if sufficient resources are available to proceed with the data development project.

Identifying stakeholders

For the data development process to be successful, it must reflect the shared interests of those who use or rely on the data. The feasibility study should identify relevant stakeholders, their information requirements, their availability to participate in data development activities, and their level of commitment to data development. Stakeholders could be service providers, service managers, national, state and territory governments or researchers.

What data currently exists?

It is important to identify existing data sources and assess their currency and usefulness for the current purpose. It is not always possible to get information about every system used for collection purposes, but obtaining a representative sample is recommended. The feasibility study should identify information gaps and data that are/might be problematic in terms of obtaining accurate information.

Where possible, data should be existing by-products of the service—that is, service providers already want or need to collect the data. When developing data for secondary uses (in relation to answering key policy questions) it may be necessary to undertake a cost–benefit analysis to determine whether it is cost-effective to collect data that are not directly required to meet service delivery needs.

As part of the feasibility study, it would be useful to assess whether the data that are currently being collected allows for comparison over time and if there are any existing standards, performance indicators or benchmarks.

How the data are currently collected

An understanding of the collection methods and how data are currently being reported assists data development. Data development that is mindful of the operational constraints of the systems and people who need to record the data results in a more usable and reliable data set. Data collection methods may include paper, various electronic media, face-to-face or telephone contact. Data collection is normally in the form of either an ongoing or snapshot collection.

Reporting arrangements and periods

Reporting arrangements should specify if there are any current reporting requirements for existing data. Reporting arrangements usually vary between organisations/jurisdictions and data developers need to be aware of the differences and challenges.

Future reporting requirements need to be established. In the case of NMDS reporting, there is normally a two-stage process—from service providers to jurisdictions and from jurisdictions to a national collection agency. Jurisdictions are provided with a national specification and agree to collect and transmit the data in that form. Some collections can be reported directly from service providers to national collection agencies.

Scope

The scope of the data set should at least specify:

- the population that is the target of data development
- the services, agencies or organisations to be included and excluded
- information to be included in the data set (such as interventions, conditions or services).

The scope of data development may initially be limited depending on whether the data required already exists and the time that is available for data development. However, future expansion of the data set to include other priority data should be considered and planned.

This would enable data that are currently difficult to collect or which requires a longer time to develop, to be included.

Statistical Unit

Where the purpose of the data collection is for statistical analysis, it is important to give some consideration to the statistical unit by which the data will be grouped. The statistical unit determines the level at which data will be collected and/or reported and is fundamental to a data set. For example, for a client-based collection the data could be collected at the client level, at the service contact level, the service episode level or at the service provider level (all clients to which a provider has provided services). A decision also needs to be made about whether the data needs to be aggregated or reported for each client.

Timeframe

Based on an understanding of the expected information requirements and current data availability, the study should provide an estimate of the time required to develop the necessary data.

Resources

The feasibility study should provide an indication of resources that will be required to undertake the development of the data set. Costs include salaries, project management fees, costs associated with data set and data standards development, consultations with stakeholders, endorsement of data standards, travel and associated costs, meetings costs, stationery, publication fees and pilot testing costs.

At this stage, it may not be possible to provide an indication of the cost of implementing the data set as more detailed analysis of the data for development by the data working group will need to occur.

Establish a project team and data working group

Once feasibility has been accepted and a decision made to proceed with data development, a project manager (or team) is recruited to lead and manage the data development. It is worthwhile forming a dedicated project team to lead and manage the data development project, rather than trying to undertake data development as part of an ongoing committee’s work. This is because a data development project is resource intensive and requires considerable time and effort to lead, coordinate and promote. One of the first tasks of the project team is establish a data working group, if one does not already exist. The data working group should bring together stakeholders, who may include service users, service providers, policy makers, data experts and other key advocates. The data working group is usually established under the auspices of an authoritative body in the field which has recognised the need for, or is financially supporting, the data development work.

The role of the project team is to manage the data working group in relation to the data development project and liaise with the patron or funding body responsible for the project. The project team not only leads the data development work, but also promotes and coordinates other related activities.

In addition to the working group, it may also be necessary to establish a reference group to seek specialist subject-area advice and feedback throughout the data development process. The reference group, while providing valuable advice and feedback, is not expected to do the actual data development work. Reference group members include stakeholders in the area for which data are being developed. A range of expertise is important to ensure that a robust data set that is useful from both a data collection perspective (primary use) and a policy perspective (secondary use) is produced.

Some of the tasks for the project team and the data working group would include:

- developing a project plan or work program
- collaborating and consulting with stakeholders
- identifying core data that needs to be collected in collaboration with stakeholders
- leading technical discussions on the data development
- developing data standards using data elements to support the data set
- circulating draft data specifications for public comment
- considering views expressed and comments made
- organising field tests and analysing data and feedback from the tests
- estimating cost of implementing the data set
- liaising with data committees regarding endorsement of the data set.

3.2.3 Consultation and collaboration

Consultation with stakeholders enables all interested parties to have input into the data development. Requirements or comments provided inform the final decisions to be made about the data set.

Consultation should occur from an early stage and continue throughout the data development process. Interested parties then become acquainted with the work and make contributions towards it, while amendments can still be made and comments taken into account. More importantly, consultation facilitates the uptake and dissemination of the data standards. During consultation with stakeholders, agreement is reached about the data elements, including the meaning of the data elements and how they are to be represented and collected, to ensure consistency and comparability across organisations.

If the data collection is of national significance, and is to become part of the existing national health information system, it will need to be endorsed by the key national committees concerned with obtaining agreement on national health information (see Chapter 6). The Health Data Standards Committee (HDSC) and the Statistical Information Management Committee (SIMC) should be informed of the project and given an opportunity to provide input and/or advice. These committees should be regularly consulted and updated regarding the progress of the data development project.

Procedures should be established so that adequate time and opportunities are provided for written comments. Essential information in relation to the data development project should be easily accessible to all interested parties. Openness, democracy and involvement in this process by all stakeholders are important.

Combined Top-down / Bottom-up approach

Whether data development is policy driven (top-down) or initiated by service providers (bottom-up), the data working group needs to adopt a combined top-down/bottom–up approach in order to identify data that are to be developed and included in the data set.

Data development that is initiated by policy requirements needs to be mindful of the operational constraints of the systems and people who need to record and process the data. These constraints are particularly important when the data needed for management reporting is a by-product of service delivery.

Similarly data development that is initiated by service providers, or those who will be collecting the data, should consider the statistical impact of the data and how it can be used to measure service performance.

Top-down approach

Data development that is initiated from the top down is usually in response to information required to inform policy initiatives, service utilisation, program outcomes, to support planning decisions or measure performance, or to inform about cost of service provision. The data that are required to support these information needs are usually aggregated data, such as cost of service, staffing requirements, number of clients requiring service or total sales. A major challenge for data development is to ensure that data used within the service delivery setting can be accurately and meaningfully aggregated to produce statistics that inform policy and decision making.

The top-down approach leads to an understanding of how the required information will be used, such as in benchmarking or governance, and will provide guidance in relation to what we want to know or measure, the data needed to be able to achieve the measures, how the data should be represented and the corporate reports that would include these data.

Bottom-up approach

Data development that is initiated from the bottom-up is usually in response to information required to assist service providers to provide high-quality services more quickly and efficiently. For example, this would include data that are required to alert service providers when clients are due for services, or when services cannot be provided to clients for specific reasons, data for decision-support systems or data to facilitate electronic delivery of services.

This type of data usually has more detail about services provided and clients (for example, data that uniquely identifies clients and their service needs).

The bottom-up approach involves consultation with service providers who are directly involved in the collection and recording of data and have an understanding of the day-to-day issues, and experience with use of the data that are collected. The bottom-up approach provides information about the data collection and maintenance within the local or service delivery setting and leads to an understanding of the context and feasibility of data capture. It is through consultation with service providers that the data working group becomes familiar with data that are collected or required, how these data are collected and used, the terminology used, and if any difficulties are experienced in collecting any data.

A combined top-down/bottom-up approach can ensure reliable integration of requirements at service delivery and policy levels and development of data that can be appropriately aggregated for different purposes.

The top-down/bottom-up approach resolves the tensions that often exist between information that is useful for evaluation at the policy level and ease of collection at the service delivery level. Data that are easily collectable may not be comprehensive enough to undertake an evaluation of service provision and direction for policy and planning. On the other hand, if data are difficult to collect, or not perceived to be useful at the service delivery level, then it becomes time-consuming and cumbersome.

In general, if management or policy data are captured via routine administrative processes that do not require the data for the main job at hand (such as processing a payment) then data quality may be at risk.

3.2.4 Identifying data for development

An important part of data development is the detailed analysis of the data that are necessary to support business requirements, which, in turn, depends on the purpose of the data development, the sorts of questions that need to be answered or problem that needs to be addressed. Core concepts are identified and these are defined and standardised. For example, if equitable access to problem gambling support services is a focus area, then information should give an indication of equitable access to these support services to various population groups. There should be a decision made by the data working group about how to measure equitable access. For example, equitable access could be measured by the percentage of people in a population group accessing a service or the availability of interpreter services, and so on. In this way, the data to be collected are identified and each data identified is described and standardised using a data element.

As some of the required data may be less readily available, or difficult to collect consistently (such as the amount of money spent on gambling), and there is a cost and effort involved in data collection, it may be necessary for the data working group to identify a core (or minimum) set of data for collection, based on how the data can be collected in practical terms and the priority of the data. Future expansion of the data set to include data that requires a longer time to develop would set up a work program for future data development work.

Modelling data requirements

A data model is a useful tool when identifying and depicting data for development. The model provides a diagrammatic representation of the building blocks of data necessary to meet user information requirements. The model is used to provide the framework and context for the data identified and described. It is employed to help users identify and articulate data requirements and specify the business rules and relationships that exist between data. It is important that the data model is agreed to, and signed off, by stakeholders before proceeding with data development. Otherwise, the data working group may find that they have to deal with changing data requirements throughout the data development process.

In a data model the ’thing’ of interest is called an entity and data elements are used to describe ’things‘. For example, a data element ’date of birth‘ would be used alongside other data elements to fully describe a ’client‘. In this case the client is the thing of interest, which in a data model is depicted as an entity. Each entity needs to be defined and described through its associated data elements. For a specific purpose, a client may be fully described by date of birth, sex, postcode, interpreter service required and preferred language. The data model depicts all the entities and data elements that are required to meet information requirements. Identifying the data elements that are required, and agreeing to their definitions and value domains, is of critical significance to consistent data collection and use.

An example of a simple data model depicting entities and their relationships is shown in Figure 2. Entity-relationship models are useful for depicting high-level relationships amongst entities.

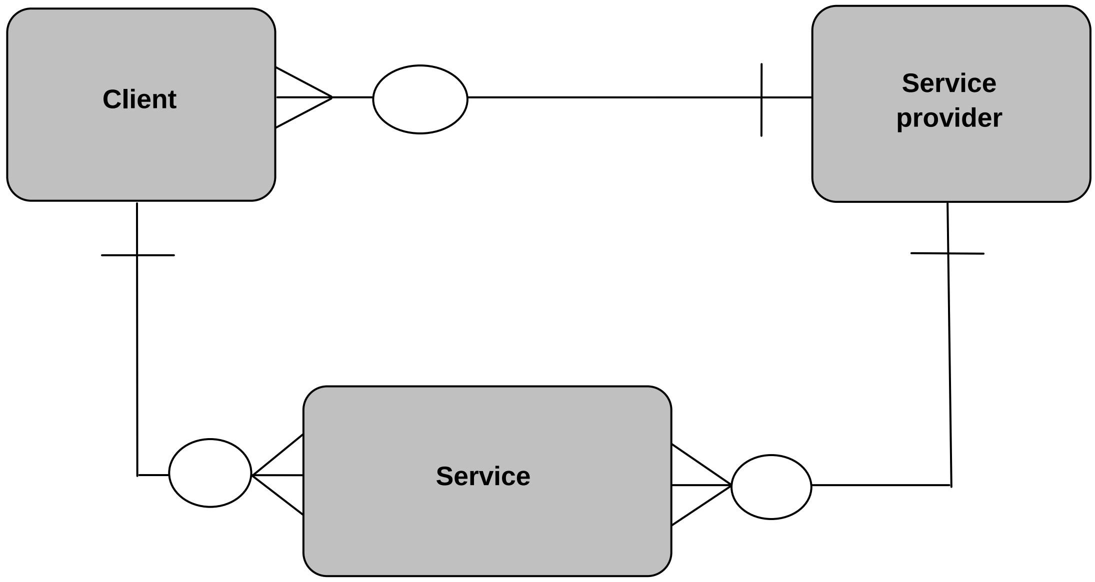

Figure 2: An entity-relationship model

Entities

The boxes in the model represent entities. The identification and definition of entities is an important part of any information.

When questions are asked such as 'How many clients receive support services?' this assumes a shared understanding of the term 'client'. Other questions such as 'What types of support services were required by clients?' suggests other entities are required, such as 'services' and 'service provider'.

The definition of entities is very important, particularly for consistent reporting, because the definition should help to standardise what is included, and what is excluded, within the scope of an entity.

Relationships

Entities have relationships (or rules of association) with other entities in the data model. The lines and arrows in the model represent relationships. There are many different conventions that describe the meaning of the relationships between entities. In Figure 3, the line into the entity *'Service provider'* is an example of an optional, one-to-many relationship which can be interpreted as 'A client must be supported by one (and only one) service provider. The line in the other direction can be read as 'A service provider may support one or more clients'. The 'O' means the relationship is optional while the vertical line indicates that the relationship is mandatory. In this instance, it is assumed that a service provider can exist without a client, but a client cannot exist without a service provider.

