AIHW | METEOR
Metadata Online Registry
meteor.aihw.gov.au

2. Principles for the development of good data standards

2 Principles for the development of good data standards

2.1 Introduction

This chapter provides the metadata development principles that lead to the development of good data standards. Data standards describe the agreed meaning and representation of data for use in a defined context. Data standards are also applied to metadata items that have been endorsed by a registration authority to become a standard. A data standard endorsed for use across Australia is referred to as a national data standard.

Good data standards improve the quality, relevance, consistency and availability of information about the health and welfare of Australians.

Creating data standards is an important part of the data development process. The aim is so that clinical, administrative, statistical and/or other information can be better understood and used across and within the health and welfare sectors. To achieve this, it is critical that the metadata used are defined and described accurately and consistently; this is done through the use of agreed-upon existing standards for compliancy and interoperability.

During the data development process, a number of characteristics of data are defined in order to aid the understanding, interpretation and use of data. The principles for the development of good data standards listed below cover the metadata attributes: 'name', 'definition', 'context', 'guide for use', 'collection methods', as well 'permissible' and 'supplementary values'. Also covered in this chapter is the principle of 'create once, use often'.

For information on data development more broadly, from the time a need for data development is identified to the endorsement of a fully developed data set, refer to A guide to data development (AIHW 2007) available on the AIHW's website.

2.2 Principle of 'create once, use often'

An important general principle for data development is the principle of 'create once, use often'. This means the availability of existing data sources should be explored and used, where possible. Data needed to support secondary (downstream) information purposes, such as reporting, policy, governance or decision support, should be derivable from primary (point-of delivery) data. Similarly, data developed and collected for mainly statistical purposes should be used to provide feedback to improve and enhance primary service delivery.

The 'create once, use often' principle also applies to the development and use of data standards in the following ways:

- When undertaking data development activity in METEOR, if a '*Standard*' exists that is fit for purpose, the standard should be used rather than creating a new metadata item. A developer should use METEOR's search function to thoroughly explore existing metadata before creating any new elements.

- To facilitate the re-use of data standards across collections, where possible, generic metadata items should be developed rather than metadata items that can only be used in a specific collection. For example, a generic data element concept can be combined with different value domains (representational values) to create data elements.

2.3 Name and definition principles overview

The name of a data standard:

1. must be unique
2. must reflect the concept being defined
3. must be stated in the singular (unless the concept itself is plural in nature)
4. must avoid the use of words that imply a preselected single instance
5. should avoid abbreviations (including acronyms and initialisms), unless they are commonly understood or widely accepted within the context of the metadata item.

An explanation of the name principles is provided in section 2.6.1.

The definition of a data standard:

6. must be unique
7. must be stated in the singular (unless the concept itself is plural)
8. must state what the concept is, not only what it is not
9. must be stated as a descriptive phrase or sentence(s)
10. must be expressed without embedding definitions of other data or underlying concepts
11. may use abbreviations (including acronyms and initialisms) provided they are first spelled out in full, or else are commonly understood or widely accepted within the context of the metadata item
12. should state the essential meaning of the concept
13. should be precise and unambiguous
14. should be concise
15. should be able to stand alone
16. should be expressed without embedding rationale, functional usage, domain or procedural information
17. should avoid circular reasoning

18. should be expressed in the present tense

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

20. should be appropriate for the type of metadata item being defined.

An explanation of the definition principles is provided in section 2.6.2.

2.4 Context, guide for use, and collection methods principles overview

The context of a data standard defines the setting within which the subject data has meaning:

1. A metadata item may exist within a specific context.

2. Context must only contain information about the environment or setting within which a metadata definition is valid.

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

An explanation of the context principles is provided in section 2.6.3.

A data standard may include guide for use information to provide advice or interpretation on how to use the data standard. Guide for use information must only be included if it provides information on the use of the metadata item or data collected:

- In a metadata item other than a data element, 'guide for use' is about how to use the metadata item itself, not about any data that may be collected or used.

- In a data element, 'guide for use' may also include information about to how to use or interpret the data.

An explanation of the guide for use principles is provided in section 2.6.4.

A data standard may include collection methods information to provide advice or instructions on how the data are to be collected or captured.

An explanation of the collection methods principles is provided in section 2.6.5.

2.5 Value domain principles overview — permissible and supplementary values

An enumerated value domain is a value domain that is specified by a list of all its permissible values. A non-enumerated (or described) value domain is a value domain that is specified by a description rather than a list of all permissible values. We use the term 'described' for our purposes to reflect the updated ISO 11179-3:2013/AMD1:2020 standard.

For example, an enumerated value domain for age ranges might include the following permissible values:

1. CODE 1: <18 years
2. CODE 2: ≥18 years and ≤64 years
3. CODE 3: ≥65 years

By contrast, a described value domain for age may simply specify the format N[NN], meaning that the permissible values for age in years should be a number between 1 and 3 digits in length.

In both enumerated and described value domains, one or more supplementary values may be required to capture 'missing' information. The following principles should be considered in the initial stages of any data development activity or data collection project. However, we acknowledge that when the metadata for a data set specification is included in METEOR, it may often be too late to take these principles into account (for example when describing data that have already been collected.)

1. Permissible values must contain value meanings that are exhaustive and mutually exclusive within the value domain.
2. Permissible values must be a true representation of the concept defined in the data element.
3. Permissible values must 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
    - may be commonly used as a supplementary value.
4. The need for supplementary values in a value domain should be considered carefully.
5. Supplementary code values that are contiguous with the last code in the permissible value sequence should be avoided.
6. Supplementary values should be used consistently in a data collection by using:
    - a default supplementary value meaning of 'not stated/inadequately described' to limit variations in the meaning within a specific data collection
    - a logical set of supplementary values if using more than one supplementary value.

7. A supplementary value code field size should be the same as the permissible value, where possible.

8. A valid permissible value must not be used as a supplementary value (and vice versa).

9. In described value domains, the supplementary value used must not be a valid permissible value.

An explanation of the value domain principles is provided in section 2.6.6.

2.6 Explanation of the principles for the development of good data standards

An explanation and generic examples of the principles for the development of good data standards are provided below. These principles generally relate to the metadata types: object class, property, data element concept, value domain, classification scheme, glossary item, data element, and data set specification.

