VA Enterprise Design Patterns
Interoperability and Data Sharing

Data Integration with Semantic Web Technologies (SWT)

OFFICE OF TECHNOLOGY STRATEGIES (TS)
OFFICE OF INFORMATION AND TECHNOLOGY (OI&T)

VERSION 1.0
DATE ISSUED: OCTOBER 2016

APPROVAL COORDINATION

Gary E. Marshall 137891

Digitally signed by Gary E. Marshall 137891
DN: dc=gov, dc=va, o=internal, ou=people, 0.9.2342.19200300.100.1.1=gary.marshall@va.gov, cn=Gary E. Marshall 137891

Date: 2016.11.09 14:02:01 -05'00'

Gary Marshall
Director, Technology Strategies, ASD

PAUL A. TIBBITS 116858

Digitally signed by PAUL A. TIBBITS 116858
DN: dc=gov, dc=va, o=internal, ou=people, 0.9.2342.19200300.100.1.1=paul.tibbits@va.gov, cn=PAUL A. TIBBITS 116858
Reason: I am approving this document.
Date: 2016.11.15 17:39:00 -05'00'

Paul A. Tibbits, M.D.
DCIO Architecture, Strategy, and Design

REVISION HISTORY

Version	Date	Approver	Notes
1.0	10/20/2016	Aaron Ibarra	Final version for TS leadership approval and signature, including all applicable updates addressing stakeholder feedback and Section 508 Compliance.

CONTENTS

1 Introduction....4
1.1 Business Problem....4
1.2 Business Need....5
1.3 Business Case....5
2 Current Capabilities and Limitations....7
2.1 Disconnected, Discordant, and Deficient Technology....7
2.2 Data Integration in the SOA Framework....9
2.3 Vista Evolution....10
2.4 VINCI....10
3 Future Capabilities....11
3.1 PROVISO....11
3.2 Principles of Semantic Applications Development....11
3.3 Alignment to the One-VA Technical Reference Model (TRM)....15
3.4 Alignment to Veteran-Focused Integration Process (VIP)....15
4 Use Cases....16
4.1 Use Case: Semantic Data Integration....16
4.1.1 Purpose....16
4.1.2 Assumptions....17
4.1.3 Use Case Decription....18
Appendix A. Scope....19
Appendix B. SWT Design Principles....21
Appendix C. Definitions....25
Appendix D. Acronyms....27
Appendix E. References, Standards, and Policies....29
Appendix F. Standard SWT and Tools....30
Appendix G. Current SWT Entries in TRM....34

Table 1: Business Benefits....6

Figure 1: Discordant Schemas and Silos....8
Figure 2: Cross-linked Semantic Models....13
Figure 3: Notional Use of a Semantic Data Layer Based on VA Data Instances....16

QUICK JUMP

Select an icon to skip to a section.

Current Capabilities

Future Capabilities

Use Cases

One-VA Technical Reference Model

The Veteran-Focused Integration Process

Enterprise Design Pattern Scope

1 INTRODUCTION

The Department of Veterans Affairs (VA) has established standards for data exchange, but these standards do not account for machine-readable linkage of datasets to support data integration and analytics. As VA integrates dispersed datasets with standardized access protocols, VA needs to adopt enterprise-level data approaches and standards as a foundation for advanced IT capabilities. VA's digital transformation also includes implementation of cloud- based services that leverage industry standards for interoperable data exchange.

VA must resolve inconsistencies among enterprise datasets and support its evolving "big data" needs in order to achieve its digital transformation goals. Semantic Web Technologies (SWT) provides dynamic data processing via machine-readable, semantic linkages. Basic design principles and methods of SWT are described in Section 3.2 (Principles of Semantic Applications Development) and Appendix B (SWT Design Principles).

1.1 Business Problem

VA faces systemic problems in managing and sharing diverse datasets among dispersed systems across each Line of Business (LOB).[1] The following integration challenges hinder data accessibility and interoperability across VA:

[1] FY 2013-2015 Enterprise Roadmap

- Disconnected technology – isolated or redundant datasets
- Discordant technology – syntactic and/or semantic inconsistencies
- Deficient technology – disjointed, non-linkable datasets

Currently, the Office of Information and Technology (OI&T) lacks standards regarding the explicit linkage of datasets that enable robust, dynamic data exchange and integration. This hinders VA's ability to achieve the dynamic integration of data located in varied sources via explicit, standardized, machine-readable linkages. The standards conveyed herein support adoption of SWT "building blocks" in accordance with the Enterprise Service-Oriented Architecture (SOA) Enterprise Design Pattern (EDP).

1.2 Business Need

Adoption of SWT standards will improve interoperability among VA's different systems supporting diverse business needs. Semantic linkage of datasets and machine-readable data processing may secure flexible and dynamic data integration for VA (see Section 3 Future Capabilities, Section 3.2 Principles of Semantic Applications Development, and Appendix B).

OI&T supports business needs by establishing official standards for interoperable data exchanges to overcome the following organizational barriers:

- The patchwork of data capabilities, including legacy data management systems, does not adequately meet the emerging data needs of its changing operations.
- Evolving data needs are not adequately addressed in the fragmented landscape of VA data capabilities, as analysts must resort to ad-hoc, labor-intensive processes.
- The "little data" capabilities within VA offices are not plugged into enterprise-level "big data" capabilities.

1.3 Business Case

VA will benefit from enterprise-wide adoption of machine-readable data processing, furnished by SWT.[2] SWT enables:

- Comprehensive enterprise data management beyond conventional predefined/predetermined schemas.
- Reusable, extensible, standardized, and interoperable semantic data models and tools.
- Flexible, consistent, and machine-intelligent data processing that leverages machine-readable semantic linkages among (local or remote) datasets.

[2] Semantic Web, http://www.w3.org/standards/semanticweb/

- Dynamic exchange and integration of enterprise datasets.
- Enhanced analytic power, knowledge discovery, insight, and decision support.
- Synergy with multiple technologies and data formats (e.g., relational databases, non-relational databases).
- Improved coordination, collaboration, and cooperation among VA stakeholders and external partners.

In addition to extending the framework of data processing, storage, and analytics operations described in the Hybrid Data Access (HDA), Data Analytics, and Data Storage EDPs.[3] SWT enhances interoperable data sharing and flexible integration, in harmony with SOA design principles (per the Enterprise SOA EDP) and scalable cloud-based services (per the Cloud Computing Architecture EDP).[4]

The machine-readable configuration of data in SWT models empowers access, extraction, and exchange of datasets (from diverse data stores) with flexible, elastic, and dynamic queries that produce value-added knowledge discovery decision support. Data integration from "reason-based" tools augments research and business processes, enabling efficient data sharing and enriched meta-analysis to increase analytic power, validity and generalizability. Operationalization of SWT can also support progress toward automation, machine learning, predictive analytics, and "big data" capabilities.

TABLE 1: BUSINESS BENEFITS

Business Benefits	Description
Data management	Data unification
Linked data	Machine-intelligent data processing
Data integration	Enhanced analytics, discovery, and decision support
Technology synergy	Automation, machine learning, predictive analytics, "big data"
Coordination	VA stakeholders and partners leveraging linked datasets based on common metadata standards

This EDP supports SWT-driven data integration and interoperability that can be applied throughout VA, including:

- Promoting application of SWT across all LOB data integration use cases.
- Aligning SWT applications with organizational goals and needs.

