ICGC Data Portal Deployment Guide

1. PREREQUISITES

Software: Java 1.6, Ant and SVN client
OS: major Linux distribution
Server: min. 4 GB memory, 8 GB for better performance

2. DOWNLOAD BIOMART FROM SVN

From the directory in which you wish to install BioMart, run the following command:

svn checkout https://code.oicr.on.ca/svn/biomart/biomart-java/branches/icgc3_d

3. COMPILE AND BUILD BIOMART

From the directory in which you installed BioMart, run the following command:

ant

This will compile BioMart source and build the distribution.

> *Note*: *If you do not have local data, but want to set up an ICGC Data Portal mirror site, please go directly to* **Section 9.**

4. START MARTCONFIGURATOR

Run MartConfigurator with the following command in the directory of your installation:

./dist/scripts/martconfigurator.sh

5. CONFIGURE THE REGISTRY

In MartConfigurator, go to the "File" menu, select "Open," and open *icgc3member.xml*. In the left-hand panel, labeled "Source", there is a list of data sources included in the ICGC data portal.

You will need to modify the following sources to point to your local database that contains the data loaded by MartLoader: Genes, Samples, Transcriptome, Structural Rearrangement, Simple Mutations, Copy Number Mutations, mirna, jcn.

To modify data source settings, right click on a source then select "Dataset Source", this will open a window for data source management, which contains the connection parameters for each cancer type.
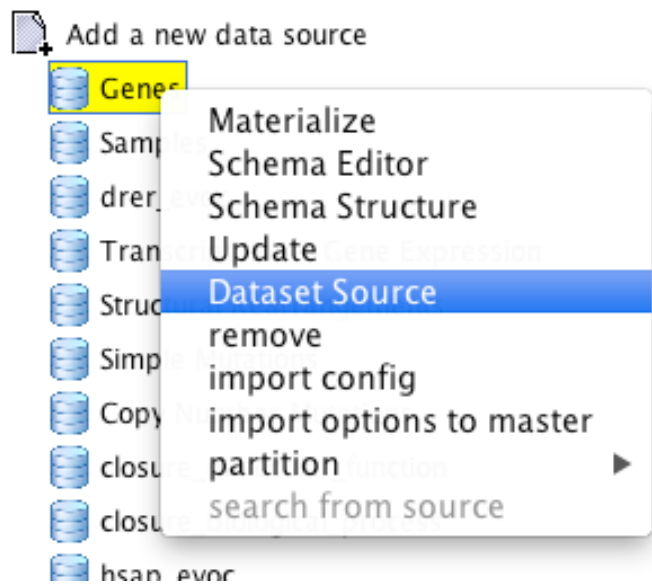
To modify the connection parameters for the QCMG entry, right click on that row in the connection column and select "edit connection".

Dataset Display Name	Dataset Connection Parameters	Database	Schema
Serous Cystadenocarcinoma (TCGA, US)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice
Small Cell Lung Carcinoma (WTSI, UK)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice
Glioblastoma Multiforme (TCGA, US)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice
Liver Cancer (NCC, JP)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice
Pancreatic Cancer (QCMG, AU)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice
Malignant melanoma (WTSI, UK)	http://dcc-test.res.oicr.on.ca	edit connection	/martservice
Pancreatic Cancer (OICR, CA)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice
Breast Carcinoma (WTSI, UK)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice
Glioblastoma Multiforme (JHU, US)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice
Colorectal Cancer (JHU, US)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice

To change the connection to point to your database, change the connection type to "DB" and enter the connection parameters for your database server.

After clicking OK, your changes should be reflected in the table:

Data source management

| Dataset Name	Dataset Display Name	Dataset Connection Parameters	Database	Schema	Hide
hsapiens_gene_ensembl_tcgaOV3	Serous Cystadenocarcinoma (TC...	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_sangerLung3b	Small Cell Lung Carcinoma (WTSI,...	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_tcgaGBM3	Glioblastoma Multiforme (TCGA,...	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_jpLiver3b	Liver Cancer (NCC, JP)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_garvanPanc3	Pancreatic Cancer (QCMG, AU)	jdbc:mysql://yourserver.address.com:3306/	8200	/biomart/martservice	false
hsapiens_gene_ensembl_sangerMelan...	Malignant melanoma (WTSI, UK)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_oicrPanc3b	Pancreatic Cancer (OICR, CA)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_sangerBreast3b	Breast Carcinoma (WTSI, UK)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_hopkinsGBM3	Glioblastoma Multiforme (JHU, US)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_hopkinsColon3	Colorectal Cancer (JHU, US)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_hopkinsPanc3	Pancreatic Cancer (JHU, US)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_jpRiken3b	Liver Cancer (RIKEN, JP)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_hopkinsBreast3	Breast Cancer (JHU, US)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_TSPLung_2	Lung Adenocarcinoma (TSP, US)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_tcgaREAD	Rectum adenocarcinoma (TCGA,...	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false

The "Database" and "Schema" columns must also be changed. For MySQL servers, these values are both the same, and are the name of the database. They can be edited by double-click in the relevant field:

Dataset Name	Dataset Display Name	Dataset Connection Parameters	Database	Schema	Hide
hsapiens_gene_ensembl_tcgaOV3	Serous Cystadenocarcinoma (TC...	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_sangerLung3b	Small Cell Lung Carcinoma (WTSI,...	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_tcgaGBM3	Glioblastoma Multiforme (TCGA,...	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_jpLiver3b	Liver Cancer (NCC, JP)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_garvanPanc3	Pancreatic Cancer (QCMG, AU)	jdbc:mysql://yourserver.address.com:3306/	my_db	mydb	false
hsapiens_gene_ensembl_sangerMelan...	Malignant melanoma (WTSI, UK)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_oicrPanc3b	Pancreatic Cancer (OICR, CA)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false
hsapiens_gene_ensembl_sangerBreast3b	Breast Carcinoma (WTSI, UK)	http://dcc-test.res.oicr.on.ca	8200	/biomart/martservice	false