All examples are provided for illustration and/or comparison purposes only.

2.6.1 Name principles

1. **A name must be unique for the metadata item type and context**

   The name given to a standard must be unique within METEOR for each metadata item type. This means that there must not be two identical metadata item types at *Standard* status (e.g., two data element standards) with the same name. This does not mean, however, that there cannot be metadata items at *Standard* status and 'superseded' status with the same name (e.g., a standard object class and a superseded object class both named 'Episode of residential care').

   It is important to consider the environment or setting in which the metadata item is applicable, as the name given to a metadata item must also be unique within the context of the item. This means there cannot be two standard metadata items (e.g., two standard object classes) with the same name and applicable in the same context. In theory, however, there could be two standard items with the same name (e.g., two glossary items) if the environment or setting in which they are applicable is different, although in practice this should be avoided to save confusion.

2. A name must reflect the concept being defined

The name of a metadata item must clearly reflect and identify the concept being defined. This not only helps in the understanding of the item, it also aids in the searching of metadata items in METEOR.

Example: Concept 'type of clinical assistance a person receives in their home' name

	Name	Reason name is poor/good
Poor name	Home help	The name is not a true reflection of the concept so it is not clear from the name that the metadata item is about clinical assistance received in the home.
Good name	Home based clinical assistance type	The name clearly identifies the concept.

3. A name must be stated in the singular (unless the concept itself is plural in nature)

To provide clarity of understanding, the name given to a metadata item must be stated in the singular, not in the plural or collective form. The exception to this is if the concept being named is plural in nature, such as in the case of a value domain. As a value domain is a set of permissible values, the name given to a value domain may be plural.

Example 1: Concept 'person' name

	Name	Reason name is poor/good
Poor name	People	Name is stated in the collective form.
Good name	Person	Name is stated in the singular.

Example 2: Concept 'person' name

	Name	Reason name is poor/good
Poor name	Service contact dates	Name is stated in the plural. Note: If a metadata item was about a (singular) set or group of service contact dates, the name 'Set of service contact dates' could be used.
Good name	Service contact date	Name is stated in the singular.

4. A name must avoid the use of words that imply a preselected single instance

Avoiding the use of words in the name that imply a preselected single instance is necessary to enable the item to apply to any occurrence of the concept.

Example: Concept 'admission date' name

	Name	Reason name is poor/good
Poor name	Date of this admission.	The name implies a specific instance of an 'admission'.
Good name	Admission date	The name indicates that the metadata item applies to any 'admission'.

5. A name should avoid abbreviations (including acronyms and initialisms), unless they are commonly understood or widely accepted within the context of the metadata item

If necessary, the name of a metadata item may contain abbreviations (including acronyms or initialisms), as long as the abbreviations used are commonly understood or widely accepted. In using an abbreviation in a name, however, it is important to note that an abbreviation that has a specific meaning in one environment, may have a different meaning in another environment. This can create confusion and misunderstandings and lead to the misinterpretation of data. In such situations, it is best to avoid ambiguity by using full words, not abbreviations, in a name.

	Name	Reason name is poor/good
Poor name	Date of this admission.	The name implies a specific instance of an 'admission'.
Good name	Admission date	The name indicates that the metadata item applies to any 'admission'.

Commonly understood or widely accepted abbreviations include those that are more readily understood than the full form of a complex term and have been adopted as a term in their own right, such as 'radar', 'laser' or 'pH'.

If an abbreviation (such as an acronym) is used in a name, it must be spelt out in full followed by the abbreviation in parentheses when the term is next used (generally this will be in the definition).

Example: Use of abbreviation in a name

	Example	Reason use of abbreviation in name is poor/good
Poor use of abbreviation in a name	Name: AUDIT frequency alcohol consumption Definition: The frequency of alcohol consumption as measured by the AUDIT.	The meaning of the 'AUDIT' in the name may not be commonly understood and it is not stated in full in the definition.
Good use of abbreviation in a name	Name: AUDIT frequency alcohol consumption Definition: The frequency of alcohol consumption as measured by the Alcohol Use Disorders Identification Test (AUDIT).	The meaning of 'AUDIT' in the name is subsequently stated in full in the definition followed by the acronym in parentheses. This makes the meaning of the 'AUDIT' clear.

In relation to the name given to a data set specification, it is acceptable to use the acronym for the data set type in the data set name.

Example: Acceptable use of a data set type acronym in a data set name

Data set type	Acronym	Name example
National minimum data set	NMDS	Admitted patient care NMDS 2024–25
Data set specification	DSS	Adoptions DSS 2022–23
National best endeavours data set	NBEDS	Prisoner health NBEDS 2022
National best practice data set	NBPDS	Person and provider identification in healthcare NBPDS

2.6.2 Definition principles

1. A definition must be unique

The definition of a standard must be unique within the context of the item. One or more characteristics expressed in the definition must differentiate the concept from other concepts.

2. A definition must be stated in the singular (unless the concept itself is plural)

To provide clarity of understanding, the concept expressed by a definition must be expressed in the singular, rather than the plural or collective form. The use of plurality in a definition is only acceptable when the concept being defined is plural in nature, such as in a value domain. The definition of a value domain may be plural as it encompasses all the permitted values for that domain.

Example 1: Concept 'article number' definition

	Definition example	Reason definition is poor/good
Poor definition	A reference number that identifies articles.	The use of the plural 'articles' in the definition is ambiguous as it could imply that an article number refers to more than one article. Note: If a single article number did in fact identify more than one article, then the definition may be correct although the name given to the metadata item would still need to be expressed in the singular (i.e. the singular name 'Article number' as used in this example).
Good definition	A reference number that identifies an article.	The use of the singular 'number' and 'article' in the definition makes it clear that an article is identified by one article number.

Example 2: Value domain definitions

	Name example	Definition example
Enumerated value domain	Age range code N	A code set representing age groups.
Described value domain	Total hours NNNN	Total number of hours.

3. A definition must state what the concept is, not only what it is not

A definition must not define a concept exclusively by stating what the concept is not.

