



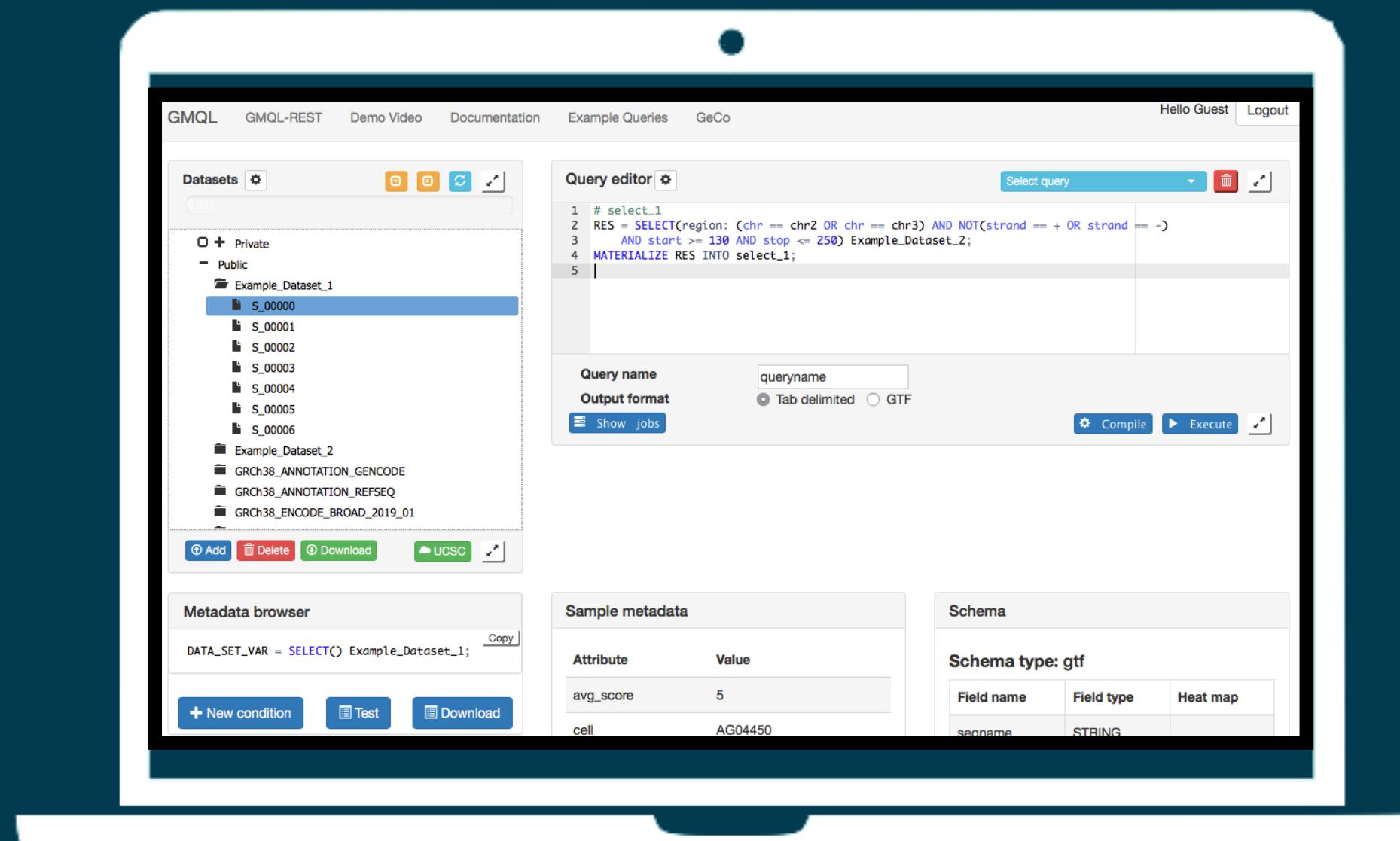
DEIB - POLITECNICO DI MILANO

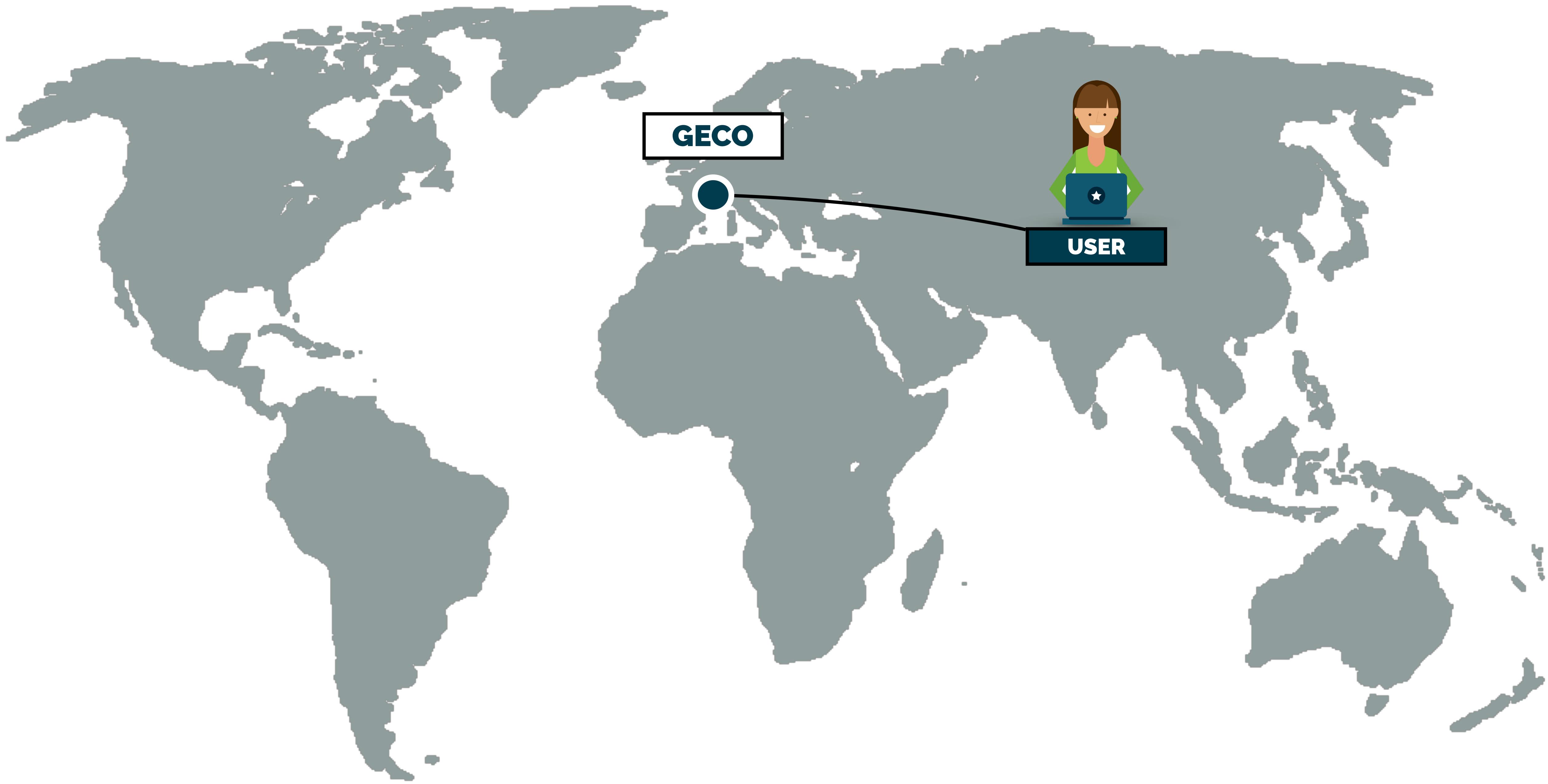
Federated GMQL

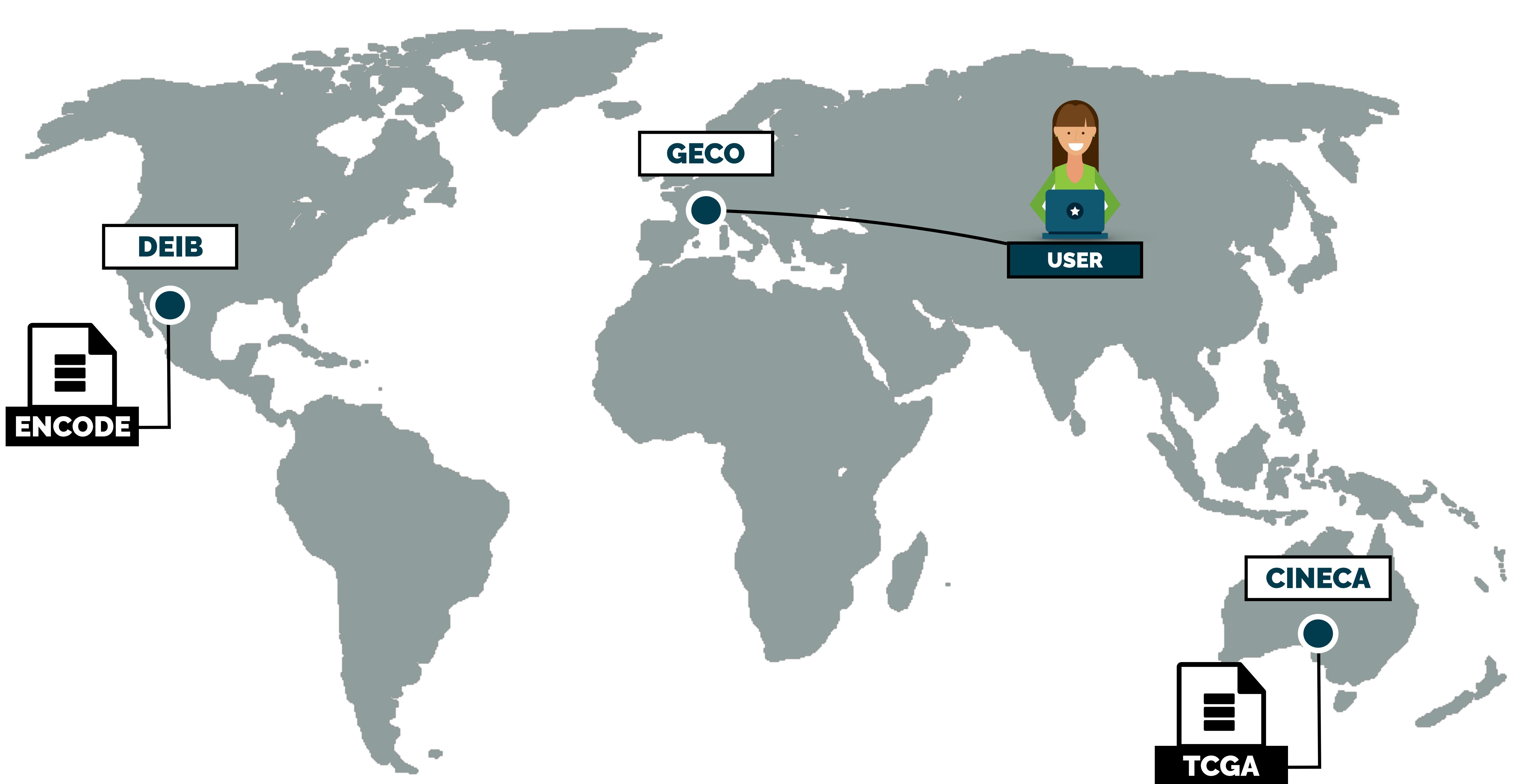
OVERVIEW

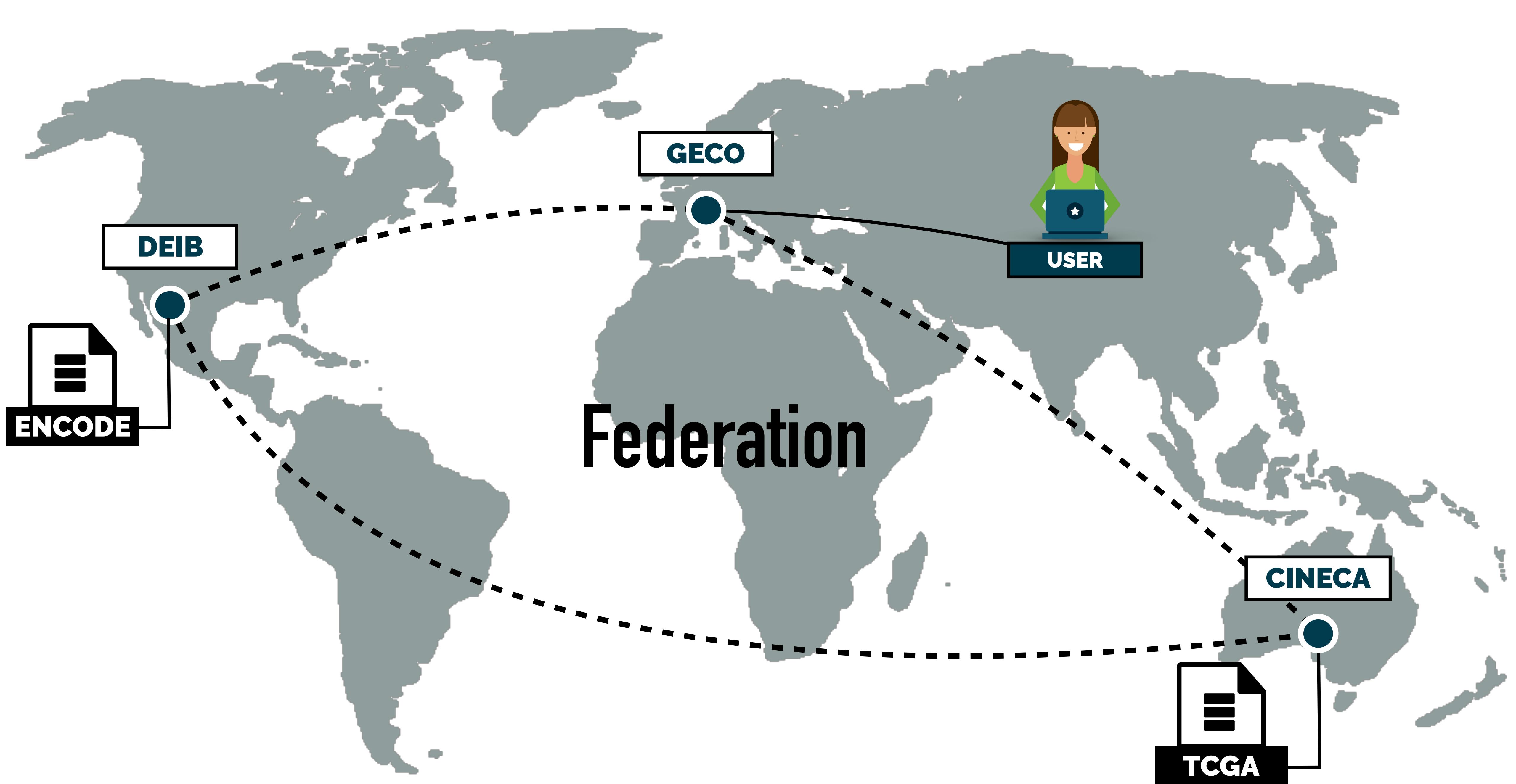
GenoMetric Query Language (GMQL)

- Open Source Project
- Enables querying **hundreds of datasets and thousands of samples**
- Runs on a cloud-computing system based on **Apache Spark**
- Publicly available through a user-friendly **web interface**
- Well maintained public repository which integrates genomic datasets from **ENCODE, TCGA/GDC and Roadmap Epigenomics**









