Task Order #9 | October 29, 2018

**SOLOR Support Services: Use Case #1**

PRESENTED TO:

Dr. Keith Campbell & Stephanie Klepacki

IDIQ: VA701-16-D-0017

TO9 PO: 776-C80148

CLIN: 2005B Proposed Standards Artifacts (Medium)

PRESENTED BY:

Team BZ

Program Manager: Jayme Welty



360 Central Ave., Suite 970 | St. Petersburg, FL 33701

O 727.378.9006 | bookzurman.com

Table of Contents

[Introduction 3](#_Toc528587188)

[Statement of the Problem 3](#_Toc528587189)

[Aim 3](#_Toc528587190)

[Significance 3](#_Toc528587191)

[Methodology 4](#_Toc528587192)

[ClinVar Knowledge Source 4](#_Toc528587193)

[ClinVar Implementation 4](#_Toc528587194)

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 desperate 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, (**S**ystem of **Lo**gical **R**epresentation) 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.

Statement of the Problem

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. 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.

Aim

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 Aim of this deliverable is to develop 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.

Significance

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.

Methodology

The ClinVar knowledge source is added to the SOLOR ecosystem using a transformation process which allows for ClinVar specific data representation within the SOLOR ecosystem.

ClinVar Knowledge Source

An analysis was conducted across several genomic data resources such as ORIEN[[1]](#footnote-1), M2GEN[[2]](#footnote-2), and ClinVar[[3]](#footnote-3) with the goal of determining the most viable resource to be used in the SOLOR genomic use case. Below is the breakdown of specific considerations that best qualified ClinVar as the selected data source for the SOLOR genomic use case:

**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[[4]](#footnote-4), which provides data in a structure similar to relational database tables, was identified as the easiest data format to be used in the SOLOR transformation process.

**Documentation** – Various ReadMe files were identified within each ClinVar release, which described in detail every data point contained within the overall ClinVar release data structure. Based on these descriptions, reliable inferences were made in creation of 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 updated more frequently results in less stale, more accurate, and more reliable health data to be incorporated into the SOLOR ecosystem.

**Data Structure** – Specific data entities, such as variant, gene, and disease, were normalized, modular, and isolated from other more complex entity relationships. These aspects for such key data entities resulted 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 ecosystem, enables proper terminology concept quality assurance and classifications to be performed on all SOLOR health data.

ClinVar Implementation

Incorporating ClinVar knowledge source into the SOLOR ecosystem requires a custom implemented transformation process, which focuses 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 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. Ultimately, a comprehensive SOLOR taxonomy was created incorporating both ClinVar and SNOMED CT concept.

1. [Online] ORIEN Homepage, Available: <http://oriencancer.org/> [↑](#footnote-ref-1)
2. [Online] M2GEN Homepage, Available: <http://m2gen.com/> [↑](#footnote-ref-2)
3. [Online] ClinVar Homepage, Available: <https://www.ncbi.nlm.nih.gov/clinvar/> [↑](#footnote-ref-3)
4. [Online] ClinVar TSV Data Repository, Available: FTP: ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/tab\_delimited/ [↑](#footnote-ref-4)