

Overview on i2b2 to Extract Clinical Data for Research

Vineeth Ramasahayam¹, Ashok Mudgapalli², PhD, Carol Geary², PhD, Ann Fruhling¹, PhD, James McClay², MD
¹University of Nebraska at Omaha, ²University of Nebraska Medical Center

Introduction

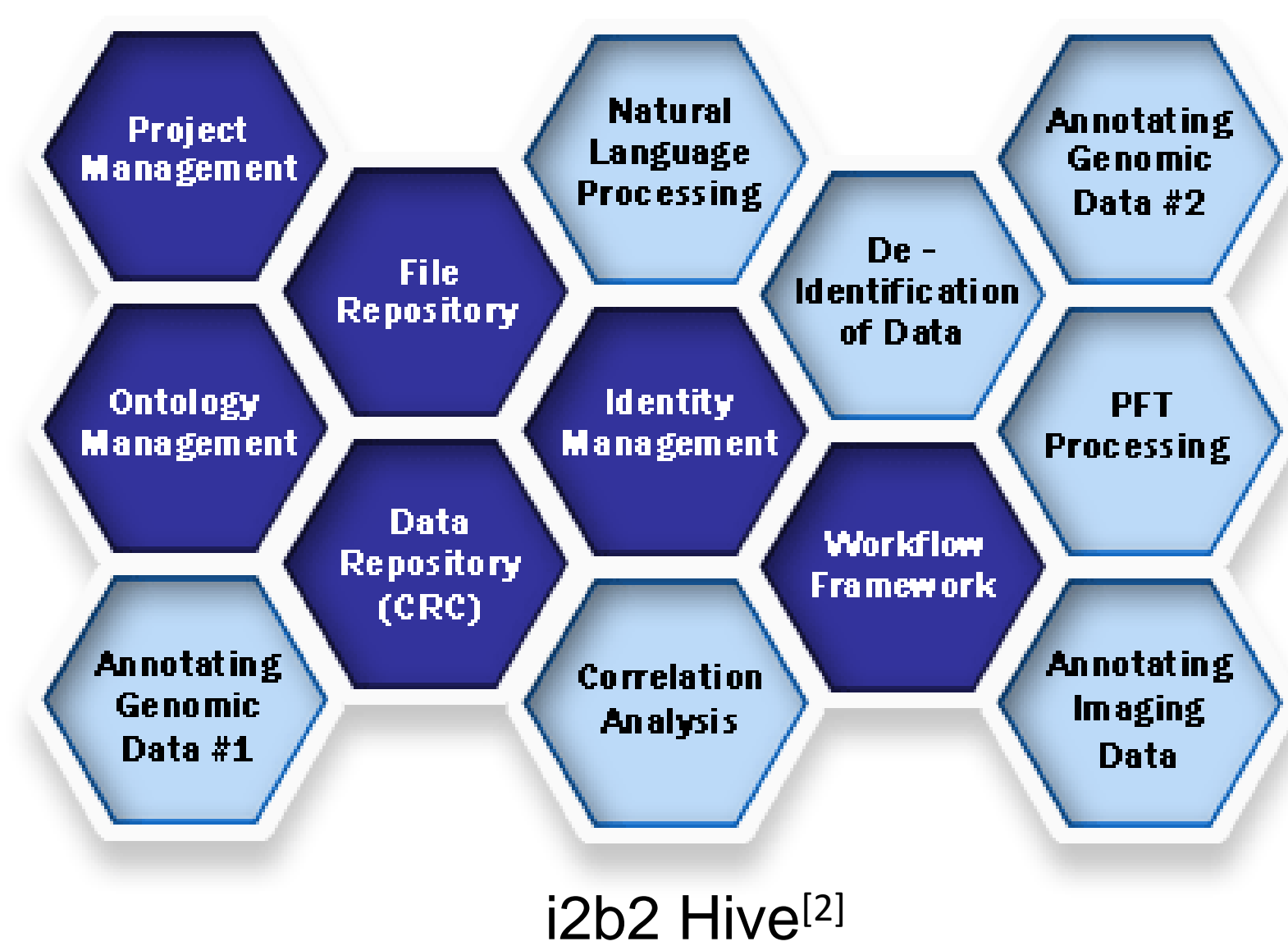
i2b2 (Informatics for Integrating Biology and the Bedside, <http://www.i2b2.org>) is an NIH-supported National Center for Biomedical Computing that developed a scalable informatics framework designed for translational research [1].

