Semantic Web and eHealth

January 2013

SEMANTIC IDENTITY

OBSERVE REASON IMAGINE
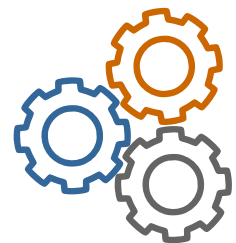
Acknowledgements

The information contained in this report represents analysis of publicly available sources believed to be reliable as of the date of this whitepaper. This document is for informational purposes only. Semantic Identity shall have no liability for errors, omissions or inadequacies in the information contained herein or for interpretations thereof, expressed or implied, in this document.

About Semantic Identity

Semantic Identity provides independent consulting, thought leadership, technical advice, research, and strategic direction for emerging ICT technologies. We focus our expertise around; Information Architecture, Semantic Web; eHealth, Social Media; Policy & Privacy; Information Modeling; Rights Management; Metadata and Identifiers and Data Analytics

http://semanticidentity.com

Table of Contents

1. Introduction

The W3C has developed a number of technologies under the Semantic Web program that aims to provide and support greater knowledge representation and manipulation. The key technologies include:

- Resource Description Framework (RDF)

- Web Ontology Language (OWL)

Both RDF and OWL provide mechanisms to model information in such as way as to enable greater semantics to be captured and inferred. The key reason behind this is that they are based on a formal model of logic that has mathematical foundations. This provides the crucial difference with existing information modelling technologies, as it provides computable guarantees that can be inferred from information models. Additionally, both RDF and OWL have well-defined syntactical encodings (eg XML) that provide the machine representation - without the need for stakeholders to develop additional technical specifications. (For an overview of the Semantic Web, see the Semantic Web Architectures Whitepaper[1].)

This white paper discusses the potential use of Semantic Web technologies as a way forward towards better eHealth information management for the Health Sector. The corollary to this outcome is the improvement in creating and managing eHealth specifications.

The Semantic Web combines a set of new technologies with grounded knowledge representation techniques to address the needs of more formal information modelling and reasoning for information services. The Semantic Web is used for many purposes; from a standardised way to markup metadata, to describing resources for the growing movement favouring the open and shared expression of common ontologies.

The benefits of the Semantic Web include:

- Consistent mechanisms to model information from simple vocabularies to complex ontologies.

- A formal modelling approach that ensures information reasoning outcomes that are computationally complete and decidable.

- Data linking opportunities aimed at supporting better user experiences, and hence, improved business outcomes.

- A groundswell of activity in the development of open-source tools to exploit Semantic Web technologies and information

- Standardised by the W3C indicating global consensus and open royalty-free specifications.

The formal aspects of the Semantic Web provide a key long-term benefit for any enterprise. Large organisation are often inefficient when it comes to data integration, sharing, and reuse and often duplicate data with unforeseen consequences. The formal modelling and reasoning pedigree of the Semantic Web reduces this impact by allowing information entities to be clearly defined, openly linked, integrated, and extended by stakeholders for greater decision support.

---

[1] http://semanticidentity.com/whitepapers/

The key messages for the Semantic Web include:

- Semantic Web technologies provide a strong base for long-term stability.

- Enterprise Architectures can be improved by adopting Semantic Web techniques in the information and data layers.

- Business improvements outcomes are likely by early adoption of an emerging technology that fit the business goals of the organisation.

The implication from this is that we are not dealing with just technical interoperability of health information, but semantic interoperability.

The core solution to support semantic interoperability is for the information to be modelled formally. There are many ways of doing this. The key is to choose a mechanism that also supports technical interoperability (i.e. over the wire representations).

2. Information Models

We will briefly review the current state eHealth Information process (based on structured information) and the target state (based on semantic information). These two modelling approaches follow fundamentally different modelling logic.

The structured approach (formally called extensional modelling logic) is based on the closed-world-assumption. This means that information classes must be extended by the authors of the model. That is, all classes and subclasses have to be explicated stated. For example, Person and Patient classes could never be subclassed together unless the author explicitly creates this relationship.

