Managing Medical Ontologies using OWL and an e-business Registry / Repository

Keywords: RDF, OWL, Ontologies, e-Health, ebXML Registry Repository, BCM Business Centric Methodology, Templates, Archetypes, Semantic Web, HL7, EHR

Carl Mattocks Esq

CEO / Founder
CHECKMi (cross checked knowledge mediation and interoperation)
Berkeley Heights, New Jersey, USA

Carl Mattocks is Founder / CEO of CHECKMi a New Jersey-based company developing agent –based software capable of linking a 1000 databases into a single compendium. Commencing with the ISO IRDS standard for Repositories - Carl is a long time contributor to the development and exploitation of metadata-centric specifications. Currently he acts as a co-chair for the Business Centric Methodology TC and the ebXML Registry Semantic Content SC. Much of Carl's wisdom was gained managing a London–based Artificial Intelligence company that pioneered the use of faceted-classification techniques and multi-national language processing for online information discovery. He has also benefited from being - one of the first Certified Systems Analysts in the UK; one of the first students in a condensed-MBA program at Cranfield School of Management; and one of the first designers to apply (ITIL) Structured System techniques on (Government / Financial Service) intelligence projects.

Abstract

This paper outlines how the knowledge embedded in OWL and medical ontologies support e-Health web service deployment that may involve the interaction of independent resources from different domains and owned by different entities. In particular, it (1) identifies how the OASIS Business Centric Methodology guidelines apply to the development and publishing challenges for medical information that is required by agents representing the e-Health stakeholders and (2) how the enterprise content management capabilities of the ISO/TS 15000 ebXML Registry / Repository are being used to manage standards based (e.g., HL7 CDA, GEHR) medical records and assist the discovery process.

1 Introduction

"*e-Health refers to the use of modern information and communication technologies to meet needs of citizens, patients, healthcare professionals, healthcare providers, as well as policy makers*." - Declaration made during the European Commission 2003 e-Health Ministerial Conference

The Semantic Web is about the autonomous discovery and assembly of distributed remote resources. e-Health services depend on the automated sharing of metadata across web applications that provide a common approach for the discovery, understanding, and exchange of semantically rich Electronic Health Record (EHR) information. This paper outlines how electronic medical records information embedded in an e-business Registry & Repository support e-health service goals that involve the interaction of independent resources from different domains and owned by different entities. In particular, it identifies Business Centric Methodology (BCM) guidelines for the development and publishing of the semantic content required by agents representing the Health Management stakeholders that use OWL, ontologies, templates and medical information archetypes to both assist the discovery process and pull health care services together.

2 e-Health Service

2.1 Service Goals

Across the world many governments are announcing e-Health service initiatives. For example, in July 2004 the USA Health & Human Services (HHS) Secretary outlined a 10-year plan to transform the delivery of health care by using electronic health records and accelerate regulations for e-prescribing drugs. Within the OASIS open source specifications body there are a number of Technical Committee (TC) groups actively contributing to the evolution of e-Health service oriented standards. Specifically, this paper references the work of TCs for (i) the Business-Centric Methodology (BCM), and (ii) the ebXMLRegistry to help explain how the application of standards based technology support the e-Health goals of:

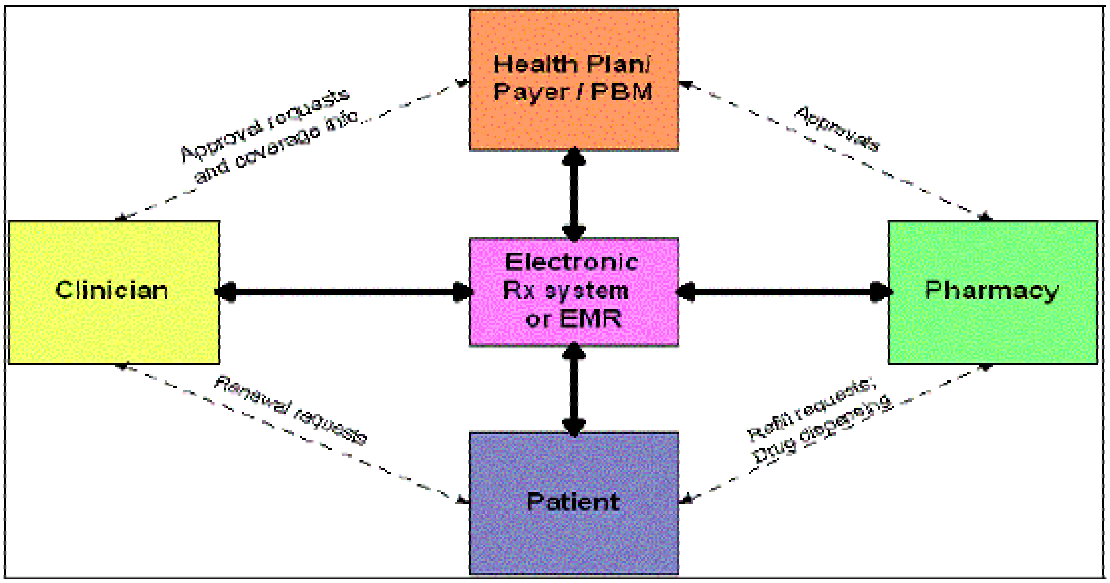
- Bringing information tools to the point of care by the EHR (Electronic Health Record) systems used in physician offices and hospitals *aka* "Inform Clinical Practice"
- Building an interoperable health information infrastructure, so that records follow the patient and clinicians have access to critical health care information when treatment decisions are being made *aka* "Interconnect Clinicians"

