Companion guide

What are the national data dictionaries?

The national data dictionaries contain standard data definitions and data elements for use in any Australian health or community services data collection. They are the authoritative source of information about endorsed national metadata standards and provide the basis for consistent national collection and reporting. The national metadata standards are approved by the Australian government and all state and territory relevant health and community services departments as well as the Australian Bureau of Statistics (ABS) and the Australian Institute of Health and Welfare (AIHW). This work is funded under the Australian Health Minister's Advisory Council (AHMAC) and the Community Services and Disability Minister's Advisory Council (CSDMAC).

Where possible, metadata standards in the dictionary are consistent with other national standard classifications to ensure overall comparability of national data. The Australian Bureau of Statistics is the source of many key socioeconomic classifications used in data collections, for example, the Australian Standard Classification of Languages (ASCL), the Australian Standard Geographic Classification (ASGC) and the Australian and New Zealand Standard Classifications of Occupations (ANZSCO).

The data dictionaries have been downloaded from METeOR, which is an online metadata registry for developing, registering and disseminating metadata based on the second edition of the international standard ISO/IEC 11179 Information Technology-Metadata registries (MDR) in 2003. For more information about METeOR see the 'METeOR' section on page The national health and community services data dictionaries are also available online at www.meteor.aihw.gov.au.

Governance

The national health and community services data dictionaries are, respectively, initiatives under the National Health Information Agreement, auspiced by the AHMAC, and the National Community Services Information Agreement, auspiced by the CSDMAC.

Under these agreements, all parties agree to ensure that the collection, compilation and interpretation of national information are appropriate and are carried out efficiently. This requires agreement on definitions, standards and rules for collecting information, and on guidelines for coordinating the access, interpretation and publication of national health and community services information.

The national health and community services data standards are agreed to by the respective state, territory and Australian Government departments of health and community services, the ABS and the AIHW.

Why use metadata standards?

Making data count

Metadata standards promote the quality, accuracy, interpretability, reliability, relevance, interchangeability, transparency, currency, accessibility, coherence and comparability of data and information.

Without metadata standards there is the potential for data to fail to measure what it is supposed to measure, or not be comparable across collections or over time.

Metadata standards enable consistent and comparable reporting of information about services, including describing what services are available, where they are located and how much they cost. Such standards also achieve consistency when reporting about people, for instance, who delivered a service and when, who received the service, and what happened to them as a result.

By making endorsed metadata standards readily available, users are assured that they can use these standards with confidence and that they will enable the maximum reuse of their data for future research: 'create once, use often'.

How are metadata standards developed?

Creating metadata standards is part of data development

Data development is the process of building a data set for a specific purpose. For example, one might wish to build a data set that supports research into cancer diagnoses, treatment and outcomes, or one that supports statistical reporting about the operations of hospital emergency departments.

Data development demands a clear understanding of why the data are needed—for example, to underpin the design and evaluation of health policy, or to assess the performance of health services, or to support research into patterns of disease.

It also requires an assessment of the practicalities of collecting the data. Building almost any data set will demand the assembly of data from multiple sources. Thus there must be an understanding of the diverse information systems from which data are drawn; and there must be an effort to apply or develop common standards for concepts, classifications, terminologies, data values and so on.

The quality of data, including its consistency and comparability, is enhanced when metadata standards are used to support the collection and use of a data set.

Metadata standards development process

The data development process is outlined below.

---

1. Project proposal stage

A submission is made to the relevant data committee outlining the following:

- the purpose and scope of the project
- a business case for adoption

- details of national consultation, including the composition of the data working group, (comprising members with expertise in the subject area of the data development project)
- the nominated project manager responsible for managing the data development project
- expected timeframes for completion of the data development project.

2. Data development stage

The project manager works with the AIHW's METeOR and Metadata Unit and the data working group to develop the metadata items that will be stored in the national metadata registry system, METeOR.

Data collected for national reporting purposes must be defined in a standardised way. The standard adopted for health data is the International Standards Organisation's metadata registry standard (ISO/CEN 11179:2003).

For more information about the data development process, see the AIHW publication *A guide to data development*, available at:
http://www.aihw.gov.au/publications/index.cfm/title/10422.