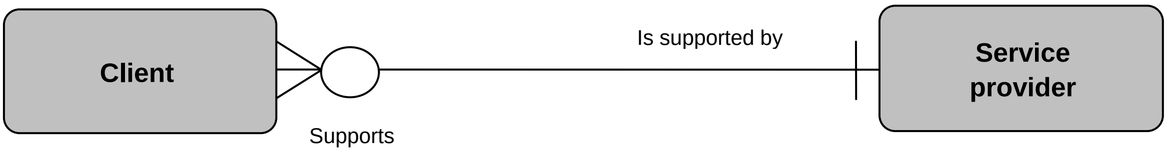

Figure 3: A data model showing relationships between entities

Decisions on relationships and consistent implementation of relationship rules are fundamental to consistent and effective informational collection and use.

Data elements

Once the essential entities have been defined, the characteristics that describe the entities are identified and defined. These characteristics are turned into data elements and are used to standardise the agreed meaning and acceptable representation of data for use within a defined context.

It is important that each data element describes only one concept and therefore it may be necessary to create more than one data element to describe the data that are required. Since there is a cost associated with data collection, every data element should have a clearly defined purpose.

The formulation of data elements is described in more detail in Chapter 4, while the next section describes when to create new data elements, when and how to modify existing data elements and when to reuse existing data elements.

3.2.5 Developing data elements

Re-using existing data elements

In some instances, the data elements that are required for a data set may already exist. Wherever possible, data developers should identify opportunities to use existing data elements from the authoritative sources, such as the National Data Dictionaries, ABS catalogue and Standards Australia. This saves unnecessary effort involved in recreating data elements and promotes alignment with existing standards. If an existing data element adequately describes the concept to be collected, then this data element should be used. However, data developers should not be limited to re-using an existing data element if it is not relevant or is too constraining for the needs of the data set.

Modifying existing data elements

If there is close alignment with an existing data element, but it is not exactly what is required, there may be scope to modify the existing data element to meet new requirements. It may be possible either to modify an existing data element without affecting its current use or to include collection-specific information that is not inconsistent with the use of the current data element. Before changing an existing data element, the data developer should consult with others using the data element.

In deciding whether an existing data element can be used or if it can be modified, the data developer needs to ensure that the existing data element is conceptually the same. Even if it is conceptually the same, the concept may still need to represented differently for a different purpose. In this case, a new data element is created, instead of re-using or modifying the existing data element. For example, if the pre-existing data element 'height of person' specifies measurement in metres and the current data set is about newborns, it may be necessary to create a new a new data element that specifies measurement in centimetres. In this case, the two data elements are the same conceptually, but because the concept is represented differently, separate data elements are required.

What to do when a modification to an existing data element is required.

If it is established that an existing data element could be used, but only with modifications, the first thing that should be done is to identify users of the existing data element. The steward of the data element may be able to provide information about current users of the data element or, if not, data working group members and stakeholders of the project are often able to assist. The project team will need to communicate with existing users (for example, by attending data working group meetings of the existing user) to collaborate on a joint solution that would be workable for existing and new users of the data element. This can sometimes be a slow and time-consuming process because it may take time for existing users to assess the impact of the change. It is therefore important to establish the priority of the data and the importance of the modifications before seeking changes to the data element.

It is important to determine if the existing data element currently conforms to agreed national, state or international standards, where they exist and how the modifications required will impact on compliance. Where modifications to the value domain are sought, it is important to ensure that it is consistent, or mappable, to existing data standards wherever possible.

It is often the case that data standards developed for national (or even state) purposes are of low specificity (less granular). This may not satisfy local needs where more detail is required. Similarly, clinical users of health data often require very granular data, while statistical users of health data are usually satisfied with less granular data. Depending on the purpose, value domains may be expressed to varying degrees of granularity as long as they are mappable to the recommended standard. A one-size-fits-all approach is not necessarily required to achieve standardisation. For example, the national standard for eye colour may include only:

1	blue
2	green
3	grey
4	brown
5	black
6	other

Locally a broader range of colours may be required. The additional range must be mappable to the national standard. For example, dark-green and light-green are mappable to 2-green (many to one), or dark-blue is mappable to 1-blue (one to one). A value such as greenish-grey is not an acceptable addition since it can be mapped either to 2-green or 3-grey (one to many). Such a case should ideally be taken to the appropriate governance body responsible for the national standard for discussion before being adopted at the local level. Similarly, local additional values must be able to be mappable to 'other' in the national standard, or to one of the other existing values as described in the previous example. More examples of compliance and mapping to existing standards are provided in the 'Value Domain' section in the next chapter.

Creating new data elements

Creating new data elements includes formulating and defining the data element and specifying associated rules for when and how the data are collected. It may be possible to re-use some of the components of existing data elements, such as existing value domains, or data element concepts for the new data element where relevant. The following chapter outlines how to create data elements.

3.2.6 Field testing

Having developed a draft set of data elements, the next step is to field test the data set. Normally field testing is conducted for the whole data set rather than for individual data elements.

The purpose of field testing, or pilot testing, is to:

- test the practicality and consistency of the interpretation of data elements and the comprehensiveness of the codes (range of possible answers)
- test the quality of reported data
- assess how easy, or how difficult, it is to collect the data
- test the transmission of data and data analysis
- identify guidelines for implementation that need to be added to ensure consistent interpretation and reporting.

A selection of service providers or agencies are identified and recruited to participate. It is desirable that the sample include a range of agency/service types, models of care and locations.

The pilot test is usually conducted over a period of approximately 2 to 4 weeks.

Pilot test participants must be able to provide feedback on the data set and the collection process, both verbally and in writing.

The outcome or findings from the field test should be used to refine the data set before it is implemented.

3.2.7 Estimating cost of implementation

There are costs associated with implementing a data set. In some instances, a business case may need to be submitted before a data set is approved for collection.

The business case must consider the following:

- The cost of the change or new data.

- The ability to make a 'generic' change to the data collection, or the collection process that would manage this change, and other related issues. If there is the ability to institute such a modification, this should be costed.
- The opportunity and timing relevant for data collection review. Where the information needs have changed, there should be a review of the whole process of information collection in order to identify the most relevant change. It is easy to simply change the data collection rules or values, but, over time, this can result in a disjointed and complex collection system. When considering the introduction of any change, it is important to consider whether it would be cheaper in the long term to restructure the data collection rather than to make further 'bit' changes. It is often significantly cheaper to recognise opportunities for change. For example, as systems move towards electronic health records, the need to develop new systems for data extraction will arise. This provides an opportunity to build table-driven extraction rules that can be readily maintained and updated without the need to change software, thus offering significant future flexibility and potential to reduce the costs of later changes.
- Cost of making no change.

Implementation costs will vary depending on a range of factors, including, the extent to which systems are already in place and data set requirements over and above current system capabilities and reporting arrangements (for example, whether the data are sent directly by data collectors to the collection agency or if it is first sent to an intermediate agency for checking and compilation). With national data set collections, data from service providers is often transmitted to a central point within state/territory jurisdictions, before it is forwarded onto the national collection agency.

The cost to data collectors (service providers) include any system changes, training and other requirements necessary to facilitate accurate collection, extraction and transmission of data. Actual costs will depend on decisions made about scope, reporting arrangements and individual data elements in the data set.

Where jurisdictions are involved in the checking and compilation of data from service providers, there are costs associated with receiving, checking and loading data from service providers, training costs to ensure new systems are used appropriately, staff costs (additional staff required to support collection, collation and data supply) and system costs. Where there is no intermediate agency, many of these costs are borne by the collection agency.

Collection agency costs include costs required to develop and maintain a system to receive, process and store data received from data collectors. There is normally an initial set-up cost, and an estimate for ongoing maintenance.

Each time a data set is expanded, an estimate of the additional costs of proposed data set changes should be prepared.

3.2.8 Obtaining authoritative endorsement

Before the data set can be implemented, it is important to obtain endorsement of the data set and its accompanying data standards from the appropriate authoritative body. This is important to ensure credibility and uptake of the data set. For example, where a data collection is to become a NMDS for use in the health sector, the data standards need approval by the HDSC and the NMDS needs to be approved by the Statistical Information Management Group (SIMC) before it can be endorsed by the National Health Information

Management Principal Committee (NHIMPC). Similarly, governance arrangements exist for the endorsement of data collections for use within the community service and housing sectors. Chapter 6 provides more detail about data agreements that exist at the national level within these sectors, and the processes in place to obtain the endorsement of national data collections and data standards.

Many state/territory departments have documented processes in place for the approval of new or modified data sets, and accompanying data standards, for use in local jurisdictions.

3.3 Risk management

There are a number of risks associated with developing data sets. Some of these risks are as follows:

- Changing information needs—as services grow and expand, and as service models and policies change, it is reasonable to expect that information requirements will also change. Accordingly, there should be mechanisms in place to ensure that data sets are periodically reviewed and improved in order to reflect the changing information needs. This ensures the continued relevance of data sets in answering questions that are of importance to relevant parties. Data that are no longer relevant may need to be dropped from the data set, while new data may need to be developed and included.
- Failure to re-use data—opportunities to re-use data that are already developed should be investigated in order to avoid costly duplication and reduce possible conflicts in data. It is important that data required for secondary purposes (for example, policy initiatives, planning and measuring performance) is obtained as a by-product of service delivery. For example, e-health information initiatives, which are primarily concerned with communication and sharing of information that is relevant to clinical care, provide an excellent opportunity for the development of new statistical collections that inform and compare service delivery and measure performance. Statistical impact analysis should be undertaken to maximise the scope for extraction of useful statistical information from data developed for mainly primary (or administrative) purposes.
- Poor-quality data collected—periodic analysis of the data collected can help to identify data quality issues. The presence of poor-quality data may be an indication that the data collected is not useful or has been poorly understood by data collectors. Training or education campaigns may be employed as a means of addressing and emphasising the value of good-quality data. However, where it is found that the data are not of value to actual service delivery, it would be worthwhile undertaking a cost–benefit analysis of the need to routinely collect the data as part of the data set.

4 Developing data elements

This section examines how to create a data element. A significant component of data development is the specification of data elements to standardise meaning and representation of the data to ensure consistency and comparability of data collected and information generated from a data collection.

During data development, a data element is specified for each concept that needs to be described or standardised within the data collection.

Before starting to develop data elements, it is important to understand the context of the collection and use for the data that are being described or standardised. This will provide the basis for the terminology to use. It also facilitates the selection or development of the metadata components (data element concept and value domain) that constitute the data element.

4.1 The components of a data element

A modified ISO/IEC 11179 model used to describe a data element and its supporting components is provided below.

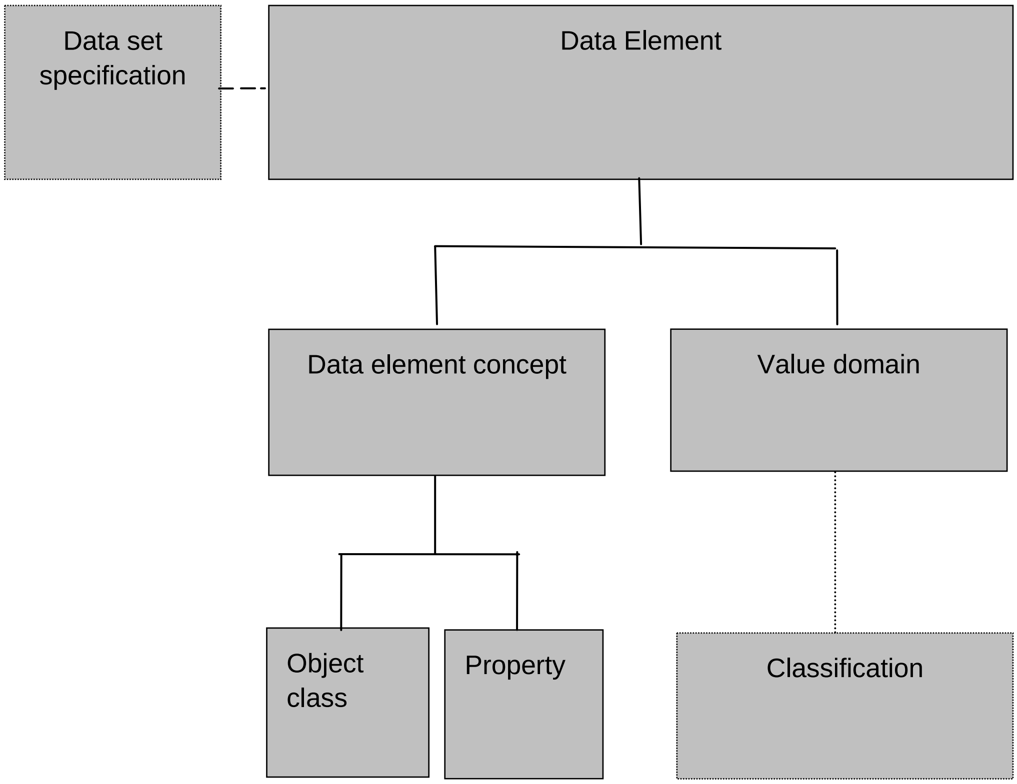

Figure 4: Model of data element and supporting components

A data element is made up of component parts. A data element is composed of a data element concept in association with a value domain. While the meaning of the data element is essentially captured within the data element concept, the value domain specifies its representation.

A value domain may itself be part of a classification, or a classification may provide the basis for value meanings of a value domain. Many examples of classifications exist, such as ICD-10-AM and the Australian Standard Classification of Religious Groups. Classifications such as these are official terminological systems, recognised and endorsed by national or international bodies that are used to classify data.

A data element concept is composed of an object class and a property.

A collection of data elements used to describe data for a specific purpose is a Data Set Specification (DSS). A NMDS is a special type of DSS—one consisting of a minimum set of data elements that are mandated for national collection. For example, the Elective Surgery Waiting Times (census data) NMDS contains the following data elements:

- Census date
- Clinical urgency
- Establishment identifier
- Extended wait patient
- Indicator procedure
- Listing date for care
- Overdue patient
- Surgical specialty
- Waiting time at a census date.

It is possible for a data element to be a component of more than one data set. For example, the data element 'Establishment identifier' is also a component of the Public Hospital Establishments NMDS.

