Veterans Health Administration | Office of Health Informatics
Knowledge Based Systems | Standards and Interoperability – Informatics Architecture

Task Order #9 | November 29, 2018

SOLOR Support Services: Use Case #1 Evaluation Plan

PRESENTED BY:

Team BZ
Program Manager: Jayme Welty

bookzurman
Transforming healthcare through technology + trust

360 Central Ave., Suite 970 | St. Petersburg, FL 33701
O 727.378.9006 | bookzurman.com

PRESENTED TO:

Dr. Keith Campbell & Stephanie Klepacki

IDIQ: VA701-16-D-0017
TO9 PO: 776-C80148
CLIN: 2005B Proposed Standards Artifacts (Medium)

A FORMATIVE EVALUATION OF THREE SOLOR EXTENSION USE CASES PROMOTING SEMANTIC INTEROPERABILTY

(CLIN 2005B_10.14, 2005B_11.14 and 2005B_12.14)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

VERSION HISTORY

DATE	VERSION	DELIVERABLE	DESCRIPTION
10/29/2018	0_0	Proposed Standards Artifacts (Medium)	Use Case
11/29/2018	0_1	Proposed Standards Artifacts (Medium)	Evaluation Plan

1. INTRODUCTION

The vision of the Department of Veterans Affairs (VA), Veterans Health Administration (VHA), Office of Informatics & Analytics (OIA), and Health Informatics (HI) is to provide timely, relevant information and data services that support improvements in Veterans' health. In meeting these goals, OIA strives to provide high quality, effective, and efficient information and data services to those responsible for providing care to the Veterans at the point-of-care as well as throughout all the points of the Veterans' health care in an effective, timely and compassionate manner. VA depends on the interoperability of information and data to meet mission goals.

To this end, VHA's informatics architecture was created to integrate disparate knowledge sources and preserve the meaning of information for the interoperability of electronic health record data (i.e., semantic interoperability) which is critical for delivering safe veteran care and leveraging standards-based clinical decision support. SOLOR, (System of Logical Representation) is the open source ecosystem of capabilities and services for assimilating disparate health knowledge sources into a consistent representation based on best practices of computer science. By doing this, SOLOR enables collaboration in health IT, unifies health terminology standards and removes ambiguity, leading to improved patient care.

1.1. Aims

The overarching objective of this body of work is to inform the development of SOLOR by exploring its extension as an ecosystem for integrating desperate knowledge sources and creating interoperability by making information meaningful and computable. The specific aims of this work are:

Aim 1: Develop use cases for the extension of SOLOR.

Aim 2: Evaluate constructs of the SOLOR use cases developed in previous aim.

2. BACKGROUND

To be completed as part of future deliverable.

2.1. The SOLOR System

To be completed as part of future deliverable.

2.2. SOLOR Knowledge Sources

2.2.1. Terminology Knowledge Sources

Terminology systems are increasingly critical components for achieving interoperability across applications in the healthcare domain. The role of standard terminologies in achieving interoperability for the purposes of advancing patient care is well documented [1]. The federal government recognizes the benefit of standard terminologies and promotes their development and use. The *Federal Health IT Strategic Plan 2015-2020* set a strategy to encourage consistent terminology standards implementation in Electronic Health Records (EHR) and encourage use through federal payment policies [2]. A standard terminology is one that has wide industry acceptance or use. Standards are obtained from a variety of efforts, cover different domains of clinical and nonclinical content relevant to the EHR, and serve various purposes. Currently, no one terminology or classification system contains everything that is needed for the medical record. Examples of standard terminologies include:

- Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT®): a comprehensive clinical terminology, maintained by the International Health Terminology Standards Development Organization (IHTSDO) [3],
- Logical Observation Identifiers, Names, and Codes (LOINC®): a terminology for laboratory tests, results, and clinical observations, developed and maintained by the Regenstrief Institute [4], and
- RxNORM: a terminology for human clinical drugs, maintained by the National Library of Medicine (NLM) and distributed via the Unified Medical Language System (UMLS) [5].