The semantic approach (formally called intensional modelling logic) is based on the open-world-assumption. This means that information classes can be inferred from the model itself. That is, all classes and subclasses do not have to be explicitly stated. For example, if the necessary and sufficient conditions to define a Person class have been created (for example, must has a name, gender, and date of birth), and another author defines a new Patient classes which has all the same necessary and sufficient conditions then a reasoner will automatically infer that Person and Patient are in the same hierarchy. This logic is very powerful in the clinical space as it supports relationships that may have escaped the attention of the modellers.

2.1. Structured eHealth Information

There are typically three steps in creating a eHealth Clinical Document specifications:

- Conceptual Clinical Models
- Logical Structures
- Implementable Architectures

A Conceptual Clinical Model is a representation of clinical concepts. It is typically represented as a structured item or components complete with data types and terminology constraints. For example, a "Medication Instruction" clinical model may include a "Medicine" and a group of "Ingredients and Form" items - two of which may be "Active Ingredients" and "Inactive Ingredients", and each of those may include "Name", "Compound", "Strength", and "Role" concepts.

A Logical Structure is a representation of a clinical document that brings together a number of Conceptual Clinical Model items with further constraints based on the domain requirements for that type of clinical document. For example, a "Shared Health Summary" logical structure may include an "Adverse Reaction" Conceptual Clinical Model and the "Medication Instruction" Conceptual Clinical Model. For the latter, the "Shared Health Summary" logical structure may declares that the "Ingredients and Form" items have been constrained from the Clinical Document structure. The constraint has to be explicitly declared otherwise the Conceptual Clinical Model structure forces their inclusion.

An Implementable Architecture is a technical guide on the implementation of a clinical document based on the items in the corresponding Logical Structures. Typically, the Implementable Architecture maps the items into and XML representation based on the HL7 Clinical Document Architecture (CDA) - incorporating the Hl7 Reference Information Model (RIM) and utilising

the HL7 Version 3 Data Types. For example, the "Shared Health Summary" Implementable Architecture specification using CDA declares that the "Medicine" item should be expressed using the <code> element (within the <manufacturedMaterial> element etc) and may use clinical terminology values (from a number of Reference Sets).

2.2. Semantic eHealth Information

The approach with Semantic eHealth Information is to focus on modeling the artefacts necessary for clinical documents. This process is a combination of the Conceptual Clinical Models and Logical Structures steps above and is a key efficiency step as the process of constraining in the Logical Structures step is not ideal and typically the reverse process of industry practice.

For example, to model the "Medication Instruction" in semantic techniques would result in separate concepts (Classes) for "Medicine" and "Ingredient" and there would not be an "aggregate class" called "Ingredients and Form" as this is a structural view and would not enable any other relationships to be asserted between the two classes. Additionally, the model of "Ingredient" would include a property indicating that this is "active" or "inactive" – not separate structures that have no relationship that they are the same concept.

A model for a Clinical Document (such as a "Shared Health Summary") could then be easily created by including the necessary concepts based on the domain requirements. In essence the Clinical Document Logical Structure is simply a collection of the appropriate clinical concepts.

The technical representation is also where efficiency gains can be obtained. This is because the semantic techniques (such as the W3C RDF/OWL semantic model) include XML serialisations as part of the model generation. This frees the developers from CDA-like issues, as the XML is completely generated from the semantic model.

2.3. Information Model Mix

Even though both models have fundamental differences, they could coexist as long as the overall information model clearly separates the two and applies them to distinct parts of the model. For example, the structured model could be applied to non-clinical information (such as information about People, Organisations, and Payments) and the semantic model to more complex clinical models (such as Prescriptions and Allergies).

However, this is not the ideal target architecture state, as the potential for improved care is lost across the models. For example, to match a person's particular clinical condition to an appropriate medicine.

3. Semantic eHealth Architectures

There are a number of differences, benefits, and risks between the two Structured and Semantic modeling approaches, outlined in the below table. One of the most fundamental differences is that semantic modelling follows the open-world assumption (OWA), and structured modelling follows the closed-world assumption (CWA).

