

Data Concept Repository Processes Standard

HISO 10014.1

Copyright information

This document has been approved as a standard for the New Zealand health and disability sector by the Health Information Standards Organisation (HISO), a sub-committee of HISAC.

The copyright owner of this document is the Health Information Strategy Action Committee (HISAC), which is a ministerial committee appointed by the Minister of Health. HISO work on this document may be reproduced in any number of copies and in any format or medium provided:

- the content is not changed;
- the material is not sold;
- the material is not used to promote or endorse any product or service;
- the material is not used in an inappropriate or a misleading context having regard to the nature of the material;
- any disclaimers included on the published information are reproduced on the material;
- a copyright acknowledgment to the Health Information Strategy Action Committee is included.

Permission to reproduce HISAC work does not extend to include any work identified in this document as the copyright material of a party other than HISAC. Authorisation to reproduce such material must be obtained from the copyright holders concerned.

Published in April 2007 by HISAC
PO Box 5013, Wellington, New Zealand

978-0-478-19115-8 (Online)
This document is available on the HISAC website:
<http://www.hisac.govt.nz>

Contents

1	Introduction	6
1.1	Purpose of the repository and business processes	6
1.2	Context of ISO/IEC 11179 for the repository	6
1.3	Objectives	7
1.4	Intended audience	7
1.5	Review of this Standard	7
1.6	Scope	7
1.7	Interpretation	8
2	Framework for the repository	8
3	Information Model	9
3.1	Base repository element	9
3.2	Concepts	12
3.3	Attributes	13
3.4	Data types	15
3.5	Relationship types	16
3.6	Relationships	17
3.7	Data items	18
3.8	Data groups	19
3.9	Data sets	19
4	Core business process	21
4.1	Development	21
4.2	Update concept or data set	23
4.3	Deprecation processes	24
	Appendix A: Glossary of Terms	25
	Appendix B: Sample form	30
	Appendix C: Data concept Repository Example	31
	Table 1 - Concept synonym examples	9
	Table 2 - Concept synonym examples	12
	Table 3 - Attribute synonym examples	14
	Figure 1 - Data Concept Repository – working framework	8
	Figure 2 - Develop, approve and implement new concept and/or data set	21
	Figure 3 - Update, approval and implement changed current concept and/or data set	23
	Figure 4 - Deprecate interim or current concept and/or data set	24
	Figure 5 - Data concept repository example	31

Committee Representation

Committee 10014-1 was responsible for the preparation of this draft document. It comprised the following representatives of nominating organisations:

Representative	Nominating Organisation
Jeffrey Cornwell	Accident Compensation Corporation (ACC)
Gina McEwen	Hawkes Bay District Health Board
Graeme Stretch	HealthLink Ltd
Peter Sergent	MedTech Software Limited
Tony Griffiths	Ministry of Health
Dave Scarrow	New Zealand Association of Pathology Practices
Douglas Kingsford	Orion Health
John Geering	Pharmac
Carole Ingley-Maraku	Te Upoko O Nga Oranga O Te Rae
Trish Anderson	The IPA Council of New Zealand (IPAC)
Dr Sue Wells	University of Auckland
Shona Wilson	Victoria University of Wellington
Patsy MacAulay	Waikato District Health Board

Related Documents

NZS Standards

SNZ HB 8169:2002 Health network code of practice

ISO Standards

ISO/IEC 11179:2004 Information technology -- Metadata registries (MDR) - Part 1: Framework

ISO/IEC 11179:2005, Information technology -- Metadata registries (MDR) - Part 2: Classification

ISO/IEC 11179:2003, Information technology -- Metadata registries (MDR) - Part 3: Registry metamodel and basic attributes

ISO/IEC 11179:2004, Information technology -- Metadata registries (MDR) - Part 4: Formulation of data definitions

ISO/IEC 11179:2005, Information technology -- Metadata registries (MDR) - Part 5: Naming and identification principles

ISO/IEC 11179:2005, Information technology -- Metadata registries (MDR) - Part 6: Registration

HISO

HISO: 10005 HPI Data Set. Wellington: Ministry of Health, 2004

HISO: 10006 HPI Code Set. Wellington: Ministry of Health, 2004

Other Standards

E-government Standards including the NZGLS (New Zealand Government Locator Service) metadata Standard

Health Level Seven Inc., HL7 Standard Version 2.4 - An Application Protocol For Electronic Data Exchange in Healthcare Environments.

W3C: XML Path Language (XPath) 2.0

New Zealand Legislation

Public Records Act 2005

Privacy Act 1993

Injury Prevention, Rehabilitation, and Compensation (IPRC) Act 2001

Other Publications

Archives New Zealand Electronic Recordkeeping Systems Standard

1 INTRODUCTION

Health care practice currently interacts with a variety of external services, e.g. in the on-line submission of electronic referral letters, ACC 45 forms and NHI (National Health Index) lookup. Most, if not all, of the interactive services currently available use different architecture and different Standards for implementing services, creating complexity and duplication of effort within the Sector. Because of the different approaches taken by the various organisations, health practitioners need to install, learn and operate a range of interfaces and techniques, which can provide a barrier to the uptake of clinical data capture.

A suite of Standards is proposed to define a mechanism for providing interactive communications between health care systems and external services.

It is envisaged that this document will define the business process required for all aspects relating to the consistent coding of data elements, including approval of new codes and deprecation of obsolete codes.

The data concept repository shall constitute a repository of metadata used for the purpose of describing data held in information systems within the New Zealand health Sector. Appendix C contains an example of a data concept repository.

1.1 Purpose of the repository and business processes

The purpose of the repository is to manage the coding of data held in New Zealand health information systems by providing a shared registry of concept definitions, coding nomenclature and minimum data set definitions. This shall include documentation sufficient to provide a common understanding of concepts and/or data sets.

The data concept repository and supporting business processes shall:

- (a) Enable consistent representation of data across the health Sector;
- (b) Inform the design of tools for centrally managing metadata and for coding health information relative to that metadata;
- (c) Facilitate acquisition and registration of agreed concepts and/or data sets;
- (d) Prevent duplication of concepts and/or data sets;
- (e) Enable consistent and easy access to metadata definitions;
- (f) Enable accurate and appropriate data/information to be collected and shared;
- (g) Communicate when concepts and/or data sets are added, altered or deprecated.

1.2 Context of ISO/IEC 11179 for the repository

The classification of clinical concepts is highly complex. It has been the subject of considerable inquiry over the years and has evolved into a specialist subject (e.g. refer to Open EHR at <http://www.openehr.org>). While the theoretical work to date has helped us to a better understanding of the complexity of the situation, it has not led to any practical implementations.

The primary health sector wants to exchange information using coded terms without the need to develop a complex dictionary or ontology.

Standards aim to facilitate the exchange of clinical information in a manner that does not require a comprehensive information model.

The recommended starting point for a repository is ISO/IEC 11179. This is a documented Standard from the International Standards Organisation (ISO) for representing metadata for an organisation in a metadata registry.

Concepts for the repository should make use of the existing coding and classifications system as much as possible and minimise the addition of customised local concepts. The order of preference should be:

- (a) International Standards (SNOMED, READ, ICD-10, etc.);

- (b) Australasian Standards;
- (c) New Zealand Standards;
- (d) Organisational Standards (Customised, e.g. ACC, NZHIS).

1.3 Objectives

The objective of a data concept repository process Standard is to:

- (a) Establish the process for defining and maintaining a core set of uniform definitions of concepts and/or minimum data sets;
- (b) Promote uniformity, availability, reliability, validity, consistency and completeness of data;
- (c) Accord with nationally and internationally agreed protocols and Standards, wherever possible; and
- (d) Promote national standard definitions by making them readily available to stakeholders.

1.4 Intended audience

The intended audience for this Standard is those in the health Sector who are defining and using data sets.

1.5 Review of this Standard

It is anticipated that this Standard will be reviewed at least every five years.