3. Metadata standards review stage

The AIHW's METeOR and Metadata Unit undertakes a quality assurance review of the proposed metadata standards. This involves assessing all metadata standards for issues such as technical integrity, compliance with ISO 11179 and consistency with other standards, and the clarity of the content.

During this stage, the project manager receives feedback, makes adjustments and reports back to the data working group if required.

4. Endorsement of metadata standards

Once a final metadata assessment has been conducted, and the METeOR and Metadata Unit has provided a certification of the quality of the metadata standards, the metadata is submitted to the relevant endorsing body for final endorsement as a national standard.

Registration status progression

The registration status is the value assigned to a metadata item as it progresses through the standards development and review process.

The registration statuses in METeOR are:

Proposed

- A developer has submitted this item for consideration by the METeOR registrar.

Recorded

- The METeOR registrar has conducted an initial technical review and determined that the item meets basic quality criteria.

Candidate

- The item has been compliance checked by the METeOR registrar and is ready for further development or review by the data developer.

Standardisation pending

- The item has undergone a final compliance check by the METeOR registrar and the data working group has recommended the item to a registration authority for approval as a standard.

Standard

- The item has been endorsed by a registration authority as a national data standard.

Superseded

- A registration authority has superseded this item with another standard.

Retired

- The item has been nominated by registration authority as retired.

Not progressed

- The item will no longer be considered by the registrar, a data committee or a registration authority.

Who benefits from using metadata standards?

Information managers use standard formats and definitions to support the receipt, transfer, storage and management of data.

Program managers use data standards as the basis for describing information requests (that is, data required under formal service or funding agreements, and contracts); measuring service activity, client flows, client characteristics and service usage; understanding demand; better planning of services; describing unmet need (need comparability of population and survey data); and understanding ways to integrate service delivery, resulting in better targeting and usage of services and ultimately cost savings.

Researchers use data standards as the common language to support survey work, as well as to integrate data from other sources.

Policy makers need aggregated information for future policy, management and funding decisions, that is, information to support comparisons across jurisdictions, programs and sectors.

Statisticians use data standards for interpreting and analysing data; analysing results; linking data sets for statistical purposes; and carrying out time series analysis (over a period of time) and longitudinal studies (over a period of time within groups).

Metadata structure

The metadata standards are based on the 2003 version of the ISO/IEC 11179 international standard for metadata registries. Part three of the standard is a model for a metadata registry and the formulation of metadata items.
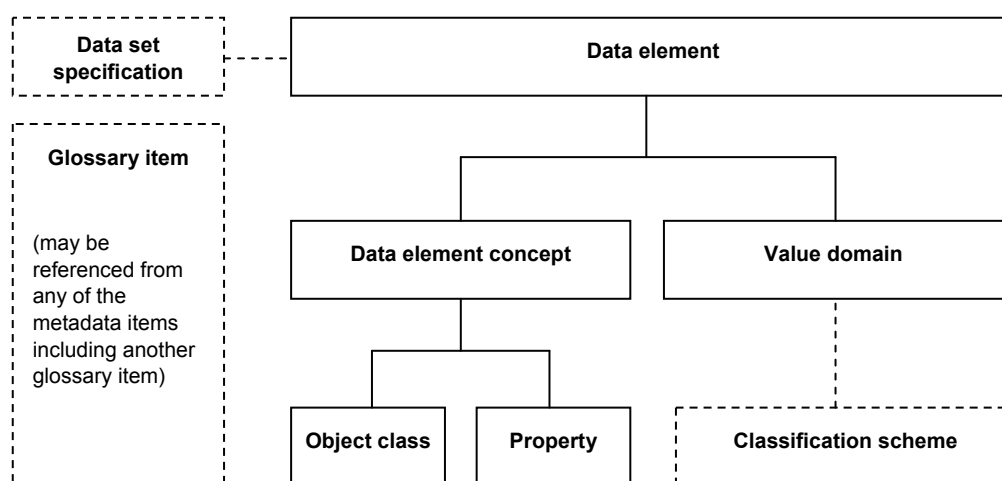
There are six types of metadata defined by ISO/IEC 11179 that have been applied to METeOR and the data dictionaries:

- object class
- property

- data element concept
- data element
- value domain
- classification scheme.