[3] Enterprise Data Analytics Enterprise Design Pattern, Hybrid Data Access Enterprise Design Pattern, Data Storage Enterprise Design Pattern

[4] Enterprise SOA Design Pattern, Cloud Computing Architecture Enterprise Design Pattern

- Reviewing existing VA data sources, selecting data processing strategies, and aligning semantic implementations with data sources.
    - Enterprise-level production implementations would most likely employ Commercial Off-The-Shelf (COTS) tools.
    - Local, smaller-scale ventures might use semantic instruments developed from open source standards (see Appendix F. Standard SWT and Tools).
- Implementing pilot SWT projects building upon a foundation of prioritized content, followed by extension of capacity to meet needs.

2 CURRENT CAPABILITIES AND LIMITATIONS

VA administrations comprise numerous lines of business, operating units, information systems, and external mission partners. VHA, for example, manages the largest integrated healthcare network in the United States, caring for Veterans, employing medical personnel, and operating hospitals, clinics, and nursing homes using the Veterans Information Systems and Technology Architecture (VistA). Currently, VHA is investigating SWT as a potential element to help reach their project aims.

Recent legislation has added greater complexity to the VHA's healthcare delivery, requiring advancements in health information interoperability, coordination, and sharing among a broader and more integrated community of providers.[5] Significant motivations compel all VA administrations to deploy more harmonized business processes and IT systems. VA faces challenges attaining standardized integration of its datasets internal and external to VA.

2.1 Disconnected, Discordant, and Deficient Technology

Modern data analyses increasingly draw on dispersed and heterogeneous sources, such as electronic health records (EHR), genomic or epidemiologic datasets, and "big data" resources. The Electronic Health Management Platform (eHMP), for example, is driving rapid growth in the volume and complexity of data that VA generates and uses. As such, VA must leverage new technologies for extracting information from free text, process genomic data and images, and analyze data from personal health monitoring devices.

Relational databases are designed to report answers to predetermined questions, according to anticipated user needs. In this approach, data is pre-categorized at the point of entry, which affects data quality, retrieval, and analysis. In addition, the majority of VA datasets persist as functionally disconnected and autonomous silos. This hinders the reuse, exchange, and

[5] VISTA EVOLUTION (VE), http://vaww.ea.oit.va.gov/health-care-transformation/vista- evolution/

integration of multisource data, resulting in the employment of non-standard, inconsistent, and locally-defined schemas, datasets, and terminologies. In relational schemas, the meaning of data (expressed by relations) is implied only by the structure of the table (rather than explicitly asserted) and is hidden from machine-readable processing (Figure 1). As a result, incompatible data structures cannot communicate effectively (see Appendix B).

Patient Table

Patient ID	Name	D.o.B.	Gender
1340	Jeff T. Bell	4-Jul-93	Male
1783	David Ham	8-Feb-87	Male
1974	Jane Doe	14-Feb-69	Female
2108	Frank J. Ford	12-Mar-62	Male
2588	Rachel Day	8-Feb-87	Female

Health Consumer Table

P_ID	F_Name	L_Name	Date of Birth	Sex
P5391	Jane	Doe	2/14/1969	F
P0437	Joe	Smith	4/16/1988	M
P4492	David	Winter	9/22/1964	M
P2644	Lisa	Jones	6/11/1972	F
P7046	Frank	Ford	3/12/1962	M

FIGURE 1: DISCORDANT SCHEMAS AND SILOS

Parallel and incompatible work can result from autonomous groups unable to leverage the knowledge assets of others. Substantial effort (e.g., Extract Transform Load (ETL), terminology harmonization/standardization, warehousing, etc.) is required to align and share data, even when performed between systems on a regular basis. Traditional data integration is difficult to modify and interoperate with other resources.

Inadequately coded terminologies and standards hinder VHA's ability to capture clinical data or measure outcomes of care, as well as hamper data sharing, aggregation, and analysis within VistA. Only capturing this small amount of machine-readable information hampers VHA's ability to adequately examine its clinical, operational, and financial performance and to exchange data among VA facilities or with third parties. Currently, Department of Defense (DoD) and VA seek to seamlessly integrate EHR data, but share only a limited amount of standardized computable data. VHA cannot also readily process electronic records that conform to industry standards.

Image information of third party clinical reports are also not well integrated into VistA; including little (if any) searchable, computable metadata about images, which hampers retrieval and

analysis. Data sharing is also hindered by the complexity of VA's IT infrastructure: multiple access layers, multiple software technologies, and multiple functional components. Managing federated health records across VistA by point–to-point applications requires complex integration schemes with DoD, Federal, and industry partners.

2.2 Data Integration in the SOA Framework

VA has launched a multiyear effort to transform legacy IT systems into standardized Enterprise Shared Services (ESS) within the SOA infrastructure. Development of these capabilities is proceeding through various initiatives,[6] some of which directly promote data integration: VistA Evolution, Customer Data Integration (CDI), and Open Data.

The shared ESS infrastructure unifies enterprise capabilities and information. CDI addresses integration across overlapping databases, enterprise-wide processes and services that manage data as an asset, and standards for data representation. The unified environment implements a consolidated data layer for on-demand data access and sharing. The shared environment establishes an infrastructure for advanced analytics capabilities (predictive analytics, context sensing, machine learning, etc.). HDA, Data Analytics, and Data Storage EDPs[7] establish capability frameworks to achieve consistent management of data across VA LOBs.

Two examples of VA implementation strategies, in healthcare delivery and healthcare research, illustrate common challenges with current approaches to enterprise data integration, and are referenced in the Use Case described in Section 4:

- VistA[8]
- VA Informatics and Computing Infrastructure (VINCI)[9]

The Use Case in Section 4 demonstrates how loosely coupled services, based on the Future Capabilities in Section 3, enhance data integration through SWT standards. SOA principles by themselves do not support machine-readable data processing, and data integration remains a challenge as data resources evolve (Appendix B).

---

[6] VA_Enterprise-Roadmap_2_FINAL_2014 0409

[7] Enterprise Data Analytics Enterprise Design Pattern, Hybrid Data Access Enterprise Design Pattern, Data Storage Enterprise Design Pattern

[8] eHEALTH VistA, http://www.ehealth.va.gov/vista.asp

[9] Health Services Research & Development VA Informatics and Computing Infrastructure (VINCI), http://www.hsrd.research.va.gov/for_researchers/vinci/

2.3 Vista Evolution

VistA Evolution (VE) will link resources and observe SOA principles to achieve interoperability and deliver an integrated health record that supports the continuum of care.[10] VE drives data standardization by applying national coding standards for sharing and interpreting health information. Implementing SOA design principles improves VistA from a decentralized legacy systems into a single set of reusable, shared services and data architecture. However, SOA design principles do not specify comprehensive, automated, and machine-readable data integration through direct linkage among datasets (see Future Capabilities).

2.4 VINCI

VA Informatics and Computing Infrastructure (VINCI) – the result of a partnership among VHA, VA Office of Informatics Analytics (OIA), and the VA Office of Information and Technology (OI&T) Business Intelligence Service Line (BISL) – provides a valuable, central, and secure computing environment inside the VA intranet with numerous enterprise-level healthcare datasets, analytic tools, data processing applications, and associated services for research studies, reporting, data analysis, epidemiology, decision support, and business intelligence.[11] VINCI has a common access point from anywhere in the VA network and hosts standardized, authoritative data available through the Corporate Data Warehouse (CDW), including extractions from VistA and DoD sources, as well as unique data sources.

The CDW is supported and supplied (daily) with data from Regional Data Warehouses (RDW), and this data includes key data collections extracted from VistA. Data gathered, cleaned, uploaded, and integrated (ETL) from multiple RDW sources are organized in joined relational tables and aggregated by subject into groups (e.g., data marts) in the central CDW repository. Data managers provide extracts from SQL and Statistical Analysis System (SAS) datasets in the CDW and assist researchers with identifying CDW data that meets research project needs.

