Semantic Healthcare

Sankhesh Jhaveri — *Kitware,Inc.*

Catherine Dumas — *Kitware, Inc., SUNY Albany*

Joshua Cope — *Kitware, Inc., SUNY Albany*

Table of Contents

1 SNOMED RT®

Systematized Nomenclature in Medicine Reference Terminology (SNOMED RT) represents the initial step towards unifying various clinical terms in healthcare. SNOMED RT was designed to complement the broad coverage of medical concepts in SNOMED with a set of enhanced features that significantly increased its value as a reference terminology for representing clinical data (Spackman et al., 1997). SNOMED RT was developed by the College of American Pathologists (CAP).

SNOMED RT is a concept-based terminology. A *concept* is a unit of thought that refers to a unique, clearly defined entity. An example is "Fundus of stomach". A *term* is a particular lexical string or expression that represents a

Concept Code	Descriptions	Status
D3-89550	Cerebrovascular accident CVA Stroke	Preferred name Synonym Synonym

Table 1: SNOMED RT - Concepts, Descriptions and Synonyms (A Y Wang, 2001, Table. 1)

concept. Terms are used in clinical information systems or other healthcare applications. In SNOMED RT, we use *description* to refer to terms that are linked to concepts in core tables. This imparts a specific, non-ambiguous meaning to each term. A single concept may have one or more associated descriptions. One description in each concept is designated the *preferred name*, and the others are called *synonyms*(See Table 1). Term and description have often been used interchangeably in the past. However, the two are being distinguished because a term can be associated with different concepts in the clinical information systems depending on context, but a description is ideally non-ambiguous and always associated with a concept.

Some of the fundamental aspects of SNOMED RT (Dolin et al., 2001) are:

- Hierarchies in SNOMED RT represent strict supertype-subtype relationships. Therefore, a child concept is necessarily always a kind of the parent concept.
- Concepts are defined by their placement in a (poly)hierarchy and by additional properties called "Relationship Types" or "Roles", whose target values are also SNOMED concepts. For example, Appendectomy (P1-57450) has an "ASSOCIATED-TOPOGRAPHY" role, whose value is Appendix (T-59200).
- SNOMED RT contains textual definitions, which are especially valuable when the underlying description logic is unable to define a procedure fully.

2 SNOMED CT®

Systematized Nomenclature in Medicine Clinical Terms (SNOMED CT) is a comprehensive, multilingual clinical terminology that provides clinical content and expressivity for clinical documentation and reporting.

It can be used to code, retrieve and analyze clinical data. SNOMED CT was formed by the merger, expansion and restructuring of SNOMED RT[1] and the United Kingdom National Health Service (NHS) Clinical Terms (also known as the Read Codes). In a nutshell, SNOMED CT consists of concepts arranged in a hierarchy, connected by relationships. The International Health Terminology Standards Development Organization (IHTSDO) owns and administers the rights to SNOMED CT.

According to (SNOMED User Guide, 2011), there are three basic components of SNOMED CT:

- Concepts
- Descriptions
- Relationships

2.1 Concepts

Concepts are clinical ideas, ranging from *abscess* to *zygote*, identified by a unique numeric identifier (*ConceptId*) that never changes and represented by a unique human readable *Fully Specified Name (FSN)*. The concepts are formally defined in terms of their Relationships with other concepts. These logical definitions give explicit meaning which a computer can process and query on. Every concept also has a set of terms that name the concept in a human-readable way. There are well over 300,000 active concepts in the terminology with differing levels of granularity linked to one another by | is a | relationships as depicted in Figure 1.

Concept identifiers in SNOMED CT are meaningless to avoid changes to reflect revised understanding of the nature of a disorder. Meaningless identifiers also enable multiple aspects of meaning to be represented in the same way.

2.2 Descriptions

Concept Descriptions are the terms or names assigned to a SNOMED CT concept. A unique DescriptionId identifies a Description. Multiple Descriptions might be associated with a concept identified by a ConceptId. There are nearly a million English Descriptions in the International Release of SNOMED CT. Each translation of SNOMED CT includes an additional set of descriptions, which link terms in another language to the same SNOMED CT concepts.

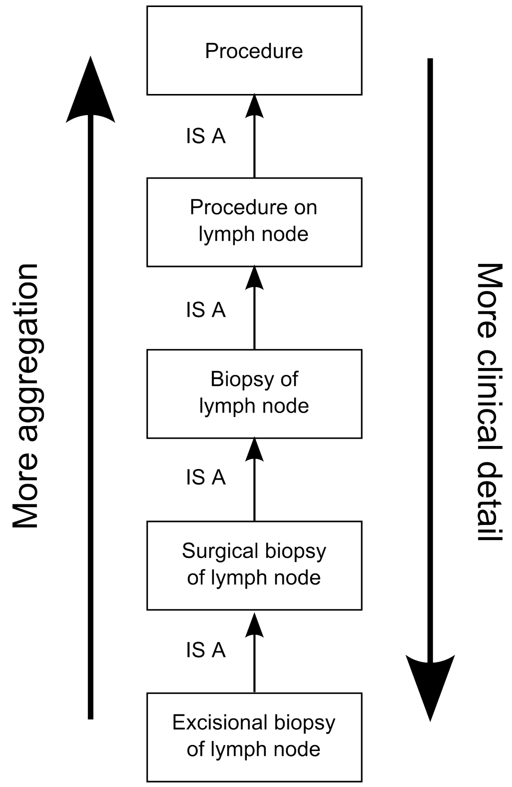


Figure 1: SNOMED CT - Multiple levels of granularity
(SNOMED User Guide, 2011, Fig. 1)

Example: Some of the Descriptions associated with *ConceptId* 22298006:

- Fully Specified Name: — Myocardial infarction (disorder) — *DescriptionId* 751689013
- Preferred term: Myocardial infarction *DescriptionId* 37436014
- Synonym: Cardiac infarction *DescriptionId* 37442013
- Synonym: Heart attack *DescriptionId* 37443015
- Synonym: Infarction of heart *DescriptionId* 37441018

Each of the above Descriptions has a unique *DescriptionId*, and all of these Descriptions are associated with a single Concept (and the single *ConceptId* 22298006).

2.3 Relationships

SNOMED CT Relationships link each concept to other concepts that have a related meaning. These relationships provide formal definitions and other characteristics of the concept. One type of link is the | is a | relationship which relates a concept to the its more general concepts. For example (Figure 2), the concept “viral pneumonia” has an | is a | relationship to the more general concept “pneumonia”. These | is a | relationships define the hierarchy of SNOMED CT concepts. Other types of relationships represent other aspects of the definition of a concept. For example, the concept “bacterial pneumonia” has a | causative agent | relationship to the concept “bacteria” and a | finding site | relationship to the concept “lung structure”. There are well over a million relationships in SNOMED CT.

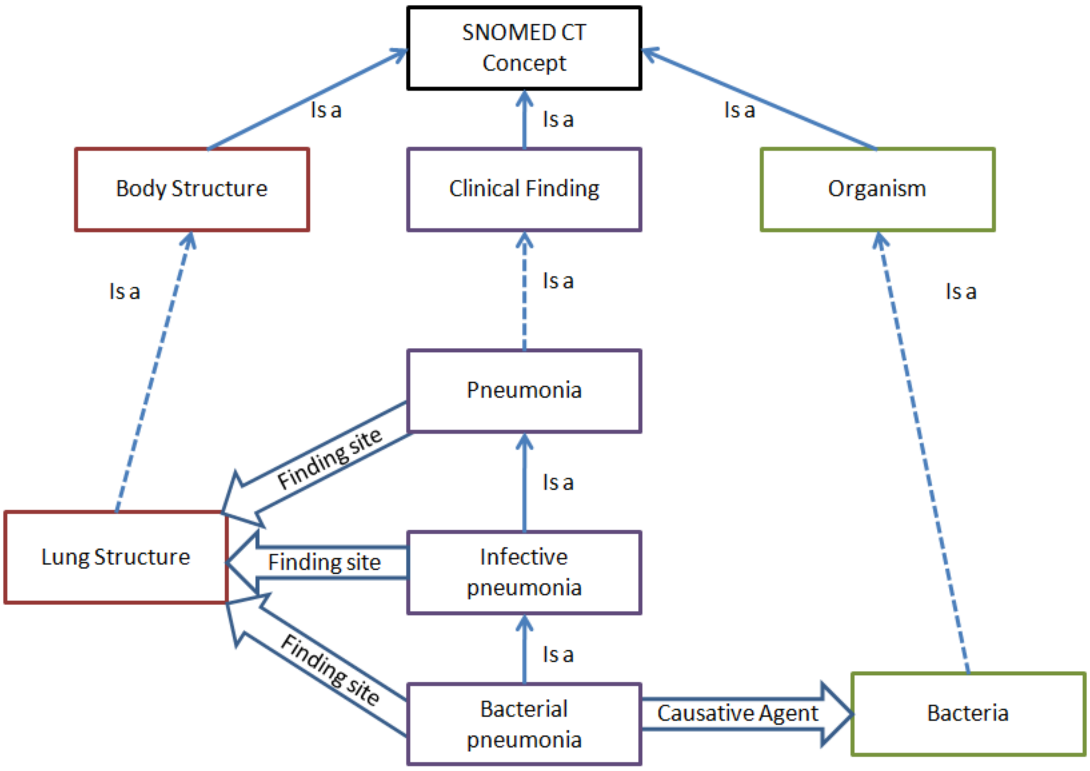


Figure 2: SNOMED CT - Illustration of Defining Relationships (SNOMED User Guide, 2011, Fig. 7)

