Using OpenCGA

Table of Contents:

RESTful Web Services

OpenCGA implements a comprehensive and well-designed REST web service API, this consists of more than 200 web services to allow querying and operating data in OpenCGA. You can get more info at RESTful Web Services page.

We have implemented **three different ways** to query and operate OpenCGA through the REST web services API:

- REST Client Libs: four different client libraries have been implemented to ease the use of REST web services, This allows bioinformaticians to easily integrate OpenCGA in any pipeline. The four libraries are equally functional and fully maintained, these are *Java, Python* (available at PyPI), *R* and *JavaScript*
- Command Line: users and administrators can use ***opencga.sh*** command line to query and operate OpenCGA.
- IVA Web Application: an interactive web application called IVA has been developed to query and visualisation OpenCGA data.

OpenCGA Demo

We have deployed a public ***demo*** installation to facilitate the testing and development for all users. We have loaded and indexed **five different datasets** organised in 3 *projects* and 5 *studies*, these cover the most typical data use cases today such as multi-sample VCF, family exomes and genomes; or cancer somatic data. All documentation examples and tutorials use this *demo* installation.

Connecting to *demo* installation

OpenCGA *demo* REST URL is available at http://bioinfo.hpc.cam.ac.uk/opencga-prod/. You can check REST API and documentation at http://bioinfo.hpc.cam.ac.uk/opencga-prod/webservices/.

We have created a read-only user called ***demouser*** with password ***demouser***. As in most OpenCGA installations where normal users are not the owners of the data, *demouser* has been given VIEW access to all *demo* user data, this is a very common configuration in OpenCGA where the owner of the data grant access to other users. In this *demo* installation the owner of the data is *demo* user, while *demouser* user is the public user created to query data.

Genomic Data

In this *demo* we have indexed 5 different genomic datasets. Data has been organised in three *projects* and five *studies*. These represents different assemblies and data types such as multi sample VCF, aggregated VCF or family genome or exome. The data is organised in 3 *projects* and 5 studies. You can find some useful information in this table:

Project ID and Name	Study ID - Name	VCF File Type	Samples	Variants
population **Population Studies GRCh38**	**1000g - 1000 Genomes phase 3**	WGS Multi sample	2,504	**82,587,763**
	uk10k - UK10K	WGS Aggregated	10,000	**46,624,127**
family **Family Studies GRCh37**	**corpasome - Corpas Family**	WES Family Multi sample	4	**300,711**
	platinum - Illumina Platinum	GWS Family Multi sample	17	**12,263,246**
cancer **Cancer Studies GRCh37**	**rams_cml - RAMS_CML** Chronic Myeloid Leukemia - Russian Academy of Medical Sciences	Somatic	11	**121,384**

Clinical Data

In order to make this *demo* more useful to users we have loaded or simulated some clinical data, this allows to exploit OpenCGA analysis such as GWAS or clinical interpretation. You can find clinical data for each study in the following sections.

1000g

We loaded the 1000 Genomes pedigree file, you can find a copy at http://resources.opencb.org/opencb/opencga/templates/demo/20130606_g1k.ped

uk10k

There is no possible clinical data in this study. This is a *WGS aggregated* dataset so no samples or genotypes were present in the dataset and, therefore, no *Individuals* or *Samples* have been created.

corpasome

We simulated two different disorders and few phenotypes for the different members of the family. To be documented soon.

platinum

To be documented soon.

rams_cml

To be documented soon.