The crucial difference is the OWA allows for new information to be added to the model (by anyone) and does not produce any conflicts. Conversely, the CWA creates a fixed model in which new information to the model cannot be added, unless the author of the original model adds it. The OWA follows more closely with the real-world model and is more flexible.

	Structured Architecture	Semantic Architecture
Description	The current architecture is based on the following series of artefacts: • Conceptual Clinical Models • Logical Structures • Implementable Architectures A Conceptual Clinical Model is created from consultation in the wider community to meet specific clinical needs (eg an "Adverse Reaction"). The Logical Structures meets a specific use case for a clinical document (eg a "Discharge Summary") and is assembled from a combination of Conceptual Clinical Models, which typically have been constrained (ie elements removed) and additional information added (such as relevant clinical terminologies). The Implementable Architectures is a mapping of the Logical Structures elements into an XML representation based on the HL7 Clinical Document Architecture (CDA).	The semantic architecture is based on modelling the required clinical artefacts, but has one significant difference in that the clinical artefacts would be modelled as a whole, not as separate models. This will enable dependencies to be established (and checked) across the entire model. The resultant model would not attempt to model the entire health domain, only the parts that are relevant to the purpose of interoperability. Since it follows the OWA, the model can grow over time, including external parties adding their own relevant models. The semantic models can then be used in specific ontologies to represent clinical documents. The links to clinical terminologies would be enhanced and exploited where needed. For example, the rules from a relevant Medicines Handbook may be incorporated. The machine encodings of the models (ie for transport between systems) would follow the predefined ontology serialisations. That is, there is no need for XML technical specifications per clinical document, only some general guidelines to support features like ordering of properties and collections of objects.

	Structured Architecture	Semantic Architecture
Benefits	• A well known process and similar outcomes • Reuse existing artefacts and tool support • No change to the industry practice	• Greater capture of semantics for the eHealth domain • OWA allows extensions by other stakeholders • Ability to support decidable reasoning (eg to fully support Medications Management) • XML encodings are automatically supported • Self documenting ontologies (specifications) • Free tool support • Keep up to pace with formally modelled terminologies (eg SNOMED CT) • Easy to reuse and extend existing semantics
Risks	• Three layer separation of Clinical Concepts/Logical Structures/ Architecture leads to potential loss of critical information • Unable to support any decidable decision support • No support for extensions and confusion over reuse of elements by stakeholders • Continued use of the complex CDA specification • Continued creation of large (ie #pages) eHealth specifications • High dependency on vendor tools • Community clinical requirements not addressed completely • Unable to exploit formally modelled terminologies • Unclear how to support the emerging HL7 Fast Health Interoperable Resources (FHIR)	• Change to industry and eHealth specification process • Increases complexity of clinical relationships • Staff expertise

4.   Scenarios

This section describes a number scenarios that outline the current issues related to Structured eHealth Information and the additional features provided by a Semantic eHealth Information.

4.1. Medication Management

The provision of Medication Management is a critical area in eHealth that is prone to errors. For example;

- prescribing a medicine to a consumer that has an allergy to one of the medicine's ingredients

- a pregnant consumer being prescribed an incorrect medicine (for their condition)

- a consumer being prescribed a medicine and dispensed with a brand-substituted alternative (that they were not aware of), that conflicts with their allergy

Let us consider the first example above. Figure 1 shows the current model used by the Structured eHealth Information approach.

FIGURE 1 - STRUCTURED EXAMPLE INSTANCE

In this example, the Consumer has two Clinical Documents (Shared-Health Summary SHS and an ePrescription) and highlights two clinical data instances that indicates that the Consumer has an Allergy (to Penicillin) and has a Prescription (to Moxacin).

With the current Structured approach, these two pieces of data are independent and isolated from each other. The structured mechanisms define these two data elements solely for the purposes of their relative Clinical Document use cases.

As we will see in the following section, the above example is not ideal for the Consumer. Since these two data elements were defined independently, there has been no consideration on the semantic relationship between Allergy and Medicine across the two clinical documents.