Example: Concept 'freight cost amount' definition

	Definition example	Reason definition is poor/good
Poor definition	Costs that are not related to packing, documenting, loading, unloading and insurance.	The definition only identifies what is not included in the meaning of the concept. By not specifying what is included in the concept the meaning is ambiguous and open to interpretation.
Good definition	Cost amount incurred by an organisation in transporting goods from one place to another.	The definition identifies what is included so clearly describes what the concept means.

4. A definition must be stated as a descriptive phrase or sentence(s)

In the English language, a phrase is necessary to form a precise definition that includes the essential characteristics of the concept. A definition must be expressed as a complete, grammatically correct, descriptive phrase or sentence(s). It is not sufficient for a definition to simply consist of one or more synonyms for the name given to the concept or to restate the words in the name in a different order.

Example: Concept 'agent name' definition

	Definition example	Reason definition is poor/good
Poor definition (1)	Representative	The definition simply uses a near-synonym of the name. This is insufficient as it does not articulate the meaning of the concept.
Poor definition (2)	The name of an agent	Although the terms used in the definition may be understood, this definition is an illustration of the words that form the name being rearranged to form a definition. This is insufficient for a definition.
Good definition	The name of a party authorised to act on behalf of another party	The definition is written as a sentence and provides a clear understanding of the meaning of the concept.

5. **A definition may use abbreviations (including acronyms and initialisms) provided they are first spelled out in full, or else are commonly understood or widely accepted within the context of the metadata item**

When an abbreviation (such as an acronym) is used in a definition, it must be spelt out in full on its first occurrence, followed by the abbreviation in parentheses.

Commonly understood or widely accepted abbreviations in a definition include those such as 'i.e.' (for 'that is') and 'e.g.' (for 'for example') and the like. It also includes abbreviations that are more readily understood than the full form of a complex term and have been adopted as terms in their own right, such as 'radar', 'laser' or 'pH'. When using abbreviations such as 'i.e.' or 'e.g.', the abbreviation and its accompanying text should appear in parentheses; for example, '... (e.g., the financial year)'.

Example 1: Concept 'tide height' definition

	Definition example	Reason definition is poor/good
Poor definition	The vertical distance from MSL to a specific tide level.	The definition is unclear because the acronym 'MSL' may not be commonly understood and some users of the data standard may need to refer to other sources to determine what it represents.
Good definition	The vertical distance from mean sea level (MSL) to a specific tide level.	The meaning of the definition is clear as the full name of 'MSL' is written in full before the abbreviation appears in parentheses.

Example 2: Concept 'blood oxygen level' definition

	Definition example	Reason definition is poor/good
Poor definition	The measurement of a SAO2 levels.	The definition is unclear because the acronym 'SAO2' may not be commonly understood and some users of the data standard may need to refer to other sources to determine what it represents.
Good definition	The measurement of the saturation level of oxygen in blood (SAO2).	The meaning of the definition is clear as the name of SAO2 is stated in full before the abbreviation appears in parentheses.

6. A definition must be expressed without embedding definitions of other data or underlying concepts

The definition of another (secondary) concept must not appear in the definition of the primary concept. When the definition of another concept, even if closely related, appears within another concept's definition, it tends to obscure the original definition. However, there may be cases where a term or related concept is considered necessary in order to convey the true meaning of the primary concept. A glossary item is recommended for any such term or concept that is sufficiently important to require defining. If the term or related concept appears more than once in the primary definition, only the first occurrence of the term or related concept should be linked to the glossary item.

Example 1: Concept 'severe hypoglycaemia history flag' definition

	Definition example (for illustration purposes the secondary embedded concept is italicised)	**Reason definition is poor/good**
Poor definition	A flag of whether a person has had severe hypoglycaemia, which is defined as hypoglycaemia requiring assistance from another party, in the last 12 months.	As well as defining what the primary concept 'severe hypoglycaemia history indicator' means, the definition has a definition for the related (secondary) concept 'severe hypoglycaemia' embedded within it. In order to understand the primary concept, it may be necessary to define what 'severe hypoglycaemia' means. If this is the case, the definition of the secondary concept should be included as a glossary item, unless defined elsewhere.
Good definition	A flag of whether a person has had severe hypoglycaemia in the last 12 months.	The definition only defines the primary concept 'severe hypoglycaemia history indicator'. In order to understand the true meaning of the definition, a relationship to a 'severe hypoglycaemia' glossary item (represented by bold in the definition) has been created. In this example, the 'severe hypoglycaemia' glossary item is defined as: 'Hypoglycaemia requiring assistance from another party'.

Example 2: Concept 'blood oxygen level' definition

	Definition example **(for illustration purposes the secondary embedded concept is italicised)**	**Reason definition is poor/good**
Poor definition	A flag of whether a person has an informal carer, such as a family member, friend or neighbour, who is providing care and assistance on a regular basis.	As well as a definition for what the primary concept 'informal carer existence indicator' means, the definition has a definition for the related (secondary) concept 'informal carer' embedded within it. In order to understand the primary concept, it may be considered necessary to define what 'informal carer' means. In such cases, the definition of the secondary concept should be included as a glossary item, unless defined elsewhere.
Good definition	A flag of whether a person has an informal carer.	The definition only defines the primary concept 'informal carer existence indicator'. In order to understand the true meaning of the definition, a relationship to an 'informal carer' glossary item (represented by bold in the definition) has been created. In this example, the 'informal carer' glossary item is defined as: 'An informal carer includes any person, such as a family member, friend or neighbour, who is giving regular, ongoing care and assistance to another person'.

7. A definition should state the essential meaning of the concept

Only the primary characteristics of the concept being defined should be included in the definition. The inclusion of non-essential characteristics or superfluous information should be avoided.

The level of detail or specificity necessary in a definition is dependent on the context in which the concept has meaning. The primary and essential characteristics that are necessary to convey the essential meaning of a concept will vary according to the level of generalisation or specialisation of the data. In one context, a more general definition may be adequate, whereas in a different (specialised) context a more detailed definition may be required.

Example 1: Concept 'date of death' definition

	Definition example	**Reason definition is poor/good**
Poor definition	The date upon which a person ceases to live as stated on their death certificate	The definition includes information about where date of death is recorded. This information is superfluous to defining the concept 'date of death'. This superfluous information is more appropriate for inclusion in the collection methods attribute.
Good definition	The date upon which a person ceases to live.	The definition states the essential meaning of the concept without any superfluous information.

