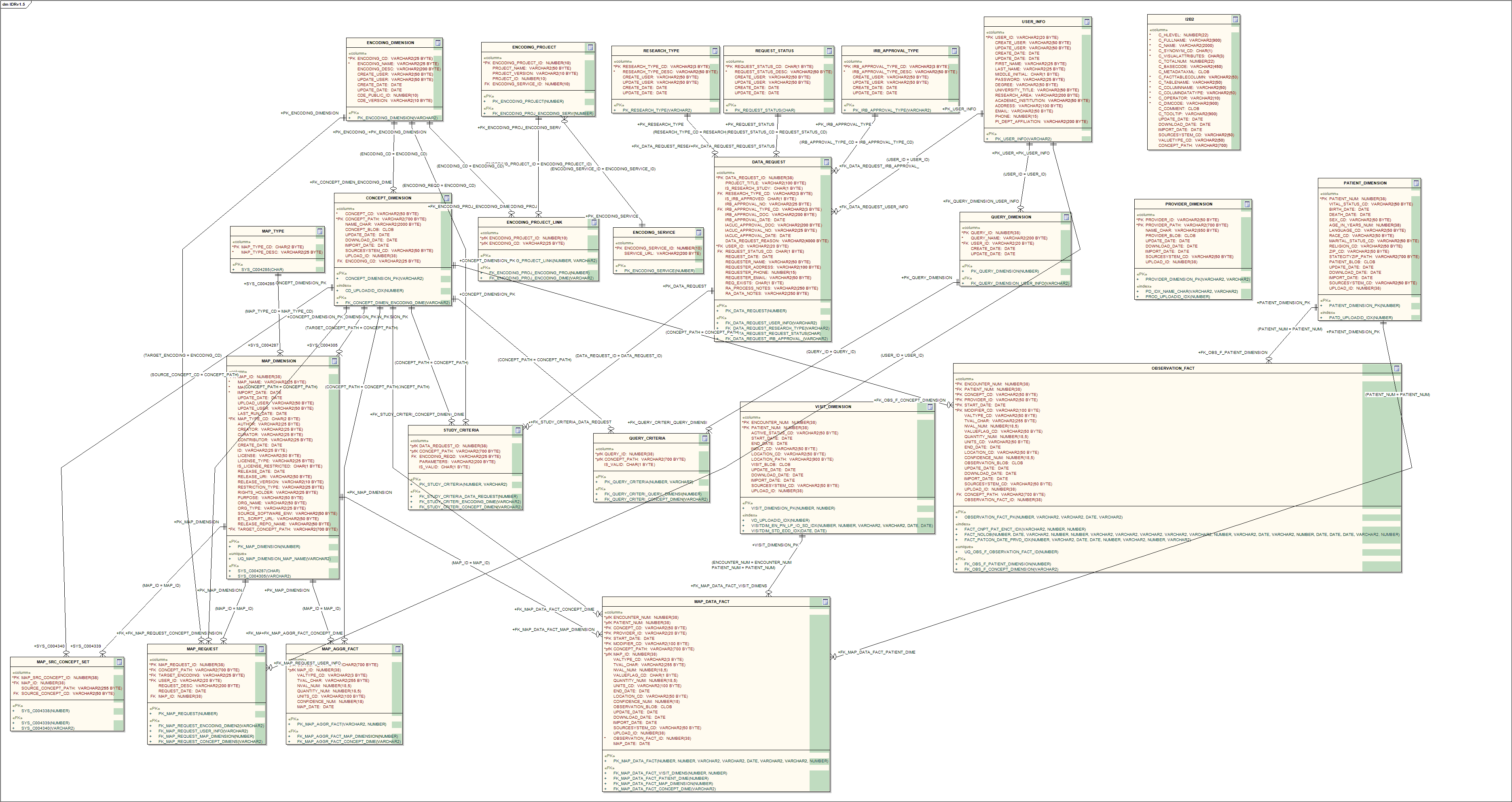