- Using health information technology to give consumers more access and involvement in health decisions *aka* "Personalize Care"
- Expanding capacity for public health monitoring, quality of care measurement, and bringing research advances more quickly into medical practice *aka* "Improve Population Health"

2.2 E-Prescribing

In the USA there are more than 3 billion prescriptions written annually and the national savings from universal adoption could be as high as $27 billion. Savings that derive from providing instant electronic connectivity between the practice, the pharmacy, health plans / PBM's (pharmacy benefit management), and other agencies, see Figure 1. Specifically, a key part of the USA HHS plan is that, internet technology will allow beneficiaries to access their personal health care information and that electronic prescribing will improve quality, efficiency, and reduce cost by the facilitation of:

- Actively promoting appropriate drug usage, e.g., following a medication regimen for lowering blood cholesterol
- Providing information about formulary-based drug coverage, including on-formulary alternatives and co-pay information.
- Speeding up the process of renewing medications
- Improving the speed and accuracy of prescription dispensing, pharmacy callbacks, renewal requests, eligibility checks, medication history, and more



Simple electronic prescribing systems include electronic communications between the electronic prescribing system and the clinician and pharmacy (fully

electronic or via fax). Alternative communications are still required with the health plan and the patient, and usually occur by phone (dashed lines). In a fully integrated system, all of these communications can be done electronically.
(source Electronic Prescribing: Toward Maximum Value and Rapid Adoption - A Report of the *Electronic Prescribing Initiative* eHealth Initiative 3/2004)

Figure 1

2.3 Care Requirements

Internationally there are a number of, professional Societies (e.g. American Academy of Pediatrics, American Medical Informatics Association) and multi-partner projects, (e.g. Mayo Foundation, ARTEMIS) assessing whether the use of electronic forms and other web service technology has the potential to transform the delivery of health care. A summary of their evolving e-Health service requirements considers both Personalized Care & Quality Care needs:

- **Timeliness** always current medical information as defined by 'whenever the patient and health professionals need it '
- **Availability** always available records reduce errors originating with partial notes captured with poor handwriting
- **Decision Support** knowledge regarding treatment options in health care must be widely published
- **Cost Effective** savings achieved through health information technology may be achieved by reducing duplicative care, lowering health care administration costs, and avoiding errors in care
- **Bench-to-Bedside** bio-informatics that accelerate biomedical research, and speed the application of findings into bedside practice thus reduce the current (estimated 17 years) time it takes for the knowledge to be fully integrated into general medical practice
- **Medically Underserved** improved access to specialty information can be especially beneficial for medically underserved areas, including inner-city and remote rural areas
- **Consumer Involvement** compliance with USA HIPAA regulations and the use of e-business security technology will allow secure PHR (Personal Health Records) to be maintained by the patient and his or her physician, insurer or others, giving the patient unprecedented access and control of the record

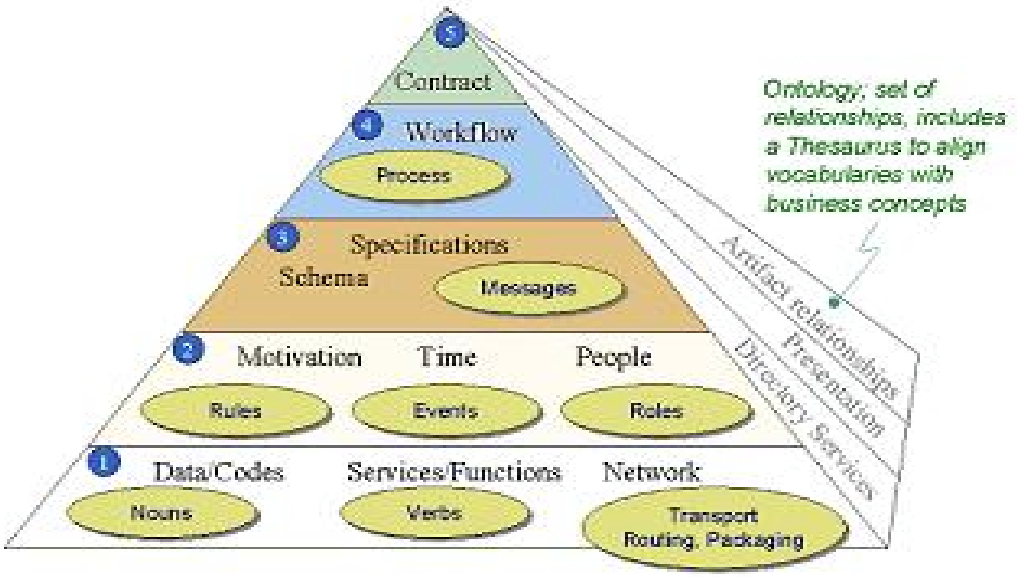
- **HIPAA Note** Under the USA federal health information privacy law patients have the power to authorize certain non-routine types of uses and disclosures made of their identifiable records

- **Bio-Threat Monitoring** Links between information networks employed by the USA-based CDC, USDA & EPA networks would enable public health agencies to better monitor disease outbreaks and act quickly in response to possible bio-threats.