A DSS specifies the conditions under which the data elements are collected. A DSS can define the sequence in which data elements are included, whether they are mandatory, what verification rules should be employed and the characteristics of the collection (for example, its scope).

4.1.1 Object Class

An object class is the identified 'thing' of interest for which the data developer is seeking to collect and store data. An object class has explicit boundaries, properties or attributes and meaning. Examples of an object class include car, person, household, employee, purchase order, client, hospital, agency, housing assistance agency, event and episode of admitted patient care.

Depending on what information is sought and what questions need to be answered, the broad entities (such as person, hospital, agency or event) may be further specialised or sub-typed to refine the 'thing' being described. For example, if a data developer were interested in grouping the population by sex, then 'person' can be further sub-typed into 'female' and 'male'. The sub-type becomes the Object Class.

A single object class can be sub-typed to more clearly describe the 'thing' of interest, where a greater degree of specialisation is required. A single object class can be sub-typed on more than one 'dimension'.

EXAMPLE—'Person (object class)'

sub-typed by age group

Child
Adult

sub-typed by Sex

Female
Male

EXAMPLE—'Hospital (object class)'

(sub-typed by primary funding source)

Public hospital
Private hospital

EXAMPLE—'Hospital'

(sub-typed by teaching status)

Teaching hospital
Non-teaching hospital

It is possible that during the course of a data development exercise other object classes are identified in addition to the ones already identified.

Question to ask when trying to identify the object class:

What is the 'thing' being described for which information is required?

4.1.2 Property

Having identified the object class, the next step is to identify the attributes, characteristics or aspects of the object class that the data developer is interested in describing or understanding. A property is the attribute, characteristic or aspect common to all members of the object class that is of interest.

For example, when considering the object class 'Client' the data developer may be interested in 'date of birth' or the 'sex' or 'postcode', and so on. For the object class 'Public hospital' the developer may be interested in the 'Identifier', 'Capital expenditure', 'Full-time equivalent staff', and so on.

In some instance it may not be easy to differentiate between a property and an object class. For example, an 'Identifier' may be defined as a property of a 'Person' (object class = person, property = identifier). In another context, the object class could be 'Identifier' and the property could be 'Number or designation', 'Name', 'Issuer', and so on.

The identification of object classes and properties is therefore contextual.

Questions to ask when trying to identify and define the property/ies.

What aspect, characteristic, property or attribute of the 'thing' does one want to describe? What does one need to know about the object class?

4.1.3 Data Element Concept (DEC)

A DEC is a concept that can be represented in the form of a data element, described independently of any particular representation. A data element concept is in fact a data element without representation.

A data element concept is a concept created by the union of an object class and a property. That is, one object class and one property are joined to form a data element concept.

Object class + Property = Data Element Concept

For example, if we are interested in the concept of whether a person lives alone or with others, the object class 'Person' could be associated with the property named 'Living Arrangement' in order to create a data element concept 'Person—living arrangement'.

As a data element, this DEC could be represented as:

Code	Description
1	lives alone
2	lives with others

Other examples of data element concepts include the day a person was born ('Person–date of birth') and the colour of a vehicle ('Vehicle–colour'). The basic characteristic of the DEC is the ability to define what we want to know, without the need to specify exactly how it should be represented.

If an existing data element is to be re-used, it is crucial that the concept that the data element needs to define is the same as that of the existing data element; otherwise a new data element should be created. In cases where the representation of the data are different for different data sets, it would be necessary to create two separate data elements. For example, an adult data set would specify measurement of a person's height in metres, while a neonatal data set would require a data element that specifies measurement in centimetres. In this case, conceptually the two data elements are the same, but, because the concept is represented differently, separate data elements are required.

Understanding the concept is crucial and this is why the data element concept and its component object class and property must be clearly defined in accordance with agreed standards in order to facilitate comparison.

4.1.4 Value Domain

A value domain provides the valid values and representation for the concept defined by a data element. The representational component of a data element is about the permitted values a data element may use. The set of these permitted values for a data element is called a value domain (VD). Each data element is only associated with one value domain.

A value domain can either be enumerated, where the value domain is specified by a list of all its permissible values (for example, 1=lives alone, 2 = lives with others, and so on), or non-enumerated, where the value domain is specified by a description rather than a list of all permissible values (for example, where values are specified by a range of values, such as age, height, weight, and so on).

An enumerated value domain contains a list of all its values and their associated value meanings. Each value and meaning pair is called a permissible value.

One of the first things to consider when developing data elements is how the concept represented by the data element is to be represented. The representation describes the form of the data, including a value domain, data type, representation class (optionally), format and, if necessary, a unit of measurement.

For example, the data element representing the concept 'date of commencement of service' may have dates in the format DDMMYYYY as valid values. This is its value domain. Or the data element representing the concept 'annual household income' may have the set of non-negative integers with units of Australian dollars, as a set of valid values. This is its value domain.

When selecting or developing value domains, it is important to ensure that they are consistent and mappable to national or international data standards, where these exist.

Where permissible values implement a classification or code set, the edition or version of the classification or code set should be referenced. Specifying the revision, edition or year of the classification enables consistent and comparable data collection, which would otherwise be left to assumptions being made about which version of the classification was or is used. For example, the edition or version of the classification should be referenced as part of the name of the value domain, such as 'Country code (SACC 1998) NNNN'

Questions to ask to help determine the value domain.

How do we want to represent the concept? What are the possible values that the concept can have?

Granularity

Depending on its purposes, value domains may be expressed to varying degrees of granularity. For example, the value domain for 'Eastern Asian Language Spoken' can be expressed as:

Eastern Asian Language Spoken

Value	***Value Meaning***
71	*Chinese*
72	*Japanese*
73	*Korean*
79	*Other Eastern Asian languages*

or

Eastern Asian Language Spoken

Value	***Value Meaning***
Chinese	
7101	*Cantonese*
7102	*Hakka*
7103	*Hokkien*
7104	*Mandarin*
7105	*Teochew*
7106	*Wu*
7199	*Other Chinese languages (includes Chang Chow, Hunan, Kan)*
Japanese	
7201	*Japanese*
Korean	
7301	*Korean*
Other Eastern Asian languages	
7999	*Other Eastern Asian languages (includes Ainu, Bhotia, Tibetan)*

The second Eastern Asian Language spoken value domain is expressed to a finer level of granularity. It is important that it is possible to map between value domains to ensure comparability, where this is required.

Alignment of requirements identified through the combined top-down and bottom-up approach to data development is frequently achieved through mapping of varying degrees of granularity of terms to recognised classifications. More detailed information is usually preferred for primary purposes, whereas less granular, but more stable, terms are often required for statistical purposes.

Compliance and consistency

In order to achieve semantic interoperability when data are exchanged between two parties or systems, the data must meet the data standards that are agreed to by the parties involved. Data exchanged between two parties is considered to be:

- **compliant**—when it meets all the requirements of the agreed data standard
- **consistent**—if the definition of the data element is the same as the agreed standard, but the Data type, Representation class, Format, Maximum field size or Data Domains are

different, and the data are still convertible to the agreed data standard without loss of the meaning.

Examples are provided below using the permissible values for the Marital Status data element. It should be noted that compliance and consistency are not just achieved when comparing domain values. In this example, the permissible values for the data standard for Marital Status are:

1	Never married
2	Widowed
3	Divorced
4	Separated
5	Married (including de facto)
9	Not stated/inadequately described

Data that uses only these coding categories, numbers and labels would be considered compliant with the data standard.

Examples of consistent and inconsistent values are:

Consistent with data standard ☑		**Consistent with data standard ☑**		**Inconsistent with data standard ☒**	
S	Never married	1	Never married	s	Single
W	Widowed	2	Widowed	w	Widowed
D	Divorced	3	Divorced		
A	Separated	4	Separated	a	Separated or Divorced
M	Married (including de facto)	5	Married (excluding de facto)	m	Married
		6	De facto		
Z	Not stated/inadequately described	9	Not stated/inadequately described		
Although the codes are not compliant with the data standard, they can be mapped (that is, converted) directly to the standard codes. A data element using these permissible values could be considered 'consistent' with the data standard.		Although codes 5 and 6 are not compliant with the data standard, the data are still 'consistent'. Code 5 and code 6 data can be mapped to code 5 of the standard data domain.		'Single' is not the same as 'Never married' and may be misconstrued for persons who are divorced and now 'single' or for persons in a de facto relationship. Code 'a' cannot be mapped to the original data domain as it combines two different standard codes in one. 'Married' does not include 'de facto', which may lead to 'de facto' being recorded under any of the other codes.	

4.1.5 Classification

Enumerated value domains may be a subset of a classification, such as ICD-10-AM. A classification 'is a set of discrete exhaustive and mutually exclusive observations which can be assigned to one or more variables to be measured in the collation and/or presentation of data' (UN Glossary 2006).

Many examples of classifications exist, for example, ICD-10-AM, *the Australian Standard Classification of Religious Groups* and ICPC2+. Classifications such as these are used to classify data and are official terminological systems, recognised and endorsed by a national or international body.

Some classifications are updated more frequently than others. For example, the Australian Standard Geographical Classification (ASGC) is normally updated annually. The International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification (ICD-10-AM), which is currently in its 5th edition, is updated every two years. Where permissible values implement a classification or code set, the edition or version of the classification or code set should be referenced. Specifying the revision, edition or year of the classification enables consistent and comparable data collection, which would otherwise be left to assumptions being made about which version of the classification was or is used.

Classifications endorsed for use within the Australia health sector belong to the Australian Family of Health and Related Classifications. The development of the Australian Family of Health and Health Related Classifications is managed by the Statistical Informational Management Committee (SIMC) and the Health Data Standards Committee (HDSC) on behalf of the National Health Information Management Principal Committee (NHIMPC).

4.2 Formulating data elements

A data element is the basic unit of identifiable and definable information. It is composed of a data element concept in association with a value domain. While the meaning of the data element is essentially captured within the data element concept, the value domain specifies its representation.

Data Element Concept (DEC) + Value Domain = Data Element

For example, if one is seeking to collect information about 'the height of people', then one could create a data element to capture this basic unit of information. The data element definition might be 'A person's height in metres, measured standing in bare feet'.

A data element concept can be represented by more than one data element. In the example above, the Data Element Concept would be 'A person's height'. This could then be instantiated as height in metres, millimetres, feet or inches. Each of these possibilities could be instantiated as a separate data element that measures a person's height.

Note: A common misconception is that a data element is a code set. This is not correct. A data element is a concept that is represented with a set of values. It is this set of values (codes) that are used for completing data entry screens or manual forms. While the codes are what most people see, they need to be supported by underlying concept definitions and information about the code's use.

4.2.1 Formulating data elements – an example

Let us reconsider the 'Living arrangement' example shown below.

EXAMPLE—'Living Arrangement'

Definition:	Whether a person usually resides alone or with others	
Data type:	Numeric	
Maximum Size:	1	
Data Domain:	1	Lives alone
	2	Lives with others
	9	Not stated / Inadequately described
Guide for use:	This item does not seek to describe the quality of the arrangements, but merely the fact of the arrangement. It is recognised that this item may change on a number of occasions during the course of an episode of care.	

Identifying the data element concept

The data element concept is the conceptual information about the data element. In this example, the 'thing' that the data relates to is the person or client (that is, the object class), depending on the terminology used in the specific context. The characteristic of interest of the person (or client) is the living arrangement of the person (that is, the property). The data element concept may be identified by the concatenation of the object class and property, which in this case is Person–living arrangement. In other words, the conceptual information of interest is the person's living arrangement, which in this case is defined as whether a person usually resides alone or with others.

Identifying the value domain

In the example above, the only characteristic of interest is whether a person lives alone or with others and the data developer wants to represent this by single digit numeric (N) codes.

Defining the data element

The resulting data element would consist of the data element concept 'Person–living arrangement' and it would be represented using a numeric code. The data element would be defined as 'whether a person usually resides alone or with others, as represented by a code'.

The name of the data element is dependent on the registry's business rules. For example, in METeOR, the name of the data element would be Person–living arrangement, code N.

Data elements should describe one concept and be appropriately represented

Where data elements are use to standardise data, a data element that describes a particular concept should not be represented or be described by values that do not represent the concept. For example, the concept 'type of fish living in an aquarium' is a totally different concept from 'the presence of fish living in an aquarium'. The former might represent data such as 'guppies', 'mollies', 'goldfish', 'swordfish', 'angel fish'…'. The latter might represent data such as 'There are no fish in the aquarium'. Neither data element on its own should be used to represent both sets of data.

Right representation		Wrong Representation	
The type of fish in an aquarium		The type of fish in an aquarium	
1	guppies	0	no fish
2	mollies	1	guppies
3	goldfish	2	mollies
4	swordfish	3	goldfish
5	angel fish	4	swordfish
8	other	5	angel fish
		8	other

Where information is required about both the presence of fish in the aquarium and the type of fish present, this should be represented by two separate data elements

Right representation		Right Representation	
The type of fish in an aquarium		The presence of fish in an aquarium	
1	guppies	1	present
2	mollies	2	absent
3	goldfish		
4	swordfish		
5	angel fish		
8	other		

Note: For the purpose of data collection and to minimise the burden on collectors the mixing of concepts is acceptable on collection forms as follows:

Types of fish in an aquarium

0	no fish
1	guppies
2	mollies
3	goldfish

4.3 Guidelines for good data standards

This section provides guidelines for creating good data standards, illustrated where possible using good and poor examples.

4.3.1 Name—guidelines

All data standards items in a metadata registry should have at least one name. Each registry establishes its own naming convention. ISO/IEC 11179-5 gives principles for naming data elements. It does not specify a mandatory naming convention.

The name should be

- unique in the registry for the metadata item type
- stated in the singular
- a reflection the concept being defined
- avoid using abbreviations or acronyms other than those widely accepted

4.3.2 Definition—guidelines

All data standards should have a definition. There are usually no 'wrong' definitions, just good ones and poor ones. ISO 11179-4 (2004) *Formulation of data definitions* has been reproduced in part to show the difference between a good definition and a poor definition. These include mandatory and recommended rules.