1.6 Scope

Inclusions to the scope of this Standard include:

- (a) Specifying key elements of concepts and/or data sets definitions;
- (b) Providing a framework for the specification of an information model;
- (c) Change management for:
 - (i) New concepts and/or data sets;
 - (ii) Altered concepts and/or data sets;
 - (iii) Deprecated concepts and/or data sets;
- (d) Communication to stakeholders when concepts and/or data sets are added, altered or deprecated.

Exclusions to the scope of this Standard include:

- (a) The actual data to be included in the repository (i.e. specific concepts or data sets);
- (b) Guidelines on the context of both collection and use of data elements;
- (c) Specifying validation rules for the concepts and/or data sets;
- (d) Specifying and maintaining a domain ontology;
- (e) Specifying repository interfaces;
- (f) Specifying interchange syntax for passing repository element structures between systems (via the repository interfaces);
- (g) Inferential semantics.

1.7 Interpretation

For the purpose of this Standard, the words 'shall' and 'will' refer to practices that are mandatory for compliance with this Standard. The words 'should' and 'may' refer to practices that are advised or recommended.

The terms 'normative' and 'informative' are used in Standards to define the application of appendices. A 'normative' appendix is an integral part of a Standard, whereas an 'informative' appendix is for information and guidance only. Informative provisions do not form part of the mandatory requirements of the Standard. Appendix A defines the terms used in this Standard.

2 FRAMEWORK FOR THE REPOSITORY

The framework for the repository shall be based on ISO/IEC 11179 Standards. Refer to Figure 1.

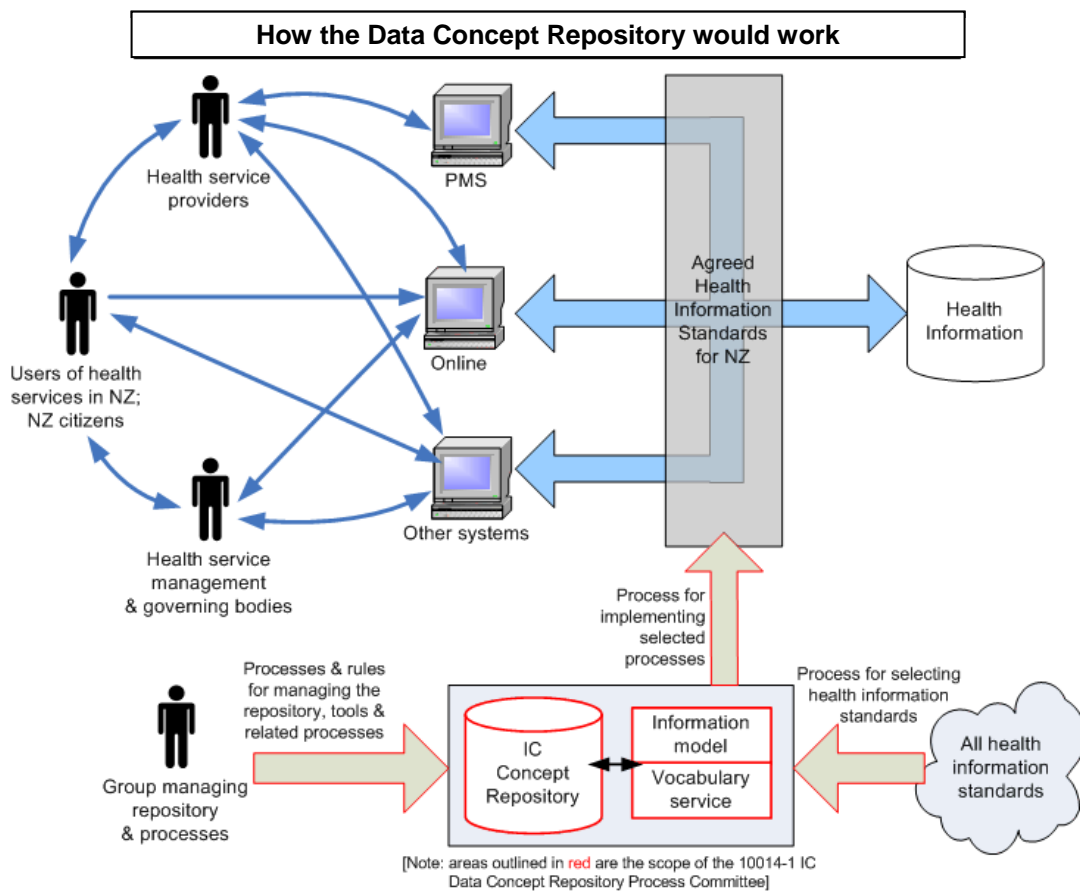


Figure 1 - Data Concept Repository – working framework

2.1.1 Purpose of ISO/IEC11179

The purpose of utilising the ISO/IEC 11179 Information Technology – Metadata Registries Standards is to provide a guideline and format able to specify and standardise data elements.

ISO 11179 consists of six parts focusing on:

- Part 1: Framework
- Part 2: Classification for data elements
- Part 3: Registry metamodel and basic attributes
- Part 4: Formulation of data definitions
- Part 5: Naming and identification principles for data elements
- Part 6: Registration

3 INFORMATION MODEL

The information model description contained in this section is informed by ISO/IEC 11179 Part 3. The intended purpose is to describe a core structure around which a more comprehensive information model could be built for actual implementation.

The data concept repository shall consist of a collection of definitions of concepts, attributes, data types, relationship types, relationships and data sets conforming to the specifications below.

An abstract item titled 'base repository element' is introduced in clause 3.1. This item contains elements common to the specification of all concepts, data types, relationship types and data sets. The specifications of concept, data type, relationship type and data set provided below should each be interpreted as including the specifications of the base repository element, in addition to the specifications listed.

All items held in the data concept repository shall be versioned. However, the mechanism for achieving this is not described in this specification.

3.1 Base repository element

3.1.1 Description

Any concept, data type, relationship type or data set is a base repository element. Therefore, the definition of concept, data type, relationship type and data set includes all elements of the definition of the base repository element.

The base repository element is an abstract construct that should not be explicitly implemented.

3.1.2 Definition (human readable)

Free text expression of the essential nature of the base repository element.

3.1.3 Synonyms

Each base repository element may have one or more human readable labels. The association between a base repository element and a synonym shall be specified by at least the following metadata:

- (a) Language/locale;
- (b) Target audience.

Refer to Table 1 for examples of synonyms.

Synonyms	Language	Target audience
Heart attack	English	Lay person
Myocardial infarction	English	Health practitioner
Food	English	Lay person
Kai	Maori	Lay person

Table 1 - Concept synonym examples

3.1.4 Active date range

The active date range of the base repository element shall be specified by the following:

- (a) Start date and time;
- (b) End date and time.

Base repository elements shall be deemed active or inactive in part based on start date and time and end date and time, where specified. Where no start date is supplied, base repository elements shall be deemed inactive. Where no end date is supplied, base repository elements shall be deemed active at any time after the start date.

Dates and times shall be recorded using the corresponding HL7 data types.

3.1.5 *Version*

Each base repository element shall have a version number, which shall be used to track changes to the base repository element.

3.1.6 *Status*

Each base repository element has a status code that indicates how it may be used when coding data.

Status codes shall include but are not limited to:

- (a) Proposed;
- (b) Approved;
- (c) Under review;
- (d) Deprecated.

3.1.6.1 *'Proposed'*

'Proposed' means that a base repository element has been submitted for inclusion in the registry but has not completed the expert review process required prior to the granting of approval. Such a code can only be used by the submitting organisation and its nominated associates, on the basis that:

- (a) These organisations share a common understanding of the meaning of the proposed base repository element;
- (b) The proposed base repository element must be reviewed by the registration authority before approval; and
- (c) During the review process, it may be determined that the submitted base repository element's definition, though unambiguously understood by the submitting organisation and its nominated associates, does not correctly describe the base repository element and requires modification prior to the issuing of approval, in order to ensure a universally understood definition of the base repository element.

