

Vision for Technical Infrastructure to Facilitate OHDSI Network Research

Nikolai Puntikov, PhD, Gleb Malikov, MSc
ODYSSEUS DATA SERVICES INC, Cambridge, MA

Abstract

OHDSI framework reveals great potential in conducting scientific research based upon statistical evidence provided by collaborators. However, the productivity of such research is adversely affected by lack of intelligent automated workflow facilitating secure access to patient data. We propose a model of system architecture that enables highly efficient observational network research within OHDSI.

Introduction

OHDSI's Vision is to allow collaborators to access a network of patient data to generate evidence about health care. To realize this vision, OHDSI needs to bring together Data Providers, Clinical Investigators, and Data Scientists (statistics, software, etc).

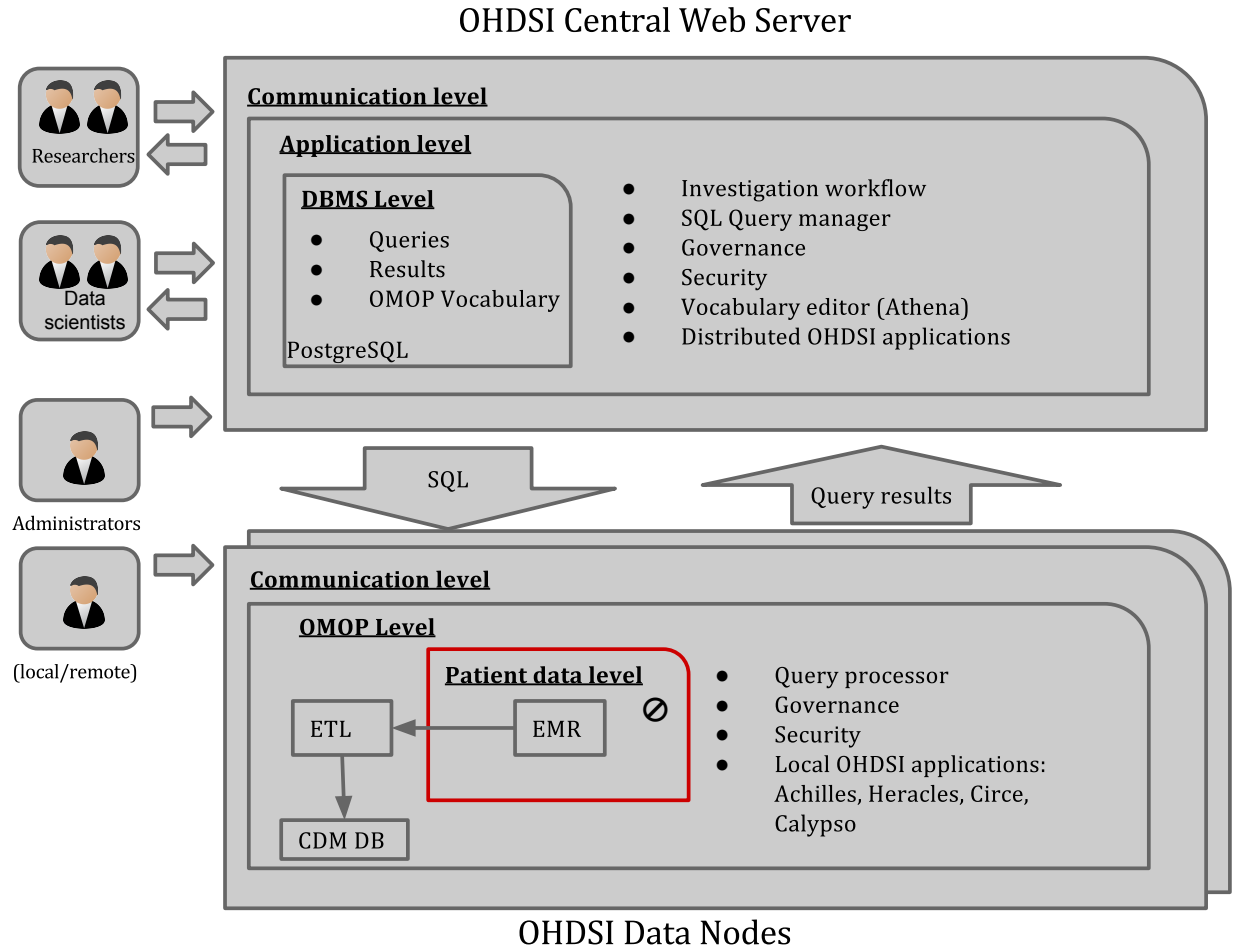
What does "bring together" mean from a technical perspective?

