**Formative Evaluation of a Solor Extension Use Case:**

**Managing Genomics Data**

**Background:** The Veterans Health Administration’s (VHA’s) Informatics Architecture division 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 (formerly the 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, therefore leading to improved patient care.

**Motivation:** One area for which Solor can assist is within the genome research domain. Even though precision medicine has become a national priority, genetic data knowledge sources are not structured or maintained in a format usable for Electronic Health Records (EHRs), clinical decision support, research, or interoperability. 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 in a standardized format – with important implications for treatment and research. 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 privately-controlled knowledge bases; open access, locus-specific knowledge bases; proprietary knowledge bases; and publicly available, centrally-managed repositories, such as ClinVar. 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.

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 EHR, enable more insightful population health statistics if in a standardized format, and contribute to more impactful research if interoperable.

**Aims:** The overarching objective of this work is to inform the development of Solor by exploring its extension as an ecosystem for integrating disparate genomics knowledge sources and creating interoperability by making information meaningful and computable. The specific aims of this work are to (1) describe a use case for the extension of Solor to support the management of genomics data from ClinVar, and (2) evaluate constructs of the Solor use case developed in the previous aim.

**Methods:**

*ClinVar Knowledge Source:* ClinVar, a publicly available central resource managed by the National Library of Medicine (NLM), represents a model wherein genome knowledge sources can upload their expertly curated knowledge into one location. 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.

*Solor Integration:* We integrated data from the ClinVar knowledge source into the Solor Editor and created a common model, allowing for a ClinVar specific data representation within the Solor ecosystem. We specifically used data from the following ClinVar data files: (1) Variant Summary with information that describes gene variants submitted to ClinVar, including NCBI gene ID; (2) Gene Specific Summary with attribute information on gene identifiers, the NCBI ID and its symbol data elements; and (3) Gene Condition Source ID with relationships between genes and correlating diseases (i.e., phenotypes, SNOMED CT ID) used in ClinVar.

*Design:* Next, we performed a formative evaluation of use case constructs with a purposive sample of informaticists with expertise in medical terminological systems and genomics. Semi-structured interviews were conducted with participants to identify key information about each expert’s background, experience with genomics data, and their insights about enabling constructs, challenges, and promising components in the genomics environment. The goal of this formative evaluation was to collect rapid subject matter expert feedback that would provide validation of use case constructs and context for future successive adaptations and improvement of the use case’s development key questions for evaluating a new proof-of-concept use case, which included: (1) does the idea provide a new and more useful capability?; (2) does it help developers better understand complex systems?; and (3) does it demonstrate by its behavior that a complex assembly of components can accomplish a particular set of activities?

*Analysis*: We used applied thematic analysis, a method for identifying and analyzing patterns of meaning in a dataset, to organize and describe the data collected from the interviews. 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. 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. The data was coded independently and reviewed as a group. Themes were identified using an immersion-crystallization approach.

**Results:** We interviewed three subjects with backgrounds in medical informatics and/or medical terminologies, with varying experience working with precision medicine and the National Institutes of Health (NIH). Participants had leadership and technical roles with 15 to 20 years of experience in the domain of precision medicine with knowledge of healthcare standards, terminologies, knowledge commons, and genomic databases. All subjects had experience with precision medicine, ranging from 1 to 13 years, with a mean of 5.67 years.

Participants agreed that ClinVar was an appropriate knowledge source and that it is a good first stop to find genomics data. However, more research must be done to validate the use of the ClinVar knowledge source compared to other existing genomic datasets. The subjects also tended to agree that the Solor Editor successfully integrated the ClinVar knowledge source to effectively demonstrate the connections between genes, variants, and disorders. However, two very important suggestions were to continue to improve the Solor extension in this use case, and to get an early adopter to use this tool in a real-world setting. All subjects agreed that the precision medicine use case for Solor is highly relevant, with many potential improvements that will make it more clinically useful. These include, but are not limited to, treatment plan support and gene-to-gene correlations. Suggestions for next steps include further refining our precision medicine extension to also assist physicians with navigating clinical decision support to better understand how genes might be correlated to each other, and how these correlations might help them in choosing more tailored treatment decisions for patients.

Figure 1 shows a screenshot from our current user interface of the Solor Editor with our build to assist the precision medicine use case. As shown in the box in the bottom left corner, Solor’s uniform visual representation helps highlight the relationships between variants to genes (e.g., ACE) to disorders (e.g., Alzheimer’s disease).

**Figure1. Current User Interface of Solor Editor, with Genomic Extension**

**Conclusion:** Our subjects agreed that Solor’s integration of the ClinVar knowledge source resolved challenges with managing information about genes, variants, and disorders. Further work with additional genomics knowledge sources and enhanced clinical decision support functionality may be required.