Typically, the Allergy from the SHS would be expressed from a clinical terminology (eg SNOMED CT) as well as the Medicine from the ePrescription (eg from the Australian Medicines Terminology AMT). SNOMED and AMT are clinical terminologies and capture and represent relationships between the concepts. In the case of SNOMED, there is an OWL ontology representation, but not for AMT.

Now lets consider the next opportunity, by assuming that we have well-defined ontologies for both SNOMED and AMT that are semantically cross-linked (which we don't really have yet). Figure 2 below shows the potential of exposing the relationship between Penicillin and Moxacin through expanding the relevant ontologies.
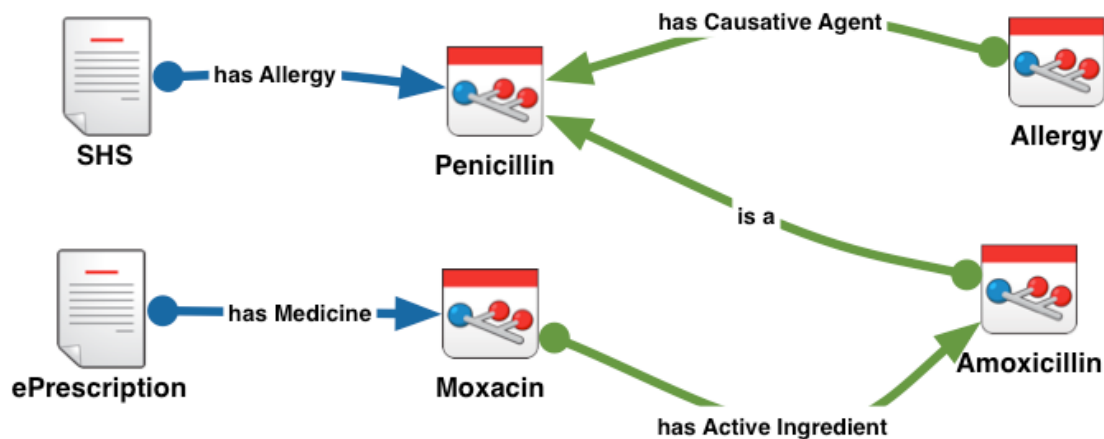
Using AMT, we can determine that Moxacin has an active ingredient of Amoxicillin. We can then use SNOMED to trace the provenance of this substance and we can infer that Amoxicillin is a type of Penicillin. Also using SNOMED, we can determine that Penicillin is a cause of Allergies.

Going through this process, we can now determine that the Consumer should not be prescribed Moxacin because of their allergy.

Now it is important to note that the current Structured approach does not require the above scenario to be evaluated (that is, to check any conflict between the Penicillin and Medicine in the separate clinical documents). In fact, there are no obvious links from the Structured mechanisms to even consider this (however, some advanced clinical decision support systems may do this).

The worse case scenario is that an ehealth developer would have to manually consider each and every element of each and every clinical document for such potential discrepancies. This is a un-scaleable solution.

Lets review our discussion and findings so far:

- Clinical Documents impose a barrier to clinical decision making as they their contents are not cross-mapped (eg for dependencies) to support a atomic-data approach (rather than a document-centric approach); and

- Clinical Terminologies must be fully expressed as complete and comprehensive ontologies with appropriate cross-linking of concepts across terminologies (eg AMT medicines linked to substances in SNOMED-CT).

The implications of these findings is that Clinical Data should be the focus of eHealth Information, and, if this needs to be represented as a Clinical Document, then that is possible (and indeed necessary in many cases) but the core clinical entities are modelled as semantically interoperable clinical data.

