Catalog and Security

Overview

Table of Contents:

- Overview
- Main Features
- Data Models
- RESTful web services

A genomic data analysis platform need to keep track of different resources such as metadata of files, sample annotations or jobs. OpenCGA Catalog aims to collect and integrate all the information needed for executing genomic analysis. This information is organized in nine main entities: *users*, *studies*, *files*, *samples*, *datasets*, *cohorts*, *individuals*, *disease panels* and *jobs*.

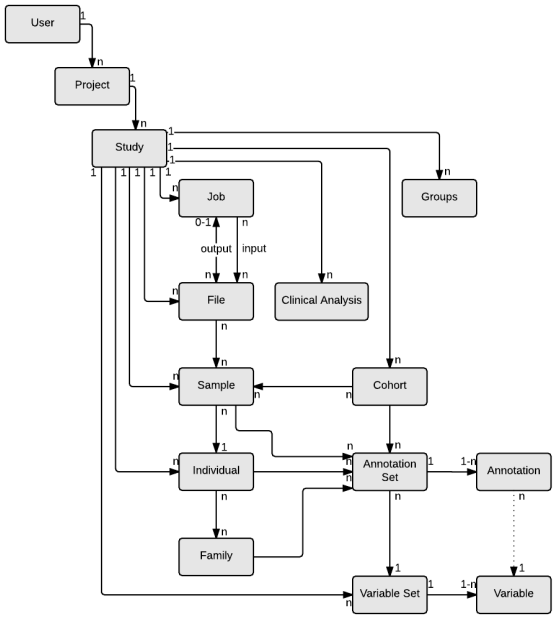
Main Features

The main tasks of Catalog are to provide:

- Authentication and authorization to the different resources.
- A collaborative environment.
- File audit to keep track of files and metadata.
- Analysis and Jobs.
- Sample, individual and cohort annotation.
- Security
- Versioning

Data Models

This section describes the most relevant entities. For more detailed information about the data models such as Java source code, examples or the JSON Schemas you can visit OpenCGA Catalog Data Models page. You can see an overview of the data model in this picture:



The most relevant entities in OpenCGA Catalog are:

- User: Contains the data related to the user account.
- Project: Contains information of a project, covering as many related studies as necessary.
- Study: Main space set environment. Contain files, samples, individuals, jobs...
- File: Information regarding a submitted or generated file.
- Sample: Information regarding the sample. Closely related to file entity.
- Individual: Contain the information regarding the individual from whom the sample has been taken.
- Cohort: Group sets of samples with some common feature(s).
- Disease panel: Define a disease panel containing the variants, genes and/or regions of interest.
- Job: Job analysis launched using any of the files or samples.

RESTful web services

All this information can be stored and retrieved using our Java and RESTful web services API.