2.4 Implementation

SNOMED CT is distributed as a set of tab-delimited text files that can be imported into a relational database. The three tables - the Concepts table, the Descriptions table and the Relationships table are commonly referred to as *Core Components* (SNOMED Implementation Guide, 2011). Supplementary tables called *Reference Sets* specify the common extensible pattern that is used to add additional information related to the core components.

2.5 Summary

SNOMED CT is a used widely to achieve semantic uniformity and consistency of health terms as well as to achieve interoperability between HL7[9] (Health Level 7) based health frameworks and other healthcare entities as shown by the works in (Ryan, 2006; Argello et al., 2009; Khan et al., 2012). Not only does it provide unique semantic identifiers to clinical concepts, SNOMED CT also describes and links different concepts in an ontogolical fashion. While SNOMED CT has emerged internationally as a leading terminology, the work of (He et al., 2012; Khare et al., 2012) delineates that the existing SNOMED CT lexicon suffers from a surprisingly huge paucity of synonyms. Efforts are underway to reduce SNOMED CT's structural complexity and provide a metathesaurus of clinical concepts with mappings to different terminologies, thereby improving semantic integrity in practical healthcare scenarios. (Lindberg et al., 1993; Wei et al., 2012)

3 CMT

Convergent Medical Terminology (CMT) is a set of clinician and patient friendly terminology, linked to US and international interoperability standards, and related vocabulary development tools and utilities. Developed by Kaiser Permananente over many years for use within its health-IT systems, CMT now includes more than 75,000 concepts. CMT is a core component Kaiser Permanente's comprehensive electronic health record *KP HealthConnect*®.

In September 2010 Kaiser Permanente, the International Health Terminology Standards Development Organisation (IHTSDO) and the US Department of Health and Human Services jointly announced Kaiser Permanente's donation of their CMT content and related tooling to the IHTSDO. The donation consists of terminology content already developed, a set of tools to help create and

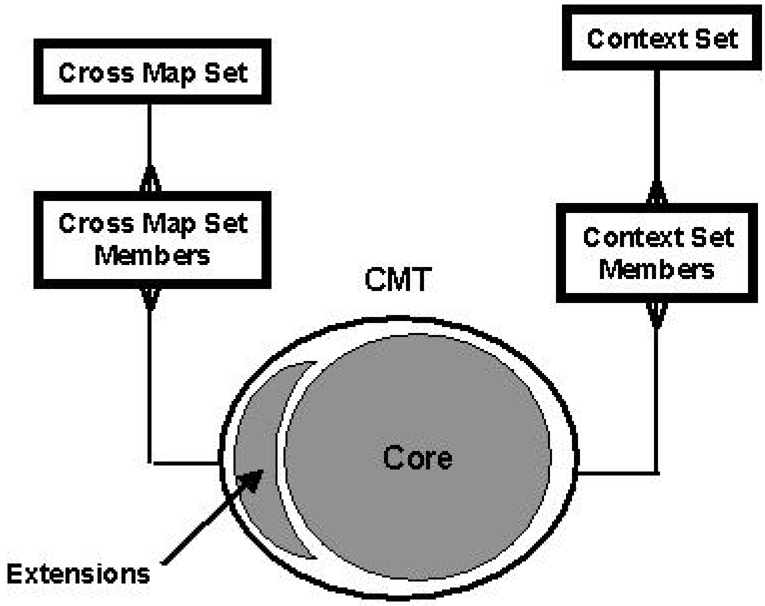


Figure 3: High-level graphical description of CMT
(Dolin et al., 2004, Fig. 1)

manage terminology and processes to control the quality of terminology that is developed. CMT also includes mappings to classifications and standard vocabularies including SNOMED CT[2].

A high-level graphical depiction of CMT is shown in Figure 3. CMT is built upon industry standard terminologies. SNOMED CT[2], laboratory LOINC[8] and First DataBank[13.1] drug terminology form the core of CMT. Core terminologies are integrated into a single poly-hierarchically structured knowledge base. A classifier organizes the CMT concepts into a poly-hierarchy, based on their definitions. The act of classifying helps identify synonymous concepts, and maintains quality and consistency across the some 400,000 concepts.

Applications can directly access CMT via a provided interface and/or CMT can provide applications with cross map sets and context sets, both of which are patterned after the SNOMED CT[2] model. Cross map sets are used to store mappings between CMT concepts and other coding schemes. Context sets are CMT subsets used within a particular context. Contexts can include a particular drop-down list or vocabulary table in an application, a field in an HL7[9] message, or any other CMT subset needed within the organization.

CMT is currently distributed within the UMLS Metathesauras (http://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/index.html).

4 RDF

R*esource Description Framework (RDF)* is a World Wide Web Consortium (W3C) standardized data model for representing semantic Web resources. It uses graphs to represent information using a triple-based notation comprising a subject, predicate and an object. All these entities can be uniquely identified by Internationalized Resource Identifiers (IRIs) (Pathak et al., 2012).

4.1 How can we use it

We can use it by evoking already existing tools such as D2R Map (Bizer, 2006) which is an open source tool thats transfers relational data into RDF format which will then allow us to easily gain better insight between data.

4.2 Why this model would be useful for our application

RDF offers a practical evolutionary pathway to semantic interoperability. It enables information to be readily linked and exchanged with full semantic fidelity while leveraging existing IT infrastructure investments. Being schema-flexible, RDF allows multiple evolving data models and vocabularies to peacefully coexist in the same instance data, without loss of semantic fidelity. This enables standardized data models and vocabularies to be used whenever possible, while permitting legacy or specialized models and vocabularies to be semantically linked and used when necessary. It also enables a limitless variety of related information to be semantically linked to patient data, such as genomic, geographic and drug interaction data, enabling more effective treatment, and greater knowledge discovery. Other reasons for adopting RDF as a universal healthcare exchange language include (Munnecke, 2013):

- Its ability to make information self-describing with precise semantics
- Its support for automated inference
- Its foundation in open standard

By using a standard language for data interchange, new research discoveries could be made more efficiently and effectively.

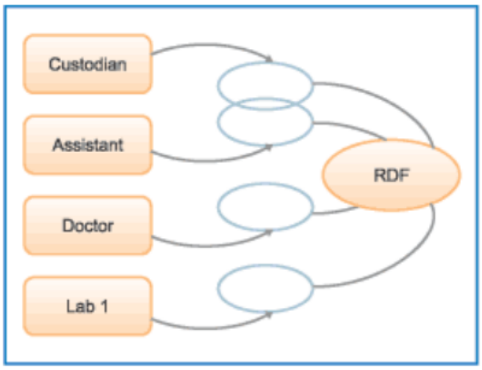
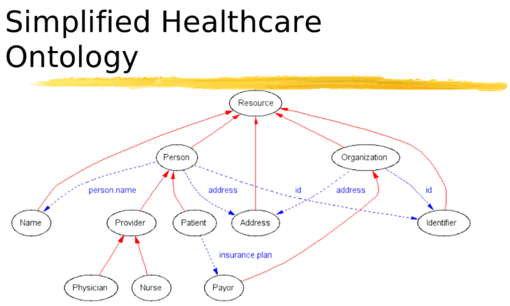


Figure 3: Data integration within a Healthcare Information System
Source: Infosys Research



(a) RDF Example 1 (Parachuri & Majumdar, 2008) (b) RDF Ontology Example (Borden, 2012)

Figure 4: RDF Examples

4.3 Comparison with other data models

RDBMS/SQL	XML /Xquery	RDF /SPARQL
Efficient and correct transactions	Transaction across organizational boundaries	Flexible information sharing
Metadata is embedded in application or database schema	XML wraps the metadata about the transaction around the data thereby increasing size	Enables semantics as well as the syntax to be embedded in documents making it machine readable

Table 1: *Advantages of Using Semantic Languages* ***Source:*** *Infosys Research*

Figure 5: Comparison of RDF, XML and SQL
(Parachuri & Majumdar, 2008, Fig. 1)

4.3.1 XML

XML is a comparable data model to RDF and in fact one way you can express RDF data is in a certain XML format. What sets RDF apart from XML is that RDF is designed to represent knowledge in a distributed world. That RDF is designed for knowledge, and not data, means RDF is particularly concerned with meaning.