Mandatory rules

A definition must:

1. be unique within a registry and within the context of the metadata item

A data definition must be unique within a specific context within any registry and for the registration authority in which it appears. One or more characteristics expressed in the definition must differentiate its concept from other concepts.

2. be stated in the singular

The concept expressed by the data definition should be expressed in the singular. (An exception to this guideline is made where the concept itself is plural.)

EXAMPLE — 'Country name'

Good definition: *The commonly known short name of a country.*

Poor definition: *The commonly known short name of countries.*

REASON *– The poor definition uses the plural word 'countries', which is ambiguous, as it implies that a 'Country name' could refer to more than one country.*

3. state what the concept is, not what it is not

When constructing definitions, the concept cannot be defined exclusively by stating what the concept is not.

EXAMPLE —'Country short name'

Good definition*: The ISO recognised short name of a country.*

Poor definition: *The ISO recognised name that is not the long name of a country.*

REASON *– The poor definition does not specify what is included in the meaning of the data.*

4. be stated as a descriptive phrase or sentence

A phrase is necessary (in most languages) to form a precise definition that includes the essential characteristics of the concept. Simply stating one or more synonym(s) is insufficient. Simply restating the words of the name in a different order is insufficient. If more than a descriptive phrase is needed, use complete, grammatically correct sentences.

EXAMPLE —'Country name'

Good definition: *The commonly known short name that identifies a country.*

Poor definition: *Name of a country.*

REASON – *The poor definition does not describe the concept, that this is the short name, not an expanded or long name.*

5. contain only commonly understood abbreviations

Understanding the meaning of an abbreviation, including acronyms and initials, is usually confined to a certain environment. In other environments the same abbreviation can cause misinterpretation or confusion. Therefore, to avoid ambiguity, full words, not abbreviations, should be included in the definition.

Exceptions to this requirement may be made if an abbreviation is commonly understood such as 'i.e. ' and 'e.g.' or if an abbreviation is more readily understood than the full form of a complex term, and has been adopted as a term in its own right such as 'radar' standing for 'radio detecting and ranging'. All acronyms must be expanded on the first occurrence.

EXAMPLE —'Unit of Density Measurement'

Good definition: *The code that represents the unit for measuring mass per unit (m.p.u.) volume.*

Poor definition: *The code that represents the unit for measuring the m.p.u. volume.*

REASON— *m.p.u. is not a common abbreviation, and its meaning may not be understood by some users.*

6. be expressed without embedded definitions

The definition of a second data element or related concept should not appear in the definition of the primary data element.

EXAMPLE — 'Severe Hypoglycaemia History'

Good definition: *A code that represents whether a person has had severe hypoglycaemia in the last 12 months.*

Poor definition: *A code that represents whether a person has had severe hypoglycaemia, which is defined as hypoglycaemia requiring assistance from another party, in the last 12 months.*

REASON— *The poor definition contains a concept definition for severe hypoglycaemia, which should be included in a glossary.*

Recommended rules

A data definition should:

1. state the essential meaning of the concept

All primary characteristics of the concept represented should appear in the definition. The inclusion of non-essential characteristics should be avoided.

EXAMPLE—'Date of birth'

Good definition: *The date a person was born.*

Poor definition: *The date of birth of the person as stated in the birth certificate.*

REASON *– The poor definition includes extraneous material. How or where the information is captured can be stated in the collection methods. This information does not serve to define the concept.*

2. be precise and unambiguous

The exact meaning and interpretation of the defined concept should be apparent from the definition. A definition should be clear enough to allow for only one possible interpretation.

EXAMPLE —'Country code'

Good definition: *The 2-character alphabetic code assigned by the International Standard Organisation (ISO) 3166-1 to represent a country.*

Poor definition: *The code that represents a country.*

REASON *– Codes are assigned by ISO 3166-1:1997, the ABS, Standards Australia, and so on. Some codes are alphabetic and others are numeric and the number of characters can also vary. The poor definition is imprecise, making it difficult to determine the source of the code.*

3. be concise

The definition should be brief and comprehensive. Extraneous qualifying phrases, such as, 'for the purpose of this metadata registry' or 'terms to be described' should be avoided.

EXAMPLE—'Family Name'

Good definition: *A name shared in common to all members of a family, as distinguished from each member's given name.*

Poor definition: A name shared in common to all members of a family for purposes of identification, as distinguished from each member's given name.

REASON – The poor definition contains the extraneous qualifying phrase 'for purposes of identification'.

4. be able to stand alone

The meaning of the concept should be apparent from the definition. Additional explanations or references should not be necessary for understanding the meaning of the definition.

EXAMPLE—'School Location City Name'

Good definition: *Name of the city where a school is situated.*

Poor definition: *See "school site".*

REASON— *The poor definition does not stand alone, it requires the aid of a second definition (school site) to understand the meaning of the first definition.*

5. avoid circular reasoning

Two definitions should not be defined in terms of each other. A definition should not use another concept's definition as its definition.

EXAMPLE— two data elements with poor definitions:

a) Employee ID Number—Number assigned to an employee.

b) Employee—Person corresponding to the employee ID number.

REASON— Each definition refers to the other definition for its meaning. Accordingly, the meaning is not given in either definition.

6. use the same terminology and consistent logical structure for related definitions

A common terminology and syntax should be used for similar or associated definitions. Where the terminology and syntax are not the same, a user might assume that there is an implied difference between related definitions.

EXAMPLE—Good Consistency

The code that represents the method to determine the vertical coordinate.

The name of the method used to determine the vertical coordinate.

The name of the method used to determine the horizontal coordinates.

EXAMPLE—Poor Consistency

The name of the method used to determine the horizontal coordinates

The code that represents the method used to determine the latitude and longitude

REASON— *As the terminology is different (horizontal coordinates versus latitude and longitude), the registry user might assume that the different terms have a somewhat different meaning, even though they are simply different representations of the same concept.*

Data element concept definitions versus data element definitions

The definition of a data element concept should not make any reference to a specific representation. However, a data element definition must make reference to the specific representation.

EXAMPLE—Data Element Concept Definition

Data element concept: *'Job Grade Maximum Salary Amount'*

Definition: *The maximum salary permitted for the associated job grade.*

Note: The data element concept makes no reference to a specific value domain.

EXAMPLE—Data Element Definition

Data element 1: *'European Job Grade Maximum Salary Amount'*

Definition: *The maximum salary permitted for the associated job grade expressed in Euros.*

Data element 2: *'U.S. Job Grade Maximum Salary Amount'*

Definition: *The maximum salary permitted for the associated job grade expressed in U.S. dollars.*

Note: Data element definitions may refer to explicit values domains, since this may be all that distinguishes two data elements.

Since the data element definition always includes representation, as a guide, the phrase that defines the data element should begin (or end) by stating the representation class for the data element and its value domain. The definite article 'the' is used, because the definition refers to a one instance of a data value.

Name:	*The name of …*
Code:	*The code that represents…*
or	*…as represented by a code*
Number:	*The number assigned by …*
or	*The number that represents*
or	*…as represented by a number*
Measure:	*The measure of the (area, mass, distance)…*
Quantity:	*The (sum, amount, capacity) of…*
Date:	*The date expressed as DDMMYYYY when …*

4.3.3 Context – guidelines

Data standards may have a context in which the definition has meaning. If the context is left blank, this will imply that the meaning imparted by the definition is valid in all contexts.

The context could be the setting in which data collection or use is valid (for example, juvenile justice or intensive care), or it could be a whole sector of service or care (such as custodial services or admitted patient care), or it could be more general, covering the whole of the health sector, or across service sectors.

Some examples of appropriate use of the context include:

- Public health
- a community aged care program
- a supported accommodation assistance program
- an emergency department.

Information about why the data element is important for collection should not be included in the context—for example, 'this item is collected for the analysis of outcome by treatment'. The purpose of justification for a data element may be included in the 'Comments' attribute or may be more appropriate as part of the data set specification (DSS).

4.3.4 Value domain—guidelines

The definition of the value domain can be plural, because it encompasses all permissible values for that domain. This is in contrast to the definition of the object class, property, data element concept and data element, where the concept is expressed in the singular

FOR EXAMPLE

The set of permissible values for……….

or

The code set representing………

Associated with a value domain are other attributes that make it easy to distinguish among data elements in a metadata registry. These include representational class, format, data type, maximum length, and so on, and are useful to help differentiate between data elements. For example, a data element categorised with a data type of 'string' is different from an element categorised as 'number'. It will not make sense to compare their contents or perform calculations. It is therefore useful to convey these representational attributes of a data element.

Representation class

Examples of representation class values and their associated meaning

Value	Meaning
Average	A numeric value representing an arithmetic mean
Code	A system of valid symbols that substitute for longer values
Date	A numeric value representing a calendar date (that is, day, month and year) or recognised part of a calendar date (that is, day, month, and/or year)
Identifier	A value which establishes identity
Percentage	Parts per hundred
Ratio	An expression of the quantity of one substance or entity in relation to that of another
Text	An unformatted, descriptive value
Time	A numeric value representing a specific instance in time
Total	A numeric value representing the sum of a set of values or an entire quantity (including monetary)

Data types

Examples of data types values and their associated meaning

Value	Meaning
Boolean	A binary value expressed using a string e.g. true or false.
Currency	A numeric value expressed using a particular medium of exchange.
Date/Time	A specific instance of time expressed in numeric form.
Number	A sequence of numeric characters which may contain decimals, excluding codes with ‘leading’ characters e.g. ‘01’,’02’,’03’.
String	A sequence of alphabetic and/or numeric characters, including ‘leading’ characters e.g. ‘01’,‘02’,‘03’.

Format

Examples of format values and their associated meaning

Value	Valid character range
A	Alphabetic character set: contains the letters a-z and A-Z and may contain special characters, but not numeric characters
N	Numeric character set: contains whole and decimal numbers and may contain special characters, but not alphabetic characters
X	Alphanumeric character set: contains alphabetic and numeric characters, and may contain blank characters
D	A numeric character representing a number of days
M	A numeric character representing a number of months
Y	A numeric character representing a number of years
H	Any numeric character representing a number of hours
M	Any numeric character representing a number of minutes
S	Any numeric character representing number of seconds

Unit of measure

Unit of measure—some value domains require that values for a data element be measured in only one unit (for example, height in centimetres). This attribute contains the name of the unit of measure for all data values for the value domain.

Examples of Units of measure classified by measure

Measurement	Unit of measure name	Unit of measure symbol
Concentration	Microgram per litre	µg/l
	Milligram per 24-hour period	mg/24h
	Nanogram per decilitre	ng/dl
Currency	Australian currency	AU$
Length	Centimetre	cm
	Millimetre	mm
Temperature	Degree Celsius	°C
Time	Second	s
	Minute	min
	Hour	h
	Day	d
Weight	Gram	g

4.3.5 Permissible values—guidelines

The value domain is specified by a list of all its permissible values. In a value domain, the permissible values must be exhaustive within the value domain and mutually exclusive with the value domain.

Exhaustive within the value domain

In the example below, the permissible values listed is not a full set of 'living thing'. Plants, microscopic animals (including some things that are both plant and animal), viruses and bacteria are missing.

<u>Example</u>—'Living thing'

Permissible values	Value	Meaning
	1	Fish
	2	Bird
	3	Mammal
	4	Reptile

'Other' should be used to ensure an exhaustive set of permissible values. See the example below. The use of the term 'Other' by itself actually signifies 'other living thing' or 'other living thing not elsewhere classified (more often recorded as "Living thing nec")'.

Permissible values	Value	Meaning
	1	Fish
	2	Bird
	3	Mammal
	4	Reptile
	8	**Other**

Mutually exclusive within the value domain

In the example below, the permissible values listed are not mutually exclusive. Marsupials are a subset of mammals. This means that a kangaroo could be counted twice, as both a mammal and a marsupial. Codes 3 and 5 are therefore not mutually exclusive.

Permissible values	Value	Meaning
	1	Fish
	2	Bird
	3	Mammal
	4	Reptile
	5	Marsupial
	8	Other

To fix this, the value domain needs to include a proviso in the ‘mammals’ value meaning, as shown below.

Permissible values	Value	Meaning
	1	Fish
	2	Bird
	3	Mammal other than a marsupial
	4	Reptile
	5	Marsupial
	8	Other

Use of the ‘Other’ permissible value

When using ‘Other’, to ensure an exhaustive set of permissible values, using a code value that is contiguous with the last code in the permissible value sequence should be avoided.

The following set of permissible values contains a code for ‘other’ that is contiguous with the other enumerated values:

Permissible values	Value	Meaning
	1	Fish
	2	Bird
	3	Mammal
	4	Reptile
	5	**Other**

This allows another enumerated category to be added to the list of permissible values without renumbering the codes or assigning a code that appears out of place in the list.

Otherwise, this would lead to a set of values as follows:

Permissible values	Value	Meaning
	1	Fish
	2	Bird
	3	Mammal
	4	Reptile
	5	**Other**
	6	Plant

or

Value	Meaning
1	Fish
2	Bird
3	Mammal
4	Reptile
5	Plant
6	**Other**

In a longitudinal data collection, renumbering of codes can be problematic as this would mean that the data for the new collection period would be different from previous periods.

If the original set of values was

Permissible values	Value	Meaning
	1	Fish
	2	Bird
	3	Mammal
	4	Reptile
	8	**Other**

it could have been easily changed to:

Permissible values	Value	Meaning
	1	Fish
	2	Bird
	3	Mammal
	4	Reptile
	5	Plant
	8	**Other**

This leaves more space for future expansion if desired.

- Avoid the use of a value for 'Other' that is commonly used as a supplementary value. This means avoiding the use of 97, 98, 99, 997, 998, 999, 9997, 9998, 9999 for the coded value of 'Other'.