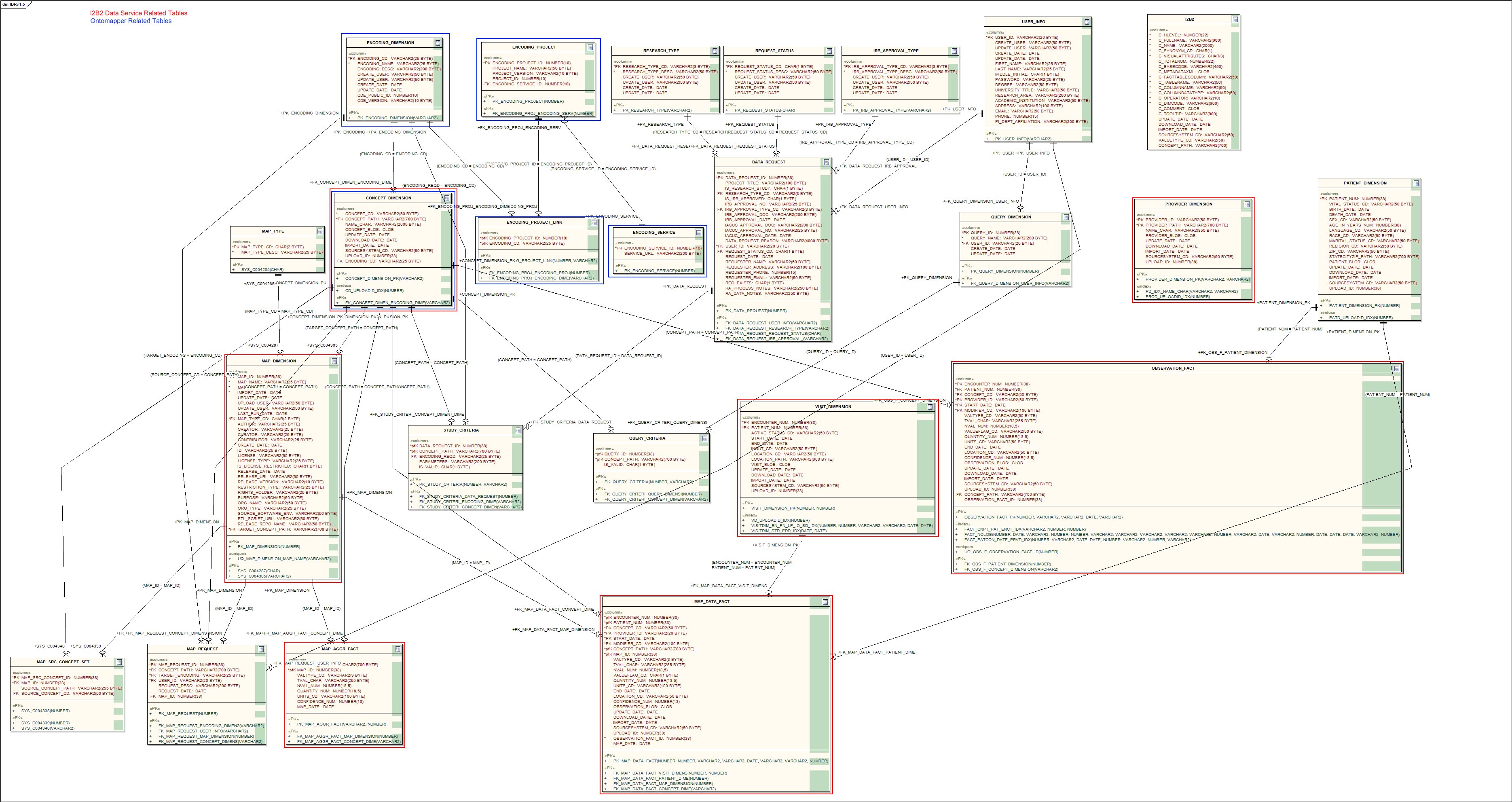
I2B2 as a caGrid Data Service