In some ways, RDF can be compared to XML. XML also is designed to be simple and applicable to any type of data. XML is also more than a file format. It is a foundation for dealing with hierarchical, self-contained documents, whether they be stored on disk in the usual brackets-and-slashes format, or held in memory and accessed through a DOM API. (Tauberer, 2008b)

4.3.2 SQL

Relational Databases such as SQL is also a comparable data model to RDF, and you can actually store your RDF data inside of a relational database. Individual statements in RDF are expressed as subject, predicate, object triples. Sets of these with a common predicate can be mapped to binary relations in the relational model, in the the common parlance, 2-column tables.

But a difference between the two is that in Relational DB's, for a certain set of values a relation is either considered either true (there is a corresponding row in the table) or false (there isn't). In the RDF model in the general case, if a set of values isn't in the "row" (i.e. you don't have a particular statement), then it's not false, just unknown. (Tauberer, 2008a)

5 Bio2RDF

Bio2RDF is an open source project that uses semantic web technologies to build and provide the largest network of Linked Data for the Life Sciences. It defines a set of simple conventions to create RDF[4] compatible Linked Data from a diverse set of heterogeneously formatted sources obtained from multiple data providers. Bio2RDF is the culmination of efforts towards addressing the pressing need for a global multisite search engine. It is defined as a system that is able to query and connect different databases available on the Internet (Belleau et al., 2008).

Bio2RDF uses RDF documents and a list of rules (*Banff Manifesto*[*]) to create URIs that will produce linked data:

Rule 1: URI's are normalized and dereferencable

[*] http://sourceforge.net/apps/mediawiki/bio2rdf/index.php?title=Banff_Manifesto

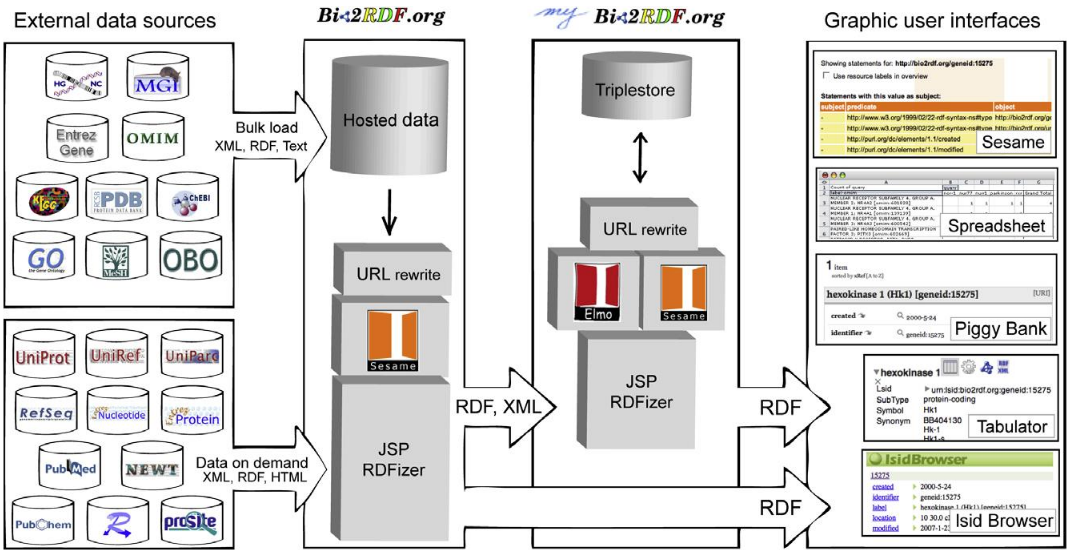


Figure 6: Bio2RDF knowledge system framework architecture (Belleau et al., 2008, Fig. 1)

Rule 2: Authoritative public namespaces are used

Rule 3: Mandatory predicates are used

Rule 4: Blank nodes are forbidden

Rule 5: RDFizer programs are open source

Rule 6: Deference-able ontologies

Figure 6 shows a schematic description of the Bio2RDF architecture. All external data sources, in different formats (XML, Text, ASN.1, KGML and RDF), are listed on the left part. Data is then made accessible either on the Bio2RDF.org server or on demand from the original source. The myBio2RDF application contains an rdfizer program and a servlet that answers Bio2RDF HTTP requests by formulating SPARQL[11] endpoints (Callahan et al., 2013a).

5.1 Naming Convention

Bio2RDF entities are named as follows:

`http://bio2rdf.org/namespace:identifier`

where, *namespace* is the preferred short name of a biological dataset and *identifier* is the unique string used by the source provider to identify the given record. For example, HUGO Gene Nomenclature Committee identifies the human prostaglandin E synthase gene (PIG12) with the accession number '9599'. This dataset's namespace is "hgnc" in Bio2RDF's dataset registry* and the corresponding Bio2RDF IRI is: http://bio2rdf.org/hgnc:9599 There are over 1800 such namespaces.

5.2 Bio2RDF Dataset Provenance Model

Provenance data for each Bio2RDF dataset is stored in a separate named graph in each corresponding SPARQL endpoint. The provenance graph URI follows the pattern:

http://bio2rdf.org/bio2rdf-[dataset]-provenance

where, 'dataset' is the preferred short name (or prefix) for a given source dataset as extracted from the Life Science Registry†.

For example, the "NLM Medical Subject Headings (MeSH)" dataset provenance graph makes use of http://bio2rdf.org/bio2rdf-mesh-provenance as its graph URI. Bio2RDF's provenance model uses the W3C Vocabulary of Interlinked Datasets (VoID), the Provenance vocabulary (PROV) and Dublin Core vocabulary. Each dataset provenance object has a unique IRI and label based on the dataset name and creation date. For example, http://bio2rdf.org/bio2rdf_dataset:bio2rdf-mesh-20120827. An example provenance graph for the MeSH dataset can be seen in Figure 7. Note that each subject IRI in the dataset is linked the date-unique dataset IRI that is part of the provenance record using the VoID 'inDataset' predicate. Other important features of the provenance record include the use of the Dublin Core 'creator' term to link a dataset to the script on Github that was used to generate it, the VoID predicate 'sparqlEndpoint' to point to the dataset SPARQL endpoint, and VoID predicate 'dataDump' to point to the data download URL.

Although Bio2RDF facilitates integration of and programmatic access to otherwise heterogeneous datasets (in both, content and format), a complete syntactic and semantic normalization across numerous datasets has yet to be fully realized. Works of (Ansell, 2011; Callahan et al., 2013b; Castro et al., 2013) demonstrate better and improved models of resolving queries to the Bio2RDF datasets.