The i2b2 platform allows an institution to integrate a wide variety of clinical data sources into a single repository, and it supports the integration of genomic data for clinical and translational research.

i2b2 was designed primarily for cohort identification, allowing users to perform an enterprise-wide search on a de-identified repository to determine the existence of a set of patients meeting certain inclusion or exclusion criteria.

The i2b2 framework comes with either a web-client or workbench application that allows the user to drag-and-drop search terms from a hierarchical ontology into an interface to generate patients' cohort. Researchers can perform analysis on the de-identified cohort and conduct research to find solutions to various problems that currently exist in the field of medicine.

Structure of i2b2:



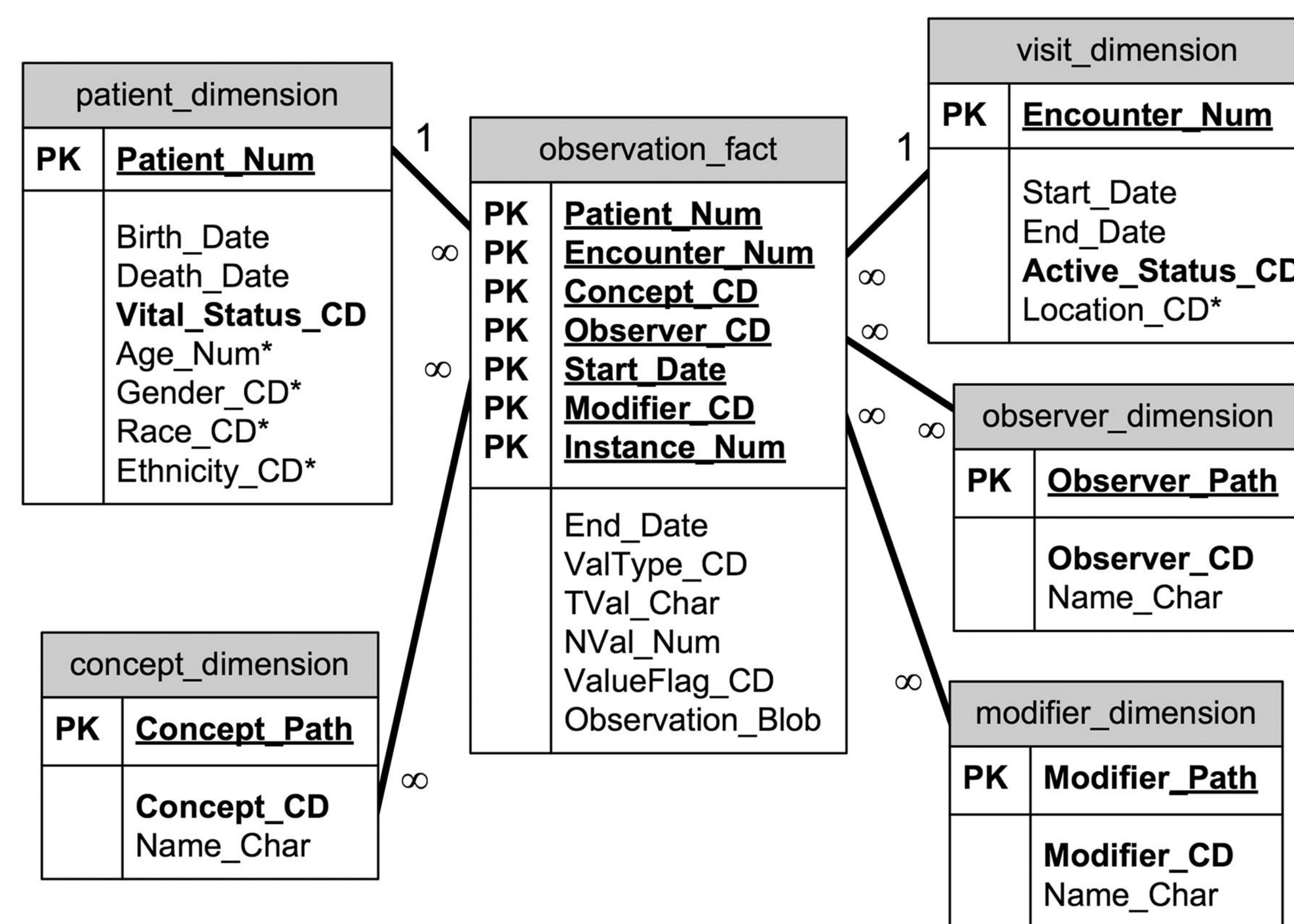
i2b2 Web Client or Workbench is connected to i2b2 Hive which in-turn is connected to the database containing the data. i2b2 Hive is a combination of micro-services, known as Cells. The i2b2 Cell is the basic building block of an i2b2 environment, and encapsulates business logic as well as access to data objects behind standard Web interfaces.

