

Templates and Archetypes: how do we know what we are talking about?

Sam Heard, Thomas Beale, Gerard Freriks, Angelo Rossi Mori, Ognian Pishev
Version 1.2, 12th February 2003

This discussion paper is addressed to the HL7 template group, the EHRcom Task Force of CEN/TC251 and the *openEHR* community (which have substantial overlap) in the interests of minimising the confusion that arises from using the terms ‘templates’ and ‘archetypes’. The problem is compounded by the *openEHR* approach proposing to need both archetypes and templates. Not only that, but that archetypes vary in the degree they express organisational convention and core clinical concepts. It is clear that we need a range of words to describe the artefacts that are required to provide ‘extensibility’ of information models in a manner that preserves their semantic interoperability - and that in current discussions at HL7 and other meetings people are *not* talking about the same things. This paper will attempt to describe the sorts of constraint models that appear to be required and how they might be named in *openEHR* and HL7.

THE WORDS

The word *template* has been used for many purposes and is defined in Webster’s as “something that establishes or serves as a pattern”. The word *archetype* has been used by Angelo Rossi Mori (actually ‘clinical archetype’) to describe the ontological basis behind the ‘clinical statements’ in CEN 13606 and by Thomas Beale and Sam Heard to describe the artefacts on which the *openEHR* approach is built. Webster’s defines archetype as “the original pattern or model of which all things of the same type are representations or copies”. From this one might conclude that ‘archetypes’ are a type of template.

How they are used in HL7 and openEHR

We propose that we use the word ‘template’ to describe any pattern or constraint - as a very generic concept. It is used to describe many things in computing. Constraints on the generic information model in HL7 are applied through the generation of RMIMs and CMETs with further constraint being applied through ‘HL7 templates’. Constraints in *openEHR* are applied through ‘archetypes’, while the set of archetypes used for a particular data collection can be set by an ‘*openEHR* template’.

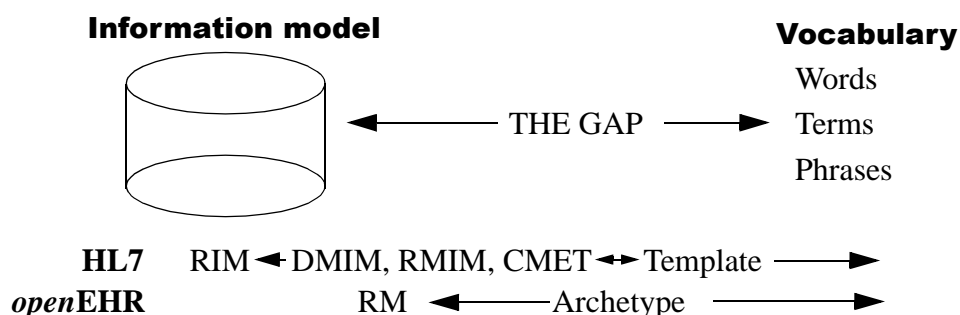
THE PROBLEM

The problem of ‘meaning’ in health information, and preserving this meaning when information is transferred between information systems (i.e. semantic interoperability) is complex. The idea that meaning is the product of context and content is helpful. The full meaning of information has traditionally been a product of the information system within which it is stored and the terminology used. The semantics of the information system are to a greater or lesser extent contained in the information model. The semantic gap between the information model and the vocabulary is usually bridged by program code and agreed conventions with users, and more recently by ‘meta-data’.

Burying the domain level semantics in the information system itself is a costly approach. The consequence is that meaning that cannot be expressed through an evolving terminology service must be dealt with by changing the information system. In health care, where development of concepts is rapid and continuous and where there is a wealth of different work settings, we have particular insight into this difficulty. Healthcare is one of the very few domains where sharing information is the norm, rather than the exception. If we are to have stable information models and stable software a new approach is required. This approach (outlined in “*Archetypes - An Interoperable Knowledge Methodology for Future-proof Information Systems*” by Thomas Beale¹) allows modelling of domain concepts external to the system information model. Many information systems today utilise metadata - often similar to the archetype approach - but the formal archetype methodology is an innovation.

HL7, as a standards organisation, concentrates on the design of generic information models that can be used in different contexts. The result is that a gap, similar to that in EHR systems, presents itself between the capabilities of the RIM derived models and the semantic requirements of a diverse set of users. As a consequence there is an even larger gap between the necessarily generic information model and the vocabulary semantics.