The purpose of the above restriction is to avoid the use of a proposed base repository element during the approval process by third parties when the definition of that base repository element is as yet unapproved. If this was permitted, ambiguous ("dirty") data could result from the code being used with its initial (incorrect) definition, which could then be changed later following expert review.

3.1.6.2 *'Approved'*

'Approved' means that a base repository element has passed an expert review process and can be used generally for the coding of data, subject to the active date range specified for the base repository element.

3.1.6.3 *'Under Review'*

'Under Review' enables a base repository element to be flagged in the repository as requiring attention, but otherwise has the same connotations as 'Approved'.

3.1.6.4 *'Deprecated'*

'Deprecated' indicates that a base repository element shall no longer be used in the coding of data. The base repository element may optionally have been replaced with another base repository element.

NOTE: *An example of 'deprecated' is where a proposed concept is submitted and is used by the submitting organisation, but during the expert review process it is discovered that the concept is already represented (under a different concept identifier) in the repository. In this situation, the proposed concept would be deprecated and redirected to the existing concept.*

Another example is where the Ministry of Health substantially changes its reporting requirements for a particular notifiable disease and has previously submitted to the data concept repository a data set defining the data required when reporting that disease. In this circumstance, the previous data set would be deprecated and a new data set defined that contained the Ministry's current reporting requirements.

Deprecation illustrates two fundamental principles of repository construction:

- (a) That no two concepts in the repository may have the same meaning; and
- (b) That a concept, once entered into the repository, cannot be deleted.

3.1.7 *Originator*

The originator is the organisation submitting a base repository element for insertion into the repository.

The originator is coded relative to a central registry of organisations.

3.1.8 *Originator associates*

The originator associates are organisations certified by the submitting organisation as understanding the correct meaning of the submitted base repository element.

The originator associates are coded relative to the same central registry of organisations used to code the originators.

3.1.9 *Replacement base repository element*

The replacement base repository element specifies another base repository element in the data concept repository that is interpreted as having the same meaning as the referencing base repository element. This optional item has no meaning unless the base repository element's status is set to 'Deprecated'.

3.1.10 *Registration authority*

The registration authority is the organisation granting approval for base repository elements to be included in the repository.

3.1.11 *Vocabulary reference*

A base repository element may be mapped to one or more external coding terminologies/vocabularies/nomenclatures (these may include but are not limited to: SNOMED-CT, ICD-10, READ 2, READ 3 and LOINC).

Vocabulary references shall be represented using one or more instances of the HL7 CE data type.

3.1.12 *Administration history*

Administration history is a log of changes made to the base repository element's representation in the repository.

Each entry shall contain at least the following data:

- (a) Date and time (coded using HL7 data types);
- (b) Descriptive notes (e.g. recording suggestions, requests, objections and review outcomes);
- (c) Identification of the person making the entry;
- (d) For changes to repository data:
 - (i) Copy of the data prior to the change;
 - (ii) Description of the change (free text);
 - (iii) Purpose of the change (free text).

3.2 Concepts

3.2.1 Description

A concept is a unique entity with an unambiguous, clear and concise meaning, containing one or more items of information uniquely coded for generic/standard/common use by software programs. A concept may have attributes and/or relationships to other concepts.

A concept is a base repository element.

3.2.2 Unique identifier

Concept identification numbers are positive 8-byte integers, non-semantic and globally unique within the issuing authority.

Concept identifiers shall have the value 1 million or greater. This creates a space between zero and 1 million that can be used by vendors for legacy codes (with the understanding that such codes will not necessarily be interoperable with other systems).

3.2.3 Definition (human readable)

Free text expression of the essential nature of the concept.

NOTE: An example of this is: 'Abscess' = 'a localised collection of pus in a cavity formed by tissue disintegration and surrounded by inflamed tissue'.

3.2.4 Specification (machine readable)

3.2.4.1 Definitional

A list of attributes, constraints on attribute values and relationships to other concepts, that together unambiguously define the concept. The representation of attributes and constraints on attribute values is described in clause 3.3 and the representation of relationships to other concepts is described in clause 3.6.

NOTE: An example of this is: 'An operation may be defined as a procedure performed by a health practitioner possessing an approved scope of practice, on a patient, at a health care facility. An operation has an operation type, operator, a patient and a location.'

3.2.4.2 Qualifying

A list of attributes, constraints on attribute values and relationships to other concepts, that together qualify or otherwise provide context to the use of the concept.

NOTE: An example of this is 'Blood pressure', which can be qualified by 'patient posture' (e.g. sitting, standing, lying), but posture is not required in the definition of blood pressure.

3.2.5 Synonyms

Examples of synonyms are given in Table 2.

Synonyms	Language	Target audience
Heart attack	English	Lay person
Myocardial Infarction	English	Health practitioner
Food	English	Lay person
Kai	Māori	Lay person

Table 2 - Concept synonym examples

3.2.6 Version

3.2.6.1 Major changes

A major change is where the meaning of a concept is changed by the addition or removal of definitional attributes, relationships or constraints.

An intention to make such changes indicates a requirement to create a new concept and optionally deprecate the original concept.

NOTE: An example of this is if 'Diabetes Mellitus' has been defined as 'Type 1 Diabetes Mellitus' + 'Type 2 Diabetes Mellitus' and a change is requested to add 'Gestational Diabetes' to the definition of 'Diabetes Mellitus'. This would change the meaning of the 'Diabetes Mellitus' concept and hence require the addition of a new concept that was a generalisation on the previous concept of 'Diabetes Mellitus'. Therefore, a diagnosis of 'not Diabetes Mellitus' after this change would have a different meaning to a diagnosis of 'not Diabetes Mellitus' before this change.

3.2.6.2 Moderate changes

A moderate change may be the addition or removal of non-definitional attributes, relationships or constraints.

These are indicated by integer increments to the version number.

NOTE: An example of this is that adding 'patient posture' (such as sitting, standing and lying) to the definition of 'Blood Pressure' does not alter the meaning of Blood Pressure. Therefore, a blood pressure measured after this change but without reference to posture still has the same meaning as a blood pressure measured before this change.

3.2.6.3 Minor changes

Minor changes are restricted to alterations such as spelling and grammatical corrections that otherwise do not alter the interpretation of the concept.

These are indicated by 0.1 increments to the version number.

3.3 Attributes

3.3.1 Description

An attribute is a single item of information about a concept or relationship, other than a relationship between two or more concepts. Such items of information may optionally be aggregated into more complex structures, represented as either structured data types, or as other data definition repository concepts.

An attribute is defined using at least the subsequent metadata and is considered to be part of the concept or relationship to which it applies.

An attribute is not a base repository element.

3.3.2 Unique identifier

Attribute identification numbers are positive 2-byte integers, non-semantic and locally unique within the concept or relation to which they are attached.

3.3.3 Definition

Free text expression of the essential nature of the concept.

NOTE: An example is 'Waist Circumference' = 'a circumferential measurement of the abdomen taken through a horizontal plane midway between the lower rib margin and the iliac crests and measured to the nearest 1cm with the person standing'.

3.3.4 Data type

This defines the format of the information held in the attribute.

Permitted data types are defined within the data concept repository as described in clause 3.7.

3.3.5 Cardinality

This specifies how many values may be concurrently chosen from the domain of the data type when specifying the value of the attribute. Cardinality shall be specified by the following metadata:

- (a) Minimum (integer);
- (b) Maximum (integer);
- (c) Interval (boolean);
 - (i) If *true* then minimum and maximum together define an *interval* within the domain of the data type;
 - (ii) If *false* then minimum and maximum define the smallest and largest (“–1” for “unconstrained”) number of permitted *discrete values* within the domain of the data type.

NOTE: Examples are: ‘Heart Rate,’ requiring exactly one value of data type ‘positive integer’; ‘Complications,’ requiring zero or more values of externally defined data type ‘SNOMED CT’; ‘Reference range,’ being an interval from 3.5 to 5 on the domain of the data type.

3.3.6 Units

Where applicable, the unit of measurement of the attribute shall be specified.

3.3.7 Synonyms

Examples of synonyms are given in Table 3.