1. Investigators propose research studies;
2. Data scientists create programs, such as SQL queries and statistical analysis;
3. SQL queries are executed against distributed, standardized Data (OMOP CDM);
4. Query results serve as evidence for investigators' research; and
5. Results of research are published.

Currently, the above is only possible using manual mechanisms. See example referenced in 2: E-mail is used to sent a query to data owner; the query is manually executed by a DBA on the Data Owner's side; results are sent back to investigator by e-mail.

We propose Arachne, an infrastructure solution for distributed query execution, including error handling and query refinement, and automated result aggregation and distribution. Arachne includes mechanisms for data protection and corporate governance, and supports a multitude of technical environments (OS, SQL engines, etc.).

Arachne System Architecture



The central OHDSI server provides several web applications:

Investigation workflow manager is an application, which is responsible for registering and initializing new research projects, followed by execution of all necessary steps required to complete the project. This application communicates with the project owner (investigator) and other collaborators involved in the project, by sending appropriate web forms, notifications, and reminders. Depending on conditions defined by the project properties, the workflow manager invokes other server applications and communicates with applications at the Data Nodes.

SQL query manager performs actual query or queries in cooperation with participating Data Nodes. It's functionality includes submission of queries by data scientists; translation (when required) of queries into all supported SQL dialects; distribution of queries between the Data Nodes; handling error messages (as may be the case) sent by the Nodes; accepting, storing and aggregating results of the query or queries preparing them for investigator or for further processing by data scientists.

Governance module is responsible for maintaining a contract between particular Data Nodes and OHDSI Network. The contract has permanent covenants, such as contacts of staff responsible for approval of the Data Node participation in any given research; and individual project information, such as agreement to run SQL query (or queries), approved level of details in the result, and conditions to approve results (automatic or manual). After the governance step is completed, the workflow manager provides information about participating Data Nodes to the SQL query manager.

Security on the server side is managed by Apache Shiro Security with Resource Based Access Control. Only authorized users can view each query result. Resulting data sets are being encrypted before sending from the Data Node, and then decrypted on the arrival to the OHDSI Web Server with the unique key, shared between project and Date Node.

The **Data Node** applications communicate with the OHDSI server and perform required actions:

Query processor receives ready-to-execute SQL query (or queries) and runs it against CDM database. The resulting tables are ready to be sent back to the server, subject to approval by the Data Node administration (automatic or manual).

Governance module provides interface to set project properties related to the particular Data Node, to verify resulting data and approve sending the results back to server.

Security at the Data Node starts from ETL, which anonymizes patient data already when converting records into OMOP CDM. The Arachne applications have no access to the EMR/EHR of the Data Node. User authorization and access control can be managed locally or remotely from the OHDSI server.

Implementation of Arachne would allow seamless deployment of both **local and distributed OHDSI software applications** that use data in CDM format. The solution provides total suite of security and governance measures and convenient user interface, allowing data scientists to focus on essential tasks at hands.

Arachne Database over time will become a valuable resource per se. Results of all queries (if permitted by the Data Node) will be stored centrally and made available to collaborators and data scientists. There is obviously potential for additional statistical analysis going beyond goals of original research. The SQL queries will also be stored, and could be run periodically offering dynamic view at the observational evidence.

Conclusions

We believe that implementation of Arachne will become a key factor in enabling efficient observational research. OHDSI already achieved great results by providing:

- Standardization of data formats and coding (vocabularies);
- Standardization of methods;
- Standardization of cohort definitions; and
- Collaborative network of renowned scientists in the areas of clinical research, statistics, method development and scaling.

Arachne aims at removing a remaining obstacle of technical execution of studies across a network of distributed databases: each contributor to a study shall be able to focus on the scientific problem, rather than dealing with SQL queries and compatibility challenges (SQL dialects, technical infrastructure, security, etc).

The fully functional Arachne:

- saves researchers significant time and effort;
- streamlines observational studies; and
- enables large scale studies that are not feasible today.

References

1. OHDSI: <http://www.ohdsi.org/>
2. OHDSI Study #1: Treatment Pathways in Chronic Disease, <http://forums.ohdsi.org/t/ohdsi-study-1-treatment-pathways-in-chronic-disease/112>
3. PopMedNet: <http://www.popmednet.org/>