

# Standards and Initiatives Related to Semantic Infrastructure 2.0

The section includes information about CBIIT internal initiatives and other standards that affect Semantic Infrastructure 2.0.

- [Office of the National Coordinator and National Health Standards](#)
- [HL7](#)
- [CDISC](#)
- [ISO 21090 Harmonized Data Types for Information Exchange](#)
- [SAIF](#)
- [caBIG@ Semantic Infrastructure v2 - Initiatives](#)
- [Terminology and Data Types](#)
  - [OMG](#)
  - [World Wide Web Consortium \(W3C\) Recommendations](#)

## Office of the National Coordinator and National Health Standards

The US Office of the National Coordinator for Health Information Technology (ONC), is developing a set of [recommendations for a nationwide health information network \(NHIN\)](#). The NHIN is a set of standards, services and policies that enable secure health information exchange over the Internet. The NIH provides a foundation for exchange of health information across diverse entities, within communities and across the country, helping to achieve the goals of the Health Information Technology for Economic and Clinical Health (HITECH) Act.

Because of the convergence of federal agencies and local, regional and state-level Health Information Exchange Organizations (HIOs), the NHIN is setting a strong precedent for semantic interoperability in the United States. Many of the recommendations are likely to become part of future meaningful use specifications. With a growing number of organizations becoming part of the NHIN, it is evident that the more Semantic Infrastructure 2.0 aligns with NHIN, the less need there will be for multiple semantic interoperability strategies.

## HL7

The Health Level Seven International Standards Development Organization (HL7) is an international community, working together towards a common goal of improving patient care through technology. HL7 interoperability protocols include messaging standards, decision support standards, clinical document standards, Electronic Health Record (EHR) functional requirements, drug product labeling standards, and more. Many of these protocols are specifically called out in the meaningful use final rules for the HITECH Act.

In addition, HL7 defines Electronic Health Record (EHR) and Personal Health Record (PHR) functional requirements, which provide a reference list of functions that may be present in an EHR. The function list is described from a user perspective with the intent to enable consistent expression of system functionality. In 2009, the HL7 EHR-System Functional Model became an [internationally recognized ISO standard \(PDF of press release on the HL7 site\)](#), setting the stage to achieve common functionality of EHRs globally.

At the heart of many HL7 specifications is the [Reference Information Model \(on the HL7 site\)](#) (RIM). An object model created as part of the HL7 Version 3 methodology, the RIM is a large, pictorial representation of the HL7 clinical data (domains) and identifies the life cycle that a message or groups of related messages will carry. It is a shared model between all domains and, as such, is the model from which all domains create their messages. The RIM is an American National Standards Institute (ANSI) approved standard.

The Clinical Document Architecture, a V3-based standard, provides an exchange model for clinical documents (such as discharge summaries and progress notes) - and brings the healthcare industry closer to the realization of an electronic medical record. CDA leverages the use of XML, the HL7 Reference Information Model (RIM) and coded vocabularies. RIM is the basis for the Clinical Document Architecture standard being adopted internationally.

Common Message Element Types (CMETs) are standardized model fragments intended to be building blocks that individual content domains can "include" in their designs. These blocks reduce the effort to produce a domain-specific design and assure that similar content across multiple domains is consistently represented.

The Model Interchange Format (MIF) is a set of XML formats used to support the storage and exchange of HL7 version 3 artifacts as part of the HL7 Development Framework. It is the pre-publication format of HL7 v3 artifacts used by tooling. It is also the formal definition of the HL7 metamodel. The MIF can be transformed into derived forms such as UML XMI or OWL.

The HL7 Version 3 Development Framework (HDF) is a continuously evolving process that seeks to develop specifications that facilitate interoperability between healthcare systems. The HL7 RIM, vocabulary specifications, and model-driven process of analysis and design combine to make HL7 Version 3 a methodology for development of consensus-based standards for healthcare information system interoperability. The HDF is the most current edition of the HL7 V3 development methodology. The HDF documents the processes, tools, actors, rules, and artifacts relevant to development of all HL7 standard specifications, not just messaging.

The growing adoption of HL7 standards (for example, the Healthcare Information Technology Standards C32 (HITSP/C32) specification called out in the meaningful use final rule) throughout the world suggests that aligning Semantic Infrastructure 2.0 around these specifications will streamline attainment of Semantic Infrastructure 2.0 objectives.

## CDISC

Clinical Data Interchange Standards Consortium (CDISC) is a global, open, multidisciplinary, non-profit organization that has established standards to support the acquisition, exchange, submission and archive of clinical research data and metadata. The CDISC mission is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare. CDISC standards are vendor-neutral, platform-independent and freely available via the CDISC website.