Data resources supplied by VINCI include: CDW extractions from VistA, Med Statistical Analysis System (MedSAS), Registries, DoD, Centers for Medicare and Medicaid Services (CMS), State claims, Radiology notes, and extracted free texts. VINCI is also partnering with research groups, such as the Consortium for Health Informatics Research (CHIR), to bring new types of data and applications into VINCI. Coupled with these data resources, VINCI employs a cohort data extraction selector, rules engine, Natural Language Processing (NLP), annotation, custom web-services applications, geospatial services, and programs that support machine learning and prediction.

---

[10] VistA 4 Product Architecture

[11] VA Informatics and Computing Infrastructure (VINCI), http://vaww.virec.research.va.gov/VINCI/Overview.htm

Integration of data from the regional data warehouses and from other sources into the CDW provides a single query engine, improved performance, consistent codes, and a consistent view of data across the enterprise infrastructure. VINCI currently does not ensure machine-readable data processing using SWT standards. Access to data prescribed by VINCI managers (currently) precludes more flexible exploration of comprehensive, dynamically-generated, and semantically integrated datasets by research investigators (see Section 3).

3 FUTURE CAPABILITIES

3.1 PROVISO

This section examines key attributes and principles regarding the use of SWT standards in advanced data integration capabilities as part of VA's digital transformation. The following planning assumptions apply:

1. Well-formed ontologies, consisting of adequate and suitable content pertaining to domains of interest, are available for public consumption.
    a. Well-formed ontologies consist of sufficient content (Classes, Properties, etc.) and logically sound organization to adequately describe a domain of interest.
2. Data sources of interest, owned by VA, DoD, Federal, or Non-governmental organization (NGO) partners, are known, as well as programmatically discoverable.
    a. Data governance, provenance, quality, and maintenance are correct and compliant with VA recognized standards.
    b. Formats (SQL, NoSQL, image, free text, etc.), datasets/types, and access protocols of data sources are also known.
3. The development, implementation, and deployment of semantic applications (prototype to production) employ iterative assembly, testing, and improvement within a test sandbox environment.
4. All software tools (SWT and otherwise) comply with the Technical Reference Model (TRM) as discussed in Section 3.3.

3.2 Principles of Semantic Applications Development

The key SWT attributes for enterprise solutions are as follows:

1. Ontology – semantic graph data model containing machine-readable statements representing objects and relations (as Classes, Instances, Properties) in a domain of interest (e.g., Immunology)

2. Ontology Editor – software application that supports addition, deletion, or modification of structured content (Classes, Instances, Properties, Restrictions, Rules, etc.) in semantic data graphs
3. Triplestore – NoSQL semantic graph database used to house and consume deployed ontologies that drive semantic applications
4. Reasoner – software application able to infer/compute logical consequences (based on first-order predicate logic) from statements asserted in an ontology
5. Broker – software application that extracts, transforms, and loads data from non- semantic resources into Uniform Resource Identifier (URI)-based RDF (Resource Description Framework)[12] stores or executes queries of a semantic data model(s) to extract data from non-semantic target data stores

Best practice recommendations consider several key issues for attention in advance of application development, including:

1. Project operational goal(s)
2. Knowledge domain(s) of interest
3. Ontology supplies
4. Triplestore operation
5. Data processing strategy

Project operational goals support the identification of the purpose, function, and technical requirements of a semantic application. Project goals might seek, for example, to annotate/tag image files (and picture content) with semantically described objects (concepts/terms) to empower comprehensive, machine-intelligent compilation (integration) and analysis of instance cases (extracted from diverse and dispersed stores) matching specific criteria for evaluations of efficacy, disparity, or quality of clinical treatment regimes.

Another project may seek integration (by distributed query to VistA, DoD, and CDW – with disparate data models) of patient medical and demographic data (e.g., signs, symptoms, tests, diagnosis, treatment, outcome, etc.) for the tracking/monitoring of cohort safety surveillance. Yet another project may seek dynamic and flexible extraction of mental health, risk factor, and public health data from diverse and dispersed resources (clinical notes, reference databases, project data, etc.) to investigate and evaluate the value of alternative predictive models of behavioral symptoms or crises. These sorts of inquiries might also investigate such issues across the context of population, time, location, gender, and so on.

[12] RDF Current Status, http://www.w3.org/standards/techs/rdf#w3c_all

Development of semantic engines originates with specification of the sundry knowledge domains of interest (to project goals) – e.g., geolocation, time, identity/family relations, demographics, medical history/diagnosis, signs and symptoms, clinical tests, pharmacology, pathology, genetics, etc. Designation of applicable knowledge domains guides specification of relevant ontologies (describing these diverse subjects) that require consideration for use in the semantic application. Ontologies (semantic data models/graphs) that drives semantic applications – with meaningful representations of objects, data, and relations in a subject domain of interest (see Appendix B. SWT Design Principles).

A semantic data model embodies a machine-readable description of knowledge content based on the underlying metadata. Entities encoded in semantic models typically include Classes, or categories of similar objects (such as "Patient"), and Instances (such as unique individuals). Cross-linkage via relations within (and between) semantic models (as modules) builds integrated knowledge graphs across the context of interrelated knowledge domains (Figure 2).

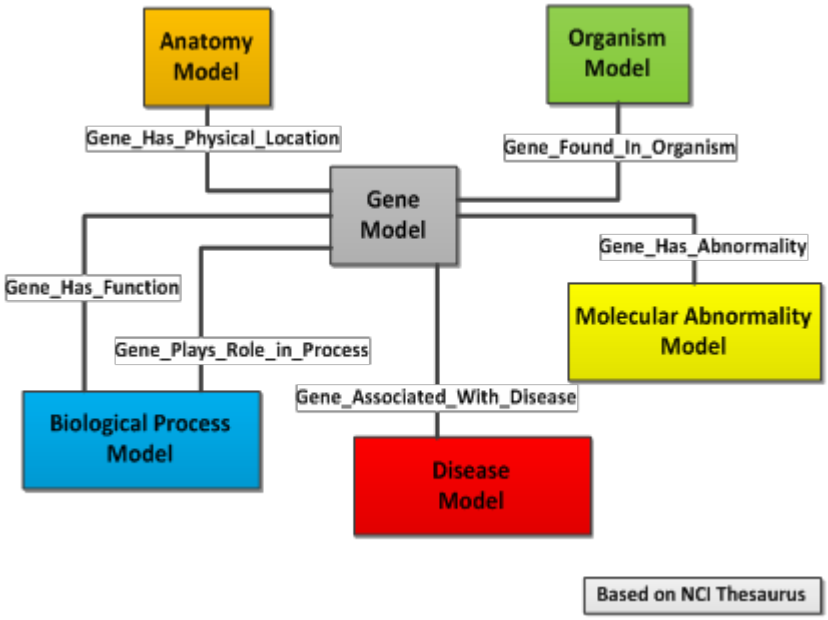

FIGURE 2: CROSS-LINKED SEMANTIC MODELS

Fortunately, published ontologies may be freely obtained from public access websites for use in development of semantic assets. Quality standard models of sundry disciplines in the biomedical domain, for example, may be acquired (downloaded) for use from sites such as The Open Biomedical Ontologies (OBO) Foundry,[13] BioPortal,[14] or Unified Medical Language System

[13] The OBO Foundry, http://obofoundry.org/
[14] BioPortal, http://bioportal.bioontology.org/

(UMLS).[15] These sites contribute standards such as SNOMED-CT[16] (medical, clinical), National Cancer Institute (NCI) Thesaurus[17] (cancer, preclinical, clinical), Logical Observation Identifiers Names and Codes (LOINC)[18] (medical laboratory), RxNorm[19] (pharmaceutical), Gene Ontology[20] (genetics), and other knowledge graphs. Similarly, models describing other common domains of interest (e.g., publication, social network, geography, government, etc.) may be obtained from other web resources, such as Linked Data[21] from the Open Data initiative. Use of specific published ontologies in the development of semantic instruments for VA IT capabilities will align with the TRM Decision Matrix.

