

# 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**

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

#### **Definitional attributes**

|                    |   |
|--------------------|---|
| <i>Definition:</i> | The age of the person in (completed) years at a specific point in time.   |
| <i>Context:</i>    | 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. |

|                              |            |
|------------------------------|------------|
| <i>Data element concept:</i> | Person-age |
|------------------------------|------------|

#### **Representational attributes**

|                                  |        |
|----------------------------------|--------|
| <i>Representation class:</i>     | Total  |
| <i>Data type:</i>                | Number |
| <i>Format:</i>                   | N[NN}  |
| <i>Maximum character length:</i> | 3      |

|                             |  |
|-----------------------------|--|
| <i>Supplementary codes:</i> | 999 Unknown/not stated   |
| <i>Unit of measure:</i>     | Year   |
| <i>Guide for use:</i>       | 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.   |
| <i>Collection methods:</i>  | 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?<br><br>What was ..... age last birthday?<br><br>What is ..... age in complete years? |

#### **Administrative attributes**

|   |  |
|---|--|
| <i>Submitting organisation:</i>                   | National Public Health Information Working Group   |
| <i>Origin:</i>                                    | Australian Bureau of Statistics, Standards for Social, Labour and Demographic Variables. Reference through:<br>< <a href="http://www.abs.gov.au/Ausstats/abs@.nsf/StatsLibrary">http://www.abs.gov.au/Ausstats/abs@.nsf/StatsLibrary</a> > |
| <i>Implementation in Data Set Specifications:</i> | 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:

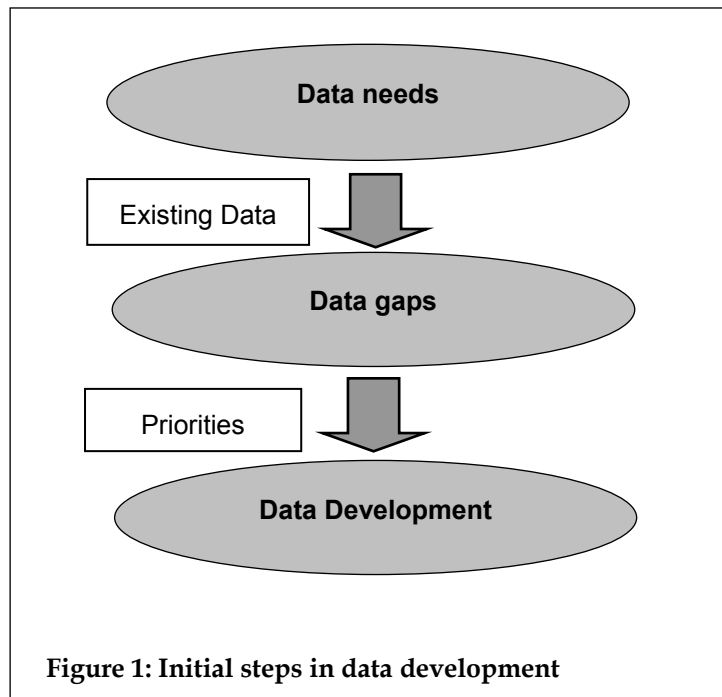
- a need for more complete information about clients and/or services
- to find out about the needs or the unmet needs of clients
- to measure services, targets and outcomes
- a need for information about the population being served
- a need for information about a new population to be serviced
- to meet 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.



### 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.

|                                       |   |
|---------------------------------------|---|
| <b>EXAMPLE – 'Living Arrangement'</b> |   |
| <b>Definition:</b>                    | Whether a person usually resides alone or with others   |
| <b>Data type:</b>                     | Numeric   |
| <b>Maximum Size:</b>                  | 1   |
| <b>Data Domain:</b>                   | 1 Lives alone<br>2 Lives with others<br>9 Not stated / Inadequately described   |
| <b>Guide for use:</b>                 | 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.