* https://docs.google.com/spreadsheet/ccc?key=0AnGgKfZdJasrdElfQzRWWWhKUFR0UnRpeG14NGZRS2c

† Hosted by Bio2RDF at https://docs.google.com/spreadsheet/ccc?key=0AmzqhEUDpIPvdFR0UFhDUTZJdnNYdnJwdHdvNVlJR1E

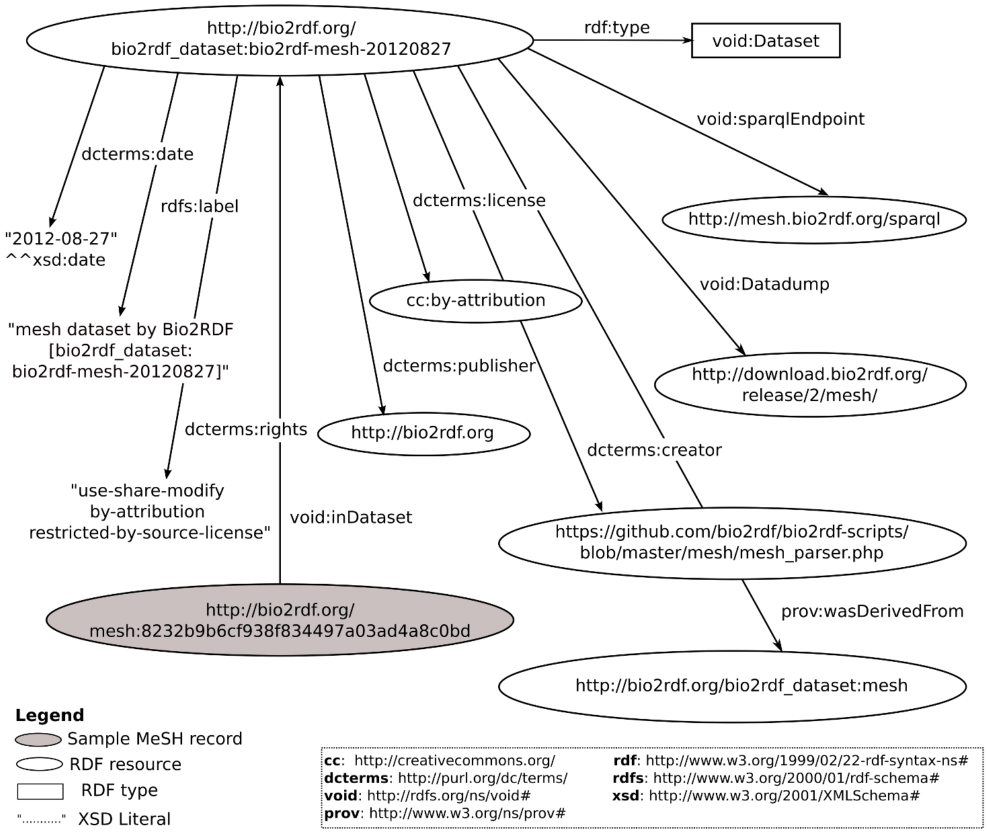


Figure 7: Bio2RDF Provenance Graph for the NLM Medical Subject Headings (MeSH) architecture (Bio2RDF Wiki, 2013)

6 3M™ HDD

Healthcare Data Dictionary (HDD) is a controlled medical vocabulary server; makes it possible to map and manage medical terminologies, integrate content and standardize healthcare data. Allows organizations to transmit and receive accurate, actionable patient data across systems and applications, regardless of where data originates (3M, 2013).

HDD incorporates and links terms from multiple clinical information systems and standard terminologies. It maps disparate medical terms to give data context and meaning; and is used to standardize data to make it more interoperable and computable. It is a concept-based vocabulary and knowledge base. (3M, 2013).

Figure 8: Key Components of HDD
(3M, 2010)

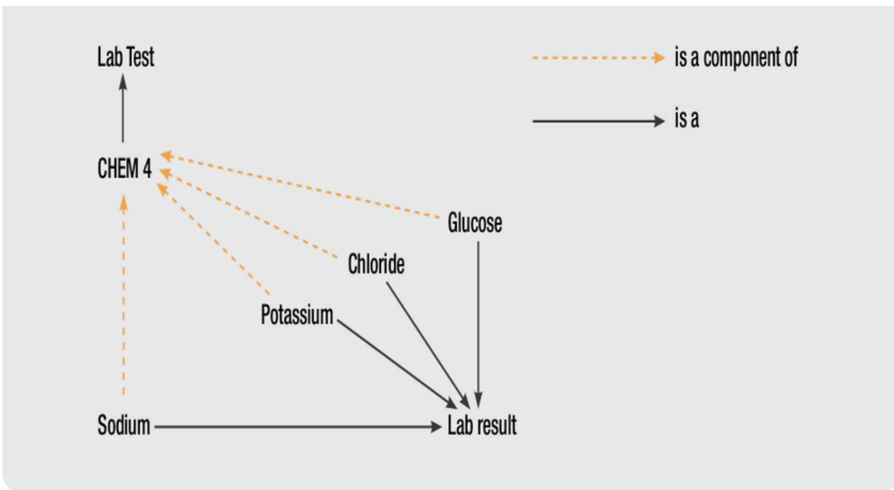


Figure 9: Sample of HDD's knowledge base as applied to CHEM 4 lab test
(3M, 2010)

In June 2012, the U.S. Department of Defense and the Department of Veterans Affairs reached an agreement with Salt Lake City-based 3M Health Information Systems to make HDD freely available as open-source content and software. HDD, which has been deployed since 1996 by 3M under an agreement with the DoD and VA, was started as a development project to standardize clinical information. All major healthcare standard technologies are mapped unto HDD: SNOMED CT[2], LOINC[8], ICD-9 and ICD-10[7]. (DeGaspari, 2013)

6.1 Information Models, Knowledge Base, Vocabulary

- **Information Models:** describes relationships among events and terminologies in a way that gives them meaning and context. IM's mediate between data gathering software and databases and are supported by terminologies.

- **Knowledge Base:** defines the domains referenced by the information models. Domains are created and populated by the web of relationships among concepts. Semantic relationships fall into two broad categories: hierarchical and non hierarchical.

  HDD's knowledge base consists of semantic networks and hierarchies that describe the complex relationships existing between concepts in the vocabulary. These relationships can be hierarchical ("parent-child" or "is-a") or non-hierarchical "is-a-component-of". Figure 6 (below) is an example of how the knowledge base can describe the relationships between the components of a CHEM 4 laboratory test. (3M, 2010)

- **Vocabulary:** Identifies medical concepts and organizes them to support synonyms, multiple surface forms, and other lexical characteristics. The HDD is a controlled medical vocabulary that follows best medical informatics practices such as concept permanence, multiple hierarchies, and meaningless identifiers. Each unique concept in the HDD is assigned a Numeric Concept Identifier (NCID) code. (3M, 2010)

6.2 How can we use it

The server is the complete repository of concepts and associations, maintained and made available for download in its most up-to-date version. Organizations can download and install the package locally to ensure constant access to its contents. Users of the HDD will have the ability to request additions to the dictionary, which the team at 3M will review and determine if they should be included. (Murphy, 2012)

ASN.1 Definition: A MedicationOrder event is defined by a set of observations

```
    MedicationOrder ::= SET {
drug              Drug,
dose              Decimal,
route             Route,
frequency         Frequency,
startTime         DateTime,
endTime           DateTime,
orderedBy         Clinician,
orderNum          OrderNumber}
```

Figure 10: Simplified ASN.1 definition
(3M, 2010)

6.3 Architecture and Platform Independence

HDD was designed to meet open architecture standards, allow for platform independence, and conforms to the following industry standards:

- Linux platform
- HL7 Common Technology Services (CTS)
- HL7 messaging
- Abstract Syntax Notation (ASN.1) information model. transferable to XML (3M, 2010)

CDR Instance Data: The ASN.1 model references the HDD when a coded element is specified in the definition and stores (NCID codes)

```
    MedicationOrder {
drug              Ampicillin (NCID 1234),
dose              500,
route             Oral (NCID 5678),
frequency         Q6H (NCID 9123),
startTime         09/01/95 10:01,
endTime           09/11/95 23:59,
orderedBy         John Doe MD (NCID 8123),
orderNum          A234567 }
```

Figure 11: Sample instance data
(3M, 2010)

6.4 Mapping Source Controlled Medical Vocabularies (CMV's) and Local Vocabularies

All industry-standard CMVs can coexist in the 3M HDD because of a process called mapping, which cross-references elements in each CMV with a concrete, unambiguous concept in the 3M HDD.
(3M, 2010)

Concept ID	Definition
Cold, #123	A sensory perception ("patient complains of feeling cold")
Cold, #569	A pulmonary diagnosis (Chronic Obstructive Lung Disease)
Cold, #784	An upper respiratory viral infection ("common cold," "cold," "flu," etc.)

Figure 12: Sample of HDD's concepts and NCIDs
(3M, 2010)

7 ICD

International Classification of Diseases (ICD) is a classification of diseases by WHO to permit the systematic recording, analysis, interpretation and comparison of mortality and morbidity data collected in different countries or areas and at different times. The ICD is used to translate diagnoses of diseases and other health problems from words into an alphanumeric code, which permits easy storage, retrieval and analysis of the data. In practice, the ICD has become the international standard diagnostic classification for all general epidemiological and many health management purposes.