3 Business Centric Guidelines

3.1 Forms & Vocabulary

An eHealth Initiative report that identified the benefits of electronic connectivity also stated that a number of enhancements in semantics are needed to improve quality, efficiency, and to facilitate interoperability between the various electronic systems especially those involved in the electronic prescribing process. In particular it noted that, unifying state (electronic) prescription-form standards, establishing a consistent "doctor-level" drug vocabulary, and standardizing formulary information are among the highest needs. As a guideline to web-based interoperability the OASIS BCM (Business-Centric Methodology) states that a proper interpretation of the business language semantics found in an e-Health SOA (Service Oriented Architecture) is essential for harnessing tacit knowledge and facilitating shared communications. Particularly, the BCM identifies that a Conceptual Layer must enable the exploitation of community-of-interest specific ontologies (as in, EHR vocabularies and code lists supporting electronic form / message validation) is a key factor in semantic interoperability. Further, the ontology capability must be rich enough to resolve all semantic (meaning & operability) conflicts over terminology used to populate the many building blocks of the Lubash Pyramid, see Figure 2.



Ontology capability spans the building blocks of the BCM Lubash Pyramid

Figure 2

While not defining a mandatory structure, BCM states that the Conceptual Layer consists of semantic relationships and controlled vocabularies that increase the meaning of schema (e.g. electronic form) metadata and provide contextual validation to items that have metadata properties (e.g. archetypes and templates). The simplest form of this being a data dictionary that contains metadata about data elements and their relationship between simple / complex data types and their valid value ranges *aka* archetypes. BCM expects that when recorded in a registry the Conceptual Layer has the role of:

- Providing trace-ability from business vision to system implementation
- Ensuring alignment of business concepts with automated procedures
- Facilitating faster information utilization between business parties
- Enabling accurate information discovery and synchronization
- Expanding the ability to integrate information by interest, perspective or requirement.

3.2 Federated Content

The BCM also identifies that a registry combined with a repository is a key factor in the management of service-oriented components. Such as, metadata registered about electronic form / electronic messaging schemas, schema archetypes (the allowable values for the data type), associations between elements and any stored artifacts. Wherein, a registry not only acts as an interface to a repository of stored content, it formalizes how

information is to be registered and shared. Since, in e-Health Services this sharing goes beyond a single enterprise or agency, this dictates that the registry catalog must be capable of supporting metadata used for federated content management. Noting that, a federated content management capability is required when there is as a need for securely managing and accessing metadata across physical boundaries. Irrespective of the boundary type (e.g. community-of-interest, system, department, or enterprise separation), federated content management enables information users to seamlessly access, share and perform analysis on information. For e-Health services this may include:

- Mappings for the critical path of information flowing across a business value chain e.g. ordering & payment of e-prescription
- Quality indicators such as statements of information integrity, authentication and certification e.g. electronic signature used for e-prescribing
- Policies supporting security and privacy requirements e.g. compliance with HIPAA regulations

4 E-Business Registry

4.1 ebXML Registry

ebXML (e-business XML) is a suite of [ISO/TS 15000] standards from the e-business community addressing the entire Service Oriented Architecture lifecycle of business to business and business to consumer activities, such as, e-Health Services. In particular, when following BCM (Business-Centric Methodology) guidelines, SOA / ebXML service description and discovery extends beyond classification and resource location towards the support of web service activities, such as, e-prescribing. Specifically, when a service is profiled within a federated ebXML registry, the meta-information identifies both the SOA business conditions and the policy rules of access per level of privilege. For instance, a successful e-prescription electronic communications that passes between USA based clinician, pharmacy service and payment provider will need to employ some form of contractual electronic exchange referencing drug labeling and drug listing information maintained by the Food and Drug Administration (FDA) and the National Library of Medicine (NLM)

Fortunately, to manage this knowledge, e-Health services can exploit the OASIS ebXML Business Process (ebXML BP) standard that defines metadata describing the service capabilities needed to support electronic business collaborations with default error state resolution. For example, an “e-contract” employing terms of workflow can be structured using Collaboration Protocol Profiles (CPPs) and Collaboration Protocol Agreements (CPAs). ebXML BP also defines and describes activities during service enactment for the transactions (including point-in-time parameters) that have to be fulfilled according to the services’ usage terms.

4.2 Registry & Repository

A foundation of the [ISO/TS 15000] suite is the EbXML Registry which specifies the components for web based registry and repository services. In terms of publishing content the ebXML Registry / Repository specification supports:

- publishing to a central registry / repository; or
- publishing to a federation of many individually many registry / repository faculties.

Its federation ability has won it many supporters including the UN/CEFACT Information Content Management Group (ICG) which has officially adopted the ebXML Registry standard and the open source code freebXML Registry as the center piece of their new Information Content Management Architecture. Version 3 of the ebXML Registry / Repository supports the following types of cooperating registry services:

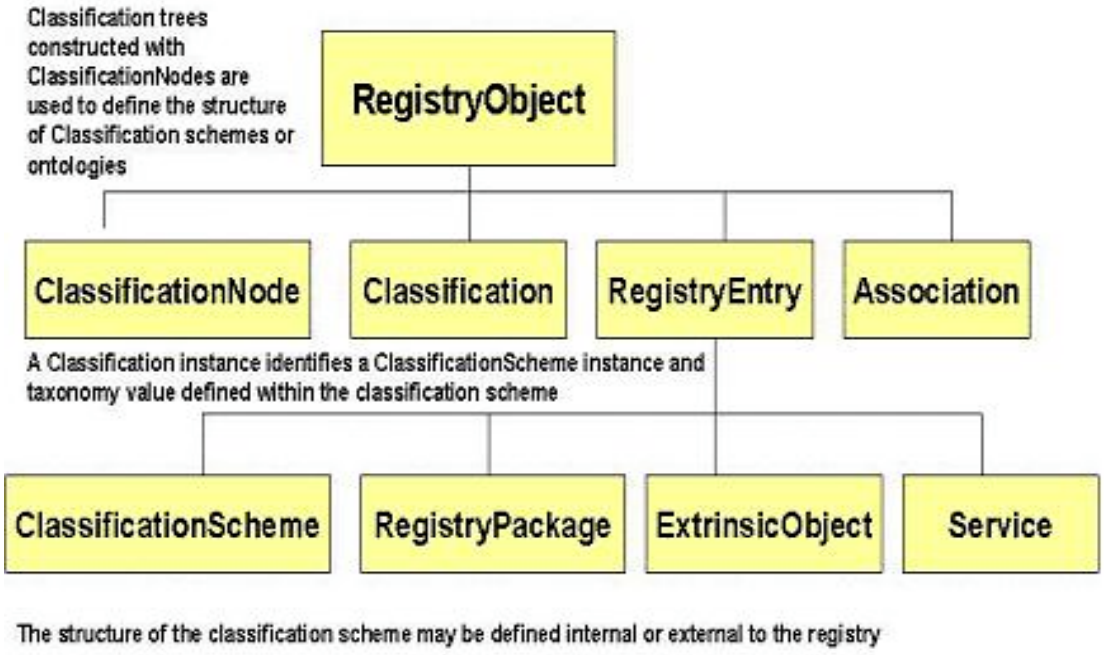
- Registration and classification of any type of object
- Objects defined by data type
- Namespaces defined for certain types of content
- Messages defined as XML Schemas
- Taxonomy hosting, browsing and validation
- Association between any two objects
- Registry packages to group any objects
- Links to external content
- Built-in security
- Event notification
- Event-archiving – enabling the production of a complete audit trail
- Service registration and discovery
- Life cycle management of objects
- Flexible query options
- Federation for inter-registry relocation, replication, references - federation metadata is stored in one registry; a registry may cooperate with multiple federations for the purpose of federated queries, but not lifecycle mgmt.

4.3 Access to Artefacts and Artifacts

Acknowledging that there are therefore two basic models of distributed information - a central repository of shared items (with individual entitiess uploading and downloading as required) or a fully distributed model (with the repository distributed over multiple facilities) the ebXML Registry specification supports a single access to many federated Registry / Repository facilities. Thus, it allows:

- logical duplication of remote federated repository items into a local federated repository to fit into local policies of information management; or
- aggregation of artifacts in the remote federated repository for creating locally defined components; or
- access to any and all federated repository items as required.

Equally, to ease discovery and deployment of the collective artifacts (*aka UK* artefacts) the ebXML Registry RIM (Repository Information Model) explicitly supports many Classification Schemes. For instance, ebXML Registry enables one or more of the objects defining an e-Health Web Service content (a "Service" class RegistryEntry) to be classified using Taxonomy Values (ClassificationNode within a ClassificationScheme). Additionally, a WSDL (Web Services Description Language) description of the service instance (i.e. as a technical specification file) may be stored as ExtrinsicObjects. Wherein, the relationship between the description files and the "Service" class is established through the "ServiceBinding" Class of ebXML. See Figure 3



ebXML Registry Classification subset of Reference Information Model

Figure 3

WSDL Note: WSDL is an XML format for describing network services as a set of endpoints operating on messages containing either document-oriented or procedure-oriented information. The operations and messages are described abstractly, and then bound to a concrete network protocol and message format to define an endpoint. Related concrete endpoints are combined into abstract endpoints (services). When managing WSDL and XSD Schemas the classification scheme provides for a number of uses:

- Find a single element from among many
- Convey semantic content that may be incompletely specified by other attributes - such as names and definitions
- Derive names from a controlled vocabulary
- Disambiguate between data elements of varying contextual classifications