Figure 1: The semantic gap



One further feature of this ‘gap’ is that different users at different sites will have a very strong preference for how information is best organised or presented, and what information will be required or collected for different purposes. This provides a strong pressure to keep domain information out of the information model itself and in the constraint model to meet the need for specialisation and extensibility of the information. It is worth noting that some data, such as the recording ‘context’ such as dates, times and participations, are best retained in the information model, as these must be controlled.

What sorts of ‘things’ are in an EHR?

Consider the sorts of aggregations that might be considered discrete and whole concepts and might appear in an EHR.

1. A collection of concepts that together form fixed attributes of a higher level concept that is not recorded as its component parts alone - e.g.:
 - a blood pressure measurement with its two pressure measurements, patient position, cuff size etc.

1. <http://www.deepthought.com.au/it/archetypes.html>

- a body weight with details about the baby's state of undress and the device used for measurement
2. A generic concept (with other fixed attributes) that is a value or a collection of values which form a subset of a larger (or very large) known set - e.g.:
 - a diagnosis - the value - with fixed attributes such as the date of onset, the stage of the disease etc
 - a laboratory battery result which includes an arbitrary set of values - the collection - with fixed attributes such as the time of sampling, or a challenge applied to the patient at the time the sample was taken (e.g. fasting).
 3. A collection of these higher level concepts that are usually measured together and might be considered themselves concepts - e.g.:
 - Vital signs - with temperature, blood pressure, pulse and respiratory rate
 - Physical examination - with for example observation, palpation and auscultation (and other findings)
 4. A collection of these aggregations which might form a record composition or a document - e.g.:
 - A clinic progress note containing symptoms, physical examination, an assessment and a plan
 - A laboratory report that contains the results as well as interpretation and details about any notifications and referrals that have been made
 - An operation report detailing the participants and their roles, a description of the operation, any complications and followup monitoring and care required

The hierarchical aggregation at higher levels, such as the level of the document or the document section, will be more informed by best practice, use-cases and convention and have less impact on the semantics. This must be true or the meaning of health information in a health record would require the precise understanding of the context of its recording before it could be safely reused. Consider having to know the context of a pulse measurement in detail before you could review a person's pulse rate over the last 2 years, or for that matter, their blood pressure. There is no doubt that knowing someone was pregnant might alter interpretation of blood pressure readings - but the measurements remain systemic arterial blood pressure values despite the altered physiological state.

It is also worth noting that default values of the lower level concepts can usefully be set in certain situations - and so might be features of the higher level aggregations. An example might be that a blood pressure is virtually always measured in the sitting position in a midwife antenatal clinic. This would not be so in a cardiology clinic.

THE SOLUTION

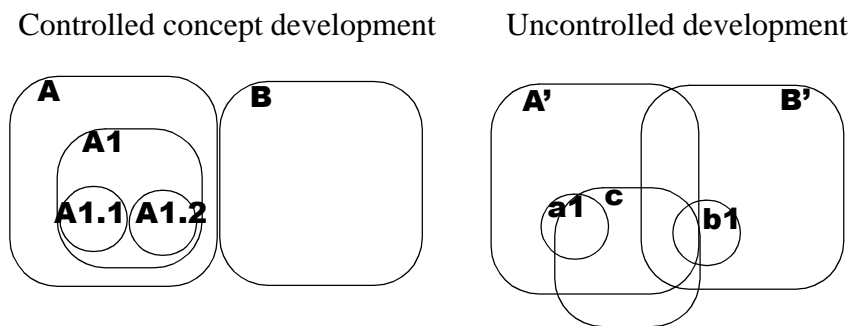
Solutions have been proposed to bridge the semantic gap and allow for the specialisation required by different users. In the *openEHR* work these are called archetypes and are the single means of achieving interoperability with a generic information model. In HL7, two means are used - a constrained version of RIM classes creating a specific information model called an RMIM and then a means of constraining this further called a 'template'. Both solutions have attempted to address the gap described above - the semantic gap and the need for specialisation and extensibility - enabling more generic, and hopefully stable, information models. Both approaches rely on vocabulary to populate these models.

