Knowledge Integration of Software for HIV Drug Resistance Research

Genaro Hernandez Jr*, Soo-Yon Rhee**, Robert W. Shafer** and Amar K. Das*

*Stanford Medical Informatics, Stanford University, Stanford, CA 94305
**Division of Infectious Diseases, Department of Medicine, Stanford University, CA 94305

Abstract
The Stanford HIVdb is implementing three applications to address the issue of HIV drug resistance. Knowledge integration of the applications is desirable. How best to integrate the software applications is the main question. We have pursued two possibilities for accomplishing knowledge integration. The most feasible method appears to use mapping ontologies. Future work is focused on defining a set of mapping relations to aid in the knowledge integration.

INTRODUCTION
The Stanford University HIV Drug Resistance Database (HIVdb) is a curated public database designed to represent, store, and analyze the divergent forms of data underlying HIV drug resistance [1]. Current efforts to accomplish the aforementioned entail the integration of three HIVdb applications. Those three applications are ChronoMiner for data mining of drug resistance patterns in HIVdb, Refdb for annotating HIV drug resistance literature, and eCARE for the treatment of patients when drug history data are available as opposed to genotype data.

How best to integrate ChronoMiner, Refdb and eCARE is the informatics challenge. eCARE is interesting in that it takes an antiretroviral therapy (ART) from a user and generates a summary of the mutations and drug resistance observed in sequences obtained from patients in HIVdb who received similar ARTs. Ordinarily one starts with an HIV genotype in order to infer drug resistance [2, 3]. eCARE is the central application since the source of its output can be either Refdb or ChronoMiner. ChronoMiner will provide a summary of drug resistance patterns to eCARE if there are a sufficient number of patients in HIVdb. Refdb will provide relevant drug resistance literature if there are not enough data in HIVdb.

Ontology is the informatics methodology we are using to integrate the three applications. This is primarily due to the use of knowledge in the three applications. ChronoMiner, Refdb and eCARE all rely on a distinct and corresponding ontology for their output.

METHOD
We are pursuing two ways of applying ontologies for the integration of ChronoMiner, Refdb and eCARE. First is possibly using a single core HIV ontology that would rely on the same HIV ontology to generate their output.

The second idea we are pursuing involves mapping ontologies. In this approach we would use an ontology-based mediation of knowledge.* Thus a set of mapping relations would have to be defined in order to bridge gaps between different ontology components and to transform instance knowledge and data between ontology components.

RESULTS AND DISCUSSION
We have discovered that a mapping ontology is the best option for integrating the knowledge used by three software applications used in HIV drug resistance research. After examining the distinct ontologies upon which the three programs rely on, a core ontology seems unfeasible because the ontologies are constructed differently. ChronoMiner for example, uses an ontology where HIV drugs are modeled as classes and subclasses. A more specific ChronoMiner example is that a drug class has a protease inhibitor subclass which in turn has protease inhibitor subclasses (e.g. saquinavir (SQV)). eCARE on the other hand uses an ontology where drugs are represented as instances of an ARV class. A more specific eCARE example is that SQV is an instance of the ARV class and SQV is stored as a type of protease inhibitor. Future work will involve the development of a set of mapping relations to map knowledge from ChronoMiner to eCARE and from Refdb to eCARE. This work will support the efforts of HIVdb to address the issue of HIV drug resistance.

References