5 Semantic Web

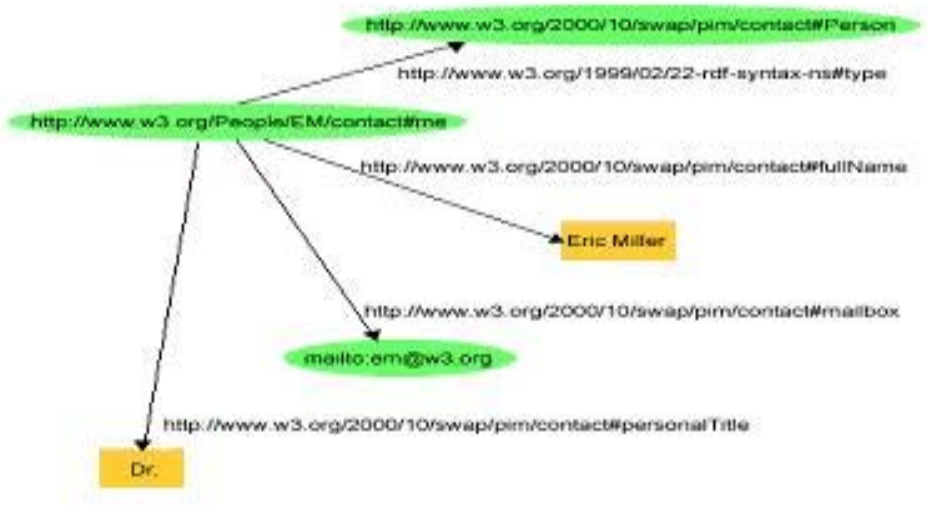
5.1 WC3 RDF

The World Wide Web was originally built for human consumption, and although everything on it is *machine-readable*, this data is not *machine-understandable*. The Resource Description Framework (RDF) is a WC3 standard based on the idea of identifying things using Web identifiers (*URIs*), and describing resources in terms of simple properties and property values. This enables RDF to represent simple statements about resources as a *graph* of nodes and arcs representing the resources, and their properties and values. For example, the group of statements "there is a Person identified by http://www.w3.org/People/EM/contact#me, whose name is Eric Miller, whose email address is em@w3.org, and whose title is Dr." could be represented as the RDF graph below Figure 4. Adding a layer of semantic information to that defined in XSD schemas RDF also provides an XML-based syntax (called) for recording and exchanging these graphs.

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
 <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me">
   <contact:fullName>Eric Miller</contact:fullName>
   <contact:mailbox rdf:resource="mailto:em@w3.org"/>
```