Use of Supplementary values

- Data that are collected for statistical purposes needs to be precisely defined and enumerated to provide meaningful statistical information. To this end, it is important that 'missing information' be understood for statistical purposes. This means that such data will nearly always need a supplementary value to capture missing values in order to accommodate statistical analysis. To limit variations in the meaning within a specific data collection, a default supplementary value meaning, such as "Not stated/inadequately described" is used.

  For a single-character field, where possible, use:
  9 Not stated/inadequately described
  For a two-character field, where possible, use:
  99 Not stated/inadequately described
  For a three-character field, where possible, use:
  999 Not stated/inadequately described
  and so on.

- When using more than one supplementary value, use a logical set. This ensures consistent use of supplementary code values and their meanings in a value domain.

  Examples

7 (or 97, or 997 and so on)	Not applicable
8 (or 98, or 998 and so on)	Unknown
9 (or 99, or 999 and so on)	Not stated/inadequately described

  If the three supplementary values shown above are commonly used in a data set, the same code/value meaning pairs should be used in every data element. If a Supplementary value of 'Unknown' is required, but a supplementary value for 'not stated' is not required, the Code/Value meaning pair '98 Unknown' should be used in a two-character field. The entry '99 Unknown' should be avoided. This would mean that in a data collection 'unknown' would never have the coded value of 99 in one data element and 98 in another.
- Data collected that is not of statistical significance does not need a supplementary value. However, there is an exception to this rule: if the collection owners want to know the reasons for missing values (such as 'data provider refused to provide the information' or 'data provider was unable to obtain the information at this time' or 'data provider did not know the information'), then supplementary values are appropriate as possible cues for further action.
- The value 'Other' is a 'Permissible value' and is not a 'Supplementary value'. Any value that is a synonym of 'Other' (such as 'Living thing not elsewhere classified' or 'Living thing nec') is also a 'Permissible value' and is not a supplementary value.
- Just as with the use of the 'Other' permissible value, the use of values that are contiguous with the last code in the permissible value sequence should be avoided. This allows another enumerated category to be added to the list of permissible values without renumbering the codes or assigning a code that appears out of place in the list.
- It is preferable that the supplementary value field size is the same number of characters as the other permissible values in the value domain. However, this may sometimes not be possible. A set of permissible values may use up all of the codes for its specified field length, for example a single numeric field length file may use all coded values from 1 to 9. In such cases, if supplementary values are required, the field size of the data element must be increased to 2 to accommodate them
- In non-enumerated value domains (that is those without defined value meanings, such as in a measurement) the supplementary value should be one that is not possible to achieve based on the definition of a value domain. If a data element collects the measurement of the height of a person in centimetres, it would be permissible to use 997, 998 and 999 as supplementary values because the probability of a person being that tall approaches zero. It would not be valid to use a supplementary value of 99 as this could be an actual value collected.

4.3.6 Guide for use—guidelines

Guide-for-use information is intended to provide advice or interpretation on how to use particular data standards.

The Guide for Use should include information about what is collected or recorded, rather than information about how the data are collected (or recorded). Information that can be included in the Guide for Use includes:

- The meaning or interpretation of values or codes; for example, 'The start date of treatment is recorded regardless of whether treatment is completed as intended or not'.

  Or

  CODE 3 Residential aged care service
  Includes nursing home beds in acute care hospitals.

- What is collected or included; for example 'Collected for radiation therapy and systemic therapy only. Date of surgical treatment is collected as a separate item'.
- What is excluded or not collected; for example, 'Does not include services provided through community health settings, such as community and child health centres'.
- Number of values to be collected; for example, 'More than one value can be recorded' or 'Each surgical procedure used should be recorded'.

4.3.7 Collection methods – guidelines

Collection methods include information about how data are to be collected. As such the collection methods attribute is only relevant to data elements. They do not apply to the object class, property, data element concept or value domain. Exceptions may be made for classifications and data set specifications.

In the collection methods attribute, include information about how the data are to collected or recorded, rather than information about what is collected, and the interpretation of codes or values. Information that may be included in the collection methods includes:

- Instruments or tools used; for example, 'data are obtained by asking the following question …' or 'data compiled from service records'.
- Who collects the data; for example, 'measurement of lipid levels should be carried out by laboratories which have been accredited by the National Association of Testing authorities'.
- Period for which the data are collated and reported; for example 'financial year ending 30 June each year'.
- How to record; for example 'The full name of the agent should be recorded if the coding manual is not available'.
- Other related data that are collected in conjunction; for example, 'if codes 1 or 2 are recorded, the dose of radiation received should also be recorded'. Or 'collected in conjunction with triage time'.

Recommended questions to ask when collecting data can also be included in the collection methods attribute. Standard questions, such as those recommended by the ABS, should be used where possible. If these are not available and questions have to be developed, the data developer should bear in mind that the wording of a question can result in different responses. Leading questions that give the impression that there is a correct response should be avoided. Asking two questions within the one question should be avoided. Questions should be kept short and simple, using language that is easily understood.

5 METeOR

5.1 What is METeOR?

METeOR is an electronically accessible up-to-date source of data standards for national data standards across the health, community services and housing assistance sectors in Australia. It is available at http://meteor.aihw.gov.au

METeOR integrates and presents information about:

- the National Health Data Dictionary
- the National Community Services Data Dictionary
- the National Housing Assistance Data Dictionary
- National Minimum Data Sets
- Data Set Specifications

METeOR provides:

- a user-friendly interface and functionality for users to search and browse the registry
- comprehensive background information, as well as help screens to assist in interpreting onscreen material
- assistance and tools for developing metadata, including private workspaces for developer groups to review and develop their draft data standards and context-specific online guidelines for developing data standards
- a tool for submitting developed data standards online to the national data committees for review, endorsement and registration
- enhanced functionality for those with the task of maintaining the metadata content
- a facility for tagging data standards such that an e-mail notification is received when the data standard is changed.

METeOR is the replacement for the AIHW's previous electronic data standards registry—the Knowledgebase. The Knowledgebase design was based on the previous version of the international standard for data element definition, ISO/IEC 11179 (1994), *Information Technology— Specification and standardization of data elements*. METeOR is based on the 2003 version of the ISO/IEC 11179 international standard for metadata registries.

METeOR stands for METadata Online Registry. The small 'e' in the middle of METeOR is for electronic.

5.2 Principles of METeOR

METeOR was developed based on a number of underlying principles. These principles include:

- promoting reusability of data standards and its supporting building blocks (such as data element concepts and value domains)
- allowing generic and specific data standards to coexist
- communicating different context of application of the same data standards
- ensuring that data standards are relevant and meaningful in specific contexts
- facilitating the integration and sharing of data standards across the health, community services and housing assistance sectors in order to cut metadata development costs and lead to comparable data among the sectors
- developing data standards based on ISO/IEC 11179 metadata standards
- providing standardised templates for data developers to develop and submit data standards for registration and endorsement
- providing an indication of the currency and authority of data standards.

5.3 METeOR metadata item types and attributes

Object class name

For example, person

Identifying and definitional attributes

Synonymous names: **A synonym or list of synonyms for the name within the specified context.** For example:

Human being

METeOR identifier: **A unique identifier within METeOR, automatically generated by METeOR.** For example:

269299

Registration status: **A status value for a metadata item indicating its stage in the registration process, automatically generated by METeOR**. For example:

NHIG, Standard 01/03/2005
NCSIMG, Standard 01/03/2005
NHDAMG, Standard 01/03/2005

Explanation:

This object class was approved by the National Health Information Group, the National Community Services Information Management Group and the National Housing Agreement Data Management Group as a national Standard on 01/03/ 2005.

Definition: **A concise statement that expresses the essential nature of the metadata item and its differentiation from other metadata items.** For example:

A human being, whether man, woman or child.

Specialisation of: **An instance of specialisation of an object class, generated by the Registrar.** For example:

Person/group of persons (group status)

Explanation:

The object class Person is a subtype of Person/group of persons, when categorised by group status (that is, whether one or more people are involved).

Collection and usage attributes

Comments: **Any additional information that adds to the understanding of the metadata item.**

This attribute may be left blank.

Source and reference attributes

Submitting organisation: **One or more organisations responsible for the submission of the metadata item for national endorsement as a standard.** For example:

Australian Institute of Health and Welfare

Steward:	**The name of the organisation that has accepted responsibility and been approved by a registration authority to provide ongoing maintenance and management of a metadata item.** For example: Australian Institute of Health and Welfare
Origin:	**Any document(s) (including websites) from which any content of the metadata item originates**. For example: Macquarie University 2003. The Macquarie Dictionary, 3rd ed. Sydney: The Macquarie Library Pty Ltd.
Reference documents:	**Significant documents that contributed to the development of the metadata item, but which were not direct sources for metadata content.** *If the developer of this metadata item did not refer to any documents other than the origin in the development of this object class, this attribute may be left blank.*

Relational attributes

Related metadata references:	**An indicator of the relationships between metadata items within a given sector (that is, health, community services or housing assistance).** For example: Supersedes Adult NHIG, Standard 01/03/2005 *Explanation:* *This object class replaced the object class adult as a standard in the health sector as of 01/03/2005.*
Data element concepts implementing this object class	**A list of the data element concepts that include this object class, automatically generated by METeOR** For example: 'Person—body height'— NHIG, Standard 01/03/2005 *Explanation:* *This object class is implemented in 'Person – body height' which was approved as a standard by the National Health Information Group on 01/03/2005.*

Property name

For example, body height

Identifying and definitional attributes

Synonymous name(s): **A synonym or list of synonyms for the name within the specified context.** For example:

Stature

METeOR identifier: **A unique identifier within METeOR, automatically generated by METeOR.** For example:

268955

Registration status: **A status value for a metadata item indicating its stage in the registration process, automatically generated by METeOR.** For example:

NHIG, Standard 01/03/2005

Explanation:

This property was approved by the National Health Information Group as a national Standard on 01/03/2005. This property has not been proposed as a standard to the community services or housing assistance sector registration authorities.

Definition: **A concise statement that expresses the essential nature of the metadata item and its differentiation from other metadata items.** For example:

The standing height or recumbent length of a body.

Property group: **The grouping of properties with similar characteristics, generated by the Registrar.** For example:

Physical characteristics

Explanation:

Body height is a physical characteristic.

Collection and usage attributes

Comments: **Any additional information that adds to the understanding of the metadata item.**

This attribute may be left blank.

Source and reference attributes

Submitting organisation: **One or more organisations responsible for the submission of the metadata item for national endorsement as a standard.** For example:

Australian Institute of Health and Welfare

Steward:	**The name of the organisation that has accepted responsibility and been approved by a registration authority to provide ongoing maintenance and management of a metadata item.** For example: Australian Institute of Health and Welfare
Origin:	**Any document(s) (including websites) from which any content of the metadata item originates**. For example: Macquarie University 2003. The Macquarie Dictionary 3rd ed. Sydney: The Macquarie Library Pty Ltd
Reference documents:	**Significant documents that contributed to the development of the metadata item, but which were not direct sources for metadata content.** *If the developer of this metadata item did not refer to any documents other than the origin in the development of this property, this attribute may be left blank.*

Relational attributes

Related metadata references:	**An indicator of relationships between metadata items within a given sector (that is, health, community services or housing assistance).** For example: Supersedes Height NHIG, Standard 01/03/2005 *Explanation:* *This property replaced the property height as a standard in the health sector as of 01/03/2005.*
Data element concepts implementing this property:	**A list of the data element concepts that implement this property, automatically generated by METeOR.** For example, Person—body height NHIG, Standard 01/03/2005 *Explanation:* *This property is implemented in the data element concept Person – body height which was approved as a standard by the National Health Information Group on 01/03/2005.*

Data element concept name

For example, person—body height

Identifying and definitional attributes

Synonymous name(s):	**A synonym or list of synonyms for the name within the specified context.** For example: Stature; Standing height; Recumbent length
METeOR identifier:	**A unique identifier within METeOR, automatically generated by METeOR.** For example: 269792
Registration status:	**A status value for a metadata item indicating its stage in the registration process, automatically generated by METeOR.** For example: NHIG, Standard 01/03/2005 *Explanation:* *This data element concept was approved by the National Health Information Group as a national Standard on 01/03/2005. This data element concept has not been proposed as a standard to the community services or housing assistance sector registration authorities.*
Definition:	**A concise statement that expresses the essential nature of the metadata item and its differentiation from other metadata items.** For example: The height of a person.
Context:	**A designation and/or description of the application environment or discipline in which the metadata item has meaning.** For example: Public health *This attribute may be left blank.*
Object class:	**The name of the object class implemented in this data element concept.** For example: Person *Explanation:* *The name of the object class forms the leftmost part of the data element concept name.*
Property:	**The name of the property implemented in this data element concept**. For example: Body height *Explanation:* *The name of the property forms the rightmost part of the data element concept name.*

Collection and usage attributes

Comments:	**Any additional information that adds to the understanding of the metadata item.** *This attribute may be left blank.*

Source and reference attributes

Submitting organisation:	**One or more organisations responsible for the submission of the metadata item for national endorsement as a standard.** For example: Australian Institute of Health and Welfare
Steward:	**The name of the organisation that has accepted responsibility and been approved by a registration authority to provide ongoing maintenance and management of a metadata item.** For example: Australian Institute of Health and Welfare
Origin:	**Any document(s) (including websites) from which any content of the metadata item originates.** For example: Macquarie University 2003. The Macquarie Dictionary 3rd edition. Sydney: The Macquarie Library Pty. Ltd
Reference documents:	**Significant documents that contributed to the development of the metadata item, but which were not direct sources for metadata content.** *If the developer of this metadata item did not refer to any documents other than the origin in the development of this object class, this attribute may be left blank.*