Example 2 (specificity): Concept 'penguin' definition

	Definition example	Reason definition is poor/good
Poor definition	A bird of the southern hemisphere that is quite often black and white in colour, with wings that are insufficient to enable it to fly but enable it to swim efficiently (often to great depths and for long periods of time) through the predominately Antarctic waters that it inhabits.	The definition includes non-essential information that is superfluous to defining the 'penguin' concept.
Good definition (1)	A flightless bird that uses its wings to swim under water.	The definition is sufficient to differentiate a penguin from other flightless birds, particularly flightless aquatic birds that do not use their wings to swim under water. However, if it was necessary to differentiate one species of penguin from another (e.g., southern versus northern hemisphere), the definition does not have enough specificity to do so.
Good definition (2)	A flightless bird of the southern hemisphere that uses its wings to swim under water.	If such data was required, the definition provides the level specificity to differentiate a southern hemisphere penguin from a northern hemisphere penguin.

8. A definition should 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 only one possible interpretation.

Example: Concept 'country of birth' definition

	Definition example	Reason definition is poor/good
Poor definition	The country a person is from.	The definition does not precisely define the concept and is ambiguous. A county a person is 'from' could be interpreted in many ways, such as the last country a person resided, which is not necessarily the county a person was born.
Good definition	The country in which a person was born.	The definition is precise and unambiguous.

9. A definition should be concise

Although a definition needs to be comprehensive, it should also be as brief as possible by avoiding the inclusion of extraneous information or superfluous words.

Extraneous information or superfluous words to be avoided in a definition include:

- statements about the rationale, functional usage, domain or procedural information for a metadata item (see principle 11 below)
- repeating the name of the item being defined.

Example: Concept 'character set name' definition

	Definition example	Reason definition is poor/good
Poor definition (1)	The name given to the set of phonetic or ideographic symbols in which data is encoded, for the purpose of this metadata registry, or, as used elsewhere, the capability of systems hardware and software to process data encoded in one or more scripts.	In the definition all the information after '... data is encoded' is extraneous.
Poor definition (2)	A character set is the name given to the set of phonetic or ideographic symbols in which data is encoded.	The definition repeats the name of the concept 'character set name' at the beginning of the definition, which is superfluous to the definition.
Good definition	The name given to the set of phonetic or ideographic symbols in which data is encoded.	The definition is concise without any superfluous information.

10. A definition should be able to stand alone

The meaning of the concept should be apparent from the definition. Additional explanations or references should not be necessary in order to understand the meaning of the definition.

Example: Concept 'school location city name' definition

	Definition example	Reason definition is poor/good
Poor definition	See 'school site'	The definition does not stand alone as it requires the aid of a second definition (i.e. school site) to understand the meaning.
Good definition	The name of the city where a school is located.	The definition stands alone. It does not require another reference or definition in order to understand the meaning.

11. A definition should be expressed without embedding rationale, functional usage, domain information, or procedural information

A definition should be expressed without the inclusion of statements about the rationale, functional usage, domain or procedural information for the metadata item because these statements contain information that is extraneous to the definition itself. If such statements are considered necessary, they may be included in other attributes as indicated below:

Placement of information superfluous to a definition:

Information type	Attribute(s) for information type
Rationale, purpose or justification for a given definition or metadata item (e.g., ' ... for the purpose of ...' or 'this item is collected for the analysis of ....').	Should be included in 'comments' attribute in a data element or 'scope' in a data set specification.
Functional usage (e.g., 'this data element should not be used for ...' or 'this item can be used in calculating the total amount ...').	Should be included in 'guide for use' or 'collection methods', whichever is most appropriate (depending on the functional usage statement).
Domain information (e.g., '... in an emergency department ...', or '... as determined by a health professional ...').	Domain information should not be included in the definition as it is about the context or setting in which the data are collected. As such, domain information should be included in the 'context' attribute, wherever possible. In some cases, domain information may be best included in 'collection methods'.
Procedural (e.g., 'this data element is used in conjunction with data element [name of data element]').	Remarks about procedural aspects should be included in the 'guide for use' or 'comments'.

12. A definition should avoid circular reasoning

To avoid circular reasoning, a definition should not use another concept's definition as its definition. Circular reasoning occurs when one concept (concept A) is defined with the aid of another concept (concept B) that is, in turn, defined with the aid of the first concept (concept A), and should be avoided.

Example: Circular reasoning in definitions

	Definition	Reason definitions are poor
Employee ID number (Concept A)	A number assigned to an employee.	Each definition refers to the other for its meaning, and as a result neither definition provides a true understanding of either concept.
Employee (Concept B)	A person corresponding to an employee ID number	

13. A definition should be expressed in the present tense

A definition should be expressed in the present tense, rather than past or future tense.

Example: Concept 'hospital admission' definition

	Definition example	Reason definition is poor/good
Poor definition	The administrative process whereby a hospital accepted responsibility for a patient's clinical services.	The definition is expressed in the past tense.
Good definition	The administrative process whereby a hospital accepts responsibility for a patient's clinical services.	The definition is expressed in the present tense.

14. A definition should use the same terminology and consistent logical structure for related definitions

A common terminology and syntax for similar or associated definitions should be used to facilitate understanding. Where the terminology and syntax are not the same, it may lead to an assumption there is an implied difference between the related definitions that may not exist in reality.

Example 1: Different terminology used in definitions for related concepts

	Definition examples (italics used for illustration purposes only)	Reason definition is poor/good
'Admitted patient'	A patient who undergoes a hospital's formal admission process to receive *hospital treatment.*	The definitions in this example use different terminology for essentially the same thing (i.e. '*hospital treatment*', '*treatment and/or care*' and '*clinical services*'). In order for each definition to be completely unambiguous, the same terminology (in this case '*treatment and/or care*') should be used in each definition.
'Formal admission'	An admission in which the administrative process records the commencement of *treatment and/or care* and/or accommodation of a patient.	
'Admission'	The administrative process whereby a hospital accepts responsibility for a patient's *clinical services*.	