```
   <contact:personalTitle>Dr.</contact:personalTitle>
  </contact:Person>

</rdf:RDF>
```



RDF Graph for DR. Eric Miller

Figure 4

The February 2004 version of the *RDF/XML* syntax specification states that in order to encode the graph in XML, the nodes and predicates have to be represented in XML terms — element names, attribute names, element contents and attribute values. RDF/XML uses XML QNames (qualified names) as defined in Namespaces in XML [XML-NS] to represent RDF URI references. In qualified Names the Prefix provides the namespace prefix part of the qualified name, and must be associated with a namespace URI reference in a namespace declaration. [Definition:] The LocalPart provides the **local part** of the qualified name.

5.2 Web Ontology Language (OWL)

RDF Schema is an extension to RDF that provides the framework to describe application-specific classes and properties. Classes in RDF Schema allow resources to be defined as instances of classes, and subclasses of classes. OWL the Web Ontology Language provides an additional layer of semantics that exploits RDFS declarations. It connects the object domain identified as members of classes described within RDFS or OWL with the

datatype domain consists of the values that belong to XML Schema datatypes. For instance, OWL describes the structure of a domain in terms of classes and properties. Wherein, the classes can be names (URIs) or expressions such as,

- owl:intersectionOf, owl:unionOf, owl:complementOf,
- owl:oneOf, owl:allValuesFrom, owl:someValuesFrom, owl:hasValue.

Plus OWL provides support for axioms enables assertions of subsumption or equivalence with respect to RDF / OWL classes or properties as in:

- rdfs:subClassOf, rdfs:subPropertyOf
- owl:sameClassAs, owl:samePropertyAs,
- owl:disjointWith, owl:sameIndividualAs,
- owl:differentIndividualFrom, owl:inverseOf,
- owl:transitiveProperty, owl:functionalProperty,
- owl:inverseFunctionalProperty.

OWL has three increasingly-expressive sublanguages: OWL Lite, OWL DL, and OWL Full. The OWL-based Web Service Ontology (OWL-S) Version 1.0 "features a number of refinements to the Service Profile and Process Model. A method of registering (in UDDI & EbXMLRegistry) and linking descriptors of Web service, WSDL and OWL-S has been outlined by researchers in Turkey. A key factor is the Service Profile which is used to concisely represent the service in terms of capabilities, provenance, and operational parameters (*e.g.* cost-of-use, quality-of-service parameters, etc), for constructing both advertisements and requests, such as:

- contact information that refers to the entity that provides the service
- functional description of the service is expressed in terms of the transformation produced by the service
  - inputs required by the service and the outputs generated
  - the preconditions required by the service and the expected effects that result from the execution of the service
- properties that are used to describe features of the service
- category of the service within a classification system
  - quality rating of the service
  - service parameters that can contain any type of information e.g. geographic availability

6 Medical Documents

6.1 Health Standards Vocabularies

RDFS & OWL semantics are domain dependent; thus to be useful domain knowledge is required. E-Health service is one of the few domains to have extensive domain knowledge exposed through standards for describing electronic healthcare records. For example, Health Level 7 (HL7) is health information standards community that has its V2 messaging standards are in wide use around the world mainly for application-level messaging and clinical document management.

The HL7 Version 3 standard is based upon a Reference Information Model (RIM) that abstractly describes messages medical events and transactions. Message content schemas are derived by a restriction process which starts from the Reference Information Model (RIM), and continues through domain information models (DIMs), restricted message information models (RMIMs), common message element types (CMETs), finally ending with hierarchical message definitions (HMDs) and generated message schemas in XML. For instance, an e-prescription would reference multiple DMIM's (Domain Message Information Model) such as Orders and Observations, Pharmacy, Medications, Patient Administration (for patient and clinician identifiers) and diagnostic indications. A feature of the Version 3 methodology is the specification of vocabularies or "value sets" that convey the payload of a specific message e.g. an e-prescription message defining the ordered drug, form, dose, route, and patient instructions.

6.2 CEN ENV 13606-2 / GEHR/ openEHR / Clinical Document Architecture (CDA)

6.2.1 CEN ENV 13606-2

Beyond HL7, e-Health services can elect to leverage the CEN ENV 13606-2 information models which provide a means to represent the original organizational structure of one or more nested electronic healthcare record entries. This standard proposes sub-categorization of the EHR into four specializations:

- ***Folder****:* High-level subdivisions of the entire EHR for a patient, usually grouping entries over long time-spans within one organization or department, or for a particular health problem.

- ***Composition**:* A set of record entries relating to one time and place of care delivery; grouped contributions to an aspect of health care activity; composed reports and overviews of clinical progress.
- ***Headed Section**:* Sub-divisions used to group entries with a common theme or derived through a common healthcare process.
- ***Cluster**:* Low-level aggregations of elementary entries (Record Items) to represent a compound clinical concept.

6.2.2 Good European Health Record

In addition, the Good European Health Record, was produced by a European Health Telematics research programme that developed a comprehensive multi-media data architecture for using and sharing electronic healthcare records, meeting clinical, technical, educational and ethico-legal requirements. Later (mostly australian) extensions on the GEHR approach uses a formal semantic model, known as the GEHR Object Model (GOM). Whereas, rather than try to model all possible clinical concepts, the GOM provides concepts at a number of levels:

- EHR and Transaction level
- Navigation level
- Content (e.g. observation, subjective, instruction) level

For example, clinical models are expressed outside the GOM in the form of XML-Schema *archetypes*. These archetypes act as *constraint definitions* and define how to create clinically valid structures out of the GOM primitives. Based on Australian researchers experience using XML based metadata it was found that archetypes are needed for:

- Transaction types, e.g. contact, summary, etc;
- Navigational headings
- Clinical content types e.g. lab-results, prescriptions, including their structure (list, table, series etc)

6.2.3 openEHR

More recently, the *open*EHR specifications incorporate a two-model approach that allows the separation of generic features of any EHR from the domain-specific (usually clinical) features needed for specific EHR instances. Specifically, *open*EHR health record consists of 'container classes' which contain information about the patient or data subject :

- EHR - this is the top level class and contains all information about the data subject.
- Extract - this class contains all information that is to be transferred to another EHR.
- Folder - this class allows information within an EHR to be organized.

  Composition (or document) - this is the class that contains information committed to the EHR by a clinician.

  Section - this class allows information within a composition to be segmented

  Entry - this class contains meaningful information that is to be processed by the machine and read by the clinician.

Note : These classes contain no clinical or demographic concepts at all - and it is this feature which differentiates the *open*EHR approach. The clinical (model) requirements are met through *archetypes* and *templates* designed for the purpose.

6.2.4 Clinical Document Architecture

The HL7 Clinical Document Architecture (CDA) is a generic model for the communication of clinical documents, very similar to the "Composition" class in the CEN 13606 specification, the "Transaction" class in *open*EHR and the HL7 record architecture. CDA documents are encoded in Extensible Markup Language (XML).