Often, published ontologies serve as a generic "knowledge framework" or reference standard (template, skeleton, or straw man) consisting of major Classes and Relations of entities in a discipline, to which additional local content data may be added (as Classes, Instances, Properties, Restrictions, Rules, etc.). Development of new structured content in published/imported or de novo semantic graphs is conducted in ontology editor applications that support the automated or manual addition, deletion, or modification of content, and the cross-linkage (e.g., matching corresponding concepts) of distinct ontologies. Ontology editor applications also export to a Triplestore to drive semantically aware assets (see Appendix F).

Configuration and implementation of semantic graph databases (including utilization of Reasoners) should adhere to source documentation. Numerous (open source or commercial) editors and Triplestores, offering assorted features and capabilities, are available for procurement across the internet (see Appendix F). Use of a particular editor or Triplestore in the semantic instruments for VA IT capabilities will align with the TRM Decision Matrix.

Before material development, the design of application data models should also consider the strategy for semantic processing of source data. Current recommendations from the World Wide Web Consortium (W3C)[22] for data processing include two Relational Databases to RDF (RDB2RDF)[23] languages: Direct Mapping[24] and R2RML.[25] Both languages map relational database (RDB) content to RDF and facilitate the development of diverse products. RDB2RDF languages can either translate (extract, transform, load) relational data into RDF in order to house it in a

[15] Unified Medical Language System, https://www.nlm.nih.gov/research/umls/
[16] IHTSDO, http://www.ihtsdo.org/
[17] NCIthesaurus, https://ncit.nci.nih.gov/ncitbrowser/
[18] LOINC, http://loinc.org/
[19] Unified Medical Language System (UMLS) RxNorm, https://www.nlm.nih.gov/research/umls/rxnorm/
[20] Gene Ontology Consortium, http://geneontology.org/
[21] Linked Data - Connect Distributed Data across the Web, http://linkeddata.org/
[22] World Wide Web Consortium Main Page, https://www.w3.org/2001/sw/wiki/Main_Page
[23] RDB2RDF Relational Databases to RDF (RDB2RDF), https://www.w3.org/2001/sw/wiki/RDB2RDF
[24] Direct Mapping, https://www.w3.org/2001/sw/wiki/Direct_Mapping
[25] R2RML, https://www.w3.org/2001/sw/wiki/R2RML

Triplestore or generate mapping virtual service interface that can be queried by SPARQL Protocol and RDF Query Language (SPARQL)[26] and translated into SQL queries of the target relational data (Figure 3).

Direct Mapping transforms metadata from an RDB to RDF without controlling the structure of the resulting RDF graph. R2RML generates customized mappings of existing relational data to a "final" RDF graph for an application. R2RML can generate a virtual SPARQL endpoint over the mapped relational data, an RDF dump, or a Linked Data interface. Selection of RDB2RDF language depends on such factors as development overhead, performance, project goals, and other elements. Interfaces for NoSQL to RDF translation require customized virtual mappings.

3.3 Alignment to the One-VA Technical Reference Model (TRM)

All projects will leverage approved tools and technologies located in the VA TRM[27] to comply with the architectural standards and guidance provided in this EDP. Appendix G contains the current entries in the TRM that apply to SWT. Decisions about what approved standards and tools to support SWT will be based on a survey of the current landscape of SWT standards and tools, as provided in Appendix F. These standards are segregated between official standards and emerging standards that will be evaluated for future inclusion into the TRM. Appendix F and G are designated as the official standards profile that informs and constrains the IT capabilities that support SWT. This EDP will be updated to reflect the latest industry standards and emerging trends that are gaining acceptance by the SWT community.

3.4 Alignment to Veteran-Focused Integration Process (VIP)

All projects subject to VIP will use only TRM-approved COTS products that are verified by the TRM team to support the established technical standards referenced in Appendix F. Future projects may leverage the products listed in Appendix G and incorporate new products as they are approved in the TRM. All products are approved based on evaluations by the Office of Information Security (OIS) that they can operate in accordance with Federal and Departmental security policies. Any COTS product that cannot provide the full range of SWT standards will be evaluated as "Prohibited" for use in VA projects per evaluation guidelines established by the TRM Management Group.

---

[26] SPARQL Query Language for RDF, https://www.w3.org/2001/sw/wiki/SPARQL
[27] One-VA Technical Reference Model v16.8 Home Page, http://trm.oit.va.gov/

4 USE CASES

4.1 Use Case: Semantic Data Integration

Stakeholders provided data integration scenarios in the following programs:

- Healthcare delivery (VistA)
- Healthcare research (VINCI)

Both scenarios fundamentally represent the same technical circumstance: dynamic integration of diverse patient-related data types from distributed sources, based on semantic relations/associations among datasets. Graphic depiction of the use case for inclusion of a semantic data integration service in the VA SOA environment is displayed in Figure 3.

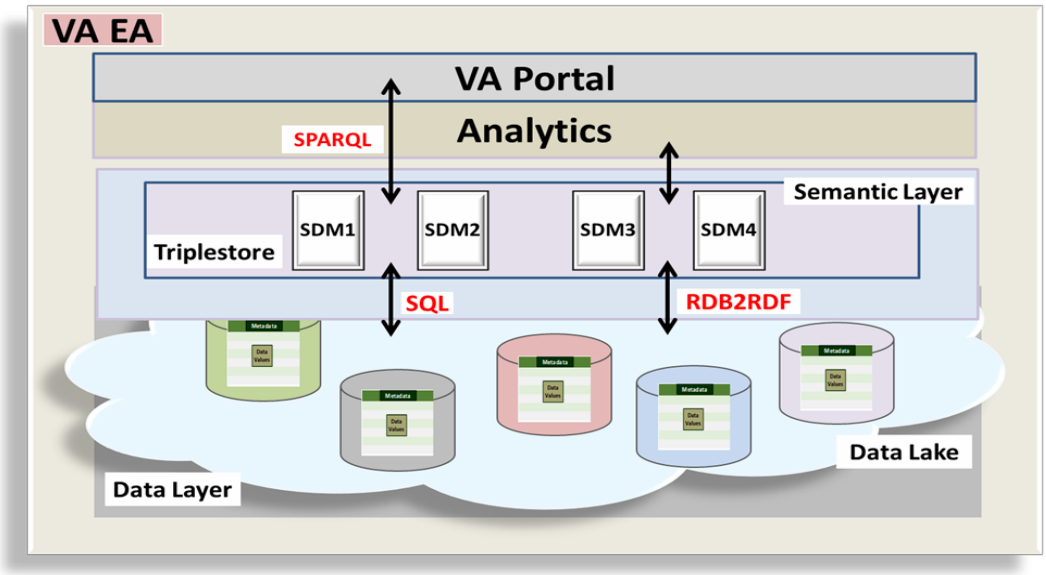

FIGURE 3: NOTIONAL USE OF A SEMANTIC DATA LAYER BASED ON VA DATA INSTANCES

4.1.1 Purpose

VistA: COHORT DATA VISUALIZATION FOR SURVEILLANCE/SAFETY

Integration/Visualization of diverse data associated with medical conditions (e.g., signs, symptoms, tests, diagnosis, treatment, outcome, etc.) from diverse sources based on relations/associations, for real-time cohort surveillance and safety.