Example 2: Comparison of syntax used for 'formal admission' and 'statistical admission' definitions

	Definition examples (italics used for illustration purposes only)	Reason definition is poor/good
Poor definitions for related concepts 'Formal admission' and 'Statistical admission'	Formal admission: The administrative process that records the commencement of treatment and/or care and/or accommodation of a patient	A different syntax is used at the beginning of each definition for these related concepts.
	Statistical admission: An admission in which the administrative process records the commencement of a new care type for a patient within one continuous hospital stay.	
Good definitions for related concepts 'Formal admission' and 'Statistical admission'	Formal admission: An admission in which the administrative process records the commencement of treatment and/or care and/or accommodation of a patient.	A different syntax is used at the beginning of each definition for these related concepts.
	Statistical admission: An admission in which the administrative process records the commencement of a new care type for a patient within one continuous hospital stay.	Both definitions use a common syntax (i.e. they both begin with 'An admission in which the administrative process records the commencement of ...').

15. A definition should be appropriate for the type of metadata item being defined

The METEOR metadata registry contains different types of metadata with each metadata type playing a different role. A definition should be appropriate for the metadata item being defined.

Definition requirements by metadata item type:

Metadata item type	Definition requirements
Object class	*An object class is the 'thing' of interest for which data is collected. It is a stand-alone concept and its definition should not contain any specific information that relates to any other metadata item (e.g., an object class definition should not refer to a property). An object class is a generic type of metadata item and the definition should reflect this. Consider the reusability of the object class when forming its definition.*
Property	A property is a characteristic of the 'thing' of interest. It is a stand-alone concept and its definition should not contain any specific information that relates to any other metadata item (e.g., a property definition should not refer to an object class). A property is a generic type of metadata item and the definition should reflect this. Consider the reusability of the property when forming its definition.
Data element concept	A data element concept is a concept that is independent of any representation. This means that a data element concept's definition must not contain reference to actual representational or data values. The definition is a fusion of the object class and property (i.e. the data element concept's 'building blocks'), although it does not need to be a one-to-one match of the object class and property definitions. However, the definition should include the essential information about the object class and property to provide meaning to the data element concept.

Metadata item type	Definition requirements
Value domain	A value domain is a set of permissible values. The permissible values may either be enumerated or expressed via a description. The definition should be specific and may need to contain information about the environment in which the value domain exists. Although a value domain definition needs to be specific, consider the potential reusability of the value domain when forming its definition.
Data element	A data element is a fusion of a data element concept and its representative value domain (i.e. the data element's 'building blocks.') N.B. a data element's definition does not need to be a one-to-one match of the data element concept and value domain definitions. However, the definition should include the essential information about the data element concept and value domain to provide meaning to the data element.
Glossary item	The definition for a glossary item does not require as strict adherence to all definition principles as other types of metadata items. The definition in a glossary item may be more detailed, contain more than one concept (if the term being defined requires it) and any other descriptive phrases that unambiguously define the concept.

2.6.3 Context principles

1. Metadata may exist within a specific context

The context for a metadata element should be closely linked to its definition. It could be the setting in which data collection or use is valid (e.g., juvenile justice or intensive care), or it could be a whole sector of service or care (such as custodial services or admitted patient care). It could even be more general, covering the whole of the health sector, or across service sectors. For example, the term 'admission' has many meanings, but within the context of hospital data, it means the admission of a patient into a hospital. If the context attribute is left blank, this implies that the definition of meaning is valid across all contexts.

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 attribute—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).

Example: Context attribute left blank

Concept	Reason why context may be left blank
Person	'Person' always remains an individual human being no matter what the context.
Height	'Height' is always a measurement of distance in the vertical plane no matter what the context.
Hospital census	'Hospital census' can be applicable in more than one context but explicitly includes environment in its definition.
Health service provider	'Health service provider' has contextual information in the name and definition.

2. Only information that is relevant to the environment or setting within which the metadata item's definition is valid must be included in the context attribute

Non-contextual information, such as the rationale or purpose of a metadata item, should not be included in the context attribute.

Example: Inappropriate information in context attribute

	Information	Reason information is inappropriate for context attribute
Context (1)	This information is required for the analysis of …	The information in context (1) and (2) is about the purpose of the data collected or the justification for the development of the metadata item, rather than the environment or setting in which each metadata item has meaning. Information about the purpose or justification for a metadata item should be included in the comments attribute if such information is necessary to provide a full understanding of the item. In a data set specification, information about the purpose of the collection should be included in the scope attribute.
Context (2)	This data is required for long term planning and policy formulation in …	
Context (3)	Public health and health care. Body Mass Index (BMI) is used as an indicator of …	The first sentence is sufficient context information. The second sentence is about the purpose or justification of the metadata item and should be included in the comments attribute.

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

Where a concept in one is metadata item is defined in terms of another, the contexts of both items must be compatible.

For example, if a 'diagnosis' glossary item has a 'health services' context, it would be problematic to reference that glossary item to help define an assessment event that is outside of the health services environment. However, it would be valid to reference 'diagnosis' in a metadata item that had a context of 'admitted patient care', 'emergency department care', 'acute coronary syndrome reporting' and so forth, as they are all subsets of 'health services.'

2.6.4 Guide for use principles

The guide for use attribute is used to provide any instructions on how to use, interpret and apply information in a data element. Use the section to describe any restrictions on how the data element is intended to be interpreted or applied which are specific to the relationship of the data element concept and the value domain. For example, describe how other data elements that should be used in conjunction with the current data element, any formulae which guide calculations, or coding guidelines.

For metadata items such as an object class, property, data element concept, and value domain, guide for use information is often not mandatory, although it may be useful to provide advice on the interpretation of codes or values in a value domain. Information about how to collect data should be in the collection methods attribute, see section 2.6.5.

Examples: The type of information that may be included in guide for use

Type of guide for use information	Example
The meaning or interpretation of codes of values	'The start date of treatment is recorded regardless of whether treatment is completed as intended or not.'
	'CODE 1 Duplicate record
Number of values to be collected	This code is used where multiple records are found for a single entity.'
	'This code set represents common sites of cancer metastasis. Where multiple sites occur, all should be recorded.'
What is collected or included	'Each surgical procedure used should be recorded.'
What is not collected or excluded	'Collected for radiation therapy and systemic therapy only. Date of surgical treatment is collected as a separate item.'
Setting the scope of specific data to be collected	'Does not include services provided through community health settings, such as community and child health centres.'
	'This value domain is not applicable for data collected about children (i.e. person's aged <18 years.)'
The organisation of the data for analysis or use	'This value domain is to be used for persons aged ≥18 years .'
	'Height data should be presented in 5cm groupings only.'