CDISC defines standards to be used or reused in the definition, development, and execution of clinical trials as well as the submission of the clinical trial data to regulatory agencies (for example, FDA) for drug, device or procedure approval. CDISC is currently focused on constructing a Shared Health and Research Electronic Library (SHARE) and would like to use the NCI Knowledge Repository as the basis for SHARE.

The CDISC SHARE project is attempting to map CDISC standards (for example, Study Data Tabulation Model (SDTM) and Clinical Data Acquisition Standards Harmonization (CDASH) to Biomedical Research Integrated Domain Group (BRIDG) concepts and then have these concepts reused by CDISC participants (for example care delivery organizations, pharmaceutical companies), who create their own local concepts. BRIDG already incorporates a number of CDISC standards (for example, SDTM) and is working to harmonize other CDISC standards.

CDISC SHARE will contain the existing CDISC standards and will provide machine readable element (variables) within those standards. This will allow a range of applications used within other organizations (for example Clinical Research Organizations (CROs), pharmaceutical companies, and other agencies) to automatically access those definitions.

## ISO 21090 Harmonized Data Types for Information Exchange

The NCI CBIT Enterprise Conformance and Compliance Framework (ECCF) document requires that, in order to achieve Computable Semantic Interoperability (CSI), each attribute of a static information model should be bound to a robust data type specification.

The ISO 21090 standard provides a harmonized set of data type definitions used for representing and exchanging basic concepts that are commonly encountered in support of information exchange in a healthcare environment. These data type definitions represent a culmination of a large scale joint effort among the standard bodies such as HL7 and ISO, and have been reviewed by experts in the field. Additionally, this standard is currently being adopted by Canada Health Infoway and Australia's National E-Health Transition Authority (NEHTA). Hence, it is important for NCI CBIT to join the adoption activity at this time and provide leadership in the area.

Guidelines on use of the emerging ISO 21090 data type standard for projects funded by CBIT have been approved by the CBIT Enterprise Composite Architecture Team (ECAT).

## SAIF

SAIF (the HL7 Service-Aware Interoperability Framework) provides HL7 with an Interoperability Framework, that is, a set of elements including but not limited to constructs, best practices and processes that enable HL7 specifications to achieve cross-specification consistency and coherency irrespective of the chosen interoperability paradigm (messages, documents, or services).

SAIF consists of four core "frameworks":

1. Information (including RIM, data types, and vocabulary bindings),
2. Behavior (subsuming the existing Dynamic Model),
3. Enterprise Conformance and Compliance (including existing HL7 Implementation and Conformance standards), and
4. Governance.

SAIF should be regarded as an adjunct to any EAF (Enterprise Architecture Framework) which focuses on Working Interoperability (WI). It is a framework which brings from service-oriented architecture (SOA) practice, two critical constructs which significantly enhance the path to WI, that is:

1. Separation of concerns (static versus behavioral semantics)
2. Formal notion of contracts

SAIF is open enough to be implemented for all or one or a few types of interoperability paradigms (messages, documents, or services).

Intended use of SAIF include enterprise interoperability projects including those building large-scale integrated health IT infrastructures at the national level.

The main benefits that will be derived from developing specifications using SAIF include:

- Consistency of specifications
- Enhanced ability to manage loosely-coupled complex interactions between multiple trading partners
- Increased cross-organization reuse of architecture primitives (realizing the value proposition of common message element types (CMETs) and extending that proposition to include behavioral as well as static semantic constructs).

Implementations include:

- National Cancer Institute (NCI)
- Canada Health Infoway (CHI)
- Open Health Tools (OHT) Architecture Project team

## caBIG® Semantic Infrastructure v2 - Initiatives

The initiatives proposed below are intended to support production operation of caBIG® semantics while evolution toward a national-scale capability begins. In addition to its legacy obligations, caBIG® and caGrid are expected to inform the design and initial implementations supporting personalized medicine (BIGHealth), improvement of health care quality, value and affordability through the National Health Informatics Network (NHIN), and more immediately support cancer-oriented initiatives such as the cancer aware extension to the national standard electronic medical record (EMR), and a number of other rapidly emerging initiatives that will require capabilities well beyond our traditional focus on cancer and on static data semantics.

The need for all semantic metadata to be formally recorded in a single central repository would limit or preclude application of the semantic infrastructure to very large, diverse communities such as national health care. Distributed, federated metadata resources will clearly be required. Also, support for behavioral semantics describing the business context of messages and services contracts, metadata creation via processing of line-of-business artifacts, support for ontology based semantics and support for semantically aware service oriented architecture (SOA) capabilities and some form of grid-to-grid semantic interoperability all appear to be capabilities that the emerging semantic infrastructure will have to provide in order to scale to national levels.