# Database

The complete i2b2 + ontomapper database schema:



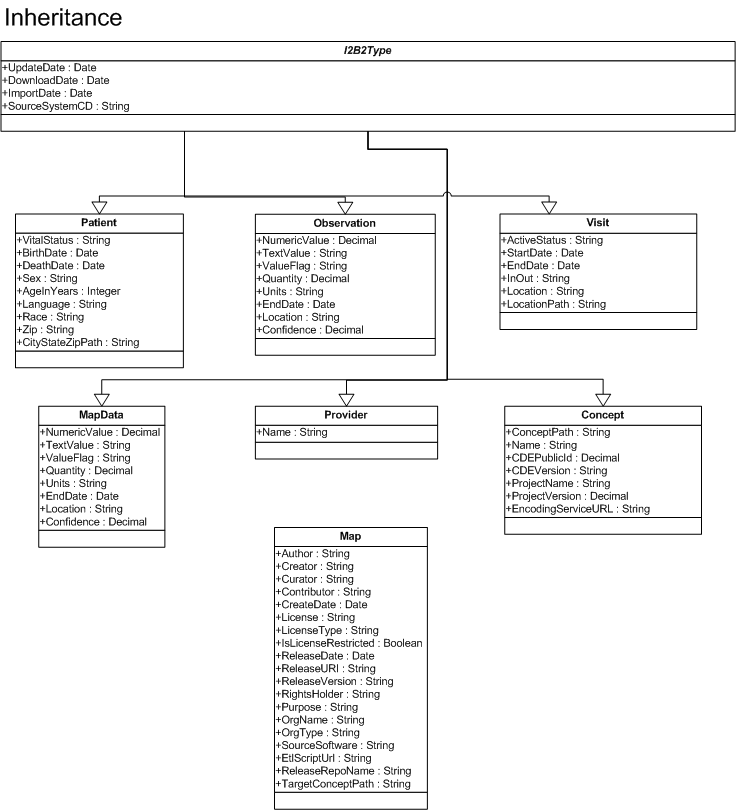
Not all of these tables are relevant to the proposed i2b2 data service model.



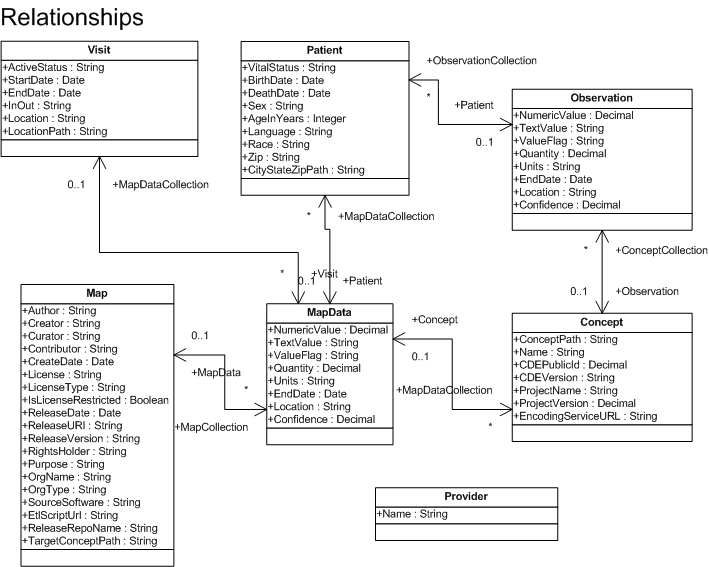
Tables highlighted in red represent tables which pertain to the i2b2 data service domain model and should be exposed via CQL to the grid. Tables in blue pertain to the Ontomapper. The CONCEPT\_DIMENSION table is highlighted with both colors, and provides a means of connecting the concept mapping capabilities of the Ontomapper with the clinical data in I2B2.

# Object Model

The Inheritance hierarchy for the I2B2 data service data model:



The class relationship model:



Many of the tables in the i2b2 database schema collapse into single datatypes by way of 1:1 join relationships in the model. For example, the Concept type draws data from the Concept\_Dimension, Encoding\_Dimension, and Encoding\_Service tables to fully describe a concept path, a corresponding CDE, and that CDE’s origin.

*I2B2Type* is an abstract base type which encapsulates some basic information about a data instance, such as when it was loaded into the datastore, and the source it came from.

TheMapData type allows for additional arbitrary data to be stored which pertains to one of the associated data type instances. The specific type of data stored in the MapData instance is specified via its association to Concept.

# Ontomapper considerations

The Ontology Mapper software by UCSF provides the facility to map caBIG CDEs to I2B2 concept paths and query and retrieve data accordingly. In cases where the Ontomapper is not used and a standard I2B2 installation is present, the Concept datatype will contain no data, and queries against it will simply return an empty set. Determining it usage could be done with a configuration property of the data service, or even discovered from the database itself with a handful of appropriate probing SQL statements. When the Ontomapper is used, a new and powerful means of querying becomes available as a client can formulate statements using semantic concepts in addition to standard logical operators. For example, a client, knowing the caBIG concept code for “cancer” could formulate a query to determine the exact visit date when a patient was diagnosed. Such knowledge might be useful for identifying a cohort for a clinical trial.