Name:	Gregory Klebanov
Affiliation:	Odysseus Data Services, Inc.
Email:	gregory.klebanov@odysseusinc.com
Presentation type (s):	• Software Demonstration • Poster

ARACHNE –
Distributed OHDSI Research Network and Study Workflow Orchestration

Gregory Klebanov, MSc / CTO; Nikolai Puntikov, PhD / CEO; Pavel Grafkin, MSc / Technical Lead
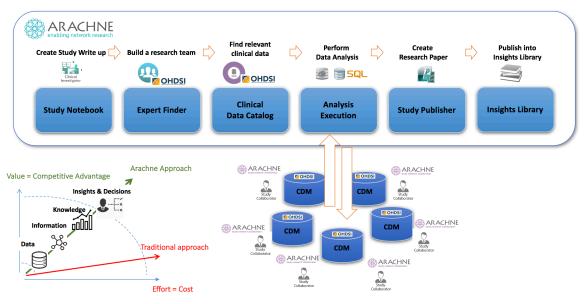**Odysseus Data Services, Inc., Cambridge, MA**

Abstract

OHDSI community, standards and platforms enable great potential in conducting scientific research based upon statistical evidence provided by collaborators. However, the consistency, quality and productivity of such research is currently adversely affected by lack of a scalable solution that enables conducting federated observational studies, including life cycles and workflow management, analysis execution and remote secure and compliant access to patient level data. Odysseus has developed ARACHNE, a platform for efficient observational research across a network of participants within the OHDSI community and other life sciences, healthcare, academic, payer and other organizations working with patient level data.

Introduction

OHDSI's mission is to improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care. To realize this mission, OHDSI brings together data providers, investigators and data scientists (statistical analysts etc.) through multicenter collaborative observational studies. A typical study states a hypothesis requiring testing, a protocol outlining design, parameters and process, a number of data sets containing sufficient number of patient level data, analytical code developed to interrogate data and compute statistical evidence and a final research paper published containing study insights and conclusions.

Up until ARACHNE, a process of conducting such study has been quite manual and inconsistent, requiring a significant effort and a number of different, unlinked communication channels, including email exchanges, forum and chats. Once in the execution phase, even with a great adoption of OMOP Common Data Model (CDM) by participating organizations, adopting code to different environments often requires tweaking, with changes and activities are typically not tracked. The statistical code and related results are not linked, and often stored in personal folders on scientist's file system.

ARACHNE was built to solve these challenges. It establishes a consistent, transparent, secure and compliant observational research process. It brings participating organizations, e.g. data providers, investigators, sponsors and data scientists, into a single, collaborative study team and facilitates an end-to-end observational study.

A typical observational study workflow includes:

- Identification of a hypothesis
- A study proposal
- Collaborators invited to serve as investigators, sponsors, data providers and data scientists
- Identification of appropriate patient level data sources
- Generation of study code (typically R and SQL)
- Execution of the study code, via iterative series of study analyses
- Collection of statistics and inference of evidence
- Publication of the final research paper

The ARACHNE enables such workflow through the following components:

- Study Notebook
- Expert Finder
- Data Catalog
- Analysis Execution, including remote Execution Engine
- Study Publisher and Insights Library

The following are a set of key features which makes execution of network studies efficient and elegant in ARACHNE:

- Complete study lifecycle and workflow management
- Discovery of federated or local databases in the Data Catalog
- The assembly of a study team
- The execution of analyses within organization or across organizational boundaries using a secure, compliant and trusted remote data access
- The exchange, store and annotation of analysis code and results
- Publication of study outcomes e.g. insights, evidence and conclusions, into Insights Library
- Remote, scalable job execution
- Support for R, SQL of the OHDSI supported dialects and complex packaged analyses
- Integration with other OHDSI Platforms (ATLAS/WebAPI, Achilles)
- Support for OMOP CDM, including database level statistics (Achilles) in Data Catalog

The ARACHNE Execution Engine is one of the most critical component that enables a creation of complete, standards based R and SQL execution environment on a fly. It is a vision to enable support for other analytical languages, such as Python.

ARACHNE is taking advantage of Docker technology and Docker-based deployments that is aiming to make an installation and configuration to be a simple, lightweight effort.

ARACHNE can be used internally within an organizational network and firewall boundaries (*ARACHNE Workflow Suite*), or across firewalls accessing databases in other organizations (*ARACHNE Research Collaboration Network*) deployed as a cloud based SaaS platform.

It is the OHDSI and Odysseus vision to eventually connect private and public network across the globe into a single global network-of-networks (NeoNe) information exchange superhighway to enable a large-scale patient data analytics.

Conclusion

ARACHNE is a key to providing an efficient, consistent, secure and compliant observational research.

OHDSI has already achieved impressive quality and speed of observational research by providing

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

ARACHNE aims at removing the remaining impediments by enabling a consistent and compliant observational study process workflow while utilizing all of the standards above, as well as streamlining the execution of studies across a network of distributed databases:

- Standardized study lifecycle and workflow
- Standardized distributed execution
- Full lineage between hypothesis, protocol, code, analytical results, data sources and study insights
- An easy to navigate, secure, compliant and integrated into the process data catalog of patient-level data (data market place)
- Insight library containing searchable study outcomes linked to archived studies

References

1. OHDSI: The research community. Accessed June 20, 2016. Available from: http://www.ohdsi.org/
2. Nikolai Puntikov, Gleb Malikov "Vision for Technical Infrastructure to Facilitate OHDSI Network Research" http://www.ohdsi.org/web/wiki/lib/exe/fetch.php?media=resources:arachneposterabstract.pdf