The ICD can be used to classify diseases and other health problems recorded on many types of health and vital records. Its original use was to classify causes of mortality as recorded at the registration of death. Later, its scope was extended to include diagnoses in morbidity. It is important to note that, although the ICD is primarily designed for the classification of diseases and injuries with a formal diagnosis, not every problem or reason for coming into contact with health services can be categorized in this way. Consequently, the ICD provides for a wide variety of signs, symptoms, abnormal findings, complaints and social circumstances that may stand in place of a diagnosis on health-related records. It can therefore be used to classify data recorded under headings such as ‘diagnosis’, ‘reason for admission’, ‘conditions treated’ and ‘reason for consultation’, which appear on a wide variety of health records from which statistics and other health-situation information are derived.

7.1 Family of Diseases

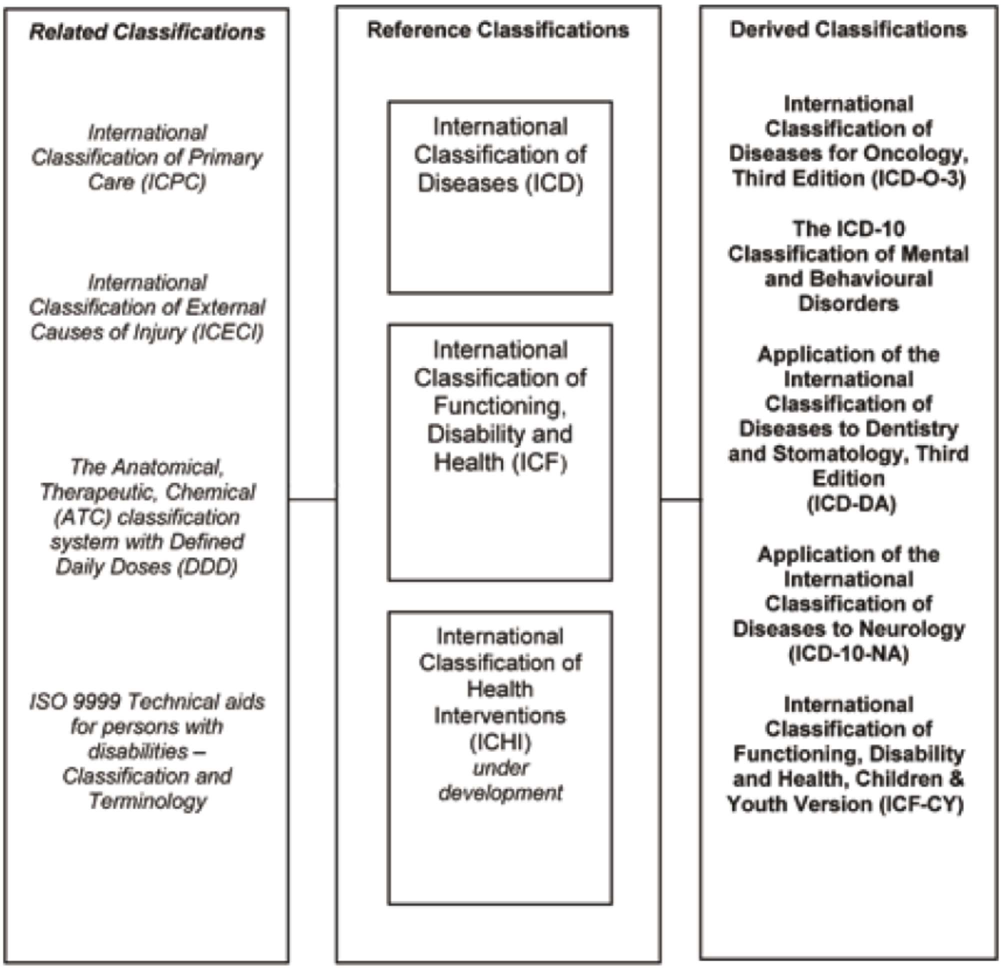


Figure 13: Schematic representation of the WHO-FIC (ICD10 Instruction Manual - Volume 2, 2010)

Although the ICD is suitable for many different applications, it does not serve all the needs of its various users. It does not provide sufficient detail for some specialties and sometimes information on different attributes of health conditions may be needed. The ICD also is not useful to describe functioning and disability as aspects of health, and does not include a full array of health interventions or reasons for encounter. In order to overcome these shortcomings, the concept of *family of diseases* was developed. Currently 'family' designates a suite of integrated classification products that share similar features and can be used singularly or jointly to provide information on different aspects of health and health-care system. For example, the ICD as a reference classification is mainly used to capture information on mortality and morbidity. Additional aspects of health domains, functioning and disability have now been jointly classified in the International Classification of Functioning, Disability and Health (ICF). The WHO Family of International Classifications (WHO-FIC) attempts to serve as the framework of international standards to provide the building blocks of health information systems. Figure 13 represents the types of classifications in the WHO-FIC.

Compared with ICD-9, ICD-10 has (DiSantostefano, 2009):

- Expanded detail for many conditions (e.g., viral hepatitis has been expanded from ICD-9 070, a single 3-digit category, to ICD-10 B15-B19), five 3-digit categories.
- Transferred conditions around the classification (e.g., hemorrhage has been moved from the circulatory chapter to the symptoms and signs chapter).
- Used alphanumeric codes instead of numeric codes (e.g., code for diabetes mellitus was 250.XX in ICD-9 and is E10-E14 in ICD-10).
- Modified coding rules (e.g., the "Old pneumonia, influenza and maternal conditions" and "Error and accidents in medical care" coding rules have been eliminated).
- Modified the tabulation lists (e.g., the U.S. ICD-10 113-clause list replaces the U.S. ICD-9 72-cause list).

A comprehensive list of resources can be accessed on the WHO-ICD website*. One of the interesting tools in an online application designed for classroom training as well as self-training on ICD-10.

*http://www.who.int/classifications/icd/en/

8 LOINC®

Logical Observation Identifier Names and Codes terminology (LOINC) is a definitive standard for identifying clinical information in electronic reports. The LOINC database provides a set of universal names and ID codes for identifying laboratory and clinical test results in the context of existing HL7, ASTM E1238, and CEN TC251 observation report messages. One of the main goals of LOINC is to facilitate the exchange and pooling of results for clinical care, outcomes management, and research. LOINC codes are intended to identify the test result or clinical observation. Other fields in the message can transmit the identity of the source laboratory and special details about the sample. (Vreeman, 2010)

Examples of fully specified LOINC names:

Sodium:SCnc:Pt:Ser/Plas:Qn

Sodium:SCnc:Pt:Urine:Qn

Sodium:SRat:24H:Urine:Qn

Creatinine renal clearance:VRat:24H:Ur+Ser/Plas:Qn

Glucose^2H post 100 g glucose PO:MCnc:Pt:Ser/Plas:Qn

Gentamicin^trough:MCnc:Pt:Ser/Plas:Qn

ABO group:Type:Pt:Bld^donor:Nom

Body temperature:Temp:8H^max:XXX:Qn

Figure 14: Example of LOINC format (McDonald et al., 2013)

8.1 Global use

Since then, many others have contributed translations. Currently, there are translation efforts underway in 18 countries to translate LOINC into 12 different languages, with translations into nine languages included in the most recent public LOINC release.

The freely available translations into many languages allows LOINC to be able to establish a global presence, and benefits the medical community by interoperable health information exchange around the world.(Vreeman et al., 2012)

8.2 Why would we use it for our application

We should use it because of it's popular use currently around the world and the ability to adhere to community standards.