VINCI: INTEGRATION OF DOD/VBA/CDW DATA

Integration/Visualization across time of patient demographic data by distributed query to DoD, Veterans Benefits Administration (VBA), and CDW network sources (with disparate data models) based on relations/associations among datasets.

4.1.2 Assumptions

1. Hypothetical technical configuration for illustration purposes only.
    a. The agile SWT data framework enables a variety of architectural configurations, depending on data processing goals, performance requirements, and so forth.
2. Healthcare investigation is of an individual subject/patient or a population.
3. Single user, single query: real-time semantic exploration of dispersed datasets in data layer/lake.
    a. Bulk automated semantic exploration/investigation of datasets is executed by similar programmatic and technical elements.
4. Technical knowledge of SWT by user is not required.
    a. Query support available by visualization of semantic model(s).
5. Transaction (query input and data access, mining, and retrieval) by a semantic data service runs over the VA Wide Area Network (WAN) network (or web) to diverse data sources.
6. RDF/OWL (Web Ontology Language) model(s) housed in an enterprise-level Triplestore.
    a. Well-formed (high quality) local and/or public data model(s), sufficiently cross linked.
    b. Sufficiently modeled objects/properties/restrictions.
        i. Medical conditions (e.g., signs/symptoms, clinical tests, diagnosis, treatment, outcome, etc.), demographics, patient identity, timestamps, etc.
7. Network/cloud/internet infrastructure and service/data protocols implemented as necessary.
    a. User (person/non-person) permission/authentication, mobile device access, secure messaging protocols in place.
    b. Semantic service connected to network/cloud/internet infrastructure.
        i. Semantic service accessed via VA web portal.
8. Query of specified structured (e.g., relational/SQL) data endpoints, whether internal or external.
    a. Query of alternative data/file formats (e.g., NoSQL, image, video, text) executed with programmatic adjustments to metadata, language, or NLP protocols.
    b. Query of SPARQL endpoints (internal or external) executed with programmatic adjustment.

4.1.3 Use Case Decription

1. User formulates/inputs query through semantic data service.
    a. User specifies/selects:
        i. Data source(s)
        ii. Search criteria
            (1) Patient(s), conditions, demographics, time (point(s)/period)
2. Query automatically translated by semantic data service into semantic configuration.
3. Semantic data service converts the source data from user-specified endpoint(s) into a triplestored RDF resource or virtual map.
    a. Query submitted to semantic model(s) in the enterprise Triplestore(s).
    b. Search executed across ontologies for objects/properties satisfying query specifications.
        i. Machine reasoning optional
    c. Virtualized queries are routed to source endpoints.
    d. Queries may also be submitted by semantic service to external data sources.
4. Query results sent back to the user.
    a. Similar/matching datatypes are integrated and displayed by semantic data service for user evaluation.
        i. Optional:
            (1) Data submitted by semantic service to analytics package(s)
            (2) Analytic results displayed
    b. Output (and analysis) is saved in a user-designated store.

APPENDIX A. SCOPE

This EDP provides an enterprise-level view of the "As-Is" and "To-Be" capabilities relevant to SWT used in VA applications and standard processes. The document will refer to, rather than duplicate, lower-level solution guidance associated with these capabilities. This EDP provides guiding principles and best practices that support the adaptation of SWT standards for VA systems and services. These standards enable dynamic integration of data – located in varied sources – by leveraging explicit, standardized, machine-readable linkages among entities.

This EDP comprises:

- A use case example that illustrates adoption of SWT and linked data standards to optimize the integration of distributed and diverse VA data sources with adoption of machine-readable data processing
- SWT standards able to perform on a variety of IT platforms commonly used by VA systems and services
- What application development and deployment capabilities will need to be considered to adopt machine-readable data processing
- Guidance that ensures a framework for seamless data integration based on SWT and linked data standards applicable to internal VA application development and to third party application developers

This EDP is a follow on to the Hybrid Data Access, Utilizing Enterprise Identities, Enterprise Data Analytics, and Data Storage EDPs. The EDP document is generally applicable across all VA Lines of Business (LOB) and describes:

- "As-Is" VA SWT capabilities
- VA SWT infrastructure
- Processes to be used by the developers
- Enterprise-level SWT constraints, strategic guidance, and terminology

This EDP document does not address detailed technical solution guidance for implementing specific SWT applications. It will only provide the constraints to drive VA SWT programs towards development of solutions that effectively meet the specific goals of their initiatives.

Topics that are out of scope for this EDP, but may be referenced, are:

- Data messaging security, authenticity, and mechanisms for securing the enterprise environment
- Minimal performance requirements and specifics of applications/services in use/support of SWT
- Network functionality, infrastructure, and hardware design specifications
- Technology criteria and baselines already covered by the TRM
- Architecting and applying next-generation analytics technologies (e.g., streaming analytics, machine learning)
- Vendor-specific products/technologies

Document Development and Maintenance

This EDP was developed collaboratively with internal stakeholders from across the Department and included participation from VA's Office of Information and Technology (OI&T), Enterprise Program Management Office (EPMO), Office of Information Security (OIS), Architecture, Strategy and Design (ASD), and Service Delivery and Engineering (SDE). Extensive input and participation was also received from VHA, VBA, and National Cemetery Administration (NCA). In addition, the development effort included engagements with industry experts to review, provide input, and comment on the proposed pattern. This document contains a revision history and revision approval logs to track all changes. Updates will be coordinated with the Government lead for this document, which will also facilitate stakeholder coordination and subsequent re-approval depending on the significance of the change.

APPENDIX B. SWT DESIGN PRINCIPLES

SWT is a platform-independent, open standard of the W3C.[28] It synergizes/enriches other technologies, security standards, and data formats (SOA, Cloud, RDBMS, Big Data/NoSQL, Data Lake, Machine Learning, Natural Language Processing (NLP), digital images, etc.) and has gained widespread recognition/adoption across many functional domains (e.g., Health and Medicine, Business Operations, Social Media, etc.).

Unlike conventional designs, SWT operates on the meaning (semantics) of data/information, derived from explicit machine-readable linkages and semantic descriptions of datasets encoded in semantic data models (ontologies). SWT leverages explicit and machine-readable cross-linkages among datasets (local or remote) to promote efficient data harmonization, sharing, and integration. Ontologies (semantic graphs)are readily reusable and extensible/scalable with the addition of new connections and data, supporting simpler maintenance, coordination, and evolution in response to emerging or changing needs.

Modifications to the content of implemented semantic models do not prompt extensive revision/adjustment of allied instruments. SWT can extend standardized, harmonized terminologies – such as SNOMED-CT, Health Level-7 (HL7),[29] and others – to machine-readable semantic data models; enabling dynamic meaningful exchange and integration of diverse data repositories on either a community or global scale.

Importantly, ontologies function according to an "open world" framework. As a result, machine-readable semantic properties of SWT applications/ontologies support the use of reasoning algorithms ("Reasoners"). Reasoners can infer unstated, but logical consequences of data (based on First Order Logic) contained in the schema, which expands the value of datasets beyond those of conventional digital resources that employ predefined/predetermined schemas.