The caBIG® Semantic Infrastructure and its operational model will be extended to provide integrated support for:

- Initiative 1. Distributed, federated metadata repositories and model repositories and operations
- Initiative 2. Automated generation of metadata from line-of-business artifacts
- Initiative 3. Rules management and contracts support (behavioral semantics)
- Initiative 4. Semantics support for W3C service oriented architecture resources
- Initiative 5. HL7 Common Terminology Services 2 (CTS 2)/ Object Management Group Model Interchange Format (OMG and HL7 Model Interchange Format (MIF)) compliant federated terminology services
- Initiative 6. Controlled biomedical terminology, ontology and metadata content
- Initiative 7. Assessment of semantic unification of compositional and derivational models

## Terminology and Data Types

Terminology and data type standards referred to during requirements gathering processes have included:

**SNOMED CT** (Systematized Nomenclature of Medicine--Clinical Terms), a comprehensive clinical terminology, originally created by the College of American Pathologists (CAP) and, as of April 2007, owned, maintained, and distributed by the International Health Terminology Standards Development Organisation (IHTSDO), a non-for-profit association in Denmark.

SNOMED CT is one of a suite of standards designated for use in U.S. Federal Government systems for the electronic exchange of clinical health information and is also a required standard in interoperability specifications of the U.S. Healthcare Information Technology Standards Panel. SNOMED CT is also being implemented internationally as a standard within other IHTSDO member countries.

**UCUM** (Unified Code for Units of Measure), a code system intended to include all units of measures being used in contemporary international science, engineering, and business. The purpose is to facilitate unambiguous electronic communication of quantities together with their units. The Unified Code for Units of Measure is inspired by and heavily based on ISO 2955-1983, ANSI X3.50-1986, and HL7 extensions called "ISO+".

WHO Drug Dictionary, an international classification of medicines created by the WHO Programme for International Drug Monitoring and managed by the Uppsala Monitoring Centre. WHO Drug Dictionary is used by pharmaceutical companies, clinical trial organizations and drug regulatory authorities for identifying drug names in spontaneous adverse drug reaction (ADR) reporting, pharmacovigilance, and clinical trials.

MedDRA (Medical Dictionary for Regulatory Activities), a clinically validated international medical terminology used by regulatory authorities and the regulated biopharmaceutical industry throughout the entire regulatory process, from pre-marketing to post-marketing activities, and for data entry, retrieval, evaluation, and presentation. MEDDRA is also the adverse event classification dictionary endorsed by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). MedDRA is used in the US, European Union, and Japan. Its use is currently mandated in Europe and Japan for safety reporting.

The US Food and Drug Administration (FDA) has committed to keeping current on MedDRA, and it has become the standard for adverse event reporting in the United States.

RxNorm terminology released by the National Library of Medicine (NLM). RxNorm provides normalized names for clinical drugs and links its names to many of the drug vocabularies commonly used in pharmacy management and drug interaction software, including those of First Databank, Micromedex, MediSpan, Gold Standard Alchemy, and Multum. By providing links between these vocabularies, RxNorm can mediate messages between systems not using the same software and vocabulary.

RxNorm now includes the National Drug File - Reference Terminology (NDF-RT) from the Veterans Health Administration, which allows for classification of the RxNorm terms into their chemical and physiological effect categories. NDF-RT is a terminology used to code clinical drug properties, including mechanism of action, physiological effect, and therapeutic category.

Unique Ingredient Identifier (UNII). The overall purpose of the joint FDA and United States Pharmacopeia (USP) Substance Registration System (SRS) is to support health information technology initiatives by generating unique ingredient identifiers (UNIIs) for substances in drugs, biologics, foods, and devices.

The UNII may be found in:

- NLM's Unified Medical Language System (UMLS)
- National Cancer Institute Enterprise Vocabulary Services
- USP Dictionary of United States Adopted Name (USAN) and International Drug Names (future)
- FDA Data Standards Council website
- VA National Drug File Reference Terminology (NDF-RT)
- FDA Inactive Ingredient Query Application

NDC (National Drug Code), a unique product identifier used in the United States for drugs intended for human use. The Drug Listing Act of 1972 requires registered drug establishments to provide the FDA with a current list of all drugs manufactured, prepared, propagated, compounded, or processed by that establishment for commercial distribution. Drug products are identified and reported using the NDC.

FDA MedWatch, the Food and Drug Administration reporting system for adverse events, founded in 1993. An adverse event is any undesirable experience associated with the use of a medical product. The MedWatch system collects reports of adverse reactions and quality problems, primarily with drugs and medical devices, but also for other FDA-regulated products (for example, dietary supplements, cosmetics, medical foods, and infant formulas).

NLM DailyMed, which provides high quality information about marketed drugs. Drug labeling and other information in the Structured Product Labeling (SPL) is what has been most recently submitted by drug companies to the Food and Drug Administration (FDA) as drug listing information.

