Australian Government

Australian Institute of Health and Welfare

National data dictionaries

Explanatory booklet 2008

NATIONAL HEALTH DATA DICTIONARY

Version 14

NATIONAL COMMUNITY SERVICES DATA DICTIONARY

Version 5

Community and Disability Services Ministers' Advisory Council

Australian Health Ministers' Advisory Council

The Australian Institute of Health and Welfare is Australia's national health and welfare statistics and information agency. The Institute's mission is *better information and statistics for better health and wellbeing*.

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

National health data dictionary
ISSN 1329-4555
ISBN 978 1 74024 799 3

Suggested citation
Health Data Standards Committee 2008. National health data dictionary. Version 14. Cat. no. HWI 101. Canberra: Australian Institute of Health and Welfare.

National community services data dictionary
ISSN 1440-8872
ISBN 978 1 74024 800 6

Suggested citation
National Community Services Data Committee 2008. National community services data dictionary. Version 5. Cat. no. HWI 102. Canberra: Australian Institute of Health and Welfare.

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

Australian Institute of Health and Welfare
GPO Box 570
Canberra ACT 2601
Phone: (02) 6244 1000

Published by the Australian Institute of Health and Welfare

Contents

The fourteenth version of the National Health Data Dictionary (NHDD V14) and the fifth version of the National Community Services Data Dictionary (V5) are stored on two separate CD-ROMs. Each CD-ROM contains all of the existing, new and modified nationally endorsed data standards in a format based on the international standard ISO/IEC 11179 edition 2.[1].

1 ISO/IEC 11179-3:2003(E), Information technology—Metadata registries (MDR)—Registry metamodel and basic attributes, Geneva: International Standards Organisation

The content of the CD-ROMs

Each CD-ROM contains:

- an introductory page and a table of contents
- the data dictionary containing all standard metadata items including data elements, classifications and glossary items
- a summary of the changes to the dictionary since its last publication
- details of the data collections such as the data set specifications and National Minimum Data Sets that use these standards
- this explanatory booklet.

How to use the CD-ROMs

To use a CD-ROM:

- Insert the CD-ROM into your computer. It should automatically open to a Start page.
- Choose a document by clicking on the buttons.
- The documents are in PDF format and can be navigated using the Adobe functions.

The majority of hyperlinks within the document link internally to other parts of the document; however, some hyperlinks may take you to pages outside the document itself and into METeOR, the online metadata registry, in which case a new web browser window will open. This will occur where references are made to metadata items that have been superseded and/or links to other websites such as the Australian Bureau of Statistics, to view a specific classification scheme. To return to the data dictionary, just close the new window.

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 data standards and provide the basis for consistent national collection and reporting. National data standards are approved by the Australian Government and all state and territory relevant health and communities 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 Ministerial Advisory Council (AHMAC) and the Community Services and Disability Ministerial Advisory Council (CSDMAC).

Where possible, data standards in the dictionaries are consistent with other national standard classifications to ensure overall comparability of national data. The ABS is the source of many key socioeconomic classifications used in data collections, such as the Australian Standard Classification of Languages (ASCL), 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 development, registration and dissemination of 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 24. The data dictionaries are also available online at meteor.aihw.gov.au.

Governance

The national data dictionaries are an initiative under the National Health Information Agreement and the National Community Services Information Agreement. These are auspiced by the respective Australian Health Ministers' Advisory Council (AHMAC) and the Community Services Disability Ministers' Advisory Council (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 will require agreement on definitions, standards and rules of collection of information and on guidelines for the coordination of access, interpretation and publication of national health and community services information.

These data standards are endorsed and approved for use by all relevant health and community services state, territory and Commonwealth departments, the ABS and the AIHW.

Why use data standards?

Making data count

Data standards promote the quality, accuracy, interpretability, reliability, relevance, inter-changeability, transparency, currency, accessibility, coherence and comparability of data and information.

Without data standards there is the potential for data to be of poor quality. Data may fail to measure what it is supposed to measure, or not be comparable across collections or over time. Decisions based on poor quality data affect us all—whether it is hospital services, or delivery of services in the community.

Data standards enable consistent and comparable reporting of information about services and people, including describing what services are available, where services are located, to whom they are delivered, by whom are they delivered and when, how much they cost and what happened as a result of delivering services.

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

How are data standards developed?