Synonyms	Language	Target audience
First name	English	General
Given name	English	General
Ingoa ko tahi	Māori	General

Table 3 - Attribute synonym examples

3.3.8 Valid values

Constraints may optionally be applied in order to restrict the permitted values of the attribute.

- (a) In the case of attributes with ordered domains, a restricted range of possible values may be specified, by way of the following metadata:
 - (i) Minimum permitted value; and
 - (ii) Maximum permitted value.

In the case of all attributes other than binary large object, an enumeration of possible values may be specified as a list of values.

Where constraints also apply to the attribute’s data type, permitted values are those consistent with both the attribute constraints and the data type constraints.

3.3.9 Vocabulary reference

Attributes may be mapped to one or more external coding terminologies/vocabularies/nomenclatures (these may include but are not limited to: SNOMED-CT, ICD-10, READ 3 and LOINC).

Vocabulary references shall be represented using the HL7 CE data type.

3.4 Data types

3.4.1 Description

A data type of an attribute is the set of values that the attribute may assume. Formal specification of a data type includes a mathematical model of the data, together with various operations defined on that model. The latter is beyond the scope of this Standard.

A data type is a base repository element.

3.4.2 Unique identifier

Data type identification numbers are positive 4-byte integers, non-semantic and globally unique within the issuing authority.

3.4.3 Specification (machine readable)

This defines a mathematical model for the data type.

A data type is (or is derived from) the following simple data types:

- (a) Integer;
- (b) Real;
- (c) Character;
- (d) Character string;
- (e) Boolean;
- (f) Date Time;
- (g) Binary large object;
- (h) Concept identifier (referencing a concept in the data concept repository);
- (i) Enumeration (a specified list of possible values);
 - (i) The enumeration data type is to be used where there is no specific data type underlying the enumeration (for example, the values {mild, moderate, severe}). Where it is intended to constrain the domain of another data type (such as specifying a range of permitted ICD-10 codes), enumerations should instead be specified as constraints on the permitted values of attributes (described in Clause 3.3.4).
 - (ii) The enumeration data type shall be interpreted as an integer within the Data concept repository
 - (iii) The domain of an enumeration shall be defined as one or more permitted values specified by the following metadata:
 - (i) Index (integer);
 - (ii) Label (string).
 - (iv) The enumeration shall further be defined by the following metadata:
 - (i) Ordered (boolean) – whether the enumeration is an ordered (true) or unordered (false) list.
- (j) Externally defined data type (a data type defined outside the Data concept repository, such as SNOMED CT or ICD-10 codes)
 - (i) In the case of externally defined data types, specification of the external domain shall be represented using the HL7 CE data type and the data type shall be interpreted as a character string within the data concept repository.

Constraints may optionally be applied in order to restrict the permitted values of the data type.

- (a) In the case of data types with ordered domains, a restricted range of possible values may be specified, by way of the following metadata:
 - (ii) Minimum permitted value; and
 - (iii) Maximum permitted value.
- (b) In the case of all data types other than binary large object, an enumeration of possible values may be specified as a list of values.

3.4.4 *Version*

3.4.4.1 *Major changes*

A major change is where the meaning of a data type is changed by altering its underlying simple data type (such as 'integer' to 'real'), or by altering its structure (such as 'single value' to 'list').

An intention to make such changes indicates a requirement to create a new data type and optionally deprecate the original data type.

3.4.4.2 *Moderate changes*

A moderate change may be the addition or removal of elements from an enumerated value list, or alteration of value constraints on a continuous simple data type.

These are indicated by integer increments to the version number.

3.4.4.3 *Minor changes*

Minor changes are restricted to alterations such as spelling and grammatical corrections that otherwise do not alter the interpretation of the data type.

These are indicated by 0.1 increments to the version number.

3.5 Relationship types

3.5.1 *Description*

A relationship type provides an unambiguous, clear and concise meaning to a relationship between two or more concepts. Relationship types in the data concept repository correspond to verbs in a language.

Examples include 'Is A', 'Ingredient Of' and 'Causes'.

A relationship type is a base repository element.

3.5.2 *Unique identifier*

Relationship type identification numbers are positive 4-byte integers, non-semantic and globally unique within the issuing authority.

3.5.3 *Specification (machine readable)*

3.5.3.1 *Attributes*

The specification of a relationship type can optionally include a list of attributes and constraints on attribute values that together qualify or otherwise provide context to the use of the relationship type. The representation of attributes and constraints on attribute values is described in clause 3.3.

NOTE: An example of this is that 'Causes' can be qualified by 'latency/incubation period'.

3.5.3.2 *Roles*

A role is the part a concept plays in a relationship between two or more concepts.

The specification of a relationship type must include a list of the roles played by the participants in the relationship. Each role shall be specified by at least the following metadata:

- (a) Reference to a concept representing the role, specified as a concept identifier; and
- (b) Cardinality (the number of concepts that may take this role in the relationship).

NOTE: An example of this is that in the relationship 'Dr Smith did an appendectomy on John Smith', Dr Smith has the role 'surgeon', John Smith has the role 'patient' and there is only one patient but there may be one or more surgeons in such a relationship.

3.5.4 Processing instructions

The 'meaning' of a relationship type is in part conveyed by how it is handled by a software application processing content from the data concept repository (e.g. if 'Concept 1' 'Is A' 'Concept 2', then Concept 1 should have the characteristics of Concept 2, in addition to other characteristics specific to Concept 1.

In order to ensure that processing software applications each interpret the relationship type in the same way, a free-text description of the intended interpretation of the relationship type should be provided.

NOTE: For example, 'The 'Is A' relationship type should be interpreted as specifying a generalisation hierarchy within which multiple inheritance can occur'.

3.5.5 Version

3.5.5.1 Major changes

A major change is where the meaning of a relationship type is changed by altering the meaning of its processing instructions, or by changing or removing existing roles.

NOTE: For example, changing the intended interpretation of a relationship type called 'Risk Factor For' from 'there is a known association between the subject (a risk factor) and the object (a disease)' to 'the subject causes the object' (a much stronger statement, with different inferential implications), constitutes a major change.

An intention to make such changes indicates a requirement to create a new relationship type and optionally deprecate the original relationship type.

3.5.5.2 Moderate changes

A moderate change may be the addition or removal of attributes or attribute value constraints.

These are indicated by integer increments to the version number.

3.5.5.3 Minor changes

Minor changes are indicated by 0.1 increments to the version number. These are restricted to alterations such as spelling and grammatical corrections that otherwise do not alter the interpretation of the relationship type.

3.6 Relationships

3.6.1 Description

A relationship is an association between two or more concepts. The meaning of a relationship is represented by its relationship type and by the roles of the participating concepts.

A relationship is *not* a base repository element.

3.6.2 Unique identifier

Relationship identification numbers are positive 8-byte integers, non-semantic and globally unique within the issuing authority.

Relationship identifiers shall have the value 1 million or greater. This creates a space between zero and 1 million that can be used by vendors for legacy codes, with the understanding that such codes will not necessarily be interoperable with other systems.

3.6.3 *Specification (machine readable)*

3.6.3.1 *Relationship type*

The identifier of the relationship type specifying the meaning of the relationship.

3.6.3.2 *Concepts*

A list of concepts participating in the relationship. Each participating concept shall be specified by at least the following metadata:

- (a) Participating concept (specified by way of a concept Identifier);
- (b) Role (specified by way of a concept Identifier).

3.6.3.3 *Attributes*

A list of attributes and constraints on attribute values that together qualify or otherwise provide context to the relationship. The representation of attributes and constraints on attribute values is described in clause 3.3.

3.7 **Data items**

3.7.1 *Description*

A data item is a single, irreducible piece of data found in a data group (defined in clause 3.8), or data set (defined in clause 3.9).

NOTE: An example of this is 'height' and 'weight'.

A data item is not a base repository element.

3.7.2 *Unique identifier*

A data item identifier is a label that is unique within the data set containing it and describes the data item in the context of that data set.

NOTE: An example of this is 'NHI Number'.