Relational attributes

Related metadata references:	*An indicator of relationships between metadata items within a given sector (that is, health, community services or housing assistance). For example:* Supersedes Adult—height NHIG, Standard 01/03/2005 *Explanation:* *This data element concept replaced the data element concept Adult—height as a standard in the health sector as of 01/03/2005.*
Data elements implementing this data element concept:	**A list of the data elements that include this data element concept, automatically generated by METeOR given the permissions of the user**. For example: Person—body height (measured), total centimetres NN[N].N NHIG, Standard 01/03/2005 Person—body height (self-reported), total centimetres NN[N] NHIG, Standard 01/03/2005 *Explanation:* *This data element concept is implemented in two data elements, which were approved as national Standards by the National Health Information Group on 01/03/2005.*

Value domain name

For example, Diagnosis code (ICD-10-AM 3rd edition) ANN{.N[N]}

Identifying and definitional attributes

Synonymous names:	**A synonym or list of synonyms for the name within the specified context.** *This attribute may be left blank.*
METeOR identifier:	**A unique identifier within METeOR, automatically generated by METeOR.** For example: 270714
Registration status:	**A status value for a metadata item indicating its stage in the registration process, automatically generated by METeOR.** For example: NHIG, Standard 01/03/2005 *Explanation:* *This value domain was approved by the National Health Information Group as a national standard on 01/-3/2005. This value domain has not been proposed as a standard to the community services or housing assistance sector registration authorities.*
Definition:	**A concise statement that expresses the essential nature of the metadata item and its differentiation from other metadata items.** For example: The ICD-10-AM (3rd edition) code set representing diagnoses.
Context:	**A designation and/or description of the application environment or discipline in which the metadata item has meaning.** For example: Public health *This attribute may be left blank.*
Classification:	**The name of the classification implemented in this value domain.** International Statistical Classification of Diseases and Related Health Problems. Tenth Revision. Australian Modification, 3rd ed *Explanation:* *This value domain implements the diagnosis codes ICD-10-AM (3rd edition).*

Representational attributes

Representation class:	**The class of representation of a value domain.** For example: Code
Data type:	**A set of distinct values, characterised by properties of those values and by the operations on those values.** For example: String

Format:	**A template for the presentation of values, including specification and layout of permitted characters, the maximum and minimum size and precision.** For example: ANN{.N[N]} *Explanation:* *This value domain has a representation of one alphabetical character followed by two numeric characters, followed by either a decimal point and one numeric character, a decimal point and two numeric characters, or nothing.*
Maximum character length:	**The maximum number of characters permitted to represent the values.** For example: 6
Permissible values:	**A list of codes and code descriptions representing values specified on a primary collection form.** For example: The diagnosis codes located within ICD-10-AM (3rd ed.) form the set of 'permissible values'. For value domains that implement classifications, this attribute should be left blank.
Supplementary values:	**A list of codes and code descriptions representing values produced in the data cleaning process (that is, they were not specified on the data collection form).** *Supplementary values within this value domain, or this attribute should be left blank.*
Unit of measure:	**The item(s) to be measured.** *This value domain is of representation class code. It is therefore not associated with a unit of measure. In such circumstances, this attribute should be left blank.*

Collection and usage attributes

Comments:	**Any additional information that adds to the understanding of the metadata item.** *Any additional comments about this value domain or this attribute may be left blank.*

Source and reference attributes

Submitting organisation:	**One or more organisations responsible for the submission of the metadata item for national endorsement as a standard.** For example: Australian Institute of Health and Welfare
Steward:	**The name of the organisation that has accepted responsibility and been approved by a registration authority to provide ongoing maintenance and management of a metadata item.** For example: Australian Institute of Health and Welfare
Origin:	**Any document(s) (including websites) from which any content of the metadata item originates.** For example: International classification of diseases 2002. Tenth revision. Australian modification, 3rd ed. Sydney: National Centre for Classification in Health.

Reference documents:	**Significant documents that contributed to the development of the metadata item, but which were not direct sources for metadata content.** *If the developer of this metadata item did not refer to any documents other than the origin in the development of this object class, this attribute may be left blank.*

Relational attributes

Related metadata references:	**An indicator of relationships between metadata items within a given sector (that is, health, community services or housing assistance).** *Relations between this value domain and other metadata items within METeOR, or this attribute may be left blank.*
Data elements implementing this value domain:	**A list of the data elements that implement this value domain, automatically generated by METeOR.** For example: Person—body height (measured), total centimetres NN[N].N NHIG, Standard 01/03/2005 Person—hip circumference (measured), total centimetres NN[N].N NHIG, Standard 01/03/2005 Person—waist circumference (measured), total centimetres NN[N].N NHIG, Standard 01/03/2005 *Explanation:* *This value domain is implemented in 3 data elements, all approved by the National Health Information Management Group as a national standard on 01/03/2005.*

Data element name

For example Person—body height, total centimetres NN[N].N

Identifying and definitional attributes

Synonymous name(s):	**A synonym or list of synonyms for the name within the specified context.** For example: Stature; Standing height; Recumbent length
METeOR identifier:	**A unique identifier within METeOR, automatically generated by METeOR.** For example: 270361
Registration status:	**A status value for a metadata item indicating its stage in the registration process, automatically generated by METeOR.** For example: NHIG, Standard 01/03/2005 *Explanation:* *The National Health Information Group approved this data element as a national Standard on 01/03/2005. This data element has not been proposed as a standard to the community services or housing assistance sector registration authorities.*
Definition:	**A concise statement that expresses the essential nature of the metadata item and its differentiation from other metadata items.** For example: A person's measured height.
Context:	**A designation and/or description of the application environment or discipline in which the metadata item has meaning.** For example: Public health *This attribute may be left blank.*
Data element concept:	**The name of the data element concept implemented in this data element.** For example: Person—body height *Explanation:* *This data element concept is the union of the object class 'Person' and the property 'Body height'.* *The name of the data element concept forms the leftmost part of the data element name.*
Value domain:	**The name of the value domain implemented in this data element.** For example: Total centimetres NN[N].N *Explanation:* *This data element is a measurement in centimetres, which accepts numeric values in the format of NN.N and NNN.N*

The name of the value domain forms the rightmost part of the data element name.

Collection and usage attributes

Guide for use: **Comments, advice or instructions for the interpretation or application of the metadata item.** For example:

Use code 999.9 Not measured when body height is not measured by the clinician, including self-reported values.

Collection methods: **Comments, advice or instructions for the actual capture of data.** For example:

All equipment, whether fixed or portable, should be checked prior to each measurement session to ensure that both the headboard and floor (or footboard) are at 90 degrees to the vertical rule. With some types of portable anthropometer it is necessary to check the correct alignment of the headboard during each measurement by means of a spirit level. Within- and, if relevant, between-observer variability should be reported. They can be assessed by the same (within-) or different (between-) observers repeating the measurement of height, on the same subjects, under standard conditions after a short time interval. The standard deviation of replicate measurements (technical error of measurement (Pederson & Gore 1996)) between observers should not exceed 5 mm and be less than 5 mm within observers.

Comments: **Any additional information that adds to the understanding of the metadata item.** For example:

For some reporting purposes, it may be desirable to present height data in categories. It is recommended that 5 cm groupings are used for this purpose.

This attribute may be left blank.

Source and reference attributes

Submitting organisation **One or more organisations responsible for the submission of the metadata item for national endorsement as a standard.** For example:

Australian Institute of Health and Welfare

Steward: *The name of the organisation that has accepted responsibility and been approved by a registration authority to provide ongoing maintenance and management of a metadata item.* For example:

Australian Institute of Health and Welfare

Origin: **Any document(s) (including websites) from which any content of the metadata item originates.** For example:

Pederson D & Gore C 1996. Anthropometry measurement error. In: Norton K and Olds T (eds). Anthropometrica. Sydney: University of New South Wales Press, 77–96

Explanation:

This is the full reference, using AIHW referencing guidelines, for the document cited within the collection methods attribute.

Reference documents:

Significant documents that contributed to the development of the metadata item, which were not direct sources for metadata content. For example:

Norton K, Whittingham N, Carter L, et al. 1996. Measurement techniques in anthropometry. In: Norton K and Olds T (eds). Anthropometrica. Sydney: University of New South Wales Press, 25–75

Explanation:

This is the full reference, using AIHW referencing guidelines, for the document used in the development of this data element concept.

Relational attributes

Related metadata references:

An indicator of relationships between metadata items within a given sector (that is, health, community services or housing assistance). For example:

Supersedes Height - measured, version 2, DE, NHDD, NHIMG, Superseded 01/03/2005.pdf (28.7 KB) NHIG, Standard 01/03/2005

Is used in the formation of Adult—body mass index (measured), ratio NN[N].N[N] NHIG, Standard 01/03/2005

Explanation:

This data element replaced the National Health Data Dictionary version of this data element as of 01/03/2005.

This data element is used in the calculation the data element Adult—body mass index (measured), ratio NN[N].N[N], which was approved as a Standard by the National Health Information Group on 01/03/2005.

Data set specifications including this data element:

A list of the data set specifications that include this data element, automatically generated by METeOR given the permissions of the user. For example:

Diabetes (clinical) NHIG, Standard 01/03/2005

Explanation:

This data element is included in the Diabetes (clinical) data set specification, which was approved as a Standard by the National Health Information Group on 01/03/2005.

Classification name

For example, International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification 3rd edition

Identifying and definitional attributes

Synonymous name:

The abbreviated title of the classification. For example:

ICD-10-AM 3rd edition

METeOR identifier:

A unique identifier within METeOR, automatically generated by METeOR. For example:

270546

Registration status:

A status value for a metadata item indicating its stage in the registration process, automatically generated by METeOR. For example:

NHIG, Standard 01/03/2005

Explanation:

This classification was approved by the National Health Information Group as a national standard on 01/03/2005. This property has not been proposed as a standard to the community services or housing assistance sector registration authorities.

Definition:

A concise statement that expresses the essential nature of the metadata item and its differentiation from other metadata items. For example:

The National Centre for Classification in Health classification for diseases, related health problems and procedures.

Context:

A designation and/or description of the application environment or discipline in which the metadata item has meaning.

Record the setting within which the classification can be applied or alternatively this attribute may be left blank.

Classification structure:

The underlying structure of a classification, such as the number and type of scales or axes within the classification. For example:

ICD-10-AM 3rd edition is composed of five volumes:

1. Tabular list of diseases. The tabular list of diseases contains the disease classification itself at the three-, four- and five-character levels. Two appendices are specified: Morphology of neoplasms; and Special tabulation lists for mortality and morbidity.
2. Alphabetic index of diseases. The index to diseases contains many diagnostic terms that do not appear in Volume 1.
3. Tabular list of procedures (ACHI). The tabular list of procedures contains the procedure classification itself. Two appendices are specified: Mapping table and ACHI codes listed in numerical order.

4. Alphabetic index of procedures (ACHI). The index of procedures contains many procedure terms that do not appear in Volume 3.
5. Australian Coding Standards. The Australian Coding Standards contains the national standards developed by the NCCH, which provide guidance in the application of ICD-10-AM codes.

Collection and usage attributes

Guide for use:	**Comments, advice, or instructions for the interpretation or application of the metadata item.** *This attribute may be left blank.*
Collection methods:	**Comments, advice, or instructions for the actual capture of data.** *This attribute may be left blank.*
Comments:	**Any additional information that adds to the understanding of the metadata item.** *Provide any additional comments on this classification or this attribute may be left blank.*

Source and reference attributes

Submitting organisation:	**One or more organisations responsible for the submission of the metadata item for national endorsement as a standard.** For example: National Centre for Classification in Health
Steward:	**The name of the organisation that has accepted responsibility and been approved by a registration authority to provide ongoing maintenance and management of a metadata item.** For example: Australian Institute of Health and Welfare
Origin:	**Any document(s) (including websites) from which any content of the metadata item originates.** For example: National Centre for Classification in Health (NCCH) 2002. The international statistical classification of diseases and related health problems. Tenth revision. Australian modification, 3rd ed. Sydney: NCCH, Faculty of Health Sciences, The University of Sydney
Reference documents:	**Significant documents that contributed to the development of the metadata item which were not direct sources for metadata content.** *This attribute may be left blank.*
Revision status:	**The status of the classification in terms of formal revisions.** For example: ICD-10-AM 3rd edition supersedes the 2nd edition of ICD-10-AM. ICD-10-AM was developed by the National Centre for Classification in Health (NCCH). During the development, the NCCH was advised by members of the NCCH Coding Standards Advisory Committee and the Clinical Classification and Coding Groups, consisting of expert clinical coders and clinicians nominated by the Australian Casemix Clinical Committee.

Relational attributes

Related metadata references:

An indicator of relationships between metadata items within a given sector (that is, health, community services or housing assistance). **For example:**

Supersedes ICD-10-AM 2nd edition NHIG, Standard 01/03/2005

Explanation:

This classification replaced the classification ICD-10-AM 2nd edition as a standard in the health sector as of 01/03/2005.

Value Domains based on this classification:

A list of the value domains that implement this classification is automatically generated by METeOR given the permissions of the user. **For example:**

Admitted patient code (ICD-10-AM 3rd edition) ANN{.N[N]} NHIG, Standard 01/03/2005

Explanation:

This classification is implemented in Admitted patient code (ICD-10-AM 3rd edition) ANN{.N[N]} which was approved as a standard by the National Health Information Group on 01/03/2005.

6 Agreements and governance arrangements for national data standards

The effective management of information requires good governance. Governance of the data sets is importance because it provides a coordinated approach for developing and managing data standards and associated data sets. It is also the only way to minimise data redundancy and unnecessary costs associated with developing data that already exists.

National, state and local agencies need to develop appropriate governance arrangements around the ongoing development, management and maintenance of their data standards to ensure that the capacity to re-use and share information is strengthened.

In Australia, national information agreements provide a framework for national data infrastructure activities in the areas of health, community services and housing (including indigenous housing) assistance. The agreements that are currently in place include:

- National Health Information Agreement (NHIA)
- National Community Services Information Agreement (NCSIA)
- National Housing Data Agreement (NHDA)
- Agreement on National Indigenous Housing Information (ANIHI)

The national information agreements are formal agreements that form the basis for collaboration and commitment to cooperate.