Figure 3 below shows an example of modelling clinical data to ensure that correct relationships can be expressed and validated.
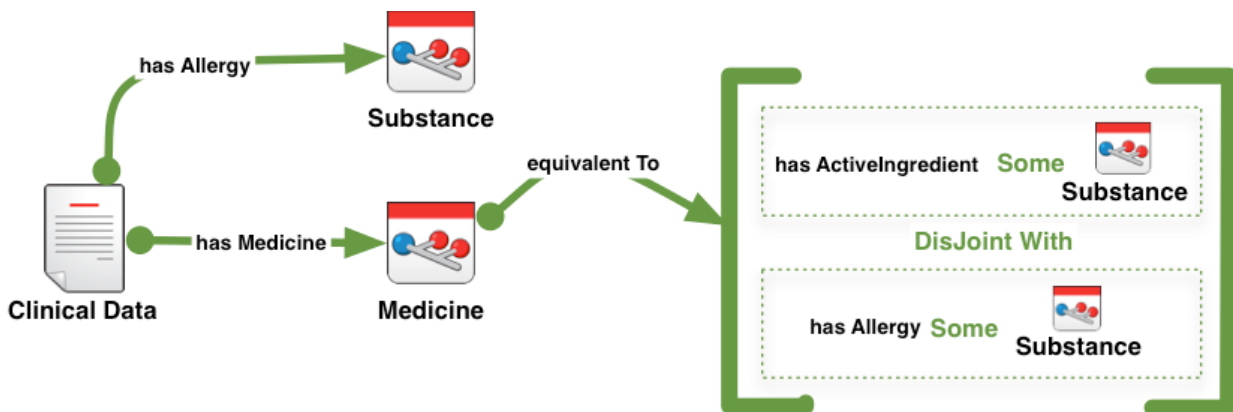
The key to this model is the equivalence constraint applied to the Medicine class. What this effectively does is define the rule that must be applied to any instance of a Medicine. This rule (in the big brackets) defines a class of possible instances that can only be applied to Medicine individuals (instances).

The detail in the rule, in essence, says that the Class Medicine is equivalent to any individual instance where there is no overlap between the Substance having an active ingredient and allergy (at the same time for the consumer).

The implications of modelling clinical data following a semantic mechanism is that we can define explicit rules that match clinical expectations at the core clinical data level, and these will be evaluated correctly when utilised in different contexts (eg SHS versus ePrescription).

4.2. Information Mapping and Reuse

As noted earlier, ehealth creates voluminous amounts of information. This creates a significant problem with creating clinical models to ensure that information is not repeated and that maximal reuse is achieved. This is exacerbated over time when older (but still relevant) information models are ignored or overlooked in newer information model designs.

In the Structured eHealth Information approach, when an information element of a detailed clinical model is created, it is given a local identifier as well as a unique identifier (eg an Object Identifier OID) together with the definition and usage metadata. Because of the volume of clinical models, there are some instances of where that new element definition may already exist. An example is the "Medication Action Instruction" (from the Medication Instruction Clinical Concepts) that has the exact same semantics as the "Label Instruction" (from the Dispense Record Logical Structure).

In the structured approach, there is no mechanism to indicate that these elements represent the same concept. Even the two different OIDs allocated to them are not able to be linked.

In the Semantic eHealth Information approach, because it is based on the open-world-assumption, it is relatively straight forward to assert the similarity of concepts. The top of Figure 4 shows this technique using the semantic "same as" property relationship. With this relationship added, semantic clinical systems will treat the two properties as interchangeable (with the same semantics). It is important to note that this new relationship could be added by a different modeller and it is not necessary for the original modeller to even be aware of this, as the software reasoners will mange it.
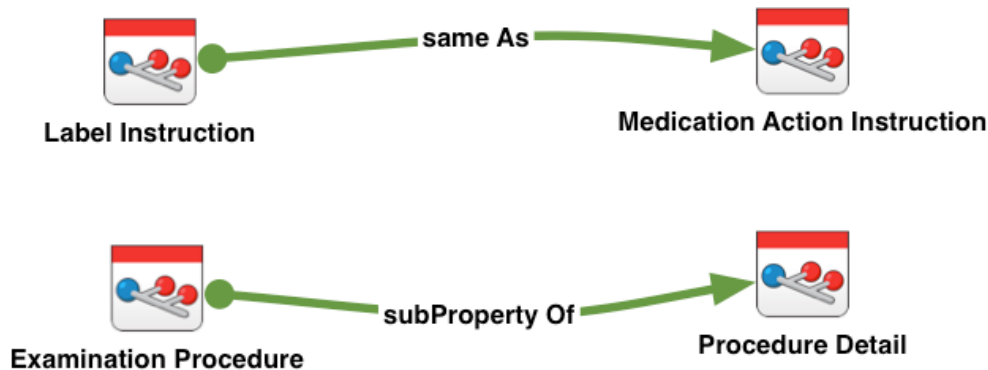