Identifying the need for data standards

Data standards are developed when a clear need for standards is identified from the sector, subject experts, program managers or policy makers. For example, a national cancer centre and the Australian Association of Cancer Registries conducted a review of data items used in cancer registries, and found many inconsistencies among registries in coverage and data collection comparability for a number of data items. This resulted in collaboration between a national cancer centre and the Australian Association of Cancer Registries, clinical specialists, government and non-government organisations to revise the items and to develop a preliminary NMDS. This resulted in agreement on a NMDS of 17 data items recommended as essential for long-term breast cancer control monitoring, evaluation and planning.

Data standards development process

The data standards development process is outlined below.

1. **Proposal stage**

   A submission is made to the relevant data committee, which includes:

   - origins (or background material)
   - the rationale for the proposal
   - a business case for adoption
   - details of national consultation (including details of experts and/or others involved or consulted during development).

continued overleaf

2. Development stage

The development process is based on a number of data development principles and a methodology. [2]

3. Assessment stage

An assessment is conducted on the degree of compliance with accepted data development standards (adherence to 11179/ METeOR business rules), the degree of overlap with, or the confounding of, existing data standards, the degree to which the data elements impact on existing metadata, systems and collections, and the clarity of the content.

4. Data committee approval

Once an initial assessment has been conducted, the metadata is submitted to the relevant data committee for comment and approval.

5. Registration authority endorsement

If a submission is approved by the data committee, the metadata items are then forwarded to the relevant registration authority for final endorsement as a national standard.

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

Registration status progression

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

The registration statuses in METeOR are:	
Proposed	A developer has submitted this item for consideration by the registrar.
Recorded	The registrar has determined that the item meets basic quality criteria and is ready for consideration by the relevant data committee.
Candidate	The item has been reviewed by a data committee and has been accepted onto their work program.
Standardisation pending	A data committee 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 data standards?

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

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

Researchers—use data standards as the common language to support ad hoc survey work, as well as integrating 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 data and analysis of results, linking data sets for statistical purposes, time series analysis (over a period of time) and longitudinal studies (over a period of time within groups).

Metadata structure

The data 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 11179 that apply to METeOR and the data dictionaries:

- object class
- property
- data element concept
- data element
- value domain
- classification scheme.

The structure underlying a data element is illustrated in Figure 1 (differences from the ISO/IEC standard are shown with dashed lines).

Types of metadata

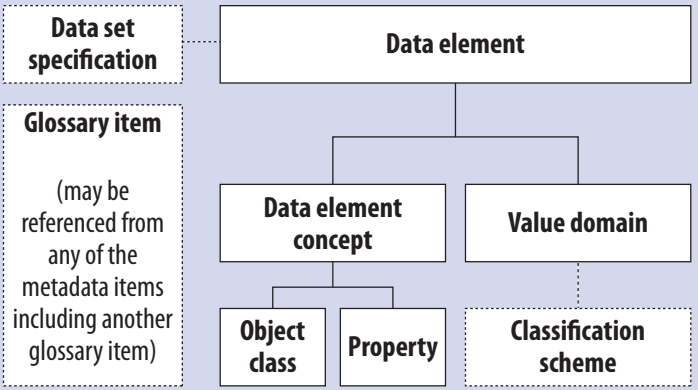

Figure 1: The 11179 model of metadata representation.

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 that all instances of the object class have in common is known as a property. A property 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, the data element concept: 'Person—date of birth'.

As can be seen from the diagram, 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 3 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 Australian Standard Classification of Languages (ASCL) 2005, or all activity codes listed in the International statistical classification of diseases and related health problems, Tenth revision, Australian modification, 5th edition.

A glossary item defines the meaning of a term within a specific context. Within METeOR examples of glossary items include 'Adoption' and 'Family'. These things of interest are not currently defined as object classes, but their meaning must be understood for data to be collected.

Data set specifications

A collection of these data elements that describe things about which we want to know has been termed a data set specification (DSS). A DSS specifies a group of data elements and the conditions under which this group is 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 (such as its scope).

National Minimum Data Set

A National Minimum Data Set (NMDS) is a special type of DSS. An NMDS is 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. Data elements may appear in more than one National Minimum Data Set.

The Data Dictionaries support data elements from a variety of national data sets, For example, the Commonwealth State Territory Disability Agreement (CSTDA) NMDS, Community Mental Health Care NMDS, and Admitted Patient Care NMDS. A full list is available on the CDs.

Non-mandatory data set specification

All data set specifications that are not mandated for collection are metadata sets recommended for collection as best practice. It is recommended that, if collecting data for the purposes of primary patient care, planning or analysis, the entire DSS be collected.