Terminology systems typically consist of the following elements:

- Coded Concepts – the discrete units of knowledge managed within the terminology. They typically consist of numeric codes and textual preferred names, synonyms, and descriptions.
- Concept Hierarchies – the logical organization of concepts into parent-child and ancestor-descendant relationships that express the semantics of generalization and specialization. The hierarchical organization of a terminology may be explicitly expressed through stored parent-child and ancestor-descendant links, or it may be implicitly expressed through the logical definitions of individual concepts that a computer can use to infer parent-child and ancestor-descendant relationships.
- Value Sets – named lists of individual concepts that represent more abstract categories useful in decision-support logic.

New applications and new medical knowledge constantly call for expansion and enhancement of existing terminologies. However, since terminology systems are often non-static, incomplete and under specified, inconsistencies may be introduced [6]. Therefore, quality assurance is an indispensable part the terminology management lifecycle.

2.2.2. Genome Variant Knowledge Sources

A key part of the work in the genome research domain is to identify genome variants and assign a clinical impact, if known. A genome variant knowledge source is a repository of known genome variants and associated clinical interpretations of that variant. There are many types of genome variant knowledge sources, which include (1) privately-controlled knowledge bases, such as the Human Gene Mutation Database (HGMD) [7]; (2) open access, locus-specific knowledge bases, such as those created using the Leiden Open Variation Database (LOVD) [8]; (3) proprietary knowledge bases, typically owned and managed by genetic testing laboratories, who maintain exclusive access [9]; and (4) publicly available, centrally-managed repositories, such as ClinVar [10]. Typically, when a new variant is discovered, or new information about a known variant is made available, this information will be recorded in one or more of these knowledge bases. Furthermore, curators may monitor publications and reports in order to update a knowledge base accordingly.

ClinVar, which is a publicly available central resource managed by the National Library of Medicine, represents a model wherein genome knowledge sources can upload their expertly curated knowledge into one location. Previously, genome knowledge consumers may have had to use several different genome variant knowledge bases and pay to access particular knowledge. Furthermore, with a open collaborative approach to genome variant annotation, ClinVar may become a more robust and extensive knowledge base than any single locus-specific or laboratory-managed knowledge bases. Open access, locus-specific knowledge bases tend to be curated and maintained on a volunteer basis, making the knowledge available limited. While laboratory-managed knowledge bases contain the best variant

knowledge, they are also (1) limited by the number of unique variants observed by that laboratory and (2) may have tightly controlled access to the variant knowledge in order to maintain a competitive advantage over other testing laboratories [9]. Nevertheless, if ClinVar is embraced by the diagnostic laboratory community with the support of the ClinGen effort [11], the laboratory knowledge bases will likely serve as one of the most important sources of variant annotations. Additionally, several characteristics of ClinVar make it attractive for our type of work:

Format – ClinVar maintains a health data repository available via FTP download in several release formats (e.g TSV, XML, and VCF). In particular, the tab separated values release format, which provides data in a structure similar to relational database tables, is the easiest data format to be used in the SOLOR transformation process.

Documentation – Robust ReadMe files within each ClinVar release, describing in detail every data point contained within the overall ClinVar release data structure. Based on these descriptions, reliable inferences can be constructed for the SOLOR transformation process.

Release Cycle – Within the ClinVar release data tables, there exists variations (e.g. daily, weekly, monthly, etc) of update frequency amongst individual data entities. Variant data is updated weekly, whereas phenotypic data is updated daily. Creating a SOLOR transformation process around data entities that are frequently updated results in more current variant data for the SOLOR system.

Data Structure – Specific data entities, such as variant, gene, and disease, can be normalized, modular, and isolated from other more complex entity relationships. These aspects for such key data entities result in a less complex, more straightforward implementation of the SOLOR transformation process.