The structure underlying a data element in the dictionary is illustrated in Figure 1.

Types of metadata

Dashed lines show metadata types that are additional to or different from the ISO/IEC standard.

Figure 1: The 11179 model of metadata representation.

Components of data elements

The things that we want to know about include ideas (knowledge), persons, organisations, the environment and events. These things are termed *Object classes*. Some examples of object classes are 'Person', 'Dwelling' and 'Service provider organisation'. A characteristic of the object class is known as a *Property*. It is normally the item of interest. For example, the object class 'Person' can have properties such as 'Sex' and 'Date of birth'.

A *Data element concept* is defined as a concept created by the union of an object class and a property. Only one object class and one property can be joined for each data element concept, for example, 'Person—date of birth'.

As can be seen from Figure 1, a *Data element* is formed when a data element concept is represented in the real world by a set of values (a *Value domain*).

A value domain specifies how something is to be represented. A value domain can specify:

- the range of permitted values, for example, a measure of weight in grams represented by three numeric characters
- all permissible values as a set of codes, for example, 'Code 1 Female', 'Code 2 Male'
- the values referenced from a nationally or internationally endorsed classification, such as all codes in the ASCL 2005, or all activity codes listed in the International statistical

classification of diseases and related health problems, Tenth revision, Australian modification, 5th edition (ICD-AM).

A *Glossary item* defines the meaning of a term within a specific context. Examples of glossary items include 'Adoption' and 'Family'. These are not currently defined as object classes but their meaning must be clarified as part of the data development process.

Data set specifications

A data set specification (DSS) is a grouping of a number of data elements and the conditions under which they are collected. A DSS defines:

- the scope of the collection, that is, the population that is the target of the data development
- the level at which the data will be collected and reported
- whether the inclusion of each data element is mandatory, optional or conditional, and the sequence in which data elements are included.

National minimum data sets

A national minimum data set (NMDS) is a type of DSS, made up of a minimum set of data elements agreed by the national health or community services information groups for mandatory collection and reporting at a national level. An NMDS may include data elements that are also included in another NMDS.

An NMDS is contingent upon a national agreement to collect uniform data and to supply it as part of the national collection, but does not preclude agencies and service providers from collecting additional data to meet their own specific needs.

An example of an NMDS is the Juvenile Justice NMDS, which is used to obtain information about young people under juvenile justice supervision in Australia.

Non-mandatory data set specifications

DSSs are data sets that are not mandated for national collection and reporting, but are recommended as best practice.

For example, the Diabetes (clinical) Data Set Specification is a non-mandatory data set which aims to ensure national consistency in relation to defining, monitoring and recording information on patients diagnosed with diabetes.

Integration of data elements in data dictionaries

The national health and community services data dictionaries contain about one hundred integrated data items that can be used consistently across the health and community services sectors. This is especially important for services that straddle sector boundaries such as aged care, mental health, drug and alcohol services, and services for people with a disability.

Examples of integrated data items are:

Address data items

Address line, text [X(180)]

Australian postcode, code (Postcode datafile) {NNNN}

Australian state/territory identifier, code N

Street name, text [A(30)]

Suburb/town/locality name, text [A(50)]

Personal and demographic data items

Activity and participation life area, code (ICF 2001) AN[NNN]

Age, total years N[NN]

Person—country of birth, code (SACC 2008) NNNN

Date of birth, DDMMYYYY

Family name, text X[X(39)]

Given name, text [X(40)]

Indigenous status, code N

Informal carer existence indicator, code N

Interpreter service required, yes/no code N

Labour force status, code N

Main language other than English spoken at home, code (ASCL 2005) NN{NN}

Marital status, code N

Mother's original family name, text [X(40)]

Occupation (main), code (ANZSCO 1st edition) N[NNN]{NN}

Person identifier, XXXXX[X(14)]

Preferred language, code (ASCL 2005) NN{NN}

Proficiency in spoken English, code N

Sex, code N

Year of first arrival in Australia, date YYYY

Service provider organisation data items

Organisation end date, DDMMYYYY

Organisation name, text [X(200)]

Organisation start date, DDMMYYYY

Date items

Accuracy indicator, code AAA

Formulation of good data standards

1 A metadata item must have, at least, a name and a definition.

2 A metadata item can have an explicitly stated context within which the definition has meaning.

3 The name of the standard version of the metadata item must follow certain criteria:

- the name must be unique within the context of the metadata item
- the name must be stated in the singular
- the name must reflect the concept being defined
- the name must avoid the use of abbreviations or acronyms other than those widely accepted (for example radar, laser, pH)
- the name should avoid the use of words that imply a preselected single instance.