Unlike SWT data models, conventional (e.g., relational) data representations cause the meaning/context (significance) of data to be hidden from machine-readable processing, which constrains data sharing and integration. In relational schemas, entities/objects, properties, and instance data values are represented in a set of prescribed structural matrices: fixed and defined table configurations (Figure 1). Significantly, the relations expressed among datasets in relational tables are simply, and only, signified by implication within the configuration of the table architecture, rather than explicitly asserted in the data model: e.g., something that is a "Patient"

[28] Semantic Web, http://www.w3.org/standards/semanticweb/ Main Page, https://www.w3.org/2001/sw/wiki/Main_Page
[29] Health Level-7 International, http://www.hl7.org/

has something that is a "Name" that has something that is a value of "Jane Doe" (only because of the spatial arrangement of these elements within the table). The meaning of these elements (Patient, Name, and Jane Doe), contained within the table architecture, is readily human-intelligible (inferred by contextual knowledge). However, the meaning of those elements remains largely inaccessible to machine-readable processing. Deployment of a SOA infrastructure does not, of course, alter the design principles of such IT resources.

Semantically modeled data, instead, are encoded (by semantic editing tools) with standardized data representation languages, such as RDF[30] or OWL,[31] a more powerful extension of RDF. In contrast to conventional representations, semantically modeled data explicitly declare machine-readable linkages between entities, in the form of "triple statements:"e.g., Patient – hasName - Jane Doe. Thus, ontologies are fundamentally a series of triple statements that explicitly represent real objects and their cross-linking meaningful relations to other entities. Each modeled entity may link by meaningful relations to multiple additional entities (objects, data values, etc.) within the same semantic model, or to entities in other (local or remote) semantic models – creating "knowledge networks" for deep searching (Figure 2).

Entities encoded in semantic models typically include Classes (categories of similar objects: Patient) and Instances (unique Individuals: Jane Doe). The Classes are often organized in a hierarchical classification (e.g., taxonomy) of Superclasses, Classes, and Subclasses; specifying a span of conceptual abstraction from general to specific types. Subclasses are specializations (subtypes) of Classes; Superclasses are generalizations. Wolf, Coyote, and Dog are all Subclasses of Canine, for example; Dog may be further subdivided (e.g., Spaniel, Hound, Terrier, etc.). Mammal is a Superclass of Canine (and of Primate, etc.). Instances are unique non-divisible Class members (e.g., "Fido", "Jane Doe", "New York City", "Washington Nationals", etc.).

All Superclasses, Classes, Subclasses, and Instances in a hierarchical semantic model are appropriately linked (as triple statements) by an explicit, machine-readable, standardized Subtype relationship. Additional machine-readable relations (e.g., hasName, parentOf, measuredBloodLDL, etc.) may be expressed in a semantic model to describe further meaningful properties of objects that also cross-link Superclasses, Classes, Subclasses, Instances, and data values within or between models (Figure 2).

A semantic graph model comprises a logic-based and machine-/human-intelligible representation and description of real objects, their properties, and their meaningful cross- linking relations to other objects. The semantic data model then resides in a so-called "Triplestore" that constitutes a

[30] RDF Current Status, http://www.w3.org/standards/techs/rdf#w3c_all
[31] OWL Web Ontology Language Current Status, http://www.w3.org/standards/techs/owl#w3c_all

("semantically aware") database.[32] These data models may represent objects in any subject of interest (e.g., demographics, geospatial knowledge, user identity, medical test results, genetic features, etc.) to drive diverse operations; from the execution of an individual software application to coordination of multiple enterprise level resources. Algorithms (RDB2RDF) can convert legacy data (in a Relational Database) to semantic format, or a semantic model may be deployed as an "executive" query engine linked to legacy data (Figure 3).

Semantic data are flexibly queried (based on the structure of triple statements) with the SPARQL language to easily retrieve matches to any combination (simple or complex) of Classes, Instances, or Properties; i.e., from multiple diverse perspectives according to the various elements of the model.[33] The linked configuration of semantic data enables retrieval of not only Instance data (as in other data systems), but also flexible and elastic query of hierarchically described categories at any desired level of abstraction within the graph. Queries for data about all subtypes of a Class, for example, may be easily executed by querying the Class, rather than specifying each individual Subclass; the specificity of a query may be easily and elastically stipulated by exploring and specifying the desired hierarchy level.

Data may also be easily retrieved at will based on user-specified relations/associations among datasets. Thus, in addition to instance data values, semantic knowledge models (ontologies) may also be queried for information about sets of entities (Classes) and their relations (i.e., knowledge). Queries may stipulate multiple datasets as well, empowering highly complex and dynamic queries to retrieve integrated data from one or many models, local or remote (so- called "SPARQL endpoints"). Retrieved data can then be interrogated within analytic pipelines.[34]

Machine-readable linkages within and across RDF/OWL models provide semantic description and integration of data content. Linkages rendered between equivalent entities in distinct models, for example, enable interoperable connection and data integration between ontologies and between systems (Figure 2). As a result, standardized semantic connections among datasets enable more efficient, flexible, and comprehensive queries of heterogeneous data to improve data mining, aggregation, exchange, and analysis, as well as to enhance knowledge discovery and decision support. Thus, semantic relations are not perfunctory mechanical linkages between data entities, but *meaning*-based, intelligent, and machine-readable connections that support improved data processing and apprehension.

---

[32] LargeTripleStores, https://www.w3.org/wiki/LargeTripleStores
[33] SPARQL Current Status, http://www.w3.org/standards/techs/sparql#w3c_all
[34] SparqlEndpoints, https://www.w3.org/wiki/SparqlEndpoints
Describing Linked Datasets with the VoID Vocabulary, http://www.w3.org/TR/void/

Importantly, RDF/OWL data graphs are exchangeable, reusable, and extensible. RDF/OWL models of genomics, diseases, pharmaceutics, clinical knowledge, and other domains for example are published for public consumption at various websites[35] and can be implemented as flexible synergistic modules to assemble highly sophisticated semantic graph networks. Selected published models can be implemented (as needed) in an information system and multiple graphs can be linked simultaneously to integrate data from heterogeneous and disparate sources (Figure 2). Content and relations can be added, removed, or modified relatively easily in locally imported public semantic models to suit user needs. As with any technology, implementation of semantic technologies requires a knowledgeable workforce and prudent implementation. Care should be exercised, for example, in the assembly of ontologies to ensure accurate computation.

[35] The OBO Foundry, http://obofoundry.org/, BioPortal, http://bioportal.bioontology.org/, Unified Medical Language System, https://www.nlm.nih.gov/research/umls/, Linked Data - Connect Distributed Data across the Web, http://linkeddata.org/

APPENDIX C. DEFINITIONS

This appendix provides definitions for terms used in this document, particularly those related to databases, database management, and data integration.

Key Term	Definition
Dataset	A collection of data. A dataset contains individual resources as well as metadata. Metadata is the "Who, What, When, Where, Why" of each dataset. Most commonly a dataset corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question.
Linked Data	A method of publishing structured data so that it can be interlinked and become more useful through semantic queries.
Machine Readable	Information or data that is in a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost.
Open Data	Accessible, machine-readable public Government datasets per OMB M-13-13 (Managing Information as an Asset). The metadata schema selected for Open Data uses Data Catalog (DCAT), which is an RDF vocabulary for linking data catalog metadata. RDF is a fundamental building block for the Semantic Web.
Ontology	A formal, explicit specification of a shared conceptualization. Conceptualization refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used, and the constraint on their use are explicitly defined. Formal refers to the fact that the ontology should be machine-readable. Shared reflects the notion that an ontology captures consensual knowledge, that is, it is not private of some individual, but accepted by a group.