- CDA documents derive their meaning from the HL7 Reference Information Model (RIM) and use the HL7 Version 3 Data Types.
- The CDA specification is richly expressive and flexible. Document-level, section-level and entry-level templates can be used to constrain the generic CDA specification. There are many kinds of HL7 Templates that might be created. Among them, two are particularly relevant for clinical documents: (1) those that constrain the document sections based on the type of document (section-level templates); (2) those that constrain the entries within document sections (entry-level templates). Modifications could include limiting the levels of nesting; constraining vocabulary and sequence, for example requiring that a section with a LOINC (Logical Observation Identifier Names and Codes) code for "Subjective" initiate the document body and be followed by a section coded "Objective". These modifications could be expressed in W3C Schema or as Xpath statements within the local schema. Instances that validate against this constrained, local version of CDA are, by definition, also valid CDA instances.

Examples of LOINC a set of more than 10,000 names and codes developed for use as observation identifiers in standardized messages exchanged between clinical computer systems

LOINC_NUM	COMPONENT **(Type of Service)**	SYSTEM **(Setting)**	METHOD_TYPE **(Subject Matter Domain and/or Training / Professional Level)**
34128-9	SUBSEQUENT EVALUATION NOTE	OUTPATIENT	DENTISTRY
34901-9	SUBSEQUENT EVALUATION NOTE	OUTPATIENT	GENERAL MEDICINE
34132-1	SUBSEQUENT EVALUATION NOTE	OUTPATIENT	PHARMACY

7 Clinical Information

7.1 Cross Enterprise Clinical Documents Sharing (XDS)

Integrating the Healthcare Enterprise (IHE) is an initiative designed to stimulate the integration of the information systems that support modern healthcare institutions. It has numerous sponsors and supporting organizations in different medical specialty domains and geographical regions. The IHE Integration Profile known as Cross Enterprise Clinical Documents Sharing is document-content neutral - it will support any type of document without regard to content and format. XDS is focused on providing a standards-based specification for managing the sharing of documents that healthcare entities have decided to explicitly share, such as documents containing simple text, formatted text, images or structured and vocabulary coded clinical information.

Based on the ebXML Registry specification (and implemented with the freebxml registry) the XDS defines the (document) Registry as an actor that maintains metadata about each registered document in a document entry. It also enforces some healthcare specific technical policies at the time of document registration The registry metadata includes a link to the (document) Repository where the actual document is stored which in turn assigns and maintains a unique identifier for each document, to allow Document Consumers to retrieve them. Some of the key XDS concepts are:

- A XDS Affinity Domain is made of a well-defined (federated) set of Document Repositories and Document Consumers that have agreed to share the clinical documents.
- A shared clinical record is called an EHR-LR (Longitudinal Record)
- The care delivery systems used within an enterprise for managing episodes of patient care are called the HER-CR (Care Delivery Record)
- An XDS Document is the smallest unit of information that may be stored in a Document Repository
- An XDS Document is a composition of clinical information that contains observations and services for the purpose of exchange with the following characteristics: Persistence, Stewardship, Potential for Authentication, and Wholeness (characteristics that are well defined in the HL7 CDA)
- An XDS Document shall be associated with Document Registry MetaData defined by the Document Source and associated with the Community of Care entity *aka* Clinical Affinity Domain to which the Document Repository belongs.
- An XDS **Folder** provides a collaborative mechanism for several XDS Document Sources to group XDS Documents for a variety of reasons. XDS Documents may be placed into an existing Folder at any time, as long as they relate to the same patient.
- The custodianship for the clinical information contained in a registered document remains with the Source of the EHR-CR.
- A Document Source may only contribute documents with Document Codes and Health Facility Codes that draw from a Vocabulary Value Set that is approved by the Affinity Domain.

7.2 NIST HL7 Experimental Registry

The USA National Institute of Standards and Technology (NIST) HL7 Experimental Registry tool (also using the freebXML Registry is part of a collaborative effort to determine the appropriate environment, policies and procedures for the development, submission, storage and retrieval of HL7 artifacts. Each artifact has Associations that link it to the organizations that submitted it or are responsible for its maintenance. If an artifact references one or more CMETs in its specification, then there is a Uses association to identify the link from that artifact to each of the referenced CMETs. Further, if an artifact specification is a refinement of a vocabulary domain over any of the structural attributes of its classes, then that artifact is classified by the node in the domain hierarchy to which it is constrained. It is NIST's intent to always have available the most recent DMIMs, RMIMs, HMDs, and MessageTypes from each of the HL7 technical development committees, including:

- HL7-specific classification schemes, especially the code hierarchies for the structural attributes in the RIM.
- RMIM static models from technical domains, including finance, Patient Administration, Scheduling, Laboratory Orders, Research Trials, Pharmacy, Medical Records, Common Message Types, Message Control, Master File, and Clinical Documents
- RMIM static models from the COCT domain (i.e. CMETs), each with a an association to the DMIM it is derived from (using external identifiers to the HL7 identification schemes) with external links to diagrams and descriptions, and a number of classifications by specific codes that are fixed by constraints on the RMIM.
- Each registered RMIMs leads to one or more derived HMDs and MessageTypes.
- Each registered artifact has ExternalLinks to its base UML diagrams, long html descriptions, and other visual display aids for presentation of base classes, attributes, relationships, and constraints.
- OWL-related template artifacts submitted by HL7 participants active in the Templates technical committee

8 Ontology Management

8.1 Mapping & Annotating Ontologies

As required, e-Health service component developers can employ the CEN ENV 13606-2 / GEHR / openEHR / Clinical Document Architecture (CDA) as standards for exchanging

information. Wherein, e-Health service information can be managed by two different e-Health service entities using different message structures. To help automate that interoperation the members of the ARTEMIS project are providing a standard way of accessing the data by registering & storing (1) ontologies based on existing healthcare standards and (2) the semantic mapping between these ontologies. Given, when the e-Health (web) services are annotated with these ontologies, it becomes possible for healthcare organizations conforming to different standards to invoke each others web services by semantic mediation.

Further, to be able to discover the services stored in a Registry, the ARTEMIS project leaders have identified the need for semantic service registry query mechanisms that leverage previous research linking OWL to the Registry Information Model objects, see Figure 5.

Linking OWL to EbXMLRegistry *objects*

Source : Enhancing ebXML Registries to Make them OWL Aware ASUMAN DOGAC et al

Figure 5

OWL	ebXML
owl:Class	ClassificationNode
individual	RegistryObject
rdf:Property	Association
rdfs:domain	sourceObject
rdfs:range	targetObject
owl:equivalentTo owl:samePropertyAs owl:sameAs	An association with a predefined association type of "EquivalentTo".
owl:differentFrom	An new association type of "differentFrom" is defined
owl:AllDifferent owl:distinctMembers	A new association type "distinctMembers" is defined to add members to a Registry Package.