4 Definition rules and guidelines

A definition must:

- be stated in the singular
- state what the concept is, not only what it is not
- be stated as a descriptive phrase or sentence(s)
- contain only commonly understood abbreviations
- be expressed without embedding definitions of other data or underlying concepts.

A data definition should:

- be expressed without embedding rationale, functional usage, domain information or procedural information
- state the essential meaning of the concept
- be precise and unambiguous
- be concise
- be able to stand alone
- avoid circular reasoning
- use the same terminology and consistent logical structure for related definitions
- contain information appropriate for the type of metadata item being defined
- use a preferred term to represent the definition of a concept specified elsewhere in the document
- pass the substitution test.

5   Context should be closely linked to definition.

6   Information must be included in a metadata item attribute only if it is appropriate for that attribute or metadata item.

- Context
  - Metadata can exist within a specific context.
  - Only information that is relevant to the environment or framework within which the definition for the metadata item is valid must be included in the Context attribute.
  - The contexts of two metadata items must be compatible when the definition of one metadata item references a term defined in another.
- Guide for use

  Guide for use information must be included in any metadata item only if it is intended to provide advice or interpretation on how to use the particular metadata item or data collected using the metadata item.
  - In metadata items other than Data elements, Guide for use information should be about how to use the item itself and not about any data that can be collected or used.
  - Data elements can also include information about how to use or interpret the data in the Guide for use.
- Collection methods

  The Collection methods attribute must only include information about how data is to be collected. The following metadata items must not have a metadata attribute of Collection methods
  - Object class
  - Property
  - Data element concept
  - Value domain
  - Glossary item.
- Permissible values in a Value domain must
  - be exhaustive within the set
  - be made into an exhaustive set of values by adding an 'Other' to aggregate all other possibilities not covered by the stated set of values
  - be mutually exclusive within the set
  - be a true representation of the concept defined in the data element.
- When allocating code values
  - avoid the use of a code value for 'Other' that is contiguous with the last code in the sequence of permissible values or that, in any other way, do not provide for inclusions in the future
  - wherever possible, avoid the use of a coded value for 'Other' that may be commonly used as a supplementary value.
- Supplementary values
  - Supplementary values must not be included in a Value domain unless it is necessary.

- Do not include valid Permissible values in the Supplementary values attribute of a Value domain.
- Avoid the use of values that are contiguous with the last code in the permissible value sequence.
- To limit variations in the meaning within a specific data collection, use a default supplementary value meaning of 'Not stated/inadequately described'.
- When using more than one supplementary value, use a logical set.
- It is appropriate that the Supplementary value field size is the same number of characters as the permissible value.
- Consistent use of supplementary codes/values across the data elements in a data set should be applied.
- In non-enumerated Value domains (that is those without defined value meanings such as in a measurement) the supplementary value used should not be a valid permissible value.

7 Always use a standard format for referencing publications and not a mixture of referencing methods.

METeOR

METeOR is the AIHW's online metadata registry.

METeOR currently 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 includes the following tools:
- data search and browse tools that allow navigation of data standards of varying levels of endorsement and across the health, community services and housing assistance sectors
- data view, collation and download tools
- data development tools, including areas in which multiple data developers may collaborate on the development of data standards
- data submission tools that enable data developers to submit draft metadata standards for consideration as national standards
- data management tools that allow the registrar to change the registration status of metadata standards under authorisation of one or more registration authorities
- comprehensive guidelines for developing and reviewing metadata.

METeOR is based on the 2003 version of the ISO/IEC 11179, titled 'Information technology—metadata registries (MDR)'. This standard was applied to provide a detailed registry architecture in which metadata standards can be better defined, navigated and managed throughout the data development lifecycle. METeOR is an Internet-based application accessible through the following web address: www.meteor.aihw.gov.au .