Variant Identifier – ClinVar utilizes the Human Genome Variation Society (HGVS) specification for naming genomic variants contained within each release. Leveraging approved standards, as part of key data elements being transformed into the SOLOR system, enables proper terminology concept quality assurance and classifications to be performed on all SOLOR health data.

2.3. Ecosystem

To be completed as part of future deliverable.

3. MATERIALS AND METHODS

3.1. Aim 1

3.1.1. Precision Medicine Use Case (CLIN 2005B_01.14)

Use Case 1 develops a Precision Medicine use case for SOLOR where variants which occur within genes are assessed for clinical impact using the curated genome variant knowledge base ClinVar. ClinVar, which is a publicly available central resource managed by the National Library of Medicine, represents a model wherein genome knowledge bases and laboratories can upload their expertly curated knowledge into one location [ref.]

Genetic data knowledge sources are not structured or maintained in a format usable for the Electronic Health Records (EHR), clinical decision support, research, or interoperability despite the fact that precision

medicine has become a national priority [Ref needed]. The market cost of genetic testing continues to decrease, while at the same time, the number of known genetic variants and number of genetic tests available continue to increase. Consequently, genetic information is becoming a more common addition to an individual's health records with important implications for treatment and research.

It is critical that individual genetic information is incorporated into electronic records in a consistent way so that clinicians and computer decision support systems (CDSS) alike can realize its benefits without errors or ambiguities. Accessible and standardized genetic-based test results and data sets have the potential to help clinicians provide better patient care if integrated into the electronic health record, enable more insightful population health statistics if in a standardized format and contribute to more impactful research if interoperable.

3.1.1.1. Genome Data Acquisition and Database Storage

The ClinVar knowledge source was added to the SOLOR ecosystem using a transformation process which allows for ClinVar specific data representation within the SOLOR ecosystem. Incorporating the ClinVar knowledge source into the SOLOR ecosystem required a custom implemented transformation process, which focused specifically on transforming the ClinVar tab separated value data format into the SOLOR common model format. Below describes the three data entities and the specific data elements used in the ClinVar to SOLOR transformation process:

Variant Summary – Contains attribute information that further describes gene variants submitted to ClinVar. The specific name of each variant in the HGVS format and the particular National Center for Biotechnology Information (NCBI) gene ID is used in the SOLOR transformation process.

Gene Specific Summary – Contains attribute information to further describe individual NCBI managed table of genes, specifically focusing on both gene's identifiers, the NCBI ID and its symbol data elements.

Gene Condition Source ID – Contains all relationships between genes and correlating diseases (phenotypes) used in ClinVar. This data entity contains not only the NCBI gene ID, but also identifiers of external phenotypic terminology concepts. For example, a specific gene ID is correlated with a potential SNOMED CT concept and the associated SNOMED CT Identifier (SCTID).

All variants and genes found in ClinVar were de-duplicated and loaded into the SOLOR model as unique SOLOR concepts. Each concept contained both a fully qualified name, based on either the variant's name and or the gene's symbol, as well as String identifiers that were based off the variant's HGVS ID, or the gene's NCBI ID. In addition, parent-child (supertype-subtype) relationships between concepts for variants to concepts for genes, and concepts for genes to SNOMED CT concepts, were encapsulated as logic graph axioms, visualizing a stated (modeled) view of the concepts as well as the view after classification, and assigned to each respective SOLOR concept. Lastly, a comprehensive SOLOR taxonomy was created incorporating both ClinVar and SNOMED CT concept.

3.1.2. Medical Device Interoperability Use Case (CLIN 2005B_02.14)

To be completed as part of future deliverable.

3.1.3. Use Case 3 (CLIN 2005B_03.14)

To be completed as part of future deliverable.