## OMG

Object Management Group (OMG) is a consortium, originally aimed at setting standards for distributed object-oriented systems, and is now focused on modeling (programs, systems and business processes) and model-based standards.

Model-driven architecture (MDA) is a software design approach for the development of software systems. It provides a set of guidelines for the structuring of specifications, which are expressed as models. Model-driven architecture is a kind of domain engineering, and supports model-driven engineering of software systems.

NCI chose to use MDA because it is a set of standards that have worked well in other areas. The advantage to this approach is that MDA allowed NCI to use available tools to automatically generate some of the code, while giving the flexibility to tailor the code to specific needs. In the future, when the system's requirements change, NCI can update models and quickly regenerate the appropriate code.

MDA is related to multiple standards, including Unified Modeling Language (UML), Meta-Object Facility (MOF), and XML Metadata Interchange (XMI). Of particular importance to model-driven architecture is the notion of model transformation. A specific standard language for model transformation has been defined by OMG called Query View Transformation (QVT).

The specifications produced by HL7 target multiple facets of the interoperability challenge, include specification of information models, data types, and vocabularies; messaging, clinical documents, and context management standards; and implementation technology, profile, and conformance specifications. Despite the diversity in depth and scope of HL7 specifications a common thread is the use of a model driven methodology and the derivation of specifications and interim work products from a common set of reference models. The models used in the Healthcare Development Framework (HDF) development methodology, use the Unified Modeling Language (UML) as the preferred syntax. The HDF closely aligns the underlying meta model governing well-formed HL7 models with the meta model of UML and applies the model driven process to all of the technical specifications of HL7, not just messages.

The Ontology Definition MetaModel (ODM) is an Object Management Group (OMG) specification to make the concepts of model-driven architecture applicable to the engineering of ontologies. It links Common Logic (CL), the Web Ontology Language (OWL), and the Resource Description Framework (RDF).

The Object Constraint Language (OCL) is a declarative language for describing rules that apply to any Meta-Object Facility (MOF) Object Management Group (OMG) meta-model, including UML. The Object Constraint Language is a precise text language that provides constraint and object query expressions on any MOF model or meta-model that cannot otherwise be expressed by diagrammatic notation. OCL is a key component of the OMG standard recommendation for transforming models, the Query View Transformation (QVT) specification.

HL7 Static models are not strictly UML diagrams. Although they have some similarity in presentation, they are actually statements of constraints against a formal UML model, which is the RIM. In a static model, each of the "clones" is actually a statement of constraints on a properly defined RIM class (properly defined in the RIM sense).

OCL constraints express additional rules that apply to the clone. In HL7 static models, OCL statements are applied to the clone - in other words, OCL statements are applied to the RIM class only when it is subject to the constraints defined in the clone. In order to use OCL with HL7 static models, a new binding mechanism must be described by HL7. This is an expected extension to the OCL language.

The Common Terminology Services 2.0 (CTS2) Specification will be an extension the HL7 CTS Specification.

## World Wide Web Consortium (W3C) Recommendations



### Additional Information

A W3C recommendation is the final stage of a ratification process of the World Wide Web Consortium (W3C) working group concerning the standard. This designation signifies that a document has been subjected to a public and W3C-member organization's review. It aims to standardise the Web technology. It is the equivalent of a published standard in many other industries.

XML Schema (XSD), published as a W3C recommendation in May 2001, is one of several XML schema languages. Like all XML schema languages, XSD can be used to express a set of rules to which an XML document must conform in order to be considered 'valid' according to that schema. However, unlike most other schema languages, XSD was also designed with the intent that determination of a document's validity would produce a collection of information adhering to specific data types.

The Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model. It has come to be used as a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax formats.

RDF Schema (variously abbreviated as RDFS, RDF(S), RDF-S, or RDF/S) is an extensible knowledge representation language, providing basic elements for the description of ontologies, otherwise called Resource Description Framework (RDF) vocabularies, intended to structure RDF resources.

The Web Ontology Language (OWL) is a family of knowledge representation languages for authoring ontologies. The languages are characterized by formal semantics and RDF and XML-based serializations for the Semantic Web. OWL is endorsed by the World Wide Web Consortium (W3C) and has attracted academic, medical and commercial interest.

SPARQL (pronounced "sparkle") is an RDF query language; its name is a recursive acronym that stands for SPARQL Protocol and RDF Query Language. It was standardized by the RDF Data Access Working Group (DAWG) of the World Wide Web Consortium, and is considered a key semantic web technology.

SWRL (Semantic Web Rule Language) is a proposal for a Semantic Web rules-language, combining sublanguages of the OWL Web Ontology Language (OWL DL and Lite) with those of the Rule Markup Language (Unary/Binary Datalog).



### Help Downloading Files

For help accessing PDF, audio, video, and compressed files on this wiki, go to [Help Downloading Files](#).