2.6.5 Collection methods principles

The collection methods attribute is used to show how data is collected. Use this section to provide instruction for the capture of data and outline any guidelines for the collection of the data element which are specific to the data element concept and the value domain. For example, data collection formats, minimum data collection requirements, requirements for supportive material, and how missing or not stated data is to be treated. While the obligation to complete this attribute is conditional, it is good to include relevant information if it will help other users gain insight into the data.

Information about interpretation of codes and values should be in the guide for use attribute, see 2.6.4. Information about purpose or rationale for collection is more appropriate in the comments attribute.

Examples: The type of information that may be included in collection methods

Type of guide for use information	Example
Data collection instruments or tools	'Standardised height measurement equipment should be used.'
	'This information should be sought from the patient's medical record.'
Who collects or records the data.	'Measurement of lipid levels should be carried out by laboratories which have been accredited by the National Association of Testing authorities.'
How or when to record the data.	'The full name of the agent should be recorded if the coding manual is not available.'
	'This item is completed when the person with cancer has been offered and accepted clinical trial entry.'
Period for which the data are collated and reported.	'Financial year ending 30 June each year'.
Other related data that are collected in conjunction with the data.	'If codes 1 or 2 are recorded, the dose of radiation received should also be recorded.'
	'Collected in conjunction with triage time.'
Recommended questions to ask.	'The permissible values for this data element are used to form the response categories to the question:
Note: Standard questions, such as those recommended by the Australian Bureau of Statistics, should be used where possible. If these are not available and questions have to be developed, it is important to 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 also be avoided. Questions should be kept short and simple, using language that is easily understood.	'Which one of the following reasons best describes why the person was not provided assistance?'.

2.6.6 Value domain principles — permissible and supplementary values

1. Permissible values must contain value meanings that are exhaustive and mutually exclusive within the value domain

An enumerated value domain is specified by a list of all its permissible values. For statistical purposes, the permissible values must be exhaustive and mutually exclusive.

Exhaustive set of values:

If a set of permissible values is not exhaustive of all possible values, there would be no place for data that does not fall within any of the categories that make up the set of values. Where necessary, permissible values should be made into an exhaustive set of values by adding an 'other' value category to aggregate all other possibilities not covered by the set of values. See example 1 below for exhaustive set of values and principle 3 for more information on the allocation of a code value for 'other'.

In some cases, however, it may not be appropriate to include an 'other' value category to make a set of values exhaustive (e.g., if the data being collected only requires a specific (non-exhaustive) set of values). In such cases, the name given to the value domain, and its associated definition, must reflect the concept defined by the non-exhaustive representation values. Another option is to include a statement in the 'guide for use' attribute that only the values listed are of interest for the data element.

For example, in a collection related to older people, a value domain representing age groups might have permissible values commencing with CODE 1: '55–64.' This value domain could dispense with an 'other' value category for people aged <55, provided the name of the value domain included the words 'older people' (e.g., 'Older person age range code N[N]') and a 'guide for use' attribute gave the instruction: "This element is only suitable to collect the age range of persons aged over 55 years."

Example 1: Exhaustive set of permissible values for 'mode of contact'

	Representation values and meanings	Reason set of permissible values is poor/good
Poor set of permissible values	1. By phone 2. By email	Note: This example assumes various modes of contact are possible (and are of interest) not just the ones listed. The list of permissible values are not exhaustive, as the list does not represent a full set of possible modes of contact. For example, there is no value for contact made in person. However, if 'By phone' and 'By email' were the only specific modes of contact of interest for a particular data collection, it may be appropriate for this to be reflected in the value domain name and/or definition; or a statement could be included in the Guide for use attribute that these two modes of contact are the only modes of contact that are of interest.

	Representation values and meanings	Reason set of permissible values is poor/good
Good set of permissible values	1. By phone 2. By email 5. Other	The permissible values are exhaustive with the inclusion of code 5 'Other'. The use of 'other' by itself in this example signifies 'other modes of contact' or 'other modes of contact not elsewhere categorised'. Note: The reason code 5 has been allocated to 'other' in this example (rather than code 3 which is the next code value in the code value sequence after code 2 'By email'), is explained in principle 3 below.

Mutually exclusive set of values:

A set of permissible values must be mutually exclusive in order for the value meanings to be distinct from one another without any overlapping meanings. Permissible values that are not mutually exclusive create uncertainty around data collection (e.g., if a respondent is recorded against more than one value category). See example 2 below.

Being mutually exclusive, however, should not be confused with the ability, where applicable, for more than one value meaning to legitimately apply in an implementation of a value domain. For example, in a set of permissible values representing the type of treatment for a particular disease, a patient may undergo more than one treatment type so more than one permissible value could legitimately be recorded for a patient.

Example: Mutually exclusive set of permissible values for 'after school carer'

	Representation values and meanings	Reason set of permissible values is poor/good
Poor set of permissible values	1. Parent 2. Relative 3. Neighbour 5. Other	The permissible values are not mutually exclusive because two of the categories overlap — i.e. code 1 'Parent' could be considered a subset of code 2 'Relative'. As a result, it is not clear which value applies to a parent and could lead to a parent who provides after school care incorrectly being counted twice; once against code 1 and once against code 2.
Good set of permissible values	1. Parent 2. Relative – other than child's parent 3. Neighbour 5. Other	The permissible values are mutually exclusive as code 2 'Relative' includes the qualification that it does not include a child's parent. This means there is no overlapping value categories.

2. Permissible values must be a true representation of the concept defined in the data element

It is possible that a question may exist on a data collection instrument that has a set of possible responses combining two concepts (e.g., the presence of disease and the type of disease). While not best practice, having mixed concepts on a data collection instrument does sometimes occur as a means of minimising the burden on data collectors. However, it is not acceptable to mix concepts when a data element is used to standardise data.

