

2 The development of Version 2

2.1 Introduction

2.1.1 Supporting material

Version 2 of the National Health Information Model (NHIM) is included in diagrammatic form, as an A3-size poster, at Appendix 1 of this working paper. Definitions of NHIM entities are provided in Appendix 2.

2.1.2 Status of this publication of Version 2

This stand-alone publication of the NHIM V2 is a working document of the National Health Information Management Group (NHIMG). The NHIMG has decided to publish the NHIM V2 only as a working document to ensure that progress made since 1997 can be incorporated into the model and the planned merging with the National Community Services Information Model (NCSIM) can be undertaken before formal publication of a new version.

Since the development of the NHIM V2 there has been considerable work undertaken by States and Territories in developing Enterprise Information Models (EIM) and Data Warehousing, which has progressed the understanding of generic health information requirements beyond that which was current at the time the NHIM V2 was completed. The State and Territory information models are at a greater level of specificity than the NHIM. A clear inheritance hierarchy should exist between the NHIM and these EIM. This hierarchy should be maintained in both directions so that the NHIM informs the development of the lower level models and in turn these models inform the next iteration of the NHIM. Late publication of the NHIMV2 would make it appear that the process of reviewing the NHIM in light of the lower level models did not take place when, in most cases, they were not available to inform the development of this version.

2.1.3 Changes from Version 1

The main differences between Version 2 and Version 1 are:

- change from an entity-relationship model to a high-level, relationship-free, multi-business framework
- deletion of the CLASSIFICATION SYSTEMS entity
- adoption of a formal process of ongoing management of the NHIM by the National Health Data Committee (NHDC) on behalf of the National Health Information Management Group (NHIMG)
- inclusion of some new entities, and changes to some existing entities; some definitions of entities have been expanded following consultation with a wide range of people working in Australia's health sector both nationally and at State and Territory level
- clarification of the treatment of DATE and TIME.

Version 2 has been in use (as an earlier draft form) since 1997 as the organising framework for the National Health Data Dictionary (NHDD) and the Knowledgebase. It was decided, however, to publish Version 2 as a working document at this time because of the importance of formally documenting and reviewing the transition to Version 2 that occurred in 2000.

Rather than being part of a specific development project, as was the case with Version 1, this latest draft version of the Model evolved through the deliberations of the NHDC and the NHIMG. This evolution took into account:

- developments in the field of 'information modelling'
- decisions made about the role of the model as a 'conceptual' information model, including the experience of Australian health information specialists in national and international standards development projects
- use of Version 1 of the NHIM in the Australian health and welfare sectors. This included its testing and development through its use as an architecture for the data elements in the NHDD and the resultant feedback from this use.
- decisions made about the use of entities and relationships for the future depiction of the NHIM.

The following sections review these aspects.

2.1.4 Use of established techniques and methods

In 1995 the NHIM publication of Version 1 stated that the project had used 'information engineering methods' and, in particular, 'information modelling' as the means of depicting the scope of information within the Australian health sector. The aim of using information engineering methods was to introduce a discipline to information development efforts, at least at a national level.

In selecting the entity-relationship diagramming technique to represent the NHIM, the project adopted one of the better known conventions in a field where a number of different conventions, techniques and methodologies persisted.

Between 1995 and 2000 it was reasonable to hope that the field of information modelling would have developed and 'crystallised' into a smaller number of more developed techniques. In reality, this was not the case with object oriented techniques, such as the Unified Modelling Language (UML), now being developed and offered, but with established information engineering techniques remaining relevant and hence persisting.

However, the important consideration is that an established, rigorous approach is still employed. Put simply, the difference is between using a 'model' with semantic interpretations that are consistently and inherently understood, and 'diagrams' which become pictures requiring interpretation by their viewer.

2.1.5 Issues with the entity-relationship approach

While the approach used to develop the NHIM has been criticised by some, it is the top-down method of developing the E-R model that has most likely drawn the assessment that it lacks formalism, not the E-R approach itself. After the development of the NHIM Version 1, the main area of discussion has been over the nature and applicability of the relationships.

In his original work, Peter Chen recognised that an entity could be strong or weak. A strong entity has no relational dependencies and, therefore, can exist in the absence of any relationships. A weak entity is dependent on the existence of another entity via an identifying relationship, however, these entities could be represented conceptually as a single entity.

Work by Codd and Date on the relational data model identified rules by which relationships may be expressed through foreign keys, resolving many-to-many (or weak) relationships by introducing 'intersection' entities. As the NHIM is not a fully attributed model, and is not intended to be, application of the relational model [applying normalisation techniques] cannot be fully applied. Relationships in this context may therefore be best omitted.