FIGURE 4 - MAPPING EXAMPLE

Another example is when two clinical elements are similar and form a hierarchy (similar to a class hierarchy). From the "Procedure" Clinical Models there is an element called "Procedure Detail" which is defined as "further information about the procedure". In the "Imaging Examination Result" Clinical Models there is an element called "Examination Procedure" which is defined as "additional details of imaging examination methodology followed".

What is clear from these two elements is that Examination Procedure is a more specialised version of Procedure Detail. Again, in the structured approach, there is no mechanism to indicate that these elements represent similar concepts.

In the semantic approach, it is relatively straight forward to assert the similarity of these concepts. The bottom of Figure 4 shows this technique using the semantic "subproperty of" relationship.

4.3. Types of Participants

The "Participation" Clinical Models includes a range of elements that comprehensively describes either a Person, Organisation, or Device. There are a common set of elements for all three types of participants, and a few specific to each. All of these elements are aggregated into one large structure.

When the Participation Clinical Model is used for the Subject of Care in a clinical document scenario, for example a Discharge Summary, the rules for use can only be expressed as human text (eg "..SHALL be instantiated as a PERSON").

With a semantic approach, the Participation entity would be modelled to reflect more of the real-world, and hence, when a Person is required in a clinical document scenario, then that can be explicitly asserted (as shown in Figure 5). That is, the Person class can be used directly as the Subject Of Care in the Discharge Summary, and the relevant semantic relationships to Participation are automatically included.
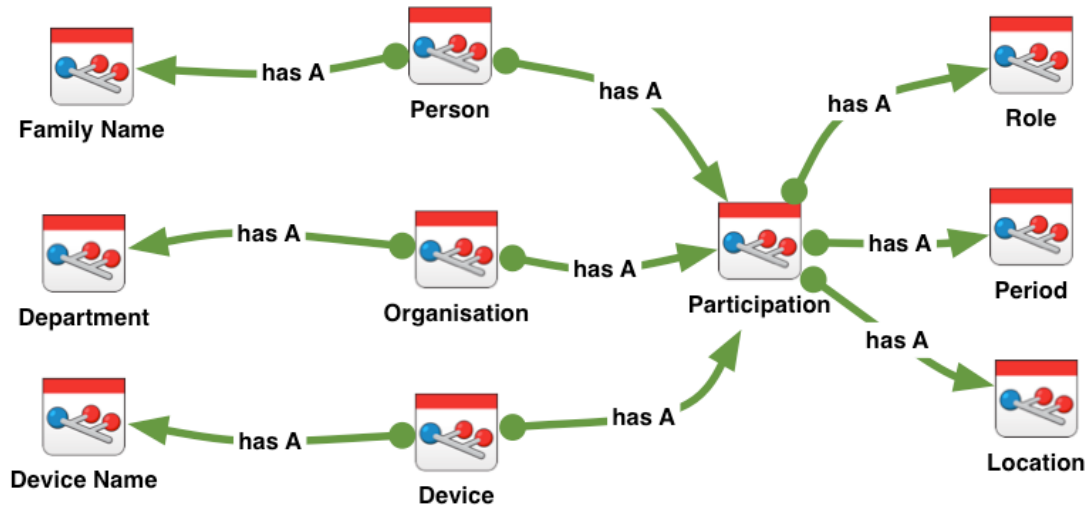
FIGURE 5 - SEMANTIC PARTICIPATION MODEL

Additionally, with the aggregated Participant Clinical Model structure, the definitions of some of the elements end up as variable, and this is far from ideal from a metadata definitions management aspect (eg ISO 11179). For example, the "Qualifications" element has two different definitions, depending on if this is used in the context of a Person/Organisation or as a Device.