3.2. Aim 2 (CLINs 2005B_04.14, 2005B_05.14 and 2005B_06.14)

3.2.1. Evaluation Design

We will perform a formative evaluation of use case constructs. Formative studies are particularly useful for applied work, where it is more important to understand the process by which things happen in a particular situation than to measure outcomes rigorously or to compare a given situation with others [12]. Formative evaluation is a common approach for improving the quality of a program being developed by identifying weaknesses throughout the design and development efforts so that it will be as likely as possible to achieve the objectives for which it was designed [13,14]. A formative evaluation aims to help develop and improve programs from an early stage, when opportunities for influence are likely to be greatest, and to identify promising components [15]. Innovative programs provide an ideal environment for use of formative evaluation findings, with key stakeholders generally much more willing to make adjustments at an early stage than when a program is well established [16].

The goal of this formative evaluation is to collect rapid feedback from subject matter experts that would provide validation of use case constructs and context for future successive adaptations and improvement of the use case's development. Having said that, key questions for evaluating a new proof-of-concept include: Does the idea provide a new and more useful capability?; does it help developers better understand complex systems?; and does it demonstrate by its behavior that a complex assembly of components can accomplish a particular set of activities? Our formative evaluation research questions are shown in Table 1.

3.2.2. Evaluation Participants

We combined both purposeful expert sampling and snowball sampling to create an interview strategy to gather knowledge from individuals that have particular expertise[17,18]. We first identified key informants (someone knowledgeable about health informatics) to begin the process of interviewing and we then asked for the names of subject matter experts (individuals especially knowledgeable and experienced with medical terminological systems). In addition, it was also important that participants were available and willing to contribute, and able to effectively communicate their experiences.

3.2.3. Methods Used for Data Collection

This work will use as its primary data gathering method a semi-structured interview approach, as described by Steinar Kvale in Doing Interviews [19]. It's a fairly open approach where a guide is used, with questions and topics to be covered. The evaluator has some discretion with the order in which questions are asked, but the questions are standardized, and provided to ensure that the researcher covers the correct material. Unlike the structured interview where the questions are fixed and they are asked in a specific order, questions or topics can be further developed on the basis of responses from the interviewee. Semi-structured interviews allow for in-depth encounters in which focused, conversational, two-way communication is used to elicit detailed narratives and are often used by evaluators wanting to delve deeply into a topic and to thoroughly understand the answers provided.

This approach aligns with the approach for conducting semi-structured interviews described in the RAND Corporation report "Data Collection Methods: Semi-structured Interviews and Focus Groups" [20]. An overview of the important aspects of semi-structured interviews includes a number of steps. First, the main research questions need to be identified. In other words, what does the researcher hope to learn? Next, the researcher needs to consider the different participant types and determine the sampling. This

study used judgment/purposeful sampling where individuals were selected based on their knowledge of medical terminologies, and because their opinion was judged to be important to the research [18].

Interviews are typically personal and intimate encounters that allow for focused, conversational, two-way communication in which open, direct, verbal questions are used to elicit detailed narratives and stories[21]. This study conducted semi-structured interviews where an interview is defined as: a method of data collection in which one person (an interviewer) asks questions of another person (a respondent) either face-to-face or by telephone[22]. Although no interview can truly be considered structured, they were relatively structured and more or less equivalent to guided conversations.

We engaged participants at a single point in time, individually, using virtual meeting software, and conducted open-ended, semi-structured interviews. Participants were contacted by email to invite them to participate and a meeting time was then set at a time and day of their convenience. The total time was allotted no more than two hours for the investigators to complete the interactions. Participation in this study was voluntary and the subject matter experts could choose not to take part in the interview. The subject matter experts could also skip any question they preferred not to answer or terminate the interview without penalty. We asked each participant four demographic questions: (1) job title, (2) number of years of experience, (3) education level and (4) previous terminology experience. All demographic data gathered about the participant were free text.