Key Term	Definition
Resource Description Framework	A family of specifications for a metadata model. The RDF family of specifications is maintained by the World Wide Web Consortium (W3C). The RDF metadata model is based upon the idea of making statements about resources in the form of a subject-predicate-object expression. RDF's simple data model and ability to model disparate, abstract concepts has also led to its increasing use in knowledge management applications unrelated to Semantic Web activity.
Semantic Web	An evolutionary stage of the World Wide Web in which automated software can store, exchange, and utilize metadata about the vast resources of the Web, in turn enabling users to deal with those resources with greater efficiency and certainty.

APPENDIX D. ACRONYMS

The following table provides a list of acronyms that are applicable to and used within this document.

Acronym	Description
API	Application Program Interface
ASD	Architecture, Strategy and Design
BISL	Business Intelligence Service Line
CDI	Customer Data Integration
CDW	Corporate Data Warehouse
CHIR	Consortium for Health Informatics Research
CMS	Centers for Medicare & Medicaid Services
COTS	Commercial off-the-shelf
CRUD	Create, Read, Update, Delete
DoD	Department of Defense
EA	Enterprise Architecture
EDP	Enterprise Design Pattern
eHMP	Electronic Health Management Platform
EHR	Electronic Health Record
ESS	Enterprise Shared Service
ETA	Enterprise Technical Architecture
ETL	Extract, Transform, Load
FHIR	Fast Healthcare Interoperability Resources
HIPAA	Health Insurance Portability and Accountability Act
HL7	Health Level 7
HSR&D	Health Service Research and Development
LOB	Line of Business
LOINC	Logical Observation Identifiers Names and Codes
NCI	National Cancer Institute
NGO	Non-Governmental Organization
NLP	Natural Language Processing
NoSQL	Not only SQL
OBO	Open Biomedical Ontologies
OI&T	Office of Information and Technology
OIA	Office of Informatics Analytics
OWL	Web Ontology Language
R2RML	Reversible Rule Markup Language
RDB	Relational Database
RDB2RDF	Relational Databases to RDF
RDBMS	Relational Database Management System
RDF	Resource Description Framework

Acronym	Description
RDW	Regional Data Warehouse
SAS	Statistical Analysis System
SNOMED-CT	Systematized Nomenclature of Medicine Clinical Terms
SPARQL	SPARQL Protocol and RDF Query Language
SPML	Service Provisioning Markup Language
SPSS	Statistical Package for the Social Sciences
SOA	Service-Oriented Architecture
SQL	Structured Query Language
SWT	Semantic Web Technology
TRM	Technical Reference Model
UMLS	Unified Medical Language System
URI	Uniform Resource Identifier
VE	VistA Evolution
VINCI	VA Informatics and Computing Infrastructure
VIP	Veteran-Centric Integration Process
VistA	Veterans Health Information Systems and Technology Architecture
W3C	World Wide Web Consortium
XML	Extensible Markup Language

APPENDIX E. REFERENCES, STANDARDS, AND POLICIES

This EDP is aligned to the following VA OI&T references and standards applicable to all new applications being developed in the VA, and are aligned to the VA Enterprise Technical Architecture (ETA):

#	Issuing Agency	Policy, Directive, or Procedure	Purpose
1	VA ASD	VA Directive 6551	Establishes a mandatory policy for establishing and utilizing Enterprise Design Patterns by all Department of Veterans Affairs (VA) projects developing information technology (IT) systems in accordance with the VA's Office of Information and Technology (OI&T) integrated development and release management process, the Veteran-focused Integration Process (VIP).
2	VA OIS	VA 6500 Handbook	Directive from the OI&T OIS for establishment of an information security program in VA, which applies to all applications that leverage ESS.
3	OMB	Open Data Policy (M-13-13)	Federal policy regarding the publication of machine-readable public datasets and data catalogs following a common metadata standard. Currently the chosen metadata standard uses DCAT, an RDF vocabulary used to link together diverse data catalogs: https://project-open-data.cio.gov/v1.1/schema/

APPENDIX F. STANDARD SWT AND TOOLS

References for the following standards for W3C language recommendations and common SWT product examples (and more) may be found at the W3C website[36] as well as additional specific references provided below.

Current Approved Standards

1. W3C SWT Language Recommendations
   a. SKOS (Simple Knowledge Organization System)
      i. Common data model for sharing and linking knowledge organization systems (thesauri, taxonomies, classification schemes and subject heading systems); low-cost path for porting existing systems to the Semantic Web, or developing new systems.
   b. RDF (Resource Description Framework)
      i. Standard model for data interchange by linking relationships even if the underlying schemas differ and supports evolution of schemas without requiring dependent consumers to change.
   c. RDFS (RDF Schema)
      i. A data-modelling vocabulary of classes and properties built as an extension of the basic RDF vocabulary.
   d. RDFa
      i. Used for embedding and extraction of RDF triples in XHTML documents.
   e. OWL (Web Ontology Language)
      i. Logic-based language, part of the W3C SWT stack, designed to represent rich and complex knowledge; exploited by computer programs to make implicit knowledge explicit; and refer to or be referred from other OWL ontologies.
   f. SPARQL (SPARQL Protocol and RDF Query Language)
      i. Query language for RDF/OWL, used to query data is stored as RDF or viewed as RDF via middleware.
   g. GRDDL (Gleaning Resource Descriptions from Dialects of Languages)
      i. A W3C recommendation to obtain RDF triples from XML documents, including XHTML
   h. POWDER (Protocol for Web Description Resources)
      i. W3C protocol for publishing metadata describing Web resources using RDF, OWL, and HTTP

[36] Semantic Web, http://www.w3.org/standards/semanticweb/, Semantic Web Development Tools, https://www.w3.org/2001/sw/wiki/Tools

i. R2RML (RDB to RDF Mapping Language)
   i. Language for mapping existing relational data in RDF graphs, as a virtual SPARQL endpoint, as a RDF dumps, or offer as Linked Data
j. SWRL (Semantic Web Rule Language)
k. SPIN (SPARQL Inferencing Notation)
   i. While not a de-jure standard, it is a de-facto standard with support not only from many data integration products, but also Sesame, Ontotext, Apache Jena Fuseki, AllegroGraph, etc. Constraints and functions parts of SPIN are now being standardized as SHACL (emerging standard discussed in the following section). Rules will probably be the next step.

2. Products (open source and COTS) of common SWT functional elements
   a. ONTOLOGIES
      i. For commonly used ontologies,[37] see
         (1) The OBO Foundry
         (2) BioPortal
         (3) UMLS
         (4) Linked Data
   b. ONTOLOGY EDITOR (common examples)[38]
      i. Anzo for Excel http://www.cambridgesemantics.com/products/anzo_for_excel
      ii. Fluent Editor http://www.cognitum.eu/Semantics/FluentEditor/
      iii. Knoodl http://www.knoodl.com/
      iv. Neologism http://neologism.deri.ie/
      v. NeOn Toolkit http://neon-toolkit.org/
      vi. OBO-Edit http://oboedit.org/
      vii. OntoStudio http://www.semafora- systems.com/en/products/ontostudio/
      viii. Open Semantic Framework http://opensemanticframework.org/
      ix. OWLGrEd http://owlgred.lumii.lv/
      x. Protégé http://protege.stanford.edu/
      xi. Semaphore Ontology Manager http://www.smartlogic.com/
      xii. Semantic Turkey http://semanticturkey.uniroma2.it/
      xiii. SWOOP http://www.mindswap.org/2004/SWOOP/
      xiv. TopBraid Composer http://www.topquadrant.com/products/TB_Composer.html
      xv. Vitro http://vitro.mannlib.cornell.edu/
      xvi. VocBench http://vocbench.uniroma2.it/