3.7.3 *Specification (machine readable)*

The meaning of each data item is specified by way of reference to the concepts, attributes and relations contained within the data concept repository.

An important issue is that a concept in the data concept repository may participate in more than one other concept's definition, e.g. a concept 'address' might be part of the description of both a healthcare provider and a patient. Therefore, when referring to a concept in the definition of a data item, it is necessary to specify the 'path' through the information model that leads to the concept *in the context in which it is being used*. In the previous example, two paths to 'address' might be 'patient/address' and 'healthcare provider/address', each path providing a clear context to the use of the 'address' concept.

A data item shall be specified by at least the following metadata:

- (a) Concept or relationship unique identifier;
- (b) Attribute unique identifier; and
- (c) Path to the concept or relationship through the information model.

Paths shall be expressed using the W3C XPath 2.0 Standard.

3.8 Data groups

3.8.1 Description

A data group is a container for one or more data items. Data groups are found in data sets (defined in clause 3.9).

Data groups are to data items what sections on a form are to individual form fields; data groups are to data sets what form sections are to a form as a whole.

A data set is not a base repository element.

3.8.2 Unique identifier

A data group identifier is a label that is unique within the data set and describes the data group in the context of the data set.

NOTE: An example of this is 'Demographics'.

3.8.3 Specification (machine readable)

A data group is specified by a list of one or more data items and/or data groups contained within the data group. Each list member shall be specified by at least the following metadata:

- (a) Data item or data group unique identifier;
- (b) Minimum number of permitted occurrences of the list member in the data group (zero or more);
- (c) Maximum number of permitted occurrences of the list member in the data group (one or more).

NOTE: An example of this is 'Diagnosis', which might contain a coded list of diagnoses where there must be at least one diagnosis, but there may be no more than 10 diagnoses (e.g.).

3.9 Data sets

3.9.1 Description

A data set is a collection of data collected for some defined purpose, such as a minimum data set used to describe an encounter with a diabetic patient. This data is specified as data items and data groups.

Data sets correspond to the content of forms and messages, whereas concepts, relationships and attributes represent the *meaning* of the data items in a data set. Data set rendering information, such as a particular form layout that might be used to capture the data, has not been included in this specification.

A data set is a base repository element.

3.9.2 Unique identifier

Data set identification numbers are positive 8-byte integers, non-semantic and globally unique within the issuing authority.

Data set identifiers shall have the value 1 million or greater. This creates a space between zero and 1 million that can be used by vendors for legacy codes (with the understanding that such codes will not necessarily be interoperable with other systems).

3.9.3 Definition (human readable)

Free text expression of the essential nature of the concept.

NOTE: An example of this is 'XDHB Diabetes Minimum Data Set' = 'The set of clinical data that X District Health Board wishes to be recorded about a diabetic patient at each visit'.

3.9.4 *Specification (machine readable)*

A data set is specified by a list of one or more data items and/or data groups contained within the data set. Each list member shall be specified, as previously defined in clause 3.8.3.

3.9.5 *Version*

3.9.5.1 *Major changes*

A major change is where data items are added to or removed from the data set.

These are indicated by integer increments to the version number.

NOTE: *An example of this is if 'Blood Pressure' has been added to a data set.*

3.9.5.2 *Minor changes*

Minor changes are restricted to alterations such as spelling and grammatical corrections that otherwise do not alter the interpretation of the data set.

These are indicated by 0.1 increments to the version number.

4 CORE BUSINESS PROCESS

This section contains the minimum required processes for the repository. The repository may have additional processes for management and administration of concepts and/or data sets. The minimum required processes shall include, but are not limited to, the following:

- Introduction of new concepts and/or data sets;
- Updating and management of current concepts and/or data sets;
- Consultation and communication with the broader health Sector;
- Deprecation of discontinued concepts and/or data sets.

4.1 Development

Figure 2 demonstrates the responsibilities for development, approval and implementation of new concepts and/or data sets, detailed in this clause.

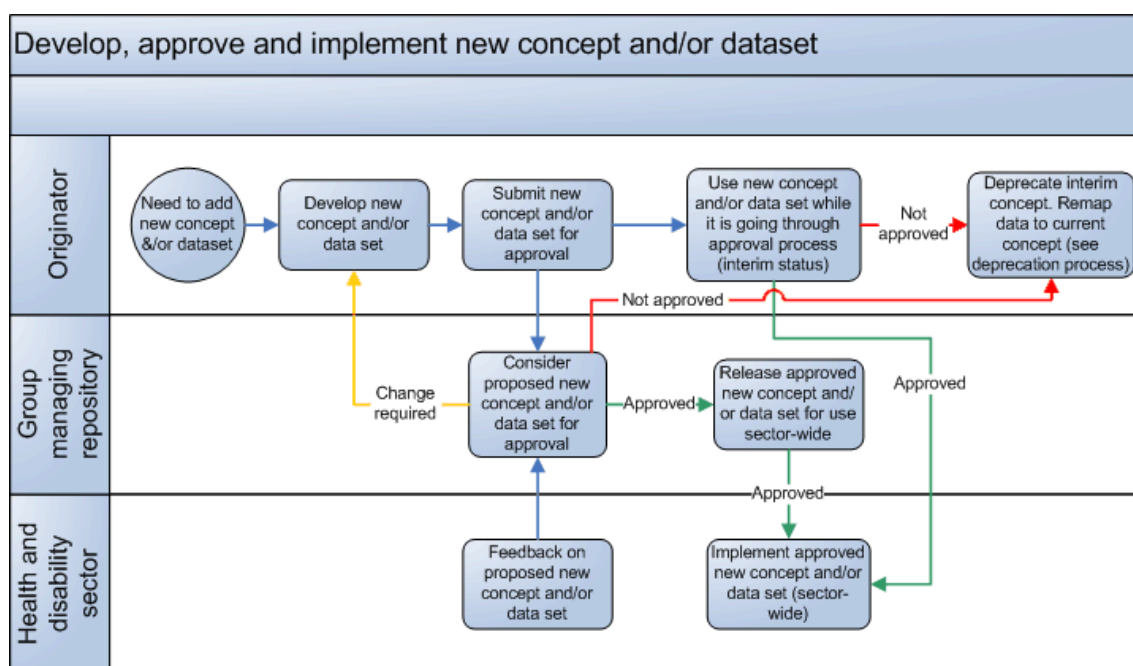


Figure 2 - Develop, approve and implement new concept and/or data set

The originator of a new concept or data set shall:

- Investigate the use of the concept or data set and the business requirements for use of the concept or data set;
- Check the existing repository for potential duplication, terminology and purpose conflicts;
- Unambiguously define the required concept or data set, using either the form in Appendix B, or an alternate process agreed to by the repository manager;
- Nominate associates to be authorised to use a concept or data set, while it has the 'proposed' status.

4.1.1 Submit

The originator of a new concept or data set shall:

- Submit the definition of the concept or data set to the repository manager, using the form in Appendix B, or any other mechanism that the governing body of the repository has specified;
- Obtain a concept or data set number from the repository manager, which shall have a 'proposed' status and be authorised for use by the originator and the originator's nominated associates only.

Upon receipt of an application for a new concept or data set, the repository manager shall:

- (a) Register the concept or data set;
- (b) Assign the concept or data set a 'proposed' status;
- (c) Notify the originator and its nominated associates of the code for the concept or data set.

4.1.2 *Use of concept or data sets with 'proposed' status*

The originator and its nominated associates may:

- (a) Use the concept or data set for the purpose of collecting and/or coding clinical data;
- (b) Review the concept or data set to ensure that it meets its intended operational requirements;
- (c) Resubmit an updated proposal with changes to the concept or data set definition prior to final approval status being granted.

Not all concepts or data sets will be granted 'approved' status. The originator and its nominated associates shall ensure the use of the concept or data set is able to be deprecated in the event that 'approved' status is not granted.