An association with a new association type is defined.	
rdfs:subClassOf owl:ObjectProperty owl:disjointWith owl:TransitiveProperty owl:FunctionalProperty owl:InverseFunctionalProperty owl:SymetricProperty	"subClassOf" "objectProperty" "disjointWith" "transitiveProperty" "functionalProperty" "inverseFunctionalProperty" "symetricProperty"
owl:DataTypeProperty	XML Schema datatypes are used by providing an external link from the registry.
rdfs:subPropertyOf owl:inverseOf	An association between associations with a new association type "subPropertyOf"/"inverseOf" is defined.
owl:intersectionOf	A registry package is created by associating the classes (i.e. the classification nodes) to be intersected through a new association type of "intersectionOf"
owl:Restriction	Since ebXML RIM associations have local scope, only the type of the Restriction needs to be specified.
A slot type is defined for the association representing a restriction.	
owl:allValuesFrom owl:someValuesFrom owl:hasValue	"allValuesFrom" "someValuesFrom" "hasValue"
An association with a new association type is defined.	
owl:cardinality owl:minCardinality owl:maxCardinality	"cardinality" "minCardinality" "maxCardinality"

8.2 Leveling Ontologies

A Community of Interest ontology is a key concept of the BCM Conceptual Layer that supports the semantic links between the building blocks that a service employs. This concept follows from the recognition (validated by formal network theory) that interacting entities, whether they be people or enterprises and their technical systems, tend to coalesce in groups with common characteristics, such as purpose, vocabulary and behavior. As recognized by the XDS Affinity Domain those communities may be defined by formal or informal organizational structures. Indeed, since one e-Health service community may be composed of many sub-communities it is recognized that there is a single Community of Interest ontology to support multiple-levels of ontologies. Wherein the scope of:

- a upper level ontology is focused on the non-volatile language and principles of a domain
- a lower ontology is focused on the knowledge specific to particular community of practice (as formulated by the recognized experts). For example, the table below is based on "a Knowledge Classification for the Clinical Medicine Domain" a definition of levels found in the medical domain produced by HL7 contributors

Scope of Ontology Level

Level Type	**Conceptual Scope**	**Examples**
Upper L 0 Basic COI principles	Vocabulary and other stable semantics of domain, facts true for all instances and all Community of Interest contexts	- SNOMED, Read, ICPC - statements about quantitative data
Lower L 1(a) COP domain	Widely used context-dependent with a common understanding by the Community of Practice	- DIM - e-prescription order
Lower L1(b) Use-case	Context-dependent defined according to particular use cases	- structures implied in LOINC codes
Lower L 2 Sectional	Structural information whose purpose is to logically segment information	- XDS Folder
Lower L 3 Storage	Relating to the physical structuring of information for messaging / storage.	- e-Prescribing DMIM

Lower L 4 Communication	Relating to the packaging of information for the purpose of sharing.	- EHR-LR

9 Choice Points

9.1 XForm & Templates

Formally, templates are considered a gauge, pattern, or molded object. The BCM guideline consider templates as a framing mechanism capable of (1) framing the details of a business process and (2) referencing the appropriate domain ontologies and other semantic information defined in the *Conceptual Layer*. To define their role as an Electronic PRocess (EPR) portal technology a BCM committee is collaborating with Norwegian researchers to specify how semantically rich XML templates, stored in a registered **Folder**, can support formal interaction among interacting communities such as e-Health, e-Building and e-Government.

The BCM e-Health effort also expects to leverage the emerging HL7 approach of employing templates as a form of constraint statement model, which is directly usable for:

- *Data Construction*: to be used at runtime to constrain the creation of data in local contexts to conform to data capture requirements;
- *Data Validation*: to be used at runtime to validate data from other sources.

Note: the following template types have been identified by researchers at the Mayo Clinic and other HL7 SIG members

- Constraint templates operate on code combinations permitted in a code phrase (a V3.0 CD data type that allows addition of data types), known as e*ncapsulation*.

Style templates allow the distribution of semantic details between name and value in name-value pairs (known as *variable vs. value* style), and the style to represent details, either by a pre-coordinated code or by a set of post-coordinated codes (known as the *molecular vs. atomic* issue).

Pattern template defining specific class of items (battery, panel, check list, data set) with some common features (e.g. procedure, sample, dates).

Document template constructed from particular *data elements*. These *data elements* are termed *containers* (sections, subsections, tables, lists), which hold, in turn, the *data values*

Profile template (definition of the relationship between templates and profiles) as a dataset of selected items that describes an existing entity, and is thus related to templates (a data structure), which can be an input form for anticipated data, following defined constraints

Meta template is an ordered aggregation of multiple other templates, such as an *EHR template* for the Electronic Health Record

Extension template which constrains or modifies a profile template

9.2 Archetypes & Templates

It is widely acknowledged that the operation of an e-Heath service will involve thousands of variants of business processes, business rules, business patterns, and data permutations. The BCM guidelines refer to the decisions that use these e-business factors as Choice Points. Further, BCM practitioners assert that the explicit identification and management of these Choice Points significantly aids to comprehensibility, alignment, while promoting tracing and accountability. A prime application of the choice point approach are e-Health templates that use archetypes, wherein:

- archetype is a computable expression of a domain level (clinical) concept in the form of structured constraint statements, based on some reference model (RMIM),
- archetypes are aligned with Affinity Domain concepts,
- archetypes all have the same formalism and may be: part of a COP ontology.
- template is used to narrow the choices of archetypes for local or specific purposes (DMIM).
    - archetype defines constraints on reference model instances which express valid structure (i.e. composition, cardinality).
    - archetype defines constraints on instances of a reference model which express valid types and values.
    - archetypes all belong to one or other ontological level.

10 Semantic To Do

10.1 EbXML Registry Semantic Content Management

To ensure that all types of e-Health Service componentry can be registered and stored in a readily accessible manner the members of the OASIS ebXML Registry are currently defining how Semantic Content Management will be facilitated in the next version of the specification. In addition to providing support for semantic links based on RDF and OWL the check list of requirements includes :

- Re-usable dictionaries of noun definitions for specific industry domains
- Re-useful dictionaries of Business Process catalogues
- Ontology searching and browsing
- Collaborative (COP) ontology development
- Classify content using OWL Ontology class hierarchies
- Discover content using semantic queries
- Attach semantic Objects to Stored Content
- Semantic reasoning …

Acknowledgements

This paper could not have been written without (1) the forward thinking of the members on the OASIS BCM Technical Committee and the ebXML Registry Technical Committee, (2) the contributors to the open source code of the freebXMLRegistry , (3) dedicated researchers Dr. Peter L. Elkin of the Mayo Clinic and Professor / Dr. Asuman Dogac of the Middle East Technical University Dept. of Computer Engineering , (4) Computer architectural innovators Paul J Alagna Chief Architect of CHECKMi ; Farrukh Najmi of SUN and David Webber the cam do man.