LOINC has been widely adopted in both the public and private sectors, within the United States and more than 140 other countries. Several countries (including Brazil, Canada, Germany, the Netherlands, Mexico, and Rwanda) have adopted LOINC as a national standard, and there are large health information exchanges using LOINC in Spain, Singapore, and Korea as well. Within the US, LOINC has been adopted by many health information exchanges, large national reference laboratories, healthcare organizations, insurance companies, research programs, and national standards. (Kroth et al., 2012)

8.3 How can we use it

There is a free software program developed by The Regenstrief Institute called RELMA (the Regenstrief LOINC Mapping Assistant) which enables browsing and searching the database, review accessory content for terms and map local terms to LOINC. (Kroth et al., 2012)

8.4 How is data stored

LOINC is available as an Microsoft Access (.mdb) database file, a tab-delimited text file (.txt), and also a comma delimited text file (.csv). (Vreeman, 2010)

8.5 Potential disadvantages

There are variations in the way LOINC is used for data exchange that result in some data not being truly interoperable across different enterprises. To improve its semantic interoperability, we need to detect and correct any contradictory knowledges. Also needed is detailed guidance on best practices for mapping from local codes to LOINC codes and for using LOINC codes in data exchange. (Lin et al., 2012)

9 HL7 FHIR™

F*ast Healthcare Interoperability Resources (HL7/FHIR)* is a is a next generation standards framework created by HL7. It defines a set of "resources" for health. These resources represent granular clinical concepts that can be exchanged in order to quickly and effectively solve problems in healthcare and related process. The resources cover the basic elements of healthcare - patients, admissions, diagnostic reports, medications, and problem lists, with their typical participants, and also support a range of richer and more complex clinical models. The simple direct definitions of the resources are based on thorough requirements gathering, formal analysis and extensive cross-mapping to other relevant standards. (HL7.org, 2012)

Resources are:

- Atomic - they are the smallest defined unit of operation and a transaction scope of their own.
- Connected - resources refer to other resources to allow for clean modular reuse of information.
- Independent - the content of a resource can be processed without having to retrieve referenced resources.
- Simple - each resource definition is easy to understand, and to implement without needing specialized tooling or infrastructure (though that can be used if desired).
- RESTful - resources are able to be used in a RESTful exchange context.
- Flexible - resources can also be used in non-RESTful contexts, such as messaging or SOA architectures and can be moved in and out of RESTful paradigms as convenient.
- Extensible - resources can be extended to allow for local requirements without impacting on interoperability.
- Webcentric - where possible and appropriate, open internet standards are used for data representation.
- Free for use - the FHIR specification itself is open - anyone can implement FHIR or derive related specifications without any IP restrictions.

In addition to the basic resources, FHIR defines a lightweight implementation framework that supports the use of these resources in RESTful environments, classic message exchanges, human-centric clinical documents and enterprise SOA architectures. Each of these approaches provides its own benefits - FHIR provides the underpinning enablement that makes the choosing one of these painless and enables enterprises to choose their own paradigm without forsaking interoperability with other approaches.

Though the resources are simple and easy to understand, they are backed by a thorough, global requirements gathering and formal modeling process that ensures that the content of the resources is stable and reliable. 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.

FHIR also provides an underlying conformance framework and tooling that allows different implementation contexts and enterprises to describe their context and use of resources in formal computable ways and to empower computed interoperability that leverages both the conformance and definitional frameworks.

The combination of the resources and the 3 supporting layers (implementation frameworks, definitional thoroughness, and conformance tooling), along with the completely free license of FHIR itself frees healthcare data so that it can easily flow to where it needs to be (across hospital production systems, mobile clinical systems, cloud based data stores, national health repositories, research databases, etc.) without having to pass through format and semantic inter-conversion hurdles along the way. (HL7.org, 2013)

10 NwHIN

Nationwide Health Information Network (NwHIN) is a set of standards, services, and policies that enable the secure exchange of health information over the Internet. (Framework, 2013a) This is currently achieved through three different initiatives:

- NwHIN Exchange
- NwHIN Direct
- Aurion Software

10.1 NwHIN Exchange

NwHIN Exchange is a more complex exchange protocol that has methods to perform universal patient lookup, document discovery and retrieval, and exchange between organizations and federal agencies (VA, DOD, CDC, SSA, plus 22 others). The organizations entering into an exchange with those federal agencies are typically sizable HIOs, HIEs or large IDNs. Participation in the NwHIN Exchange is currently limited to federal health agencies and healthcare organizations under ONC contract and other recipients of federal grants. There are technical teams devoted to the on-boarding process (validation and conformance testing), security, authentication, and adherence to the specifications standards, including producing/accepting structured data in defined formats. (Information & Communications Alliance, 2013)

Most individual providers/small practices don't have the technical capability to implement this exchange. That's where the other initiatives such as the NwHIN Direct and the Aurion Software come into play to allow for a more simpler path to achieving the goals of the program.

10.2 NwHIN Direct

NwHIN Direct known as a simpler alternative to NwHin Exchange also enables standards-based health information exchange in support of core Stage 1 MU measures, including communication of summary care records, referrals, discharge summaries and other clinical documents in support of continuity of care and medication reconciliation, and communication of laboratory results to providers. It exclusively supports cases of pushed communication between providers, hospitals and laboratories. It also consists of Authentication, certificates, vocabulary messaging standards and security.

The key difference between Exchange and Direct is messaging: Direct focuses on the technical standards and services necessary to securely push content from a sender to a receiver and not the actual content exchanged. (of Health & Services, 2013) The NwHIN Direct specifications will support unstructured messages (i.e., simple text or PDF), semi-structured text, and highly structured messages like CCD C32. NwHIN Direct is not currently capable of supporting MU exchange requirements beyond elemental Stage 1 requirements. (Information & Communications Alliance, 2013)

10.3 Aurion Software

Formally known as CONNECT, Aurion is an open source health information exchange platform that implements the Nationwide Health Information Network

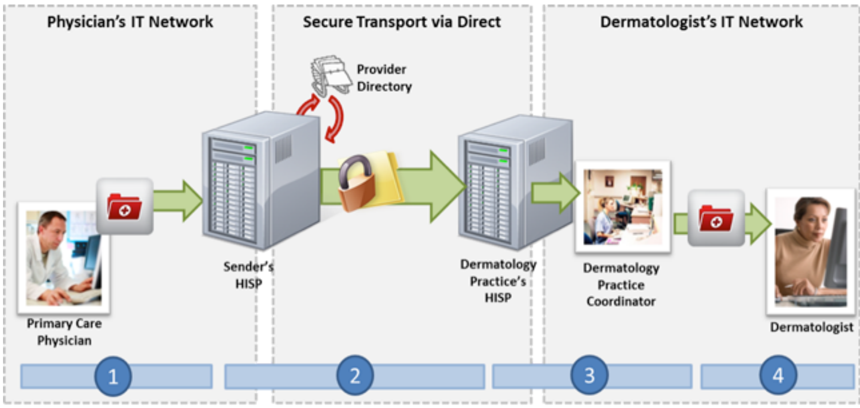


Figure 15: Transferring of Patient Record from one provider to another using DIRECT
(Framework, 2013b, Fig. 1)

standard services and content specifications. This software enables the secure exchange of interoperable health information among diverse organizations using a wide variety of technologies.

By implementing Aurion, organizations as a part of their health information exchange strategy gain the benefits of implementing nationally-recognized standards enabling data exchange with federal agencies as well as with numerous other health IT stakeholders. Aurion enables health professionals to request, send and receive medical records so critical information can follow patients as they navigate through the health care system. The software places relevant patient medical data at the doctors fingertips. It enhances security, promotes public health, and empowers patients to be more active and involved in their own care decisions.(Foundation, 2013)

11 SPARQL

S *SPARQL Protocol and RDF Query Language (SPARQL)* is a W3C recommend standard for querying RDF data. It allows one to query remote RDF resources, in a manner similar to the querying of databases using SQL. A SPARQL query is a set of graph patterns; any data triple matching these patterns is added to the query results. (Jarrar & Dikaiakos, 2008)

11.1 Why use it

By having the ability to query the RDF data, you will be able to enhance quality of clinical research and patient care by finding new insights in healthcare. Data is no good just on it's own, you need a tool available to easily analyze it and SPARQL is this tool.