4.1.3 *Approval process*

The repository manager shall ensure that the:

- (a) Proposed concept or data set is not a duplicate of any approved concept or data set currently contained in the data concept repository (noting that in the event that two proposed concepts share the same meaning, the first concept to be granted "approved" status will result in the second concept being rejected);
- (b) Definition of the proposed concept or data set is clear, concise and unambiguous for use within the wider health Sector, in order to ensure a universally understood definition of the concept or data set;
- (c) Concept or data set meets the general requirements of the broader health Sector;
- (d) Concept or data set is practical for use in the health Sector.

4.1.4 *Consultation*

The repository manager shall make concepts and data sets available to the broader health Sector and ensure that:

- (a) Consultation takes place with Sector representative organisations to ensure that the concepts and/or data sets are acceptable for use;
- (b) Feedback is considered where required and following further input and engagement with the originator, the affected concept or data set is updated.

4.1.5 *Final status review of concept or data set*

The final stage of approval of a concept or data set will be the update of the status to one of the following:

- (a) Final approval – recognises the concept or data set for use by the wider health Sector and ensures that it may enter the release process;
- (b) Change required – the 'proposed' status remains in place and a request is made to the originator and its nominated associates for further information or an updated proposal;
- (c) Deprecation – the status of the concept or data set is changed from 'proposed' to 'deprecated'. At this point, there may be a process for review. If there is no process for review requested, then the discontinuation process shall be followed.

4.1.6 *Release process*

The repository manager shall ensure that final approved concepts and data sets are:

- (a) Added to the repository for use by the wider health Sector;
- (b) Socialised through the wider health Sector;
- (c) Subjected to version control processes.

4.2 Update concept or data set

Figure 3 demonstrates responsibilities for development, approval and implementation of new concepts and/or data sets, detailed in this clause.

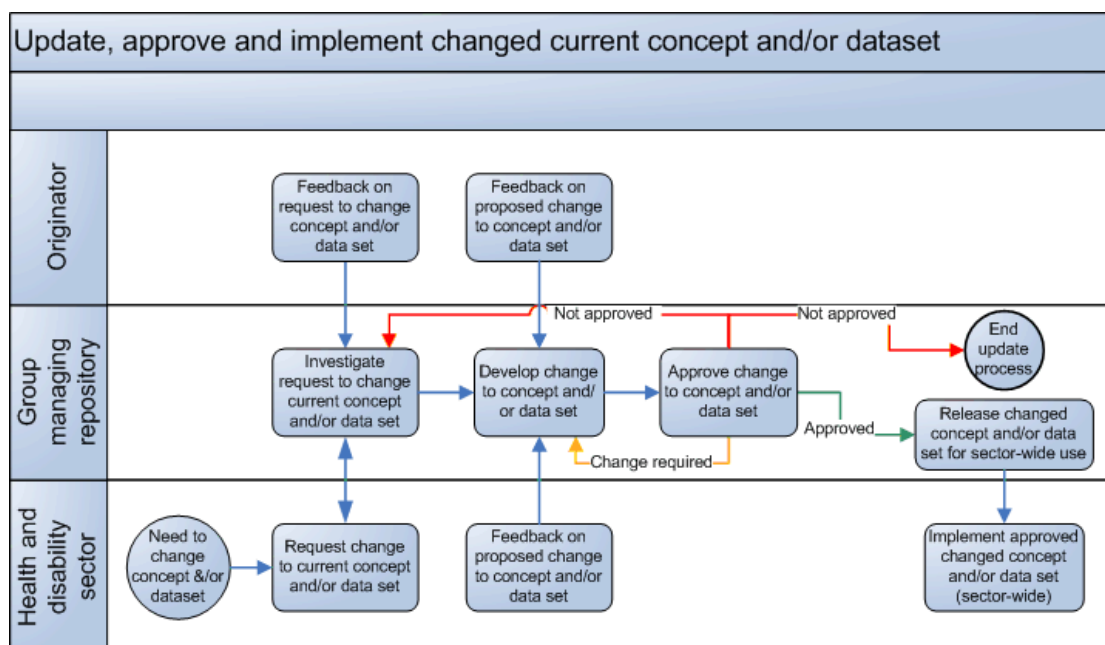


Figure 3 - Update, approval and implement changed current concept and/or data set

The repository manager shall:

- Receive and investigate feedback from the health Sector on request for change to concepts or data sets;
- Investigate the current use of concepts or data sets and the business requirements for use of concepts or data sets;
- Check the existing repository for potential duplication, terminology and purpose conflicts with proposed change requests;
- Check with the originator that the change to a concept or data set is valid within the current use of the concept or data set.

4.2.1 Consultation

The repository manager shall make proposed changes to concepts and/or data sets available to the wider health Sector and ensure that:

- Consultation takes place with sector representative organisations to ensure that the proposed change to a concept or data set is acceptable;
- Feedback is considered where required and following further input and engagement with the originator, the concept or data set is updated.

4.2.2 Approval process

The repository manager shall ensure that:

- Proposed changes to concept and/or data sets are made;
- The version numbers of updated concepts or data sets are updated.

4.2.3 Release process

The Repository manager shall ensure that final approved concepts and data sets are:

- (a) Added to the repository for use by the wider health Sector;
- (b) Socialised through the wider health Sector;
- (c) Subjected to version control processes.

4.3 Deprecation processes

Figure 4 demonstrates responsibilities for development, approval and implementation of new concepts and/or data sets.

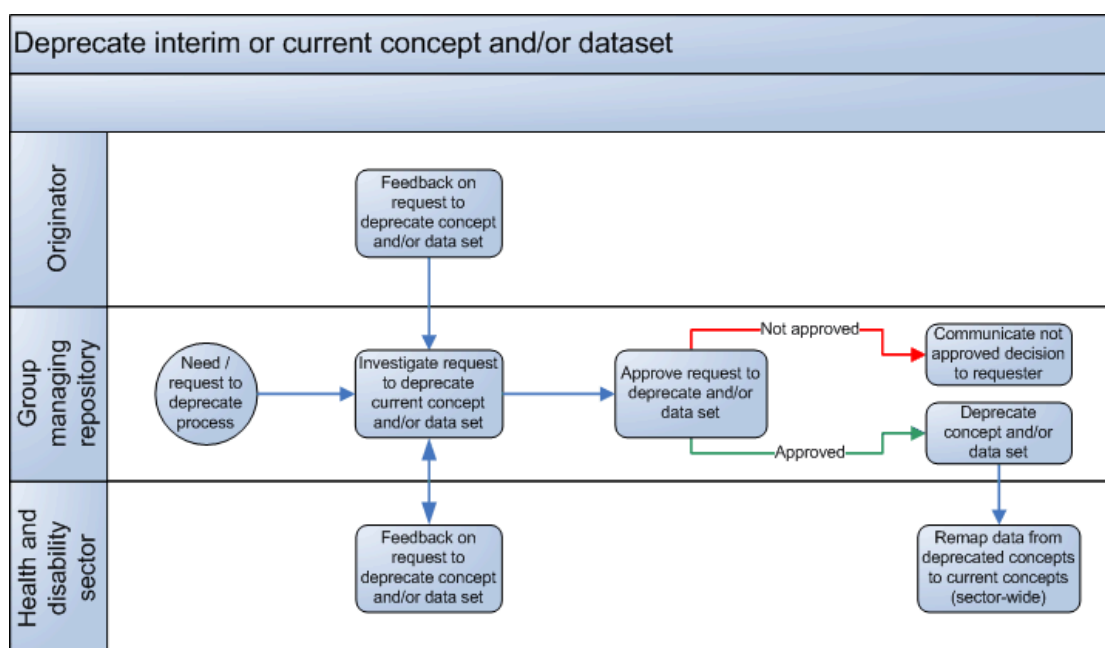


Figure 4 - Deprecate interim or current concept and/or data set

The repository manager shall ensure that deprecated concepts and data sets:

- (a) Have the status updated in the repository;
- (b) Are notified to the wider health Sector;
- (c) Are notified to the originator and their nominated associates.

Once a concept or data set has been deprecated there is no process for reactivation. If a new concept or data set which is similar is required, then the process in clause 4.1 shall be followed.