This process must be repeated for the following data sources: Genes, Samples, Transcriptome, Structural Rearrangement, Simple Mutations, Copy Number Mutations, mirna, jcn

6. UPDATE DATA SOURCES

After the connection parameters have been changed, you will need to run an update so BioMart will pick up the specific content in your data sources. To do this, right-click on the data source and select "Update".
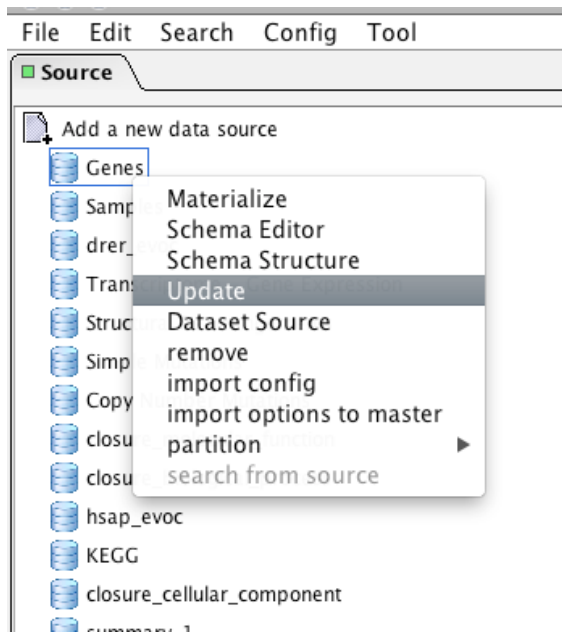
You will also be prompted if you would like to "update options", please choose "Yes".

Depending on the size of the data source, updating may take some time.

7. SAVE YOUR CHANGES

Changes to the registry can be saved by selecting "Save as" from the "File" menu, as the registry in the same file, *icgc3member.xml*. The reason you have to use "Save as" instead of "Save" is that, "Save as" will create a key file that will be used to encrypt the database account password. The key file named ".icgc3member" will be created after "Save as".

8. START UP BIOMART WEB SERVER FROM MARTCONFIGURATOR

After all the changes have been made, you can click the "deploy" button on the top right corner to quickly launch BioMart web server. This will also open up your web browser and navigate to BioMart server at http://localhost:9000/. It may take a few minutes for BioMart server to start, before it starts your web browser may show error message saying something like: could not connect to localhost:9000. Please be patient, and refresh the page until BioMart server starts.

9. RUN THE DATA STORE SERVER

You will need to run a persistent data store to hold authentication data. To run this, from the root directory of the installation change to the voldemort subdirectory:

```
cd voldemort
```

Then run the following command:

```
./bin/voldemort-server.sh config/single_node_cluster &> ./voldemort.log &
```

This will run the datastore server locally on port 6666.

10. CONFIGURE DEPLOYMENT

You can also deploy BioMart server from command line on a server.

In a text editor open the *biomart.properties* file, located in the dist subdirectory.

In the "HTTP settings" section, change the *http.host* property to 0.0.0.0 and change the *http.port* property to the port on which you would like the non-secure web server to run (default 9000). Remove "#" at the beginning of *http.url* line, and set it to the public URL:port from where your BioMart server is accessible, for example, http.url = http://your.domain.org:9000

In order to protect ICGC's controlled data, you will need to configure "HTTPS settings" section and "SSL Settings" section as following:

In the "HTTPS settings" section, remove the "#" at the beginning of the line for the *https.port* property, and change this to the port on which you would like the secure web serve to run (default 9043).

In the "SSL Settings" section, remove the "#" from the beginning of the lines for ssl.keystore and ssl.password. Set the ssl.password property to a password that you will be choosing in the next step (**Section 11**).

In the "BioMart settings" section, change the biomart.registry.file property file to point to "icgc3member.xml" instead of "default.xml". Change the biomart.registry.key.file to point to ".icgc3member" instead of ".default".

If you deployed the data store in the previous step (**Section 9**), change the "Data store settings" section to read:

```
datastore.engine = voldemort
datastore.url = tcp://localhost:6666
```

11. GENERATE AN SSL KEY

If you do not have an official SSL certificate, you can use keytool to generate a temporary one for testing.

From the root directory of your installation, change to the dist/web/etc directory with the command:

cd dist/web/etc

If a keystore file exists, delete it using the command:

rm keystore

Generate a new keystore file using the command:

keytool -keystore keystore -alias biomart -genkey -keyalg RSA

You will be prompted to enter a password, enter a password as your choice. The password you set in the ssl.password property in the previous step must be the same as what you choose here.

You will then be prompted for several pieces of information; you can leave them all blank. When prompted if the data are all correct, type "yes". Finally, you will be asked for another password; simply hit "enter" to use the same password you set earlier.

12. DEPLOY BIOMART SERVER

To deploy BioMart, from the directory of your installation run the following command:

./dist/scripts/biomart-server.sh start

To stop the server, please do:

./dist/scripts/biomart-server.sh stop

It may take several minutes before the server starts and the site is viewable.

Once server started, please navigate your browser to the proper host and port your just set up, eg, http://localhost:9000.