Figure 6 below shows the semantic approach to this problem, in which there are two different Qualification properties (with the correct single definitions for their context). Additionally, we can use other vocabularies (eg the W3C SKOS ontology) to assert the relationship between the two Qualification properties.
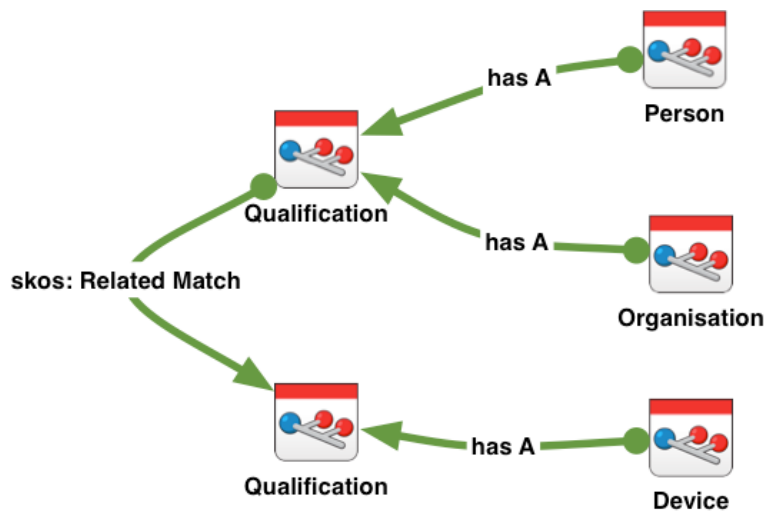

FIGURE 6 - SEMANTIC DEFINITIONS

5. Strategy

The semantic web will become a significant technology in the future of the ehealth sector. Ehealth stakeholders need to develop a strategy with semantic information modelling to enable itself to be positioned to exploit this opportunity. It is clear that a phased approach is required for this future deployment and it will require buy in from the sector and stakeholders.

5.1. Target Goal State

The target goal state for the ehealth sector would include:

- Ehealth ontologies for foundational services (eg data types and "clinical document" concepts)
- Ehealth ontologies for core clinical concepts (eg Prescriptions, Adverse Reaction)
- Technical Specification for clinical document use cases (utilising the above ontologies)
- Access to ontology reasoners for clinical validation (ie similar concept to xml schema validation)

The outcome would be improved Decision Support functionality for clinical systems and national infrastructure.

5.2. Industry Migration

The impact on Vendors and Jurisdictions and Healthcare Providers would need to be managed as a long-term issue. This would include:

- Early outreach on the benefits of semantic technologies
- A "Change and Adoption" program
- Prototype software demonstrating the benefits
- Technical migration plan
- Healthcare providers guidance on decision support advances

The technical migration plan will obviously be a critical part of the successful adoption by vendors of ehealth software. To move from the current state to the target state would be difficult as a single change-over. A more realistic approach would be a two-step process:

- Migrate the sector to an improved structured approach (for example HL7 FHIR); and then
- Migrate the sector to full semantic approach.

The key to this approach is that step 1 (eg HL7 FHIR) must use semantic information models for its clinical artefacts, but still represent this in XML (which is less of a change for the vendors.)

5.3. National Infrastructure

The potential impact on national infrastructure services would be significant and would provide some of the motivations to migrate to semantic models. A National Personal eHealth Record System would be one of the key beneficiaries of semantic models. For example, since it collects clinical documents about individuals, it is well placed to perform semantic reasoning on the data it collects. (See the Medication Management scenario as an example.)

The outcome is that national infrastructure services, like ePrescription and Personal eHealth Record services, can become more than just static document repositories, but enable additional clinical decision support services (such as the detection of abnormal events) and notify healthcare providers.

5.4. Standards Impact

A key and critical aspect of the semantic approach is adoption by the relevant eHealth standards development organisations. Clearly, HL7 is the primary option and an ideal opportunity to move towards greater semantic web adoption for eHealth specifications.