Each and every Cell has its own schema and is not connected to other Cells and doesn't know about each other. Each Cell is designed as a web service that interacts with i2b2 Workbench/Web-Client plugins.

The Cells shown in blue in the above picture are core Cells and the other Cells are the optional ones. In addition to these we can further add additional Cells to add functionalities like Web Client, CRC Plug-ins, Work Bench Plug-ins etc.

Data Storage in i2b2

All the data corresponding to the i2b2 is stored in the Star Schema format. A large central fact table usually named as "observation_fact" is surrounded by and connected to the smaller dimension tables, i.e., the patient, observer, visit, concept, modifier etc. The data in these tables is saved in a format known as Entity-attribute-value (EAV) format. The EAV format enables fast and accurate access to data.



Data Extraction

Data in Star Schema format cannot be interpreted properly by statistical analysis tools like R or SPSS as well as data mining/visualization tools for performing any computations.

The required data needs to be converted to a format called 1NF (1st Normal Form) which can be easily read by these tools for researchers to perform any computations as well as any visualizations on the data.

i2b2 has plugins available for workbench and webclient which can convert the data from star schema to 1st Normal Form.

The identified plugins suitable for this job were ExportXLS, Data Finisher, IDRT i2b2 Web Client Plugin and Generic Case Extractor.