A data element must only include representational values (i.e. the value domain with its set of permissible values) that represent a single concept as defined by a data element concept. Where data are collected on two related concepts, each concept must be represented by a separate data element (e.g., one data element for presence of disease and one data element for type of disease).

Example: Permissible values for a 'mode of contact with a client' data element

	Representation values and meanings	Reason set of permissible values is poor/good
Poor set of permissible values	1. No contact 2. Phone contact 3. Email contact 4. In person	The inclusion of the value 'No contact' does not align with the concept 'mode of contact' for the data element. Although 'No contact' is not a mode of contact, it could be an appropriate representation for a data element concept about whether or not contact is made with a client. If data were also required on such a concept, a separate data element would be required with the appropriate representation values (e.g., 'yes/no' values).
Good set of permissible values	1. Phone contact 2. Email contact 3. In person	The permissible values reflect the 'mode of contact' concept for the data element.

4. Permissible values must avoid the use of a code value for 'other' that may limit future expansion of the code set or makes use of a common supplementary value

Permissible values must 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
- may be commonly used as a supplementary value.

The value 'other' (and any other value that is a synonym for 'other'; for example, 'other modes of contact not elsewhere categorised'), is a permissible value, not a supplementary value. In allocating a code value for 'other' (or a synonym for 'other') in an enumerated set of permissible values, it is important to consider any possible future expansion of the value categories. It is also important to take into account the code values used for any supplementary values (if applicable to the value domain).

Avoiding a code value for 'other' that is contiguous with the last code:

When using 'other' to ensure a set of permissible values is exhaustive, avoid specifying a code value for 'other' that is contiguous with the last code in the list of permissible values (i.e. if the last code in the list ends with 3, 'other' should not be given the value of 4). The reason for this is to give 'space' for the code set to be expanded, if necessary, in the future. If 'other' is given a code value that is contiguous with the last permissible code value, and another category is subsequently added to the list of permissible values at some future point, the value categories would not appear in a 'logical' sequence. If the list of permissible values is renumbered as a result of its expansion to put the value categories in order, this would result in data for subsequent collection periods being coded differently from earlier periods. This in turn may create problems for time series data analysis.

Avoiding a code value for 'other' that may be commonly used as a supplementary value:

In addition to the above, a code value for 'other' that is commonly used as a supplementary value should also be avoided. This generally means avoiding the use of code values '7', '97', '997' or '9997' and so forth that are commonly used for 'not applicable'; '8', '98', '998' or '9998' commonly used for 'unknown'; and '9', '99', '999' or 9999' commonly used for 'not stated/inadequately described'. See principle 5 below for further information about supplementary values.

Example: Allocation of code value for 'other' in a 'mode of contact' set of permissible values

	Representation values	Reason set of permissible values is poor/good
Poor set of permissible values - with 'other'	1. By phone 2. By email 3. In person 4. Other	The set of permissible values contains a code value for 'other' (code 4) that is contiguous with code 3 'In person'. If it was subsequently decided to collect data on 'mode of contact' that included 'by letter', there would be no space in the code values for the additional value category to be added before 'other'. See expanded set of permissible values lists (1) and (2) below.
Poor set of permissible values – expanded list (1)	1. By phone 2. By email 3. In person 4. Other 5. By letter	In this expanded set of permissible values, code 5 has been allocated for the additional category 'By letter'. As a result, 'by letter' is listed after 'other', which makes the list of values appear out of order (i.e. not in a 'logical' sequence).
Poor set of permissible values – expanded list (2)	1. By phone 2. By email 3. In person 4. By letter 5. Other	In this expanded set of permissible values, the value category 'other' has been renumbered to enable the additional category 'By letter' to appear in order. As a result, problems may arise with comparing data over time.
Good set of permissible values - with 'other'	1. By phone 2. By email 3. In person 4. Other	The set of permissible values contains code value for 'other' (code 6) that is not contiguous with code 3 'In person'. This allows for future expansion of the set of permissible values, if required, without the values appearing out of order or the code set being renumbered. See expanded set of permissible values below.

	Representation values	Reason set of permissible values is poor/good
Good set of permissible values – expanded	1. By phone 2. By email 3. In person 4. By letter 5. Other	In this expanded set of permissible values, there was space in the set of codes between code 3 'In person' and code 6 'other' for code 4 to be allocated to the additional value category 'by letter'. The value codes and categories appear in a 'logical' sequence.

6. The need for supplementary values in a value domain should be considered carefully

Data collected for statistical purposes needs to be precisely defined and enumerated to provide meaningful statistical information. For statistical purposes, it is also important that 'missing information' be understood. Supplementary values are an important means of ensuring the integrity of a statistical data collection by enabling a data analyst to understand why a valid value is missing from a field where such data is expected.

When a data collection and/or recording system includes a mandatory field, a supplementary value will be necessary to avoid errors if there is a valid reason for a value to be missing. Reasons why a value may be missing include: the information is unknown or not applicable; data have been collated from multiple sources, some of which did not collect the variable, or; the collection instrument was only partially completed (e.g., the reverse of a paper form was not filled in.)

Careful consideration needs to be given to the use of supplementary values and their meanings. Supplementary values are not required for all value domains. For example, where the element collects a binary 'Yes/No' response, analysis may only be required for responses which are explicitly 'Yes.' Alternatively, a response may be conditional on the response to another data element. When supplementary values are required, they do not fall within the scope of the value domain definition — i.e. they are not referred to in the value domain definition.

Additionally, not all potential supplementary values will be required for all value domains. Unless it is intended to analyse why data is missing, a single supplementary value of "Not stated/ inadequately described" is used, in order to limit variations in meaning. However, if the owner of a data collection wants to know the reasons for missing values (e.g., 'data provider refused to provide the information', 'data provider was unable to obtain the information as this time' or 'data provider did not know the information'), additional supplementary values could be appropriate as possible cues for further action.

7. Supplementary code values that are contiguous with the last code in the permissible value sequence are to be avoided

When using supplementary values, it is important to avoid using code values for supplementary values that are contiguous with the last code in the permissible value sequence. The reason for this is the same as when adding an 'other' value category to a set of permissible values (see principle 3 above); that is, to allow room for the code set to be expanded in the future, if necessary, without code values appearing out-of-order, or the need to re-allocate values. This can often lead to a reduction in data quality as a result of values being recorded incorrectly due to values being out of order.