The W3C currently has a Semantic Web for Health Care and Life Sciences Interest Group, in which a few key individuals are also active at HL7. The W3C HCLS IG is keenly aware of semantic technologies and their impact in various sectors, and can join the HL7 opportunity and act as a source of expertise.

As mentioned above, the timing now is critical for HL7 as it forges ahead (relatively quickly) with FHIR. If FHIR can meet some of its stated objectives:

> "The resource contents are mapped to solid underlying ontologies and models using computable languages (including RDF) so that the definitions and contents of the resources can be leveraged by computational analysis and conversion processes"

The ehealth industry would be clearly in favour of FHIR (over CDA) as it provides a cleaner XML representation. However, as it stands now, it is yet another structured XML specification.

There is also an opportunity with the IHTSDO and the SNOMED clinical terminology. As there are OWL representations of SNOMED, IHTSDO should promote this over its (structured) RF2 representation in the future, and provide guidance on the benefits from the OWL model for decision support. Additionally, Australia should consider migrating AMT to be a complete SNOMED reference set, and thus also having an OWL representation.

National standards bodies would initially have a small role, but eventually after the semantic specifications are developed, will migrate these to National Technical Standards.

The eHealth sector now faces a prestigious opportunity to provide significant influence and directions to key standards organisations in the future development of semantically-based clinical specifications.

5.5. eHealth Specification Development

The migration to semantic technologies will have a large (but positive) impact on eHealth specification development. Some typical requirements include:

- Review of the Conceptual Clinical Models to support more focussed clinical concepts (that do not require to be constrained but simply reused)

- Review of the Conceptual Clinical Models modelling techniques to support the semantic approach (ie to closely follow the open-world-assumption)

- Updated support in eHealth Modelling tools to support the new models and to export RDF/OWL output

- Creation of new Logical Structures for clinical document use cases

- The Logical Structures specifications would represent the complete information to create the relevant clinical document

- There is no longer any requirement to create Implementable Architecture specifications as the Logical Structures specification (being based on RDF/OWL) will automatically provide the over-the-wire encoding (ie encoding in XML but following the RDF model).

The last point is significant as the current CDA specifications are a constant source of confusions for vendors and users. By supporting semantic models in eHealth clinical specifications, the "XML comes free and unambiguously". The other key advantage is that resources can focus in the semantics of clinical models more (and hence improved outcomes) and a lot less (ie close to none) on the over-the-wire encodings.

6.   Summary and Conclusion

The health sector creates and manages significant amounts of information about people used to both record patient encounters and decisions about healthcare treatments. Consequently, it is a challenge to determine the subset of this information that should be standardised to enable sharing of this information with can lead to consistent outcomes in terms of technical interoperability and clinical decisions.

To date, the majority of health information has been aimed at sharing at the document-level (ie for human readability) and only a small amount at the atomic level (ie for machine processing). The industry clearly is demanding more information standardisation at the atomic level as this is where decision-support systems can aid the healthcare provider with clinical recommendations.

Without significant levels of standardisation, the risk is that vendor's software may interpret atomic data differently resulting in inconsistent outcomes and possibly triggering clinical safety issues. The challenge is to understand the range of information technology mechanisms that can reliably support the decision-support requirements of the ehealth sector as well as the implications to the existing mechanisms used to provide this information.

This position paper has provided an overview of the ehealth information challenge. More specifically, it introduced the Semantic Web information mechanism as a candidate technology that could potentially fulfil this vital role.

A number of scenarios were used to highlight the benefits of the Semantic Web and their specific clinical advantages over current technologies. The migration to the proposed solution will need to be addressed through a range of short, medium, and long term strategies.

The advantages to the ehealth sector in adopting Semantic Web technologies will provide a long-term gain for improved health decisions. The technologies are based on proven formal information models that provides decidable (ie guaranteed) outcomes.

Other advantages in adopting Semantic Web technologies for it's clinical information standards will include rapid development of specifications (as information reuse is high) and a more comprehensive and extensible end-to-end model of the healthcare community.