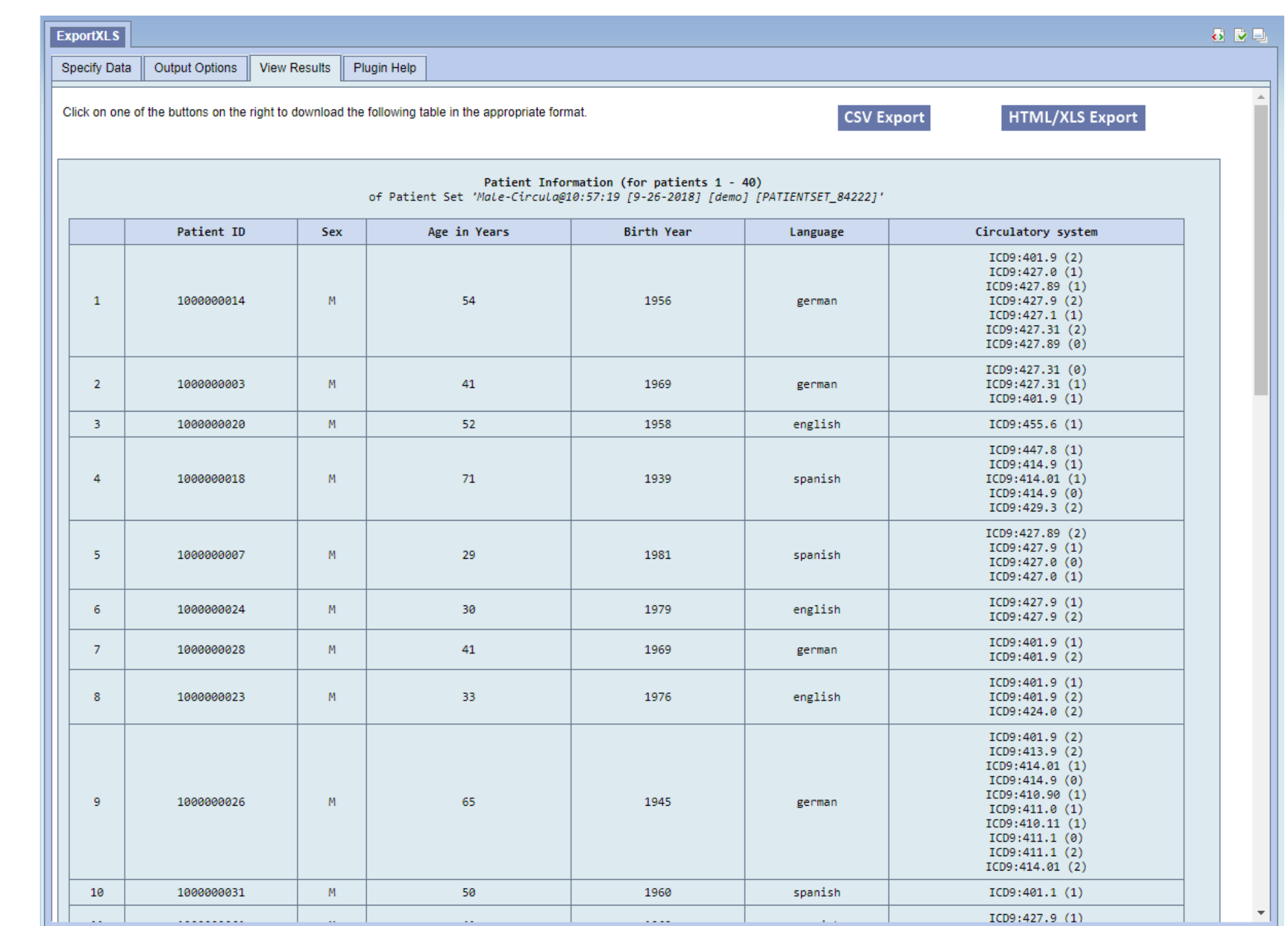
ExportXLS plugin was chosen as it was the most widely used plugin because of its simple user interface and accuracy. It has the capability of converting a patient dataset generated by the i2b2 query into CSV or HTML/XLS format

Implementation

ExportXLS plugin tabulates unidentified patient data of selectable concepts or observations and provides options to download the result to specified file type.

It can convert the data from EAV to CSV and HTML/XLS formats. ExportXLS allows multiple types of formatting depending on the requirement.

Only selected people who are approved to export the data can be given access to the plugin.



ExportXLS User Interface^[4]

Conclusion

Making the data interoperable between various platforms is one of the critical problems that must be solved for the i2b2 platform to make it more effective.

Plugins like ExportXLS will enormously save the time and effort statisticians spend on extracting the data manually from the database.

This attempt which we made will fill the gap between i2b2 and the statistical tools by converting the database files to CSV format

References

1. i2b2 Overview. Retrieved from. <https://www.i2b2.org/about/>
2. i2b2 Hive. Retrieved from. <https://www.i2b2.org/resrcs/hive.html>
3. ResearchGate: i2b2 Star Schema. Retrieved from. https://www.researchgate.net/figure/The-i2b2-star-schema_fig1_232698901
4. i2b2 Community Wiki: ExportXLS Home. Retrieved from <https://community.i2b2.org/wiki/display/ExportXLS/ExportXLS+Home>

The project described was supported by the National Institute Of General Medical Sciences, 1U54GM115458. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH.