Integration of Atlas into the Gen3 Data Commons Framework

Richard D Boyce[1], Craig C. Teerlink [2,3], Hamid Saoudian[2,3], Kyle M. Hernandez[4], Victoria Zaksas[4], Pieter Lukasse[4], Andrew Prokhorenkov[4], Noah Metoki-Shlubsky[4], Robert L. Grossman[4], Scott L. DuVall [2,3]

[1] University of Pittsburgh Department of Biomedical Informatics,[2] VA Informatics and Computing Infrastructure, [3] University of Utah, [4] University of Chicago, Center for Translational Data Science

Background

Data commons are secure platforms that enable multi-disciplinary teams to conduct research using large-scale datasets. The Gen3 platform consists of open-source software services that support the creation of cloud-based data resources, including data commons and analysis workspaces. Integrating Observational Health Data Sciences and Informatics (OHDSI) tools into the Gen3 platform would enable the tools to be made available to users in a variety of data commons. One important OHDSI application is the Atlas web application that facilitates the design and execution of clinical research analyses on standardized, patient-level, observational data stored in Observational Medical Outcomes Partnership (OMOP) common data model. Here, we propose integration of Atlas and WebAPI within the Gen3 platform to support the creation of clinical cohorts for genome-wide association studies.

Methods

This section describes the design of the integration of Atlas and WebAPI within the Gen3 platform. Within Gen3, a microservice is used to manage authorization which includes access to team projects that a user has access to. The projects are created in the microservice using an onboarding process that leads to records that associate team projects with Data Commons user accounts. In this way, users can be members of more than one team project and therefore have multiple team project roles.

Data Commons users authenticate to a custom-configured web portal which uses the Gen3 boundary as an identity provider. To enable Atlas within the Data Commons without an additional log in step, the boundary first modifies the context of the token used for authentication to hold the list of team projects that the user has access to. Second, the Data Commons web application parses the token and presents the list of team projects to the user. Third, the user selects one team project that they will usewithin Atlas. At this step, the Data Commons web application initiates the App by posting the token in the request to Atlas "home", providing the user's selected project in the body of the request. The Data Commons Atlas instance is set up to use token-based permission manager that can read the token and pull the selected project from the body of the request. A new permission manager in Atlas that can read boundary token sets the current role of the user for the session to be a role specific to the team project that was previously created by the microservice within the WebAPI security framework. A final change is that the Atlas WebAPI is modified to use a new property setting that causes it to return only concept sets, cohort definitions, or other artifacts (e.g., incidence rate analysis, characterizations, etc.) in the response provided by endpoints such as #/conceptsets, #/cohortdefinitions, etc. that a user has explicit permission to view based on the sec_role_permission and sec_permission tables in the security schema.

Results

The design outlined in the methods is being implemented to support a new Data Commons created for the United States Veterans Health Administration (VHA). The new Data Commons is designed to enable VHA researchers to conduct large-scale genome-wide association studies. Once thoroughly tested, the approach will enable the same functionality for other Gen3-based Data Commons. Changes to the WebAPI Atlas, including the new permission-based read access functionality, are being tested and integrated into public releases.

Conclusion

We are extending Gen3 to support OHDSI tools, starting with Atlas and WebAPI. These enhancements will benefit the VHA Data Commons and future Data Commons built using Gen 3. The methods might translate to other frameworks for building cloud-based data science environments.