4.3.1 Review

Where it is proposed that a concept or data set be deprecated, the process for change in clause 4.2 shall be followed.

4.3.2 Remapping of codes

In the event that a concept or data set is deprecated and another equivalent and approved concept or data set exists in the data concept repository, the deprecated concept or data set shall be mapped to the equivalent approved concept or data set.

4.3.3 Deprecation status

Once agreement has been reached that a concept or data set should be deprecated, the status of the concept or data set shall be changed to 'Deprecated'.

APPENDIX A: GLOSSARY OF TERMS

(Informative)

Term	Definition
ACC45 forms	The New Zealand Accident Compensation Corporation Injury Claim Form, which collects basic information and is completed by Medical Practitioners to certify injury and incapacity for work.
Administration history	The log of changes made to the concept's representation in the repository.
Approved concept	When a concept has passed an expert review process and can be used generally for the coding of data, subject to the activity dates specified for the concept.
Architecture	The overall design or structure of a computer system, including the hardware and the software required to run it, especially the internal structure of the microprocessor.
Archives New Zealand Electronic Recordkeeping Systems Standard	Issued under the Public Records Act 2005 as a discretionary best practice Standard for the selection of a solution designed to assist in the management of electronic records. It defines a set of functional specifications to assist public offices and local authorities in the evaluation and selection of electronic recordkeeping systems.
Attribute	A single item of information about a concept or relationship, other than a relationship between two or more concepts. Such items of information may optionally be aggregated into more complex structures, represented as either structured data types, or as other metadata registry concepts.
Availability	The accessibility of the data in a timely manner.
Boolean	Refers to items that can only have two states - like on/off, or yes/no. Many programming languages have Boolean variables which can be used with operators such as AND, OR, NOT.
Change management	The process for changing concepts (or part thereof) that have been defined and reside within the repository.
Character string	A series of characters (any symbol that requires one byte of storage) manipulated as a group. A character string differs from a name in that it does not represent anything.
Classification scheme	The descriptive information for an arrangement or division of objects into groups based on characteristics, which the objects have in common.
Clinical data	Pieces of information that have been used to describe clinical care.
Coded types	The data type, where meaning is defined in some sort of external code scheme.
Coding	The process of classification of information, where diagnoses and elements of clinical function are assigned to categories.
Coding system	Set of agreed upon symbols (frequently numeric or alphanumeric) attached to concept representation or terms in order to allow exchange of concept representations or terms with regard to their form or meaning.
Completeness	Where the concept needs no more added to it to adequately describe what it is intended to describe.
Concept	A concept is a unique entity with an unambiguous, clear and concise meaning containing one or more items of information uniquely coded for

Term	Definition
	generic/standard/common use by software programs. A concept may have attributes and/or relationships to other concepts.
Concept identifier	Concept identification numbers are positive 8-byte integers, non-semantic and globally unique within the issuing authority.
Concept under review	When a concept is flagged in the repository as requiring attention.
Consistency	The degree of uniformity, standardisation and freedom from contradiction among the data.
Contributors	The person who enters a new concept as an interim concept for approval for inclusion within the repository.
Data	The representation of information in a manner suitable for communication, interpretation, or processing by human beings or by automatic means.
Data set	A named collection of logically related data items arranged in a prescribed manner.
Data type	The format of the information held in the attribute as the set of values that the attribute may assume. Formal specification of a data type includes both a mathematical model of the data together with various operations defined on that model, but the latter is beyond the scope of this Standard.
Data type identifier	Data type identification numbers are positive 4-byte integers, non-semantic and globally unique within the issuing authority.
Definition	The free text expression of the essential nature of the concept.
Definitional attributes	Attributes that describe the semantic aspects of a data element. These attributes may be derived by inheritance from characteristics of data element concepts, objects or entities.
Definitional specification	The machine readable list of attributes, constraints on attribute values, and relationships to other concepts that together unambiguously define the concept.
Deprecation	The process for phasing-out of a concept so that it cannot be used.
Domain ontology	A systematic arrangement of all the important categories of objects or concepts that exist in some field of discourse, showing the relations between them. When complete, an <i>ontology</i> is a categorisation of all of the concepts in some field of knowledge, including the objects and all of the properties, relations and functions needed to define the objects and specify their actions.
e-Government	A way of tapping unrealised potential for high quality government in New Zealand by adapting government to the environment of the information age and the Internet. It enables government agencies to separately and collectively lift their performance and deliver better results through using information and technology in new, more collaborative ways.
e-Government Standards	The New Zealand e-Government programme is dependent upon a number of Standards, collectively termed the e-Government Interoperability Framework (e-GIF). The programme also has Standards for metadata (or data that describes data) called the New Zealand Government Locator Service.
External coding terminologies	A defined terminology that is outside the scope of the concept repository.
Government Locator Service	The official New Zealand Government Standard for creating discovery-level metadata. The Standard is based closely on two well established

Term	Definition
	Standards: the Dublin Core Metadata Element Set and the Australian Government Locator Service.
Health Practitioner Identifier Code Set	The Standard that defines the code sets for the Health Practitioner Index (HPI) in New Zealand.
Health Practitioner Identifier Data Set	The Standard that defines the data to be provided in the Health Practitioner Index (HPI) in New Zealand.
Health Practitioner Index (HPI)	A centrally managed utility that is used to collect and distribute practitioner, health worker, organisation and facility data. The HPI will facilitate the timely and secure exchange of health information, ensure the accurate and unique identification of practitioners, workers, organisations and facilities and offer operational support for health organisations that use that data and provide information of interest to the public.
Healthcare providers	All professional services that provide either clinical or social care to users of health and/or disability services.
Hierarchical types	The organisation of data segments in a tree structure.
HL7 (Health Level 7)	An international community of health care subject matter experts and information scientists collaborating to create Standards for the exchange, management and integration of electronic health care information. HL7 promotes the use of such Standards within and among health Sector organisations to increase the effectiveness and efficiency of healthcare delivery for the benefit of all.
HL7 data type	Particular type of data that is based on the HL7 Reference Information Model (RIM) and used internationally to support standardised integration and comparability of health information.
ICD-10	The latest version in the series of International Classification of Diseases classifications.
IEC	International Electro-technical Commission.
Integer	A complete unit or entity.
Interactive communications	Occurs when sources take turns transmitting messages between one another.
Interval	Where a 'value' represents a <i>range</i> .
ISO	International Organisation for Standardisation.
ISO 11179 Parts 1 – 6, Information Technology	An ISO Standard for representing metadata for an organisation in a metadata registry.
Legacy codes	Source code that relates to a no-longer supported or manufactured operating system or other computer system. The term can also mean code inserted into modern software for the purpose of maintaining an older or previously supported feature - for example supporting a serial interface even though most modern systems don't have a serial port.
Metadata	Data that defines and describes other data and processes. Data becomes metadata when used in this way. This occurs under particular circumstances and for particular purposes, as no data is <i>always</i> metadata. The set of circumstances and purposes (or perspectives) for which some data is used as metadata is called the context. So, metadata is data about data in some context.
Major change	Where the meaning of a concept is changed by the addition or removal of definitional attributes, relationships or constraints.