Example: Allocation of supplementary value code for 'location of birthmark' set of permissible values

	Representation values *(italics used for illustration purposes only)*	**Reason set of permissible values is poor/ good**
Poor set of values	Permissible values: 1. Torso 2. Face 3. Arm 4. Leg 5. Other Supplementary values: 6. Not stated/inadequately described	This set of values does not allow room for possible expansion in the future because the supplementary code value '6' (not stated/ inadequately described) is contiguous with the last permissible code value '5' (other) (which itself is contiguous with the code value '4' (leg)). If the permissible value code set required expansion (e.g., to include a code for 'foot'), there would be two options to do this, neither of which is ideal: See 'Poor set of values – expanded set of permissible values (1)'. See "Poor set of values – expanded set of permissible values (2)'.
Poor set of values – expanded set of permissible values (1)	Permissible values: 1. Torso 2. Face 3. Arm 4. Leg 5. Other 6. Foot Supplementary values: 7. Not stated/inadequately described	To expand the code set to accommodate the additional 'foot' permissible value in this example, the next available permissible value (i.e. code 6) has been allocated to 'foot'. However, this has not only resulted in a permissible values code set that appears out-of-order (i.e. 'foot' appearing after 'other), it has also resulted in the need to change the 'not stated/inadequately described' supplementary value from '6' to '7'. This situation is not ideal.

	Representation values *(italics used for illustration purposes only)*	**Reason set of permissible values is poor/ good**
Poor set of values – expanded set of permissible values (2)	Permissible values: 1. Torso 2. Face 3. Arm 4. Leg 5. Foot 6. Other Supplementary values: 7. Not stated/inadequately described	To expand the code set to accommodate the additional 'foot' permissible value and keep the values in a logical order, in this example the permissible value 'other' has been changed from code '5' to code '6' so that code '5' can be allocated to 'foot'. As per the example above, this has also resulted in the need to change the 'not stated/ inadequately described' supplementary value from '6' to '7'. The changing of permissible code values and their meanings is not ideal, particularly for time series data, nor is the changing of the supplementary code value.
Good set of values	Permissible values: 1. Torso 2. Face 3. Arm 4. Leg 5. Other Supplementary values: 6. Not stated/inadequately described	This set of permissible values allows room for expansion between codes '4' (leg) and '8' (other). In addition, code '9' has been allocated to the supplementary value 'not stated/inadequately described' so it is not contiguous to the last code in the set of permissible values.
Good set of values – expanded set of permissible values	Permissible values: 1. Torso 2. Face 3. Arm 4. Leg 5. Foot 6. Other Supplementary values: 7. Not stated/inadequately described	As the original set of values (above) provided room for possible expansion, the category 'foot' can be added between codes '4' (leg) and '8' (other) without having to rearrange codes. The addition of the 'foot' permissible value has no impact on the supplementary value.

8. Supplementary values should be used consistently in a data collection

Supplementary values should be used consistently in a data collection by using:

- a default supplementary value meaning of 'not stated/inadequately described' to limit variations in the meaning within a specific data collection
- a logical set of supplementary values if using more than one supplementary value.

Consistency in the use of supplementary values within a data collection is important.

Use a default supplementary value meaning of 'not applicable/inadequately described':

Variations in the meanings of supplementary values within a data collection should be avoided. To limit variations in the meaning of values within a specific data collection, a default supplementary value meaning, such as 'not stated/inadequately described' is recommended (see common supplementary values below.)

If using more than one supplementary value, use a logical set:

Using a logical set of supplementary values when more than one supplementary value is required enables the consistent use of supplementary code values and their meanings across a collection (e.g., a data set should avoid having the code value '99' for 'unknown' in one data element, and the code value '98' for 'unknown' in another).

Common supplementary values:

Supplementary value code	Supplementary value meaning
7 (or 97, or 997, and so on depending on the field size)	Not applicable
8 (or 98, or 998, and so on depending on the field size)	Unknown
9 (or 99, or 999, and so on depending on the field size)	Not stated/inadequately described

N.B. Where multiple supplementary value codes are used, a guide for use should be included to clarify the difference between the values, e.g.,

CODE 8: Unknown. Use this code when the respondent did not know the answer.

CODE 9: Not stated/inadequately described. Use this code when the information is not stated or otherwise inadequately described (e.g. when a response has not been recorded, or is illegible.)

9. A supplementary value code field size should be the same as the permissible value, where possible

Ideally, the length of the field size of any supplementary value code should be the same length as the permissible values in the value domain, although this is not always possible to achieve. For example, if the codes for the permissible values use a field size that is 3 digits long (e.g., 001, 002, 003, etc.) the same field size should be used for the supplementary values (e.g., 997 for 'not applicable', and/or 998 for 'unknown' and/or 999 for 'not stated/inadequately described').

A set of permissible values may use up all of the codes for its specified field length (e.g., a set of permissible values using codes 1 through to 9 in a single digit field size). In such cases, if supplementary value(s) are required, ideally, the field size of the data element should be increased to accommodate the required supplementary value(s) (e.g., a single digit field size increased to two digits, and the supplementary value of code 99 for "not stated/inadequately described').

10. A valid permissible value must not be used as a supplementary value (and vice versa)

Permissible values are not supplementary values, and vice versa. Any value that is a permissible value must not be included as a supplementary value in a value domain. As stated in principle 4 above, this includes the permissible value 'other' (and any synonyms of 'other').

11. In described value domains, the supplementary value used must not be a valid permissible value

In described value domains (i.e. those without defined value meanings, such as in a measurement), the supplementary value used should not be a valid permissible value — it should be one that is not possible to be achieve based on the definition of the value domain. For example, if a data element collects the measurement of the height of a person in centimetres, it would be possible to use '997' (for 'not applicable'), '998' (for 'unknown') and '999' (for 'not stated/inadequately described') as supplementary values because the probability of a person being that tall approaches zero. However, it would not be valid to use the supplementary values of '97', '98' and '99' because these values could be actual person height values collected.