3.2.4. Methods Used for Data Analysis

Applied thematic analysis, a method for identifying and analyzing patterns of meaning in a dataset, was used to organize and describe the data collected from the interviews [23–25]. Applied thematic analysis provided a rigorous, yet inductive, set of procedures designed to identify and examine themes from textual data in a way that is transparent and credible [26]. The procedure for performing an applied thematic analysis had the following steps: (1) collect data, (2) transcribe conversations, (3) list patterns of experience, which can come from direct quotes or paraphrasing common ideas, 4) identify data that relate to already classified patterns, (5) combine and catalog related patterns into themes, and (6) formulate theme statements and develop a summary of findings.

3.2.5. Precision Medicine Use Case

Our precision medicine use case formative evaluation questions and semi-structured interview questions are shown in Table 1 and Table 1Table 2, respectively.

Table 1: Precision medicine use case formative evaluation questions.

Use Case Construct	Formative Evaluation Questions
Knowledge Source(s)	What are the publicly available (domestic or international) non-proprietary sources of information for Genome Variant – Clinical Impact knowledge, and how can these resources be used in patient care?
SOLOR System Integration	What are the methods of SOLOR System integration for gene variant – clinical impact knowledge, and what are their constraints?
Relevance	How does gene variant – clinical impact knowledge contribute to a precision medicine extension of the SOLOR system?

Table 2: Precision medicine use case semi-structured interview questions.

Use Case Construct	Semi-Structured Interview Questions
Knowledge Source(s)	• Does the ClinVar knowledge source used here seem like it could be useful in understanding gene variant – clinical impact? • Are there any additional sources that could be utilized? • Are there any sources that should not be utilized? If so, why not?
SOLOR System Integration	• Do you think this approach to integrating ClinVar is valid? • Are there ClinVar data elements that we didn't use but should use? • Are there other clinical terminology system relationships that can be used other than SNOMED CT? • What quality assurance/control issues should be considered? (i.e., should a genomic SME perform reviews)
Relevance	• Identifying gene variant – clinical impact knowledge sources for SOLOR system precision medicine applications (1 to 5 Likert scale) • Advancing genomic interoperability (1 to 5 Likert scale) • Presenting gene variant – clinical impact information to knowledge workers (1 to 5 Likert scale)

3.2.6. Medical Device Interoperability Use Case

To be completed as part of future deliverable.

3.2.7. Use Case 3

To be completed as part of future deliverable.

4. RESULTS

To be completed as part of future deliverable.

4.1. Precision Medicine Use Case (CLIN 2005B_07.14)

To be completed as part of future deliverable.

4.2. Medical Device Interoperability Use Case (CLIN 2005B_08.14)

To be completed as part of future deliverable.

4.3. Use Case 3 (CLIN 2005B_09.14)

To be completed as part of future deliverable.

5. CONCLUSION

To be completed as part of future deliverable.

5.1. Limitations of the Work

To be completed as part of future deliverable.

5.2. Suggestions for Future Work

To be completed as part of future deliverable.

REFERENCES

1 Aspden P, Corrigan JM, Wolcott J ES, editor. *Patient safety: achieving a new standard for care.* Washington, D.C.: : National Academy Press 2003.

2 The Office of the National Coordinator for Health Information Technology, Office of the Secretary USD of H and HS. FEDERAL HEALTH IT STRATEGIC PLAN 2015 – 2020. 2015.

3 IHTSDO. SNOMED CT Starter Guide. *Snomed* 2014;:1–56.

4 Regenstrief Institute I. Logical Observation Identifiers Names and Codes (LOINC®). Log. Obs. Identifiers Names Codes. 2017.http://www.loinc.org (accessed 6 Oct2017).