11.2 Alternatives

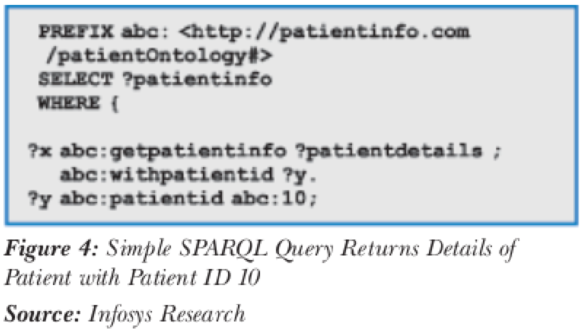
SquishQL - A simple RDF query language for beginners that allows for SQL syntax. Has very little adoption compared to the more powerful query language SPARQL (Mikhalenko, 2006)

```
prefix go: <http://purl.org/obo/owl/GO#>
prefix mesh: <http://purl.org/commons/record/mesh/>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix sc: <http://purl.org/science/owl/sciencecommons/>
prefix ro: <http://www.obofoundry.org/ro/ro.owl#>

SELECT DISTINCT ?gene ?process WHERE {
  graph <http://purl.org/commons/hcls/pubmesh>
  { ?pubmedrecord ?p mesh:D017966.
    ?article sc:identified_by_pmid ?pubmedrecord.
    ?generecord sc:describes_gene_or_gene_product_mentioned_by ?article.
  graph <http://purl.org/commons/hcls/goa>
  { ?protein rdfs:subClassOf ?res.
    ?res owl:onProperty ro:has_function.
    ?res owl:someValuesFrom ?res2.
    ?res2 owl:onProperty ro:realized_as.
    ?res2 owl:someValuesFrom ?process.
    graph <http://purl.org/commons/hcls/20070416/classrelations>
      {{ ?process <http://purl.org/obo/owl/obo#part_of> go:GO_0007166. ]
        union { ?process rdfs:subClassOf go:GO_0007166. }}
    ?protein rdfs:subClassOf ?parent.
    ?parent owl:equivalentClass ?res3.
    ?res3 owl:hasValue ?generecord. }}}
```

Fig. 1. Complex SPARQL query to retrieve all genes which are associated with b *CA1 Pyramidal Neurons* and the *signal transduction processes*

(a) SPARQL Query 1 (Stenzhorn et al., 2008)



```
PREFIX abc: <http://patientinfo.com
 /patientOntology#>
SELECT ?patientinfo
WHERE {

?x abc:getpatientinfo ?patientdetails ;
   abc:withpatientid ?y.
?y abc:patientid abc:10;
```

Figure 4: *Simple SPARQL Query Returns Details of Patient with Patient ID 10*

Source: *Infosys Research*

(b) SPARQL Query 2 (Parachuri & Majumdar, 2008)

Figure 16: SPARQL Query Examples

12 FMQL

FileMan Query Language (FMQL) is a Query Language that provides access to both FileMan data - a vital measurement of a patient - and the schema of that data - the type "Vital Measurement".

The three things that it addresses are:

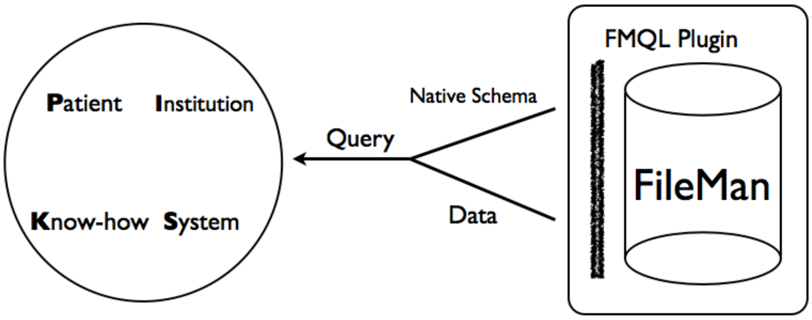


Figure 17: Querying data from Fileman (Caregraf, 2013, Fig. 1)

1. Identity: every entity and entity type in FileMan gets a unique identifier - a URI.
2. Data Formats: a consistent form of JSON, the web-friendly response format, for every type of data in the system as well as HTML for human readers and RDF for web-data practitioners.
3. Query Nuance: from the precise - SELECT - to the broader - DESCRIBE - or just COUNT, covers data hierarchies and graphical layouts, paging and filtering. (Caregraf, 2013)

13 Miscellaneous Technologies

There have been multiple efforts undertaken to unify healthcare semantics. Some of the other technologies and/or lexicons that are being used currently are summarized here:

13.1 First DataBank™

First Databank Inc.*, currently owned by Hearst Corporation, is a publisher of pharmaceutical industry market information and information technology. First Databank's proprietary knowledge base - *MedKnowledge* (First DataBank, 2013) provides prices, descriptions, and collateral clinical information on drugs approved by the US Food and Drug Administration (FDA), plus commonly used over-the-counter drugs, herbal remedies, nutraceuticals and dietary supplements.

*http://www.fdbhealth.com

References

3M (2010). 3m terminology consulting services. Online; Accessed 12-August-2013.
URL http://www.3mtcs.com/products/hdd

3M (2013). 3m health information systems. Online; Accessed 12-August-2013.
URL http://solutions.3m.com/wps/portal/3M/en_US/Health-Information-Systems/HIS/Products-and-Services/Products-List-A-Z/Healthcare-Data-Dictionary

A Y Wang, J. W. B. (2001). Mapping between SNOMED RT and clinical terms version 3: a key component of the SNOMED CT development process. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, (pp. 741–5).

Ansell, P. (2011). Model and prototype for querying multiple linked scientific datasets. *Future Generation Computer Systems*, *27*(3), 329–333.
URL http://www.sciencedirect.com/science/article/pii/S0167739X10001706

Argello, M., Des, J., Fernandez-Prieto, M. J., Perez, R., & Paniagua, H. (2009). Executing medical guidelines on the web: Towards next generation healthcare. In T. A. B. MSc, R. E. B. MSc, & M. M. AMBA, Miltos Petridis DipEng (Eds.) *Applications and Innovations in Intelligent Systems XVI*, (pp. 197–210). Springer London.
URL http://link.springer.com/chapter/10.1007/978-1-84882-215-3_15

Belleau, F., Nolin, M.-A., Tourigny, N., Rigault, P., & Morissette, J. (2008). Bio2RDF: towards a mashup to build bioinformatics knowledge systems. *Journal of Biomedical Informatics*, *41*(5), 706–716.
URL http://www.sciencedirect.com/science/article/pii/S1532046408000415

Bio2RDF Wiki (2013). Bio2RDF dataset provenance model. Online; Accessed 12-August-2013.
URL https://github.com/bio2rdf/bio2rdf-scripts/wiki/Bio2RDF-Dataset-Provenance

Bizer, C. (2006). D2r map - database to rdf mapping language and processor. Online; Accessed 13-August-2013.
URL http://wifo5-03.informatik.uni-mannheim.de/bizer/d2rmap/D2Rmap.htm

Borden, J. (2012). Xml in healthcare and the semantic web. Online; Accessed 13-August-2013.
URL www.jonathanborden-md.com/HealthcareSemWeb.ppt

Callahan, A., Cruz-Toledo, J., Ansell, P., & Dumontier, M. (2013a). Bio2RDF release 2: Improved coverage, interoperability and provenance of life science linked data. In P. Cimiano, O. Corcho, V. Presutti, L. Hollink, & S. Rudolph (Eds.) *The Semantic Web: Semantics and Big Data*, no. 7882 in Lecture Notes in Computer Science, (pp. 200–212). Springer Berlin Heidelberg.
URL http://link.springer.com/chapter/10.1007/978-3-642-38288-8_14

Callahan, A., Cruz-Toledo, J., & Dumontier, M. (2013b). Ontology-based querying with Bio2RDF's linked open data. *Journal of biomedical semantics*, *4 Suppl 1*, S1. PMID: 23735196.

Caregraf (2013). Fmql corner. Online; Accessed 13-August-2013.
URL http://vista.caregraf.info/fmql

Castro, L. J. G., McLaughlin, C., & Garcia, A. (2013). Biotea: RDFizing PubMed central in support for the paper as an interface to the web of data. *Journal of Biomedical Semantics*, *4*(Suppl 1), S5. PMID: 23734622.
URL http://www.jbiomedsem.com/content/4/S1/S5/abstract

DeGaspari, J. (2013). Open access for healthcare data dictionary. Online; Accessed 12-August-2013.
URL http://www.healthcare-informatics.com/news-item/open-access-healthcare-data-dictionary

DiSantostefano, J. (2009). International classification of diseases 10th revision (ICD-10). *The Journal for Nurse Practitioners*, *5*(1), 56–57.
URL http://www.sciencedirect.com/science/article/pii/S1555415508005230

Dolin, R. H., Mattison, J. E., Cohn, S., Campbell, K. E., Wiesenthal, A. M., Hochhalter, B., LaBerge, D., Barsoum, R., Shalaby, J., Abilla, A., Clements, R. J., Correia, C. M., Esteva, D., Fedack, J. M., Goldberg, B. J., Gopalarao, S., Hafeza, E., Hendler, P., Hernandez, E., Kamangar, R., Kahn, R. A., Kurtovich, G., Lazzareschi, G., Lee, M. H., Lee, T., Levy, D., Lukoff, J. Y., Lundberg, C., Madden, M. P., Ngo, T. L., Nguyen, B. T., Patel, N. P., Resneck, J., Ross, D. E., Schwarz, K. M., Selhorst, C. C., Snyder, A., Umarji, M. I., Vilner, M., Zer-Chen, R., & Zingo, C. (2004). Kaiser permanente's convergent medical terminology. *Studies in health technology and informatics*, *107*(Pt 1), 346–350. PMID: 15360832.