[37] The OBO Foundry, http://obofoundry.org/, BioPortal, http://bioportal.bioontology.org/, Unified Medical Language System, https://www.nlm.nih.gov/research/umls/, Linked Data - Connect Distributed Data across the Web, http://linkeddata.org/
[38] Ontology editors, https://www.w3.org/wiki/Ontology_editors, Ontology (information science), https://en.wikipedia.org/wiki/Ontology_(information_science)#Editor

c. TRIPLESTORE (common examples)[39]

i. 3store http://threestore.sourceforge.net/
ii. AllegroGraph http://franz.com/
iii. Apache Jena http://jena.apache.org/
iv. Bigdata http://www.bigdata.com/blog/
v. Garlik 4store https://github.com/garlik/4store
vi. GraphDB http://ontotext.com/products/graphdb/
vii. IBM DB2 http://www.ibm.com/analytics/us/en/technology/db2/
viii. Jena http://jena.apache.org/
ix. Kowari http://www.kowari.org/
x. MarkLogic http://www.marklogic.com/
xi. Mulgara http://www.mulgara.org/
xii. OntoBroker http://www.semafora- systems.com/en/products/ontobroker/
xiii. OpenLink Virtuoso v6.1 http://virtuoso.openlinksw.com/wiki/main/
xiv. Oracle Spatial and Graph with Oracle Database 12c https://www.oracle.com/index.html
xv. RDF gateway http://www.intellidimension.com/
xvi. RDFox http://www.cs.ox.ac.uk/isg/tools/RDFox/
xvii. Sesame http://www.openrdf.org/
xviii. Smart Content Factory http://www.mondeca.com/index.php/en/
xix. Stardog http://stardog.com/
xx. YARS2 sw.deri.org/2004/06/yars

d. REASONER (common examples)[40]

i. Reasoners are commonly included/implemented in Triplestores.
ii. BaseVISor http://www.vistology.com/basevisor/basevisor.html
iii. Bossam http://bossam.wordpress.com
iv. CLR http://reasoner.sourceforge.net
v. CLR http://reasoner.sourceforge.net
vi. CWM http://www.w3.org/2000/10/swap/doc/cwm.html
vii. Cyc www.cyc.com
viii. Drools http://www.drools.org/
ix. ELK http://elk.semanticweb.org/
x. Fact ++ http://owl.cs.manchester.ac.uk/fact++/
xi. Flora-2 http://flora.sourceforge.net/

[39] SPARQL Current Status, http://www.w3.org/standards/techs/sparql#w3c_all, List of subject- predicate-object databases, https://en.wikipedia.org/wiki/List_of_subject-predicate- object_databases
[40] SemanticWebTools, https://www.w3.org/wiki/SemanticWebTools#Reasoners_.28OWL_or_rule_based.29_and_rela ted_tools, Semantic reasoner, https://en.wikipedia.org/wiki/Semantic_reasoner

xii. Gandalf https://gndf.io/
xiii. HermiT http://www.hermit-reasoner.com/
xiv. KAON2 http://kaon2.semanticweb.org/
xv. OntoBroker http://www.ontoprise.de/en/home/products/
xvi. Oroboro http://code.google.com/p/oroboro/
xvii. OWLRL http://www.ivan-herman.net/Misc/2008/owlrl/
xviii. Pellet http://pellet.owldl.com/
xix. Pellint http://pellet.owldl.com/pellint
xx. Prova https://prova.ws/
xxi. RacerPro http://www.racer-systems.com/
xxii. SHER http://www.alphaworks.ibm.com/tech/sher
xxiii. SPARQL-DL http://www.derivo.de/en/resources/sparql-dl-api/
xxiv. SWObjects http://sourceforge.net/apps/mediawiki/swobjects/
xxv. Thea http://www.semanticweb.gr/TheaOWLLib/

Emerging Standards

1. W3C SWT Language Recommendations
   a. Shapes Constraint Language (SHACL)
      i. When shared notions of completeness and validity are lacking, one often ends up trying to make sense of an impossible big ball of data mud. The Shapes Constraint Language (SHACL), an upcoming W3C standard, promises to solve this problem and help resolve a slew of data quality and data exchange issues in Semantic Web applications.
   b. RIF (Rule Interchange Format)
      i. W3C XML language for expressing rules which computers can execute

APPENDIX G. CURRENT SWT ENTRIES IN TRM[41]

1. Apache Jena
2. Current Dental Terminology (CDT)
3. Current Procedural Terminology (CPT)
4. Health Level 7 (HL7) Application Programming Interface (API)- Fast Healthcare Interoperable Resources (FHIR)
5. Healthcare Common Procedure Coding System (HCPCS)
6. Healthcare Provider Taxonomy (HPT)
7. HL7 Clinical Genomics Pedigree Model
8. HL7 Clinical Vaccine Formulation (CVX)
9. HL7 Data Segmentation for Privacy (DS4P)
10. HL7 Healthcare Privacy and Security Classification System (HCS)
11. HL7 Identity Cross-Reference Service Functionality Model (IXS)
12. HL7 Manufacturer of Vaccines (MVX) Code Set
13. HL7 Version 3 Standard: Privacy, Access, and Security Services; Security Labeling Service, Release 1 (SLS)
14. Human Gene Nomenclature (HGN)
15. ICD-10-CM (International Classification of Diseases, 10 Revision, Clinical Modification)
16. ICD-10-PCS (International Classification of Diseases, 10th Revision, Procedure Coding System
17. IHE PCD Technical Framework Supplement Subscribe to Patient Data (SPD)
18. IHE Patient Care Device Technical Framework Volume 3 (PCD TF-3) Semantic Content
19. IHE Patient Care Devices Technical Framework Volume 1 (PCD TF-1) Integration Profiles
20. IHE Patient Care Devices Technical Framework Volume 2 (PCD TF-2) Transactions
21. IHE PCD Technical Framework Supplement Alarm Communication Management (ACM)
22. IHE PCD Technical Framework Supplement Infusion Pump Event Communication (IPEC)
23. IHE PCD Technical Framework Supplement Subscribe to Patient Data (SPD)
24. Informatics for Integrating Biology and the Bedside (i2b2)
25. International Health Terminology Standards Development Organization (IHTSDO) Workbench
26. International Classification of Diseases (ICD)
27. International Classification of Functioning, Disability and Health (ICF)
28. International Health Terminology Standards Development Organization (IHTSDO) Workbench
29. International Nonproprietary Names (INN)

[41] One-VA Technical Reference Model v16.8 Home Page, http://trm.oit.va.gov/

30. Internationalization Tag Set (ITS)
31. National Cancer Institute (NCI) Enterprise Vocabulary System (EVS)
32. National Center for Biomedical Ontology (NCBO) Annotator
33. National Drug Code (NDC)
34. National Drug File Reference Terminology (NDF-RT)
35. NCPDP (National Council for Prescription Drug Programs) Formulary and Benefits
36. Protege
37. Resource Description Framework (RDF)
38. RxNorm
39. Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT)
40. Unified Medical Language System (UMLS)
41. Web Ontology Language (OWL)

Disclaimer: This document serves both internal and external customers. Links displayed throughout this document may not be viewable to all users outside the VA domain. This document may also include links to websites outside VA control and jurisdiction. VA is not responsible for the privacy practices or the content of non-VA websites. We encourage you to review the privacy policy or terms and conditions of those sites to fully understand what information is collected and how it is used.