5 NLM. RxNorm Overview. 2013.http://www.nlm.nih.gov/research/umls/rxnorm/overview.html

6 Zhang* G-Q, Zhu* W, Sun* M, *et al.* MaPLE: A MapReduce Pipeline for Lattice-based Evaluation and Its Application to SNOMED CT. *Proc IEEE Int Conf Big Data* 2014;**29**:754–9. doi:10.1016/j.biotechadv.2011.08.021.Secreted

7 Stenson PD, Mort M, Ball E V., *et al.* The human gene mutation database: 2008 update. *Genome Med* 2009;**1**:1–6. doi:10.1186/gm13

8 Fokkema IFAC, Taschner PEM, Schaafsma GCP, *et al.* LOVD v.2.0: The next generation in gene variant databases. *Hum Mutat* 2011;**32**:557–63. doi:10.1002/humu.21438

9 Cook-Deegan R, Conley JM, Evans JP, *et al.* The next controversy in genetic testing: Clinical data as trade secrets? *Eur J Hum Genet* 2013;**21**:585–8. doi:10.1038/ejhg.2012.217

10 US Natl Libr Med. National Center for Biotechnology Information. ClinVar. 2012.www.ncbi.nlm.nih.gov/clinvar/ (accessed 28 Nov2018).

11 Riggs ER, Wain KE, Riethmaier D, *et al.* Towards a Universal Clinical Genomics Database: The 2012 International Standards for Cytogenomic Arrays Consortium Meeting. *Hum Mutat* 2013;**34**:915–9. doi:10.1002/humu.22306

12 Bickman L, Rog D. *Handbook Of Applied Social Research Methods [e-book]*. Thousand Oaks, CA: : SAGE Publications, Inc 1998.

13 Brown KG, Brown KG, Gerhardt MW, *et al.* Formative Evaluation: an Integrative Practice Model and Case Study. *Pers Psychol* 2002;**55**:951–83.

14 Salie S, Schlechter A. A formative evaluation of a staff reward and recognition programme. *SA J Hum Resour Manag* 2012;**10**:1–11. doi:10.4102/sajhrm.v10i3.422

15 Henry GT, Smith AA, Kershaw DC, *et al.* Formative Evaluation: Estimating Preliminary Outcomes and Testing Rival Explanations. *Am J Eval* 2013;**34**:465–85. doi:10.1177/1098214013502577

16 Dehar M-A, Casswell S, Duignan P. Formative and Process Evaluation of Health Promotion and Disease Prevention Programs. *Eval Rev* 1993;**17**:204–20. doi:10.1177/0193841X9301700205

17 Lavrakas P. *Encyclopedia of Survey Research Methods*. 2455 Teller Road, Thousand Oaks California 91320 United States: : SAGE Publications, Inc. 2008. doi:10.4135/9781412963947

18 Palinkas LA, Horwitz SM, Green CA, *et al.* Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Adm Policy Ment Heal* 2015;**42**:533–544. doi:10.1007/s10488-013-0528-y.Purposeful

19 Kvale S. *Doing interviews*. London: : SAGE Publications 2007.

20 Margaret C. Harrell; Melissa A. Bradley. *Data Collection Methods Semi-Structured Interviews and Focus Groups*. 2009. doi:978-0-8330-4889-9

21 DiCicco-Bloom B, Crabtree BF. The qualitative research interview. *Med Educ* 2006;**40**:314–21. doi:10.1111/j.1365-2929.2006.02418.x

22 Whiting LS. Semi-structured interviews: guidance for novice researchers. *Nurs Stand* 2008;**22**:35–40. doi:10.7748/ns2008.02.22.23.35.c6420

23 Aronson J. A pragmatic view of thematic analysis. *Qual Rep* 1994;**2**:3.

doi:10.4135/9781446214565.n17
24 Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006;**3**:77–101. doi:10.1191/1478088706qp063oa
25 Clarke V, Brown V, Hayfield N. Thematic Analysis. Qual. Psychol. A Pract. Guid. to Res. Methods. 2015;**9760**:222–47. doi:10.1037/13620-004
26 Guest G, MacQueen K. *Applied Thematic Analysis.* Los Angeles: : Sage Publications 2012.