Term	Definition
Minor change	Changes that are restricted to alterations such as spelling and grammar corrections that otherwise do not alter the interpretation of the concept.
Moderate change	The addition or removal of non-definitional attributes, relationships or constraints.
Name	The human readable label that describes each data type.
NHI	National Health Index – A unique identifier that is assigned to every person who uses health and disability support services in New Zealand.
Nomenclatures	A system of terms elaborated according to pre-established rules.
Non-definitional attributes	Attributes that do not describe the semantic aspects of a data element.
Ontology	Classification according to the network organisation of concepts meant to provide a model of some portion of the world. Consists of theories about the sorts of objects, properties of objects and relations between objects that are possible in that portion of the world.
Originator	The organisation submitting a concept for insertion into the repository.
Originator affiliates	Those organisations certified by the submitting organisation as understanding the correct meaning of the submitted concept.
Protocols	A set of guidelines.
Proposed concept	When a concept has been submitted for inclusion in the registry, but has not completed the expert review process required prior to the granting of approval.
Qualifying specification	A machine readable list of attributes, constraints on attribute values and relationships to other concepts that together qualify or otherwise provide context to the use of the concept.
READ 3	Clinical coding system which covers medical terms and procedural and administrative terms.
Real data type	A data type used in a computer programme to represent an approximation of a real number. Because the real numbers are not countable, computers cannot represent them exactly using a finite amount of information. Most often, a computer will use a rational approximation to a real number.
Registration authority	The organisation granting approval to concepts for insertion into the repository.
Relationship type	Provides an unambiguous, clear and concise meaning to a relationship between two or more concepts. Relationship types in the metadata registry correspond to verbs in a language.
Relationship type identifier	The identification numbers that identify relationships types as positive 4-byte integers, non-semantic and globally unique within the issuing authority.
Reliability	The consistency of a set of data.
Replacement data type OR Replacement concept	Specifies another concept in the metadata registry that is interpreted as having the same meaning as the referencing data type. This optional item has no meaning unless the data type's status is set to 'deprecated'.
Repository	A physical or logical compendium of data or 'information warehouse' that stores data longitudinally and in multi formats. A database of information.

Term	Definition
Representation	A value domain, data type, units of measure (if necessary) and representation class.
Role	The part a concept plays in a relationship between two or more concepts.
Semantic definition	A UML specification of the minimum set of constraints on relationships to other concepts, attributes and attribute values required to unambiguously define the concept.
Sets	Multi-valued list types.
SNOMED-CT (SNOMED Clinical Terms)	A dynamic, scientifically validated clinical health care terminology and infrastructure that makes health care knowledge usable and accessible.
Specification	The machine readable mathematical model that defines the data type.
Stakeholders	An individual or group with an interest in the success of a group or an organisation in delivering intended results. Any party that has an interest.
Standard	An output type – a consensus-based document that is developed by a balanced expert committee taking into consideration input received through public comment and engagement. Standards are generally widely applied and have high status.
Structured objects	Collections of two or more data elements.
Synonym	Where an attribute may be described by one or more human-readable labels.
Taxonomy	Classification according to presumed natural relationships among types and their subtypes. This is a hierarchical organisation based upon generalisation/specialisation and the mathematical notions of sets, subsets and set membership.
Terminology	A collection of terms with definitions for use in information systems databases, which enable comparisons to be made because the same terms are use to denote the same phenomenon.
UML	Unified Modelling Language (UML) is a non-proprietary, object modelling and specification language used in software engineering. UML includes a standardised graphical notation that may be used to create an abstract model of a system.
Uniformity	The degree of sameness of the data within a data set.
Validity	Quality of a set of data.
Vocabularies	Terms from a subject field and their definitions.
Vocabulary reference	Where data types may be mapped to one or more external coding terminologies/vocabularies/nomenclatures.

APPENDIX B: SAMPLE FORM

(Informative)

Concept Submission Form					
Unique Name	Myocardial Infarction (disorder)				?
Version	2.1	Review Prior Versions	View Change Log	?	
		Most Recent Version 2.0	Most Recent Status Approved		
Description	Death of the myocardium (cardiac muscle) resulting from interruption of blood supply to that area of the myocardium.				?
Definition					
Relations +	Role	Type	Other Concept		
	as subtype	Is A	type	Myocardial Disorder	More ?
	as subtype	Is A	type	Infarction	More
	as disorder	Occurs In	site	Myocardium	More
Attributes +	Attribute	Data Type	Units		
	Course	Course Types			More ?
	Onset	Rates of Onset			More
	Severity	Severity Types			More
Vocabulary Reference +	22298006	SNOMED-CT	?		
Synonyms +	Synonym	Language	Target Audience	Preferred	
	Myocardial Infarction	English	Health Practitioner	<input checked="" type="checkbox"/>	
	Heart Attack	English	Lay Person	<input checked="" type="checkbox"/>	
When Active	Start 17/10/06	End	?		
Originator	Ruapehu District Health Board				?
Originator Associates +	Ruapehu Community Clinic				?
	Central Island Medical Centre				

APPENDIX C: DATA CONCEPT REPOSITORY EXAMPLE

(Informative)

In Figure 5, a collection of concepts, relationships and attributes is shown to the left, a data set composed of data groups and data items is shown on the right and links are shown between data items on the right and corresponding concept attributes on the left.

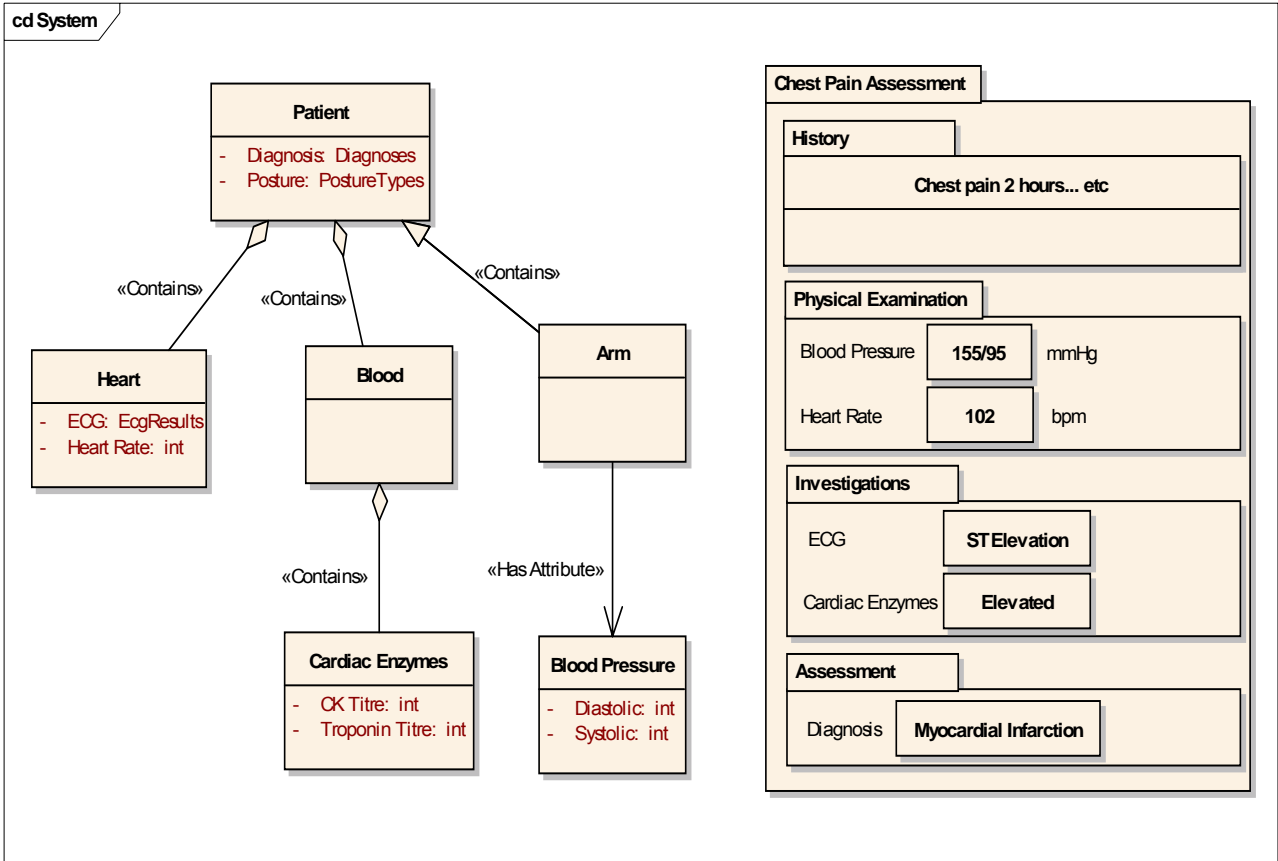


Figure 5 - Data concept repository example