What this discussion tells us is that relationships can be problematic, and are often more important at lower levels of a set of models. This was recognised in Version 1:

At the high level of the National Health Information Model, the relationships depicted are very general. However, with further development at lower levels of the NHIM, relationships become very specific and can accommodate complex representations of specific rules and associations. (p. 22)

2.1.6 Current and emerging modelling and data definition techniques and formats

The work of Peter Chen and others in the late 1970s and 1980s was drawn together by James Martin, Clive Finkelstein and others through a body of work known as Information Engineering (IE). IE introduced the depiction of 'business models' tied to systems design principles. More recent developments of 'object oriented' theory have seen information modelling evolve further, including through Unified Modelling Language, Object-role Modelling and Extensible Markup Language. These emerging techniques are briefly discussed below. Their potential significance to the future development of the NHIM and their use in major information development projects in Australia are indicated.

2.1.6.1 Unified Modelling Language (UML)

UML is a relatively recent development. It tries to bring together business concepts with information concepts in order to produce a standardisation of terminology and diagram notation for the representation of objects. UML is becoming widely used for both database and software modelling. UML Version 1.1 was adopted by the Object Management Group (www.omg.org) in November 1997 as a standard language for object-oriented analysis and design.

UML is described as a 'language for specifying, visualising, constructing, and documenting the artefacts of software systems, as well as for business modelling and other non-software systems'. (www.rational.com/uml/index.jsp)

Although UML focuses on systems and software development and is mainly an object-oriented technique, it may be of significant relevance to the future development and depiction of the NHIM. However, it would be wrong to think that established techniques, such as E-R, have become redundant:

The *logical* view describes the design's object model when an object-oriented design method is used. To design an application that is very data-driven, you can use an alternative approach to develop some other form of logical view, such as an entity-relationship diagram. (The 4+1 View Model of Architecture, Philippe Kruchten, www.rational.com/products/whitepapers/350.jsp)

Critics of UML argue that object-oriented techniques are best used in the detailed design and implementation stages of systems development, and not in the conceptual representation, analysis or initial design stages.

UML has been used in the development of the HL7 Reference Information Model (RIM). In this conceptual model only the data component of objects has been modelled resulting in a model that is characteristically similar, although diagrammatically different, to an E-R Model. Conversion is therefore possible making the model applicable for lower level modelling using either Information Engineering or Object Oriented techniques.

2.1.6.2 Object-role Modelling (ORM)

ORM originated in the mid-1970s as a semantic modelling method. One of the early versions was the Natural Language Information Analysis Method (NIAM). ORM has since been extensively revised by many researchers.

ORM aims to design database models at the conceptual level. It uses terms easily understood by users, rather than the language of information systems or data structures. ORM pictures the world simply in terms of *objects* (entities or values) that play *roles* (parts in relationships). For example, you (an instance of an entity) are now playing the role of reading, and this working paper (an instance of another entity) is playing the role of being read.

One significant distinction of the ORM technique is that it does not allow for attributes, a central part of the E-R technique and currently an important link between the NHIM and the NHDD. The ORM technique considers the difficulty of representing relationships and attributes, and dictates that all attributes should be depicted as entities and roles. For example, if in E-R terms *Date of birth* is an attribute of a person, then in ORM terminology the entity PERSON would 'have' (role) a DATE OF BIRTH (another entity).

Experience with mapping the NHDD (with its list of data elements or attributes) to the NHIM (entities) has presented the Australian Institute of Health and Welfare (the Institute) and its partners with some problems. A consideration of the ORM technique may help to solve these.

2.1.6.3 Extensible Markup Language

The rapid and unpredictable explosion of the Internet, in particular the World Wide Web, has quickly generated a number of emerging 'standards' to allow documents and data to be produced, displayed and transmitted via the web. Although this field is relatively new and still evolving, XML has emerged as a standard format with significant potential.

XML, a project of the World Wide Web Consortium (W3C) (www.w3.org/), is described as the 'universal format for document and data on the web'. It is a specific version of SGML (ISO 8879 – the Standard Generalized Markup Language (SGML)). XML Version 1.0 was accepted by W3C in early 1998. Significantly, XML is the messaging markup language used for Health Level 7 (HL7) version 3 messages which are used by a significant number of hospitals in Australia and the United States of America.

