Database working-session – Amsterdam

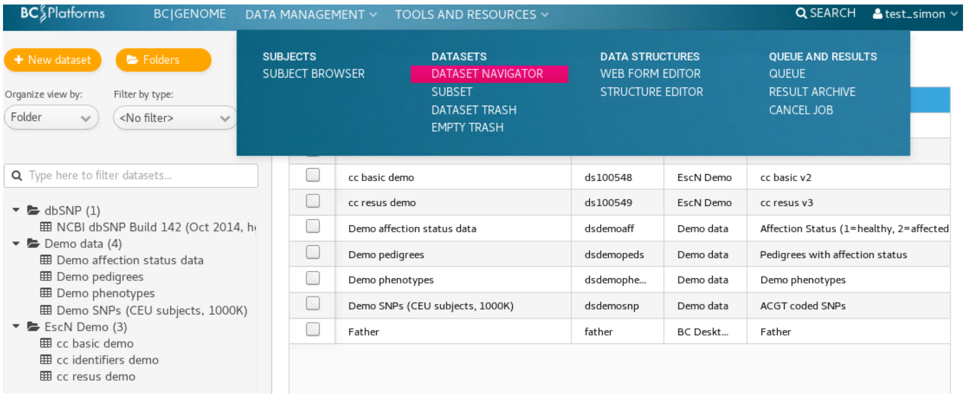
1. Follow the instructions in the SOP (https://escape-net.github.io/texts/EscapeNet_SOP.pdf) on how to login to the secure cloud 575 and the BC|GENOME app.

BC|GENOME

2. Explore the various tools and function in BC|GENOME. We can recommend you to have a look at the following chapters in the BC|GENOME user guide (https://escape-net.github.io/texts/BCGENOME_User%20Guide_Dec2018.pdf):

   a. Navigating in BC|GENOME (2.2)
   b. Working with data in BC|GENOME
      i. Creating a dataset (5.1)
      ii. Subsets (5.2)
      iii. Viewing the distribution of data values (5.3)
      iv. Running analyses in BC|GENOME (5.4) (PLINK)
   c. Result archive (6)

3. For hand-on experience purposes we have created the three datasets in the folder **EscN Demo** found in the **DATASET NAVIGATOR** in BC|GENOME. All users have read permission to the demo datasets. Additionally, demo data sets for GWAS studies are found in the folder **Demo data**, with affection status, pedigrees, phenotypes, and snp's.



Organize view by: Folder

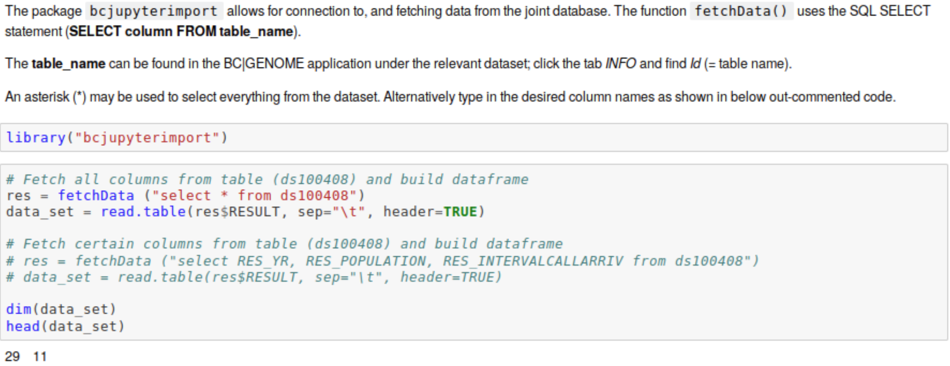
Filter by type: <No filter>

☐	cc basic demo	ds100548	EscN Demo	cc basic v2
☐	cc resus demo	ds100549	EscN Demo	cc resus v3
☐	Demo affection status data	dsdemoaff	Demo data	Affection Status (1=healthy, 2=affected
☐	Demo pedigrees	dsdemopeds	Demo data	Pedigrees with affection status
☐	Demo phenotypes	dsdemophe...	Demo data	Demo phenotypes
☐	Demo SNPs (CEU subjects, 1000K)	dsdemosnp	Demo data	ACGT coded SNPs
☐	Father	father	BC Deskt...	Father

Jupyter Notebook – analysis environment

4. Access the **Jupyter Hub** analysis environment, by typing the URL https://risoe-r12-cn518/jupyter/ in the web-browser of the virtual desktop
5. For information on what Jypyter Notebooks are and how they are used, please see https://escape-net.github.io/docs/jupyter

6. To import data sets from BC|GENOME into a R Jupyter Notebook, the library “bcjupyterimport” can be used (see code below).



The package `bcjupyterimport` allows for connection to, and fetching data from the joint database. The function `fetchData()` uses the SQL SELECT statement (**SELECT column FROM table_name**).

The **table_name** can be found in the BC|GENOME application under the relevant dataset; click the tab *INFO* and find *Id* (= table name).

An asterisk (*) may be used to select everything from the dataset. Alternatively type in the desired column names as shown in below out-commented code.

```
library("bcjupyterimport")
```

```
# Fetch all columns from table (ds100408) and build dataframe
res = fetchData ("select * from ds100408")
data_set = read.table(res$RESULT, sep="\t", header=TRUE)

# Fetch certain columns from table (ds100408) and build dataframe
# res = fetchData ("select RES_YR, RES_POPULATION, RES_INTERVALCALLARRIV from ds100408")
# data_set = read.table(res$RESULT, sep="\t", header=TRUE)

dim(data_set)
head(data_set)
```

```
29   11
```

7. Check out the Jupyter Notebook example on how to import data and carry out simple data management - found in the BC|GENOME app, under **DATA MANAGEMENT / RESULT ARCHIVE**: shared / skonig