Ontology-Driven Mapping of Temporal Data in Biomedical Databases
Preetha S. Narayanan, Martin J. O’Connor, MSc, and Amar K. Das, MD, PhD
Stanford Medical Informatics
MSOB X233, Stanford University, Stanford, CA 94305

Abstract
Biomedical databases contain considerable amounts of time-oriented data, which typically are not in a format suitable for querying complex temporal patterns. We address this problem in implementing Synchronus, a tool for ontology-driven mapping of data from an existing relational database to a database schema with a uniform temporal representation. We discuss the design of Synchronus, which consists of a schema-mapping ontology and a data-mapping algorithm that together provide general capabilities for database transformation.

INTRODUCTION
Biomedical data are frequently stored with a variety of temporal representations, which hinder the ability to perform queries over their temporal dimensions. In our prior work on Synchronus [1], we developed a set of database operators to map data with temporally heterogeneous formats into a uniform temporal schema that supports temporal querying of arbitrary complexity. This approach was limited by the use of the relational data model as the basis of the temporal metaschema (limiting its expressivity) and by undertaking data mapping only on a per query basis (leading to overhead costs in recurrent queries). We address these challenges with a new method for Synchronus that uses a formal knowledge model.

METHOD
We use the Protégé OWL tool (protege.stanford.edu) to implement a new approach for Synchronus. We evaluated the new tool using, as the source schema, the schema of the Stanford HIV Drug Resistance Database (HIV DB) [2], and, as the target schema, the valid-time format of Chronus II [3].

We first define a schema mapping ontology that provides a hierarchical representation of concepts related to the structure and properties of relational databases and temporal relational models, along with conceptual mapping of the two schemas. Each instance of these concepts corresponds to a map. An instantiation of these concepts specifies an OWL knowledgebase for every relational database that is mapped. The ontology has two parts:

1. Schema, which can encode the schemas of a relational database and the temporal relational model using classes that we have defined: RelationalDatabase, Schema, Table, Column, JDBCDataType, and Granularity

2. SchemaMapping, which describes the mapping between the schemas of a relational database and the temporal relational model, using the classes of DBSchemaMapping, TableMapping, ColumnMapping, and ColumnMappingFunction

We use knowledge encoded in this format to drive our data-mapping algorithm. This algorithm uses Jastor (jastor.sourceforge.net) to generate Java class files corresponding to OWL classes, and then uses the generated Java class files to access the instances of the OWL classes present in the knowledgebase and convert the data according to the mapping rules described in the ontology. The main features of our algorithm are a Data Conversion Manager, which goes through all the table mappings to obtain the source and target database details, and a Data Converter, which processes instances of table mappings to transform data from the source format to the target format and store in the target database.

RESULTS AND DISCUSSION
To validate our method, we created a knowledgebase, called HIVDBChronusKB, to contain instances of the HIV DB and Chronus II database schemas and the mapping between them. We applied Synchronus in batch mode to transform HIV DB tables containing from 45,425 to 126,340 rows into a Chronus II database. We tested accuracy using relational joins.

Although Synchronus was developed primarily to support the transformation of heterogeneous temporal data, its use of a knowledge model and general data-transformation algorithm provides a robust, flexible method to transform relational data from one format to another. Open-source code for our work will be available at chronus.stanford.edu.

References