Integration of data elements in data dictionaries

The national health and community services data dictionaries contain 92 integrated data items that can be used consistently across the health and community services sectors. This is especially important for services that cross the sector boundaries such as aged care, mental health, drug and alcohol services and services for people with a disability. Examples of integrated data items include:

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]

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, XXXXXX[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

Quick reference guide

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:**
   a. the name must be unique within the context of the metadata item
   b. the name must be stated in the singular
   c. the name must reflect the concept being defined
   d. the name must avoid the use of abbreviations or acronyms other than those widely accepted (such as radar, laser or pH)
   e. the name should avoid the use of words that imply a pre-selected single instance.
4. **Definition rules and guidelines**

   A definition must:
   a. be stated in the singular
   b. state what the concept is, not only what it is not
   c. be stated as a descriptive phrase or sentence(s)
   d. contain only commonly understood abbreviations

continued

Quick reference guide (continued)

e. be expressed without embedding definitions of other data or underlying concepts.

A data definition should:

f. be expressed without embedding rationale, functional usage, domain information or procedural information

g. state the essential meaning of the concept

h. be precise and unambiguous

i. be concise

j. be able to stand alone

k. avoid circular reasoning

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

m. contain information appropriate for the type of metadata item being defined

n. use a preferred term to represent the definition of a concept specified elsewhere in the document

o. pass the substitution test.

5. Context should be closely linked to definition.

continued

Quick reference guide (continued)

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

a. Context

i. Metadata can exist within a specific context.

ii. 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.

iii. The contexts of two metadata items must be compatible when the definition of one metadata item references a term defined in another.

b. 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.

i. 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.

ii. Data elements can also include information about how to use or interpret the data in the Guide for use.

c. Collection methods

The 'Collection methods' attribute must only include information about how data is to be collected. The following

continued

Quick reference guide (continued)

metadata items must not have a metadata attribute of 'Collection methods':

- object class
- property
- data element concept
- value domain
- glossary item.

d. Permissible values in a Value domain must:

i. be exhaustive within the set

ii. made into an exhaustive set of values by adding an 'Other' value to aggregate all other possibilities not covered by the stated set of values.

iii. be mutually exclusive within the set

iv. be a true representation of the concept defined in the data element.

e. Allocation of code values should

i. 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, does not provide for inclusions in the future.

ii. wherever possible, avoid the use of a coded value for 'Other' that may be commonly used as a supplementary value.

continued

Quick reference guide (continued)

f. Supplementary values

i. Supplementary values must not be included in a value domain.

ii. Do not include valid permissible values in the supplementary values attribute of a value domain.

iii. Avoid the use of values that are contiguous with the last code in the permissible value sequence.

iv. To limit variations in the meaning within a specific data collection, use a default supplementary value meaning of 'Not stated/inadequately described'.

v. When using more than one supplementary value, use a logical set.

vi. It is appropriate that the Supplementary value field size is the same number of characters as the permissible value.

vii. Consistent use of supplementary codes/values across the data elements in a data set should be applied.

viii. 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 Australian Institute of Health and Welfare's online metadata registry.

METeOR currently integrates and presents information about:

- the National Community Services Data Dictionary
- the National Health 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 allowing 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 data standards for consideration as a national standards
- data management tools that allow the registrar to change the registration status of data standards under authorisation of one or more registration authority
- comprehensive guidelines to assist metadata development and review.

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 data standards can be better defined, navigated and managed throughout the data development life-cycle. METeOR is an Internet-based application accessible through the following Internet address:

meteor.aihw.gov.au

Feedback and contact details

The development of the data dictionaries is an ongoing process, which is reliant on the support and input of a range of data development groups to expand its scope and utility.

The data committees welcome feedback on existing data standards in the data dictionaries, and also submissions (either for new data items, modifications to existing items or information on your data development activities).

Feedback

Please feel free to contact the Institute by any of the means listed below.

National Data Development and Standards Unit

For further information about the data dictionaries and for any comments and suggestions about national standards development processes, please contact the National Data Development and Standards Unit at the Australian Institute of Health and Welfare.

Phone: (02) 6244 1000
Fax: (02) 6244 1299
datadevelopment@aihw.gov.au

Data Development Hotline

Phone: (02) 6244 1222

Secretariat

For further information about the national data committees and their processes, please look up the Committees page on the AIHW website.

http://www.aihw.gov.au/committees/ncsdc/index.cfm

http://www.aihw.gov.au/committees/hdsc/index.cfm