Dolin, R. H., Spackman, K., Abilla, A., Correia, C., Goldberg, B., Konicek, D., Lukoff, J., & Lundberg, C. B. (2001). The SNOMED RT procedure model. *Proceedings of the AMIA Symposium*, (pp. 139–143). PMID: 11825170 PMCID: PMC2243437.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2243437/

First DataBank (2013). FDB MedKnowledge. Online; Accessed 12-August-2013.
URL http://www.fdbhealth.com/fdb-medknowledge/

Foundation, T. A. (2013). What is aurion? Online; Accessed 20-August-2013.
URL http://aurionproject.org/about/what_is_aurion

Framework, S. (2013a). Nwhin. Online; Accessed 13-August-2013.
URL http://www.healthit.gov/policy-researchers-implementers/nationwide-health-information-network-nwhin

Framework, S. (2013b). Transferring records between providers. Online; Accessed 13-August-2013.
URL http://www.siframework.org/scenario_toc2.html

He, Z., Halper, M., Perl, Y., & Elhanan, G. (2012). Clinical clarity versus terminological order: the readiness of SNOMED CT concept descriptors for primary care. In *Proceedings of the 2nd international workshop on Managing interoperability and compleXity in health systems*, MIXHS '12, (p. 16). New York, NY, USA: ACM.
URL http://doi.acm.org/10.1145/2389672.2389674

HL7.org (2012). Introduction - fhir v0.01. Online; Accessed 13-August-2013.
URL http://www.hl7.org/implement/standards/fhir/v0.01/print-introduction.htm

HL7.org (2013). Introduction - fhir v0.11. Online; Accessed 13-August-2013.
URL http://hl7.org/implement/standards/fhir/overview.htm

ICD10 Instruction Manual - Volume 2 (2010). International classification of diseases and related health problems 10th revision instruction manual.
URL http://apps.who.int/classifications/icd10/browse/2010/en#/A00.0

Information, N. C. H., & Communications Alliance, I. (2013). Nwhin direct and nwhin exchange. Online; Accessed 20-August-2013.
URL http://nchica.org/HIT_HIE/NHIN2/NHINdirect.htm

Jarrar, M., & Dikaiakos, M. D. (2008). MashQL: a query-by-diagram topping SPARQL. In *Proceedings of the 2nd international workshop on Ontologies and information systems for the semantic web*, ONISW '08, (p. 8996). New York, NY, USA: ACM.
URL http://doi.acm.org/10.1145/1458484.1458499

Khan, W. A., Khattak, A. M., Lee, S., Hussain, M., Amin, B., & Latif, K. (2012). Achieving interoperability among healthcare standards: building semantic mappings at models level. In *Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication*, ICUIMC '12, (p. 101:1101:9). New York, NY, USA: ACM.
URL http://doi.acm.org/10.1145/2184751.2184868

Khare, R., An, Y., Li, J., Song, I.-Y., & Hu, X. (2012). Exploiting semantic structure for mapping user-specified form terms to SNOMED CT concepts. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, IHI '12, (p. 285294). New York, NY, USA: ACM.
URL http://doi.acm.org/10.1145/2110363.2110397

Kroth, P. J., Daneshvari, S., Harris, E. F., Vreeman, D. J., & Edgar, H. J. (2012). Using LOINC to link 10 terminology standards to one unified standard in a specialized domain. *Journal of Biomedical Informatics*, *45*(4), 674–682.
URL http://www.sciencedirect.com/science/article/pii/S1532046411001699

Lin, M., Vreeman, D., McDonald, C. J., & Huff, S. (2012). Auditing consistency and usefulness of LOINC use among three large institutions using version spaces for grouping LOINC codes. *Journal of Biomedical Informatics*, *45*(4), 658–666.
URL http://www.sciencedirect.com/science/article/pii/S1532046412000093

Lindberg, D. A., Humphreys, B. L., & McCray, A. T. (1993). The unified medical language system. *Methods of information in medicine*, *32*(4), 281–291. PMID: 8412823.

McDonald, C., Huff, S., Deckard, J., Mercer, K., Phillips, J., & Vreeman, D. (2013). Logical observation identifiers names and codes (loinc) users' guide. Online; Accessed 16-August-2013.
URL http://loinc.org/downloads/files/LOINCManual.pdf

Mikhalenko, P. (2006). Squishql: The simplest rdf query language. Online; Accessed 13-August-2013.
URL http://www.techrepublic.com/article/squishql-the-simplest-rdf-query-language

Munnecke, T. (2013). Rdf as universal healthcare language. Online; Accessed 13-August-2013.
URL http://www.osehra.org/blog/rdf-universal-healthcare-language-0

Murphy, K. (2012). What the 3m healthcare data dictionary means to interoperability. Online; Accessed 12-August-2013.
URL http://ehrintelligence.com/2012/05/22/what-the-healthcare-data-dictionary-means-to-interoperability/

of Health, U. D., & Services, H. (2013). Direct project. Online; Accessed 20-August-2013.
URL http://www.healthit.gov/policy-researchers-implementers/direct-project

Parachuri, D., & Majumdar, B. (2008). Role of sparql and semantic eii in patient record harmonization. *Improve Care Delivery with Technology*, (p. 25).

Pathak, J., Kiefer, R. C., & Chute, C. G. (2012). Applying linked data principles to represent patient's electronic health records at mayo clinic: a case report. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, IHI '12, (p. 455464). New York, NY, USA: ACM.
URL http://doi.acm.org/10.1145/2110363.2110415

Ryan, A. (2006). Towards semantic interoperability in healthcare: ontology mapping from SNOMED-CT to HL7 version 3. In *Proceedings of the second Australasian workshop on Advances in ontologies - Volume 72*, AOW '06, (p. 6974). Darlinghurst, Australia, Australia: Australian Computer Society, Inc.
URL http://dl.acm.org/citation.cfm?id=1273659.1273668

SNOMED Implementation Guide, I. H. T. S. D. O. (2011). SNOMED CT technical implementation guide.
URL http://www.ihtsdo.org/fileadmin/user_upload/doc/en_us/toc_tree.html?m=tig

SNOMED User Guide, I. H. T. S. D. O. (2011). SNOMED CT user guide. URL http://www.ihtsdo.org/fileadmin/user_upload/doc/en_us/toc_tree.html?m=ug

Spackman, K. A., Campbell, K. E., & CA?tA(C), R. A. (1997). SNOMED RT: a reference terminology for health care. *Proceedings of the AMIA Annual Fall Symposium*, (pp. 640–644). PMID: 9357704 PMCID: PMC2233423. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2233423/

Stenzhorn, H., Srinivas, K., Samwald, M., & Ruttenberg, A. (2008). Simplifying access to large-scale health care and life sciences datasets. In *The Semantic Web: Research and Applications*, (pp. 864–868). Springer.

Tauberer, J. (2008a). Comparing rdf to xml and rdbms. Online; Accessed 13-August-2013. URL http://www.rdfabout.com/comparisons.xpd

Tauberer, J. (2008b). What is rdf and what is it good for? Online; Accessed 13-August-2013. URL http://www.rdfabout.com/intro/

Vreeman, D. (2010). Loinc overview. Online; Accessed 16-August-2013. URL loinc.org/faq/getting-started/getting-started/

Vreeman, D. J., Chiaravalloti, M. T., Hook, J., & McDonald, C. J. (2012). Enabling international adoption of LOINC through translation. *Journal of Biomedical Informatics*, *45*(4), 667–673. URL http://www.sciencedirect.com/science/article/pii/S1532046412000068

Wei, D., Halper, M., & Elhanan, G. (2012). Using SNOMED semantic concept groupings to enhance semantic-type assignment consistency in the UMLS. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, IHI '12, (p. 825830). New York, NY, USA: ACM. URL http://doi.acm.org/10.1145/2110363.2110465