6.1 National Health Information Agreement (NHIA)

The National Health Information Agreement (NHIA) is an agreement between the Australian Government and state/territory government health authorities, the Australian Bureau of Statistics, the Australian Institute of Health and Welfare, the Medicare Australia and other key government agencies. The agreement was established to coordinate the development, collection and dissemination of health information in Australia, including the development, endorsement and maintenance of national data standards. It includes a commitment to cooperate through the Australian Health Ministers Advisory Council (AHMAC) agreed governance arrangements for information management and information technology.

AHMAC first endorsed the National Health Information Agreement in 1993. The current National Health Information Agreement 2004–2009, was signed by all parties and came into effect on 1 September 2004. It retains the main features and scope of the agreement developed in 1993, and incorporates amendments to reflect the significant changes to health information governance that were introduced in 2003. There is also increased emphasis on achieving consistency in data standards between the health, community services and housing assistance sectors in the current agreement.

Specifically, the objectives of the NHIA are to set out the basis on which all parties will collect, publish and use nationally consistent health information. It outlines agreed principles in relation to matters such as privacy, ownership and release of information, responsibility for the costs of developing and maintaining national health information collections and the processes for obtaining agreement on information collections, as well as the responsibilities of various parties.

The Agreement also sets out the responsibilities of the key national committees concerned with the obtaining agreement on aspects of national health information collections. These include the National Health Information Management Principal Committee (NHIMPC), the Statistical Information Management Committee (SIMC) and the Health Data Standards Committee (HDSC). Committees such as the National Advisory Group on Aboriginal and Torres Strait Islander Health Information and Data (NAGATSIHID) also have responsibility for advising the Principal Committee on information needs in relation to Aboriginal and Torres Strait Islanders and population health respectively.

6.2 National health information governance arrangements

The NHIA is administered by the National Health Information Management Principal Committee (NHIMPC), previously known as the National Health Information Group (NHIG), a standing committee of AHMAC. The Health Data Standards Committee (HDSC) is a standing subcommittee of the Principal Committee. This committee has been in operation since 1994. Prior to 2003, the committee was known as the National Health Data Committee (NHDC). In September 2004, the new National e-Health Transmission Authority (NEHTA) was established and included within the AHMAC governance arrangements. On April 2006, new governance arrangements for health information management were approved comprising of six Principal Committees and specific-purpose sub-committees to support the work of the Principal Committees. Figure 5 depicts the current committee and working group structures that operate under the AHMAC governance arrangements and which came into effect on 27 April 2006.

Any data development that is intended for national implementation needs to go through the AHMAC agreed information management governance arrangements process. Figure 5 outlines the relationships of the groups involved to each other.

The primary role of the HDSC (replacing and expanding the responsibilities of the previous committee) is coordinating the development and endorsement of national data standards used for statistical analysis, research and administrative reporting in the health sector. The HDSC works closely with the National e-Health Transition Authority (NEHTA) on those data standards that are used both for point of care (clinical) and for research and reporting purposes.

The HDSC assesses health-related data definitions proposed by data development working groups for inclusion in the Institute's online metadata registry (METeOR) and the National Health Data Dictionary (NHDD). It submits recommendations to the NHIMPC on revisions and additions to national health data definitions. All data definitions to be included in the NHDD require endorsement by the NHIMPC. All proposals that affect the collection of data mandated under the NHIA must be reviewed and endorsed by the Statistical Information Management Committee (SIMC). Under the NHIA, the Dictionary is the authoritative source of national health data definitions and related documentation.

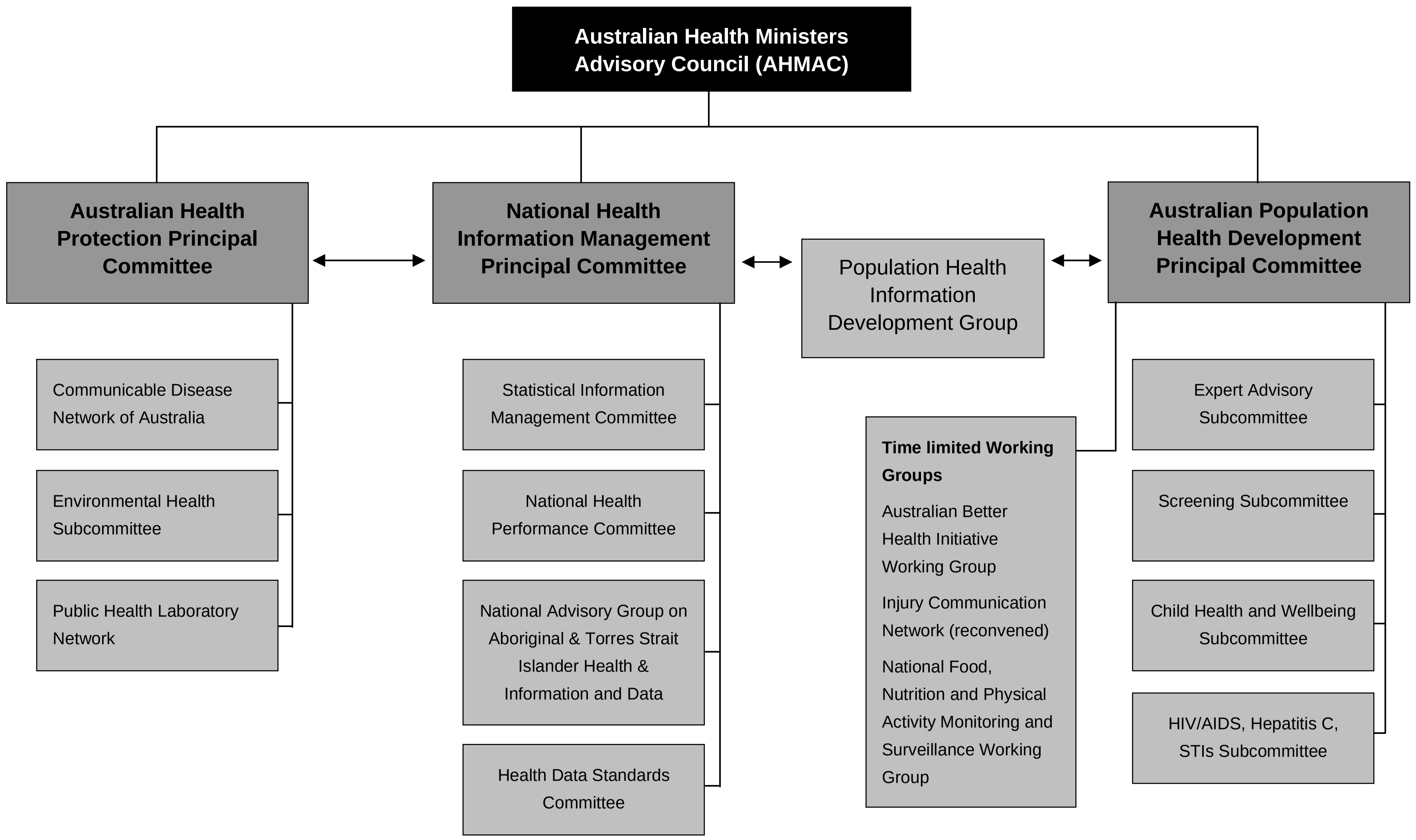

Figure 5: National health information governance arrangements

6.3 National Community Services Information Agreement (NCSIA)

The NCSIA is a multilateral agreement between the Australian Government, state and territory community services departments, Centrelink, the ABS and the AIHW to provide the framework for a cooperative approach to national, community services information development. The agreement operates under the auspices of the Community and Disability Services Ministers' Advisory Council (CDSMAC). The agreement was originally developed in 1997 and, after a major review of the agreement, it was revised and re-signed for a further 5 years in 2004.

The scope of the NCSIA includes aged care, disability services, children services, child protection and adoption, family support services, juvenile justice, gambling, homelessness services (including the supported accommodation assistance program), concessions and emergency relief services. The agreement provides a framework to support the national infrastructure and decision-making processes needed to integrate, plan and coordinate the development of consistent national community services information.

Specific objectives of the NCSIA are to ensure access to nationally consistent quality data on community services essential for policy and program development, accountability requirements and improvements in quality, efficiency and effectiveness of community services.

Core activities covered by the agreement include:

- developing and maintaining a National Community Services Data Dictionary (NCSDD) (now incorporated within METeOR) used by all data developers and collections.
- developing community services data classifications and standards.
- endorsing national minimum data sets consistent with agreed dictionary definitions and classifications.
- integrating and harmonising community services, health, housing assistance and income support data definitions, classifications and standards where possible.
- providing data linkage across community service and other sectors.
- disseminating national information.

6.4 Community services information governance arrangements

The complexity of the community services sector is reflected in the range of committees and working groups that have some influence on the development of community services information. While the two central committees established under the NCSIA are the National Community Services Information Management Group (NCSIMG) and the National Community Services Data Committee (NCSDC), a range of other committees focused on particular types of community services are also represented.

Figure 6 highlights and summarises the relationship and linkages between the NCSIMG and other organisations and groups in the national community services sector.

The NCSDC is responsible for developing and maintaining the National Community Services Data Dictionary (NCSDD) and promoting national data consistency in the community services sector. The NCSIMG is the endorsing body for the data definitions and standards recommended by the NCSDC.

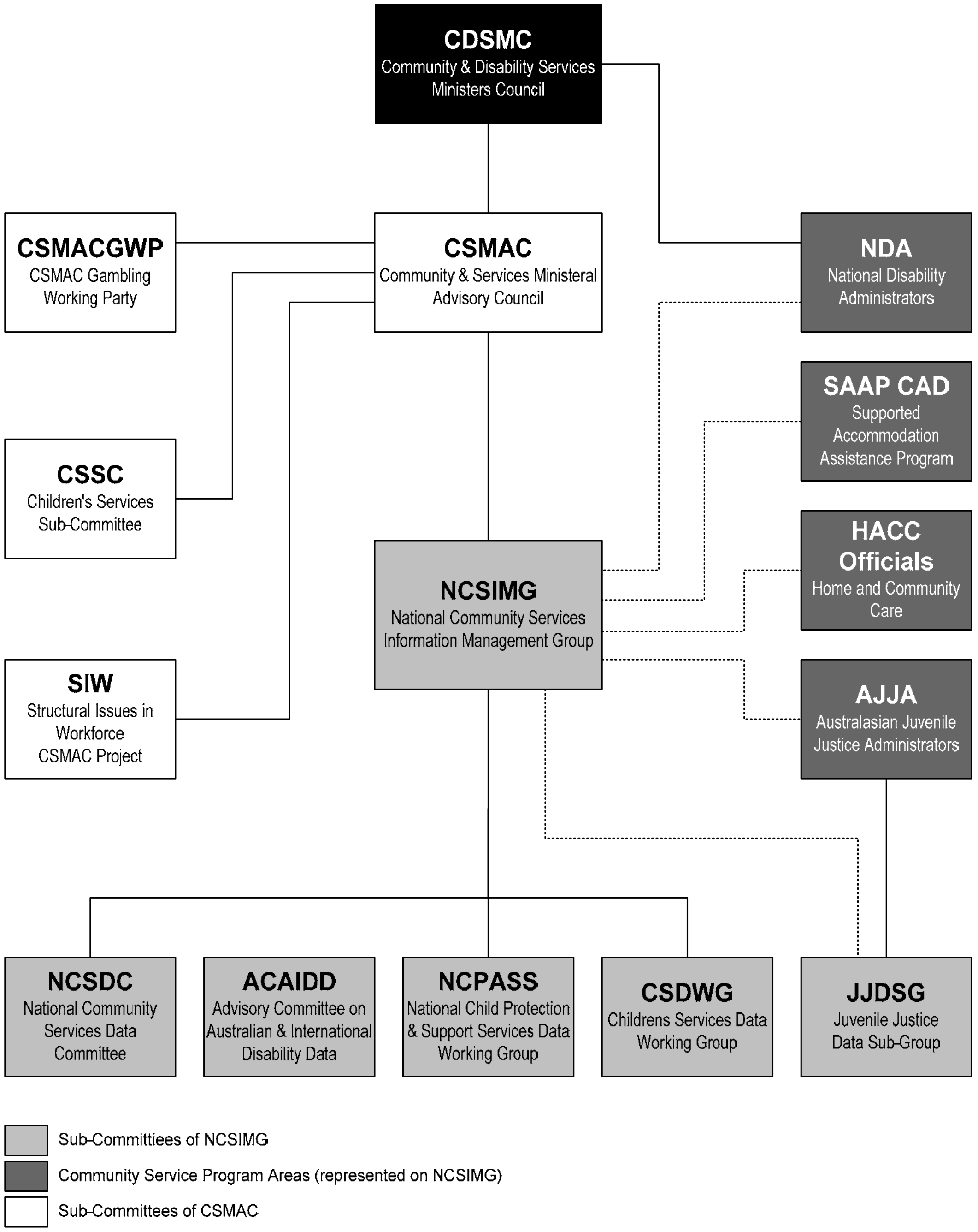

Figure 6: Community services information governance arrangements

6.5 National Housing Data Agreement (NHDA)

Parties to the NHDA are the ABS, the AIHW, the Australian Government Department of Family and Community Services, Department of Health and Human Services (TAS), Department of Housing (NSW), Department of Housing (NT), Department of Housing (QLD), Department of Human Services (SA), Department of Human Services (VIC), Department of Urban Services (ACT), Ministry of Housing (WA) and the NSW Aboriginal Housing Office. The Agreement provides the framework for a cooperative approach to national housing information development. The long-term goal of this Agreement is not only to have a workable means of obtaining nationally compatible housing data, but also to do so in a way that is compatible and consistent with other information initiatives in Australia, such as the NCSIA and the NHIA. This enables the sourcing of nationally relevant data from locally useful data and reduces the respondent burden on data collectors. The broad objectives of this agreement are to:

- develop and improve the quality of, and access to, national housing data
- reduce duplication of data development and data collections
- ensure the availability and use of compatible national data among agencies across state, territory and Commonwealth authorities and with other community and human service sectors including health.