HTML, the markup language most frequently used on the World Wide Web, is another version of SGML. SGML is often referred to as the 'mother tongue', with XML and HTML being specific implementations of it. XML was designed to provide significantly more flexibility and power than traditional HTML in implementing SGML on the web.

The promise of XML, as it applies to health information, is that it can use pre-defined data structures within web-enabled applications and documents, either for displaying or transmitting data via the Internet. XML includes a number of techniques, including Document Type Definitions (DTDs) and XML Schemas, to specify text-based document content in terms of data structures.

XML is not the same—and does not provide the same functionality or power—as the schema of traditional database management systems. It is mainly a text markup language, although it obviously does have significant application as the basis for software development and data definition.

In terms of Australian health information, Model entities and NHDD data element definitions could form the basis of XML data definitions, providing a standardised approach for the web-based transmission of health information. This field of work is likely to be the subject of considerable interest and research over the next few years.

2.2 Use of information modelling in the Australian health and welfare sectors

2.2.1 NHIM and NCSIM

In 1997, the NHIMG endorsed the NHDC recommendation to adopt Version 2 as the organising framework for printed versions of the NHDD. However, it reserved judgment about the Model's role in specific information development projects, and asked that this be considered in more depth.

In 1998, the National Community Services Information Management Group (NCSIMG) adopted a National Community Services Information Model (NCSIM) Version 1 as the organising framework for the initial edition of the National Community Services Data Dictionary. NCSIM Version 1 was based on Version 2 of the NHIM with some adaptations for the community services sector.

As outlined in the following sections, the development and use of a range of other models has formed the basis of the development of Version 2.

2.2.2 Experience in the development and use of other models

The increase in the number of entities in Version 2 reflects the further refinement of the entities presented in Version 1, as well as contributions flowing from the subsequent development of several sector-specific contextual models (definitions of entities in the NHIM are set out in Appendix 2). Each of the following national modelling projects has either already influenced the development of the entities in the NHIM, or is likely to do so in the near future:

- the National Community Services Information Model (NCSIM)—a conceptual modelling project directed at establishing an information framework for Australia's community service, welfare and housing sectors
- the Disability and Aged Care Model—a contextual modelling project that developed data items associated with the disability, disability services and aged care fields
- the Primary and Community Health Services National Information Model—a contextual modelling project that covered services delivering assessment and care, early identification and intervention, and health intervention/promotion activities to people (clients) in facilities located in the community
- the National Institution-based Ambulatory Care Model—a contextual modelling project that covered service delivery interventions for ambulatory patients in outpatient clinics and emergency departments

- the Community Health Information Model—a community health data model developed through extensive consultation with clinicians as part of the Community Health Information Management Enterprise initiative.

The State/Territory health departments are at various stages of developing enterprise information models (EIM) that are being designed to help develop their information architecture and system development processes. The lead has been taken by the New South Wales Department of Health, that developed an EIM in 1996 and has made this publicly available. Several States/Territories have based their developments on this EIM. For example, Western Australian Department of Health has recently completed development of an EIM which was initially based on the NSW model. Some divergence between the NSW Health EIM and the WA Department of Health EIM has inevitably occurred during the development process. There are now significant opportunities for the harmonisation of models, both between the jurisdictions' EIM and between the NHIM and these EIM. This harmonisation should result in the establishment of an accepted generic base model which would obviate the continued development of development models.

The jurisdictional models are intended to provide an intermediate framework that standardises corporate (State) data and the business rules that pertain to this data. These business rules are depicted in the relationships between entities. The NHIM V1 was used as a conceptual framework for the NSW model, therefore, it can be mapped upwards to the NHIM, and the data elements map to the NHDD. The development of these EIM will have an important influence in designing Version 3 of the NHIM. The high-level conceptual model should be harmonised with a number of enterprise information models and their differing business rules (relationships). The harmonisation of this notional hierarchy would require on-going maintenance, as the structure grows in both breadth, with the addition of new jurisdictional EIM, and depth, with the creation of more detailed subject area models.

2.3 Decisions on relationships at a national level

2.3.1 Describing relationships

The E-R technique specifies in some detail how and under what conditions one entity can be related to another. If, however, an information model is to serve as a basis for national (or industry-level) consensus, it must be capable of accommodating:

- the political, administrative and cultural diversity of multiple jurisdictions and sectors
- the context models developed within a range of modelling conventions.

Accordingly, and consistent with views on conceptual modelling, Version 2 of the NHIM abandons the use of relationships between entities. This does not constitute a loss of faith in the E-R convention for use at lower levels of modelling. It does, however, reflect the importance of the consistent identification of entities at the national level, and the greater importance of relationships or business rules at lower levels.

The loss of relationships from the model emphasises the role of the model in classification or grouping of items and correspondingly, reduces its potential role in systems development.

The work of the Institute and its partners has led to the conclusion that relationships between Model entities cannot be effectively described at a broad, national level. They are more appropriately described at more specific levels, such as in the context of information or applications development for a specific aspect of the health sector. This is arguably the key finding of our practical experience in applying information modelling techniques in the national health information development environment over the past 7 years.

At the national level, the focus is on the broad categories of information required for national purposes. These are, in turn, supported by more detailed information that relates to specific information requirements required for specific purposes. It is only at these more detailed levels that these relationships can be specifically identified, particularly as these relationships will differ from one specific application to another. For example, it is self-evident that there must be some kind of relationship between the NHIM entities of EVENT and STATE OF HEALTH AND WELLBEING. In a high-level model such as Version 1 of the NHIM, this relationship can be described only in fairly generic terms, and in this particular instance is described as EVENT influences STATE OF HEALTH AND WELLBEING. Clearly, the precise relationships will vary widely from case to case depending upon the particular nature of the event and the state of wellbeing and, as such, are best described within a particular context. For example, STATE OF HEALTH AND WELLBEING may be an expressed or perceived state, and may relate to an individual or a population group with particular characteristics. The EVENT may range from a routine appointment with a service provider to a highly significant life (or death), legal or physical event. Furthermore, the EVENT to STATE OF HEALTH AND WELLBEING relationship does not stand alone from other entities. This is because both these entities will each have pertinent relationships to other entities in the NHIM. Attempts to abstract these multiple, specific relationships up to a broad, national level have been counter-productive, as there are too many relationships to document effectively within a general conceptual model such as the NHIM. Furthermore, it is perceived that this could exclude particular relationships that would apply in given contexts.

This has led to the identification of a number of 'layers' of models that will be required to further the development of national health information.

- *Conceptual modelling* is a process aimed at establishing an agreed high-level structure that will prove suitable for use as an enterprise-wide information framework and, following further development, capable of supporting applications development exercises. The conceptual modelling process identifies and defines entities only, with greater emphasis given to achieving consensus about entity 'supertypes' than to the need to identify necessarily a full range of subtypes. Data elements (attributes) would not normally be present in a conceptual model.
- *Contextual modelling* is the further development of a conceptual model to accommodate the business rules and specific attributes of a particular jurisdiction, sector or specialisation. Contextual models are developed within the technical boundaries of a particular modelling notation, whether E-R, UML, ORM or other convention. A single conceptual model may accommodate several unique contextual models without loss of integrity, each one varying from the other, based on the particular sector or specialisation covered, or on the particular modelling convention used. For example, an E-R contextual model may be developed from a conceptual model by the addition of the relationships between entities that characterise the business rules of a particular jurisdiction (or context) and the enumeration of attributes belonging to entities. The process may well also define a significant range of data elements to populate the entities. The same conceptual model would also support the development of a UML contextual model for the same jurisdiction.
- *Physical data modelling* is the extension and refinement of a contextual data model to form the basis for applications development or database design. A physical data model is an attribute-rich implementation of a single contextual model and represents the most detailed implementation of the original conceptual model.

These layers have been found to be usefully interrelated – each implementation of a model at the next lower level challenges the integrity of the respective higher level(s). Although one might reasonably expect the higher level conceptual and contextual models to be more stable over time than physical data models, the interactive nature of the approach is such that the models are subject to constant and ongoing review and, where necessary, amendment.

Enterprise models may thus be developed as conceptual or as contextual models. The NHIM is an initiative within the first of these layers, i.e. it is a conceptual model.

2.3.2 Revised aims for the project

The NHIM Version 1 publication attempted to record how health information could and should be structured rather than reflecting, necessarily, how health information is currently structured. Version 2 continues to model the 'concept' of a health sector, rather than any particular system that might operate within that sector.

The aims of the NHIM are now more limited than they were previously, reflecting the specific role and niche that the NHIM has filled. The next few years, particularly with the advent of a number of major national health information initiatives, will further test and refine the Model's future role.

The aims of Version 2 are to continue to support :

- a common information management language and vocabulary for national health information projects where a data or information modelling approach is appropriate;
- a suitable framework for the development of sector-specific data models on which systems development activity might be more effectively based;
- a conceptual base for the NHDD; and
- stakeholders in conceptualising their information and data requirements.