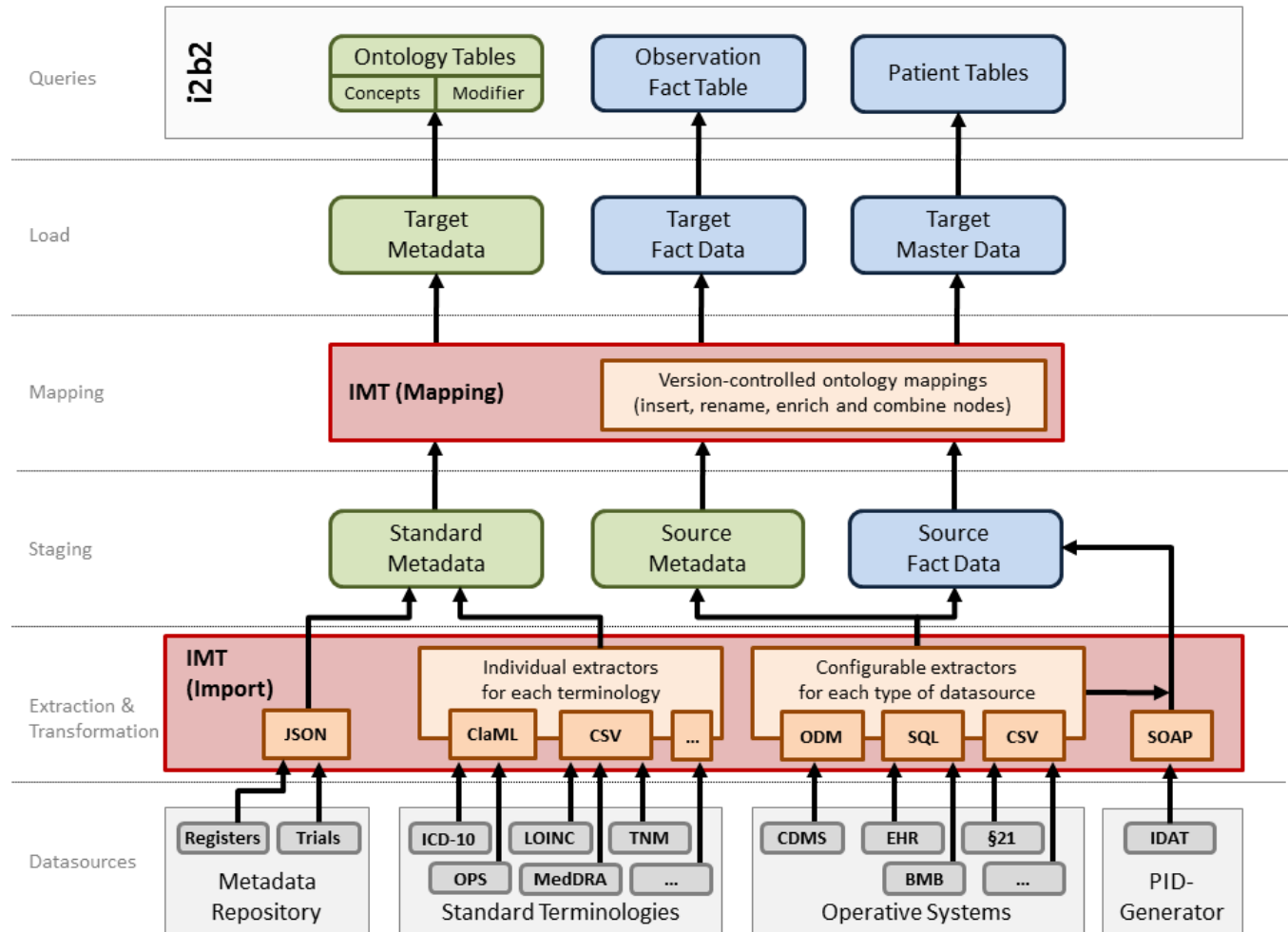


110. IDRT Platform Architecture

The Integrated Data Repository Toolkit (IDRT) was designed as an extension of the i2b2 platform, trying to fill gaps collected with a German i2b2 user survey. IDRT interacts with the established i2b2 software stack using different methods such as utilizing the [i2b2 Cell Messaging](#) or directly writing to the i2b2 database.

The architecture representation consists of six layers: from data sources (bottom) to queries. A big focus of the IDRT are the lower three layers: extracting and transforming data from a variety of data sources and storing them in the i2b2 database with the import module of the [IDRT Import and Mapping Tool \(IMT\)](#). Although all the data at this staging layer is already "i2b2 ready", the mapping and loading layer, representing the [Mapping Tool](#) of IMT, enables modification of the data and i2b2 ontologies for a better user experience at the queries layer. Underlying this data driven architecture is the [i2b2 Wizard](#) (not in representation) for installation and configuration of i2b2 installations. On top of the queries layer is the [IDRT Web Client Plugin](#).



ETL (extraction, transformation and loading)

All the import jobs of the IMT fulfill two tasks when doing the ETL (with terminologies only doing the latter): converting the patient data for filling the i2b2 database schema (*observation_fact*, *patient_dimension*, ...) and creating matching i2b2 ontologies based on data from the source, the data itself and information from configuration files. These i2b2 ontologies then are used to populate the i2b2 ontology tables (*i2b2*, *concept_dimension*, *modifier_dimension*) and define their links in the *observation_fact* table. All interactions with the i2b2 database are done with direct SQL statements and are handled by a single job (which needs updating for i2b2 version updates).

With ETL being a dominating part of the distributed development of the IDRT, data integration software was used. The Open Source software [Talend Open Studio](#) for Data Integration gives the user access to a multitude of import/export components. These can be strung together with data manipulation components to an ETL "job" in a visual interface. Talend Open Studio is a code generator that automatically converts these jobs to Java code, which can be run even outside the design software.