6.6 Agreement on National Indigenous Housing Information (ANIHI)

The Agreement on National Indigenous Housing Information (ANIHI) recognises the need for consistent, accurate, reliable and timely information about housing for Indigenous Australians. The ANIHI came into effect in December 1999.

Parties to the Agreement are the Aboriginal and Torres Strait Islander Commission, the Aboriginal Housing Office (NSW), the ABS, the AIHW, the Australian Government Department of Family and Community Services, and relevant state and territory housing departments.

The goal of the Agreement is not only to have a workable means of obtaining nationally relevant Indigenous housing data, but also to do so in a way that is compatible and consistent with other information initiatives in Australia that cover other housing data and other relevant health and community services data.

6.7 Housing information governance arrangements

Figure 7 shows the relationships between the various groups involved in the governance arrangements under the NHDA and the ANIHI.

The National Committee for Housing Information (NCHI) is a national housing data committee established by the Housing Minister's Advisory Council (HMAC). It is responsible for the development and management of the national mainstream and Indigenous housing data, as outlined in the National Housing Data Agreement (NHDA) and the Agreement on National Indigenous Housing Information (ANIHI).

The Policy Research Working Group (PRWG) and the Standing Committee on Indigenous Housing (SCIH) are responsible for the allocation of housing information budget provided through HMAC. The NCHI develops and oversees the implementation of the HMAC endorsed work plan and is directly accountable to the PRWG. The NCHI has a role in the development, implementation and endorsement of national data definitions and standards for the national housing assistance data. These data standards are included in the Institute's online metadata registry (METeOR) and the National Housing Assistance Data Dictionary (NHADD). The NCHI is also responsible for overseeing the national housing assistance data collections and national housing minimum data sets.

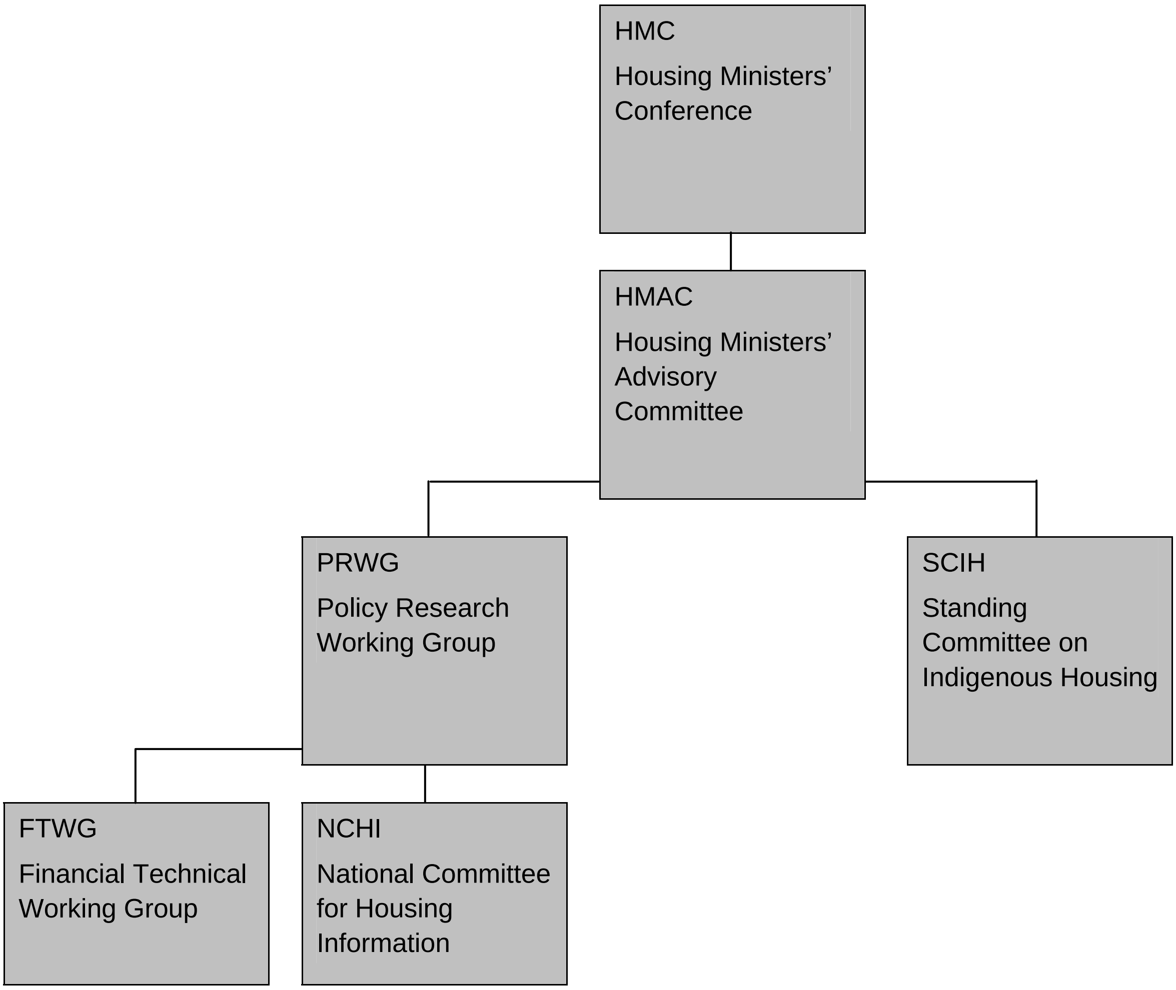

Figure 7: Housing information governance arrangements

6.8 States and territory governance arrangements

Data collections or data standards intended for use within the State or territory health service would need to be approved policy and procedures defined by the particular jurisdiction.

Many State and Territory health departments have guidelines for the development and/or implementation of data collections and information systems. This includes guidelines for developing data elements/standards and review and approval process for the development and maintenance of corporate data standards.

Figure 8 shows the approval processes that exist within Queensland Health and is an example of the data development governance process used by a State and Territory.

6.9 Local governance arrangements

Some sectors, regions or agencies may also have local arrangements for the development, approval and maintenance of data collections and data standards through metadata.

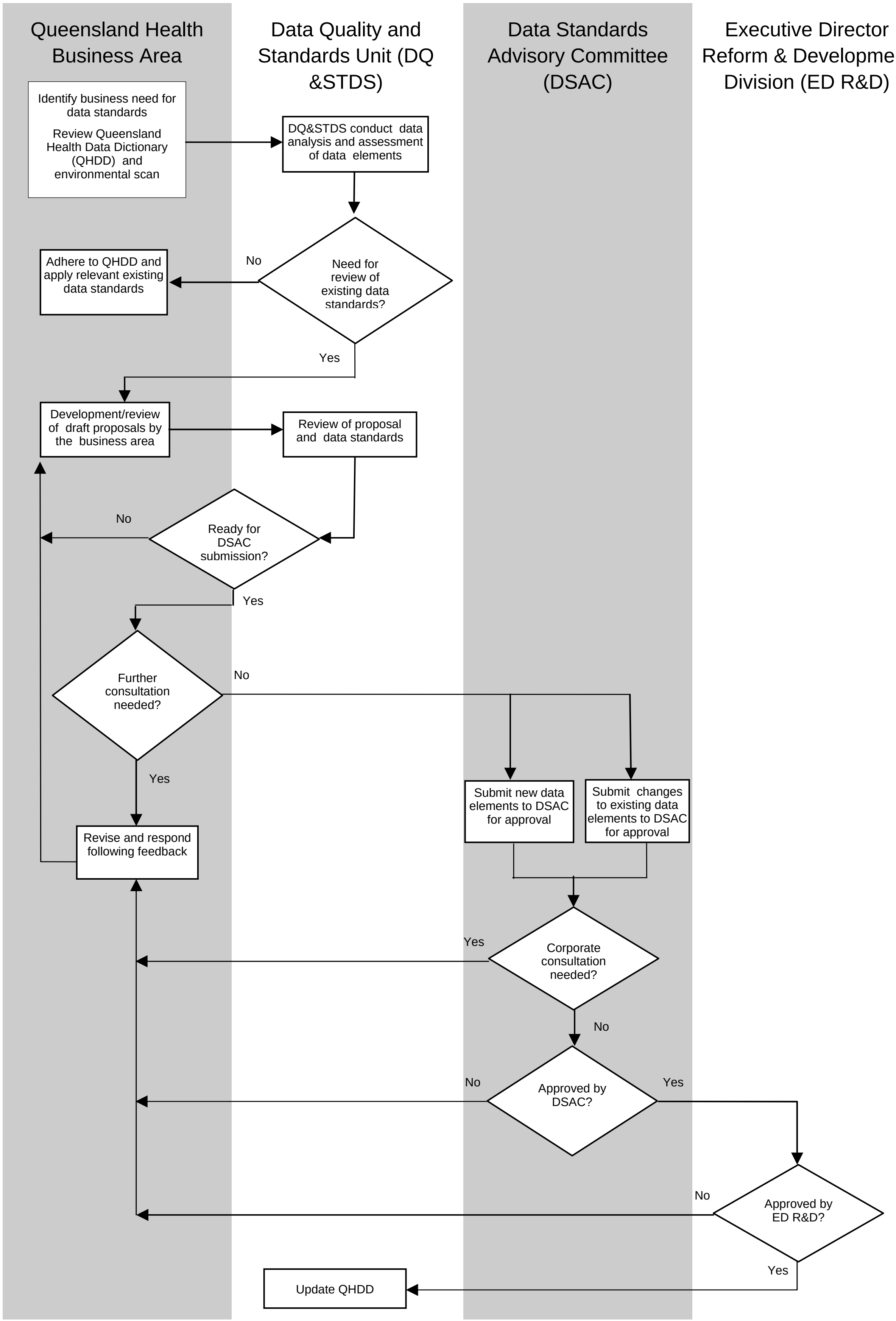

Figure 8: Queensland Health data development flowchart

7 Who to contact for help

A list of organisations that can be contacted for help in regards to data development are provided below.

For assistance with data development and the preparation of national data standards contact:

National Data Development and Standards Unit
Australian Institute of Health and Welfare
6A Traeger Court
Bruce ACT 2601
http://www.aihw.gov.au/datadevelopment/index.cfm

For more information about METeOR contact:

Metadata Management Unit
Australian Institute of Health and Welfare
6A Traeger Court
Bruce ACT 2601
http://meteor.aihw.gov.au/content/index.phtml/itemId/181162

For information about the preparation of Australian Standards and other Standards Australia and International Standards Organisation products and services contact:

Standards Australia Limited
286 Sussex Street, Sydney, NSW, 2000
GPO Box 476,
Sydney, NSW, 2001
http://www.standards.org.au/

For information about classifications contact:

Australian Bureau of Statistics
Classifications and Data Standards Unit,
Belconnen,
ACT 2616
http://www.abs.gov.au/

National Centre for Classification in Health
Faculty of Health Sciences
The University of Sydney
PO Box 170
Lidcombe NSW 1825
http://www3.fhs.usyd.edu.au/ncchwww/site/

For information about national data standards committees contact:

Health Data Standards Committee
c/o The Secretariat
Australian Institute of Health and Welfare
GPO Box 570
Canberra 2601
http://www.aihw.gov.au/committees/hdsc/index.cfm

Community Services Data Standards Committee
c/o The Secretariat
Australian Institute of Health and Welfare
GPO Box 570
Canberra ACT 2601
http://www.aihw.gov.au/committees/ncsdc/index.cfm

Community Housing Data Development Committee
c/o The Secretariat
Australian Institute of Health and Welfare
GPO Box 570
Canberra ACT 2601
http://www.aihw.gov.au/committees/index.cfm

List of tables

List of figures

References

ISO (International Organisation for Standardisation) 1995. ISO 23950. Information and documentation—Information retrieval (Z39.50)—Application service definition and protocol specification. Geneva: ISO/IEC.

ISO/IEC (International Organisation for Standardisation International Electrotechnical Commission) 2004. ISO/IEC 11179-1. Information technology—metadata registries (MDR) -- Part 1: Framework, 2nd ed. Geneva: ISO/IEC.

ISO/IEC 2000. ISO/IEC 11179-2. Information technology -- Specification and standardization of data elements -- Part 2: Classification for data elements. Geneva: ISO/IEC.

ISO/IEC 1994. ISO/IEC 11179-3. Information technology—Specification and standardization of data elements -- Part 3: Basic attributes of data elements, 1st ed. Geneva: ISO/IEC.

ISO/IEC 2003. ISO/IEC 11179-3. Information technology—metadata registries (MDR)—Part 3: Registry metamodel and basic attributes, 2nd ed. Geneva: ISO/IEC.

ISO/IEC 2004. ISO/IEC 11179-4. Information technology—metadata registries (MDR)—Part 4: Formulation of data definitions. Geneva: ISO/IEC.

ISO/IEC 2005. ISO/IEC 11179-5. Information technology —metadata registries (MDR)—Part 5: Naming and identification principles. Geneva: ISO/IEC.

ISO/IEC 2005. ISO/ IEC 11179-6. Information technology—metadata registries (MDR)—Part 6: Registration. Geneva: ISO/IEC.

Seiner, R S 2001. The importance of the data & knowledge management specialist: a commentary. The Data Administration Newsletter (TDAN.com) Issue 18, October 2001. Viewed 21 December 2005, <http://www.tdan.com/i018fe01.htm>.

Silverston, L 2001. National intelligence and the integration gap: terrorism: a call for data integration. DM Direct, 16 Nov 2001. Viewed 21 December 2005, <http://www.dmreview.com/editorial/dmreview/print_action.cfm?articleId=4290>.

Standards Australia 2005. AS 5021. The language of health concept representation. Sydney: Standards Australia.

Standards Australia 2002. AS ISO 15489.1. Records Management– General. Sydney: Standards Australia.