Semantic considerations

Semantic meaning of an artefact is derived from definitions and its relationships with other artefacts of the same type, and related artefacts derived from the same knowledge base. This approach has, until now, been predominantly the domain of terminologies, though often without explicit definitions. With the increased availability of ontological tools (such as Protégé) and understanding of the requirements for preserving semantics (such as CEN's categorial structure), more information systems are using knowledge tools and basing their systems on ontologies (and models).

Any information construct, 'archetype' or 'template', that aims to preserve semantics must be related in some formal way to an ontology. Given requirements for extensibility and specialisation mentioned above, a strict rule set must be developed which governs how this might be achieved. Further, each concept represented as a template or archetype and defined in an ontology will have to be discrete and complete. If not, there will be no limit to the complexity and relationships of these concepts.

Figure 2: Controlled and uncontrolled archetype development



The aim of these constructs is to convey meaning and as such re-use should be maximised and the number of minimised. This will curtail complexity – something that has been very difficult to manage using vocabulary alone.

We propose that the concepts described above are represented as 'strong semantic models' - and that these should be linked to an ontology and called 'primary archetypes'. They will have the following features:

- The need for these models is to provide semantic interoperability
- These models should represent whole and discrete concepts that cannot easily be dealt with by terminology alone:
 - They may require values as well as terms
 - They are compound and simplify the need for complex vocabularies as well as allowing other relevant properties to be recorded
 - They are widely held concepts that are required for automatic processing
- These models are stored, at least in their generic form, in a knowledge base that is linked to a formal ontology. This ontology may support complex queries in an information system.
- These models will be registered by a standards body in a jurisdiction (the country, the region or a sector of health care)

Use considerations – specialisation and extensibility

There are many considerations about how information is stored in an EHR that are required for specific use but do not directly affect semantic interoperability. For example, a 'diabetic care message' and a 'CDA cardiology report' might contain a blood

pressure measurement taken on the same patient. The re-use and meaning of the blood pressure is not changed substantively (and certainly not semantically) from an automatic processing point of view by its recording in these contexts. Likewise, a medication order in hospital or primary care are semantically equivalent and will often continue to be relevant after transfer from one care setting to another.

It is apparent that there are two levels that the recordings are specialised and extended - organisation and specific entries and values. I will consider each of these separately.

Organisational models

There are shared models that apply across different settings and relate to record organisation. Examples are the recording of a clinician-patient interaction in a traditional manner - history, physical examination, diagnosis, management etc and a problem oriented approach where the SOAP model might be invoked multiple times in one record under a different problem heading.

Figure 3: Traditional and problem oriented organisation

<p>History</p> <p> Headache</p> <p> Toothache</p> <p>Physical exam</p> <p> Optic Fundi normal</p> <p> Swelling under R4 molar</p> <p>Diagnoses</p> <p> Tooth abcess</p>	<p>Toothache</p> <p> S: Painful tooth on R</p> <p> O: Swelling under R4 molar</p> <p> A: Tooth abcess</p> <p>Headache</p> <p> S: Overnight - poor sleep</p> <p> O: Optic fundi normal</p> <p> A: Secondary to toothache</p>
--	---

The organisation of a diabetic care message and a CDA cardiology report may vary in different settings - although, with increasing communication, alignment will be helpful. However, it is likely that organisational models will be standardised in a particular locality or institution, and that this might - with time - extend to a wider realm.

If the information model is sufficiently generic then there is a need for an artefact to express these 'organisational' constraints, recognising that while they express some knowledge, the meaning of the 'primary archetyped' information that these models 'organise' is not altered. Such organisational models may well (and even perhaps will usually) restrict the 'primary archetyped' models that may sensibly appear at this point in the EHR. Consider the "O:" in the problem oriented example above. It would be sensible to limit what would usually appear here to observations of the patient. But where does a blood pressure reading carried out by the patient belong? As the information source of all entries is carried in the entry, the blood pressure is probably best recorded under "O:" as this is where it will be expected to be located - carrying the fact that the patient measured it within. This is certainly debatable as consistency may only be required in the local work settings. We propose that these constraints are called 'organisational' models - organisational 'archetypes'.

The features of 'organisational' archetypes are:

- The need for these models is based on use-cases – local or more general - but still has strong links with the knowledge base. Thus a section called 'physical examination' and 'examination of the abdomen' can be known to be related;
- There need be no restriction on the number of these artefacts but their relationships should be understood;

- These models carry no primary semantic content - but the efficiency of humans and machines to query or read the meaning of the finer grain content would be diminished;
- Re-use of these models is encouraged – for aiding information navigation by users;
- There is no need for registration of these models for semantic interoperability; and
- These models may not be present in the ontology.

One might consider that the sharing of organisational concepts aids ‘clinical interoperability’ - enabling health providers to find groups of information quickly and reliably. It is important to recognise that organisational archetypes will have slots that can be filled by further organisational archetypes or archetypes <ref Rossi Mori>. These organisational ‘archetypes’ will describe the ‘organisers’ in *openEHR* and the ‘sections’ in HL7 CDA. While the names of these headings do not alter the semantics of the information structures held within them they do imply some knowledge - and may be derived from the meaning of the term used to label them.

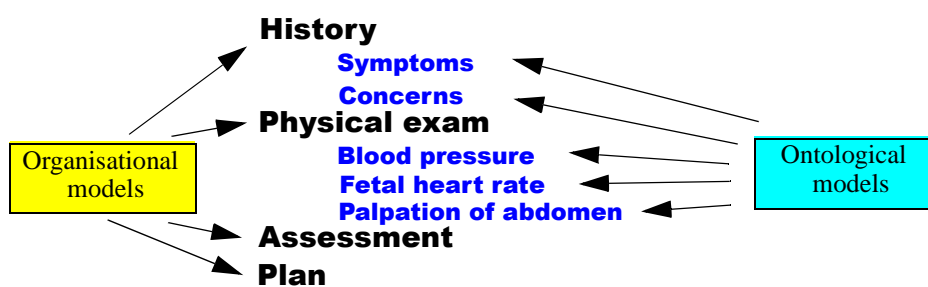
Constraint during data capture

A further level of artefact is required in EHR systems - one which specifies the use of the ‘organisational’ and ‘primary’ archetypes within a document or other record construct. This will be as specific as is required by a group of users and may be shared widely or not at all. These specifications have a firm foundation in specific environments and user preferences. The role of these specifications (which are called ‘templates’ in *openEHR* to differentiate them from archetypes) describe how the entries in the record are organised and specify what optional elements in the entries will be populated and what values and default values apply.

While these may well be registered for reuse, they will arise from consideration of best practice, specific demands for quality care, decision support systems and application requirements.

Consider an *openEHR* template for recording an antenatal care contact.

Figure 4: *openEHR* antenatal template example



The *openEHR* ‘template’ or constraint specification will show which organisational models are used, in which order, and which ‘primary’ archetypes these will contain. It will also set appropriate default values in the primary archetypes if required and describe, in the case of the antenatal examination, how the palpation ‘archetype’ is specialised to enable the position of the fetus in the uterus to be described.¹

1. It is worth noting that the *openEHR* approach would allow a generic ‘palpation’ archetype to be specialised for use in pregnancy only - and perhaps called ‘palpation of the pregnant abdomen’.

The new use of 'Primary' and 'Organisational' archetypes and templates

In summary, we propose that the term 'template' is used in HL7 and *openEHR* to describe a constraint specification for a message, a document or a fragment of this - always in terms of primary and organisational archetypes - and in addition in HL7 RMIM classes and CMETs. The term archetype is used to describe concepts that are recorded in the EHR which can be usefully linked to an ontology.

Table 1: Naming considerations

Name	Description	Current openEHR	Current CDA	Current HL7 messages
Primary archetype	Discrete whole concepts recorded in EHR	Entry archetypes	Entry templates and RMIM classes and CMETs	RMIM classes and CMETs
Organisational archetypes	Represent concepts used to organise EHR and may constrain the contained primary archetypes	Organiser archetypes	Section templates	?
Templates	Constraint specifications expressed in terms of organisational archetypes and primary archetypes	Templates	?	?

During detailed discussion, the terms templates, organisational archetypes and primary archetypes should be used. The HL7 and *openEHR* communities would need to agree that:

- The semantics of the HL7 may reside partly in the RIM, partly in the RMIM (and CMETs) and partly in the primary archetypes. Matching the semantics with the *openEHR* archetype may prove difficult in the message world - but should be easier in the somewhat aligned CDA world.
- The semantics of the *openEHR* archetypes reside in the archetype alone - which is linked to an ontology.
- The need to register primary and organisational archetypes is absolute - how many are required for use with the CDA is not clear as the power of the semantics expressed in the current version 2 RMIM has not been tested. *openEHR* is likely to require a few hundred to achieve reasonable interoperability.
- The primary archetypes are at the level of the 'ENTRIES' in the CDA RMIM and the *openEHR* reference model - providing an opportunity for convergent work.

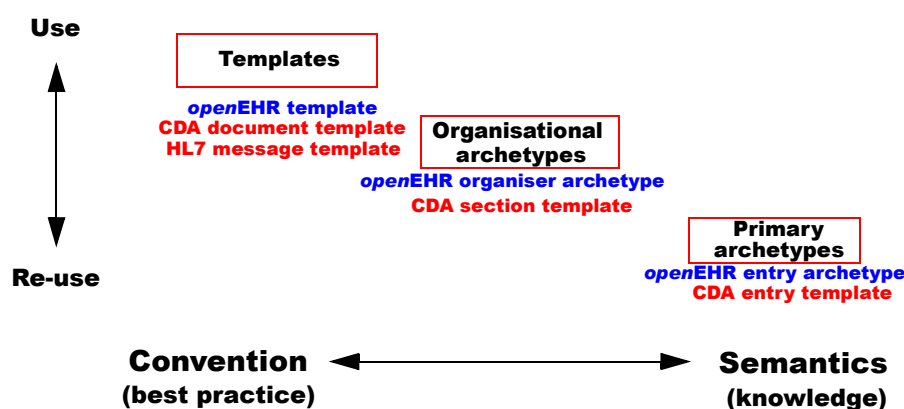
For the HL7 community, organisational archetypes will apply to CDA *SECTIONS*, and some messages. Primary archetypes will apply to CDA *ENTRIES* where the model is required for semantic interoperability and the full semantic is not provided by the

RMIM. In *openEHR*, ‘organisational’ archetypes will apply to *ORGANISERS* and ‘primary’ archetypes will apply to *openEHR ENTRIES*.

The ‘templates’ with the weakest links to the knowledge base and the strongest link to quality and practice - will apply to predominantly to HL7 CDA documents, *openEHR* transactions or compositions and to complete messages. The guide is that these are only expressed in terms of organisational and primary archetypes.

In the current proposed approach by the Templates SIG, primary archetypes would also apply to RMIM segments where there was a need for semantic differentiation.

Figure 5: The relationship of the artefacts



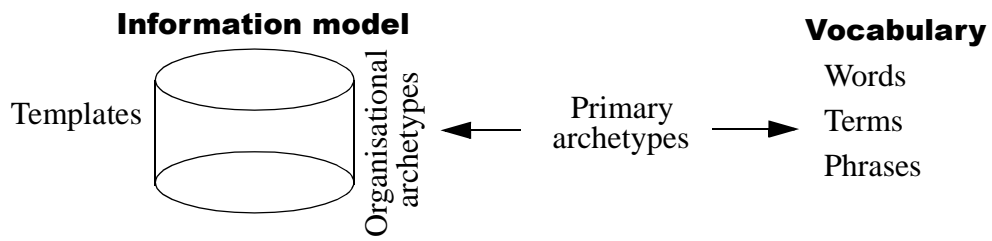
The proposal to use both the terms templates and archetypes in a specific manner, has considerable advantages for the following reasons:

- The EHR SIG and HL7 Templates group can use the same language to describe different artefacts shared by both communities.
- The essentially different artefacts will be differentiated by name, thus avoiding confusion when using the term template generically.
- The technology to express both ‘organisational’ and ‘primary’ archetypes can be the same in each setting.
- A shared ontology of key concepts – ‘primary archetypes’ - that fill the semantic gap between information models and terminology can be developed in a controlled environment.
- A shared registry of ‘organisational’ archetypes can grow in a less controlled manner in response to specific use environments.
- A shared registry of templates that allow localisation and extensibility can grow in an open and collaborative manner.

The resulting view of the information space involves 4 key components:

- The information model (RIM/RMIM or *openEHR* reference model)
- Templates which express the data entry requirements in a particular situation
- The organisational archetypes - supporting navigation and style or ‘clinical interoperability’
- The primary archetypes - supporting semantic interoperability
- The vocabulary - supporting semantic interoperability

Figure 6: The semantic space



“How and where things are stored does not reflect on what they mean - but returns from a query are usually based on the meaning...” Alan Rector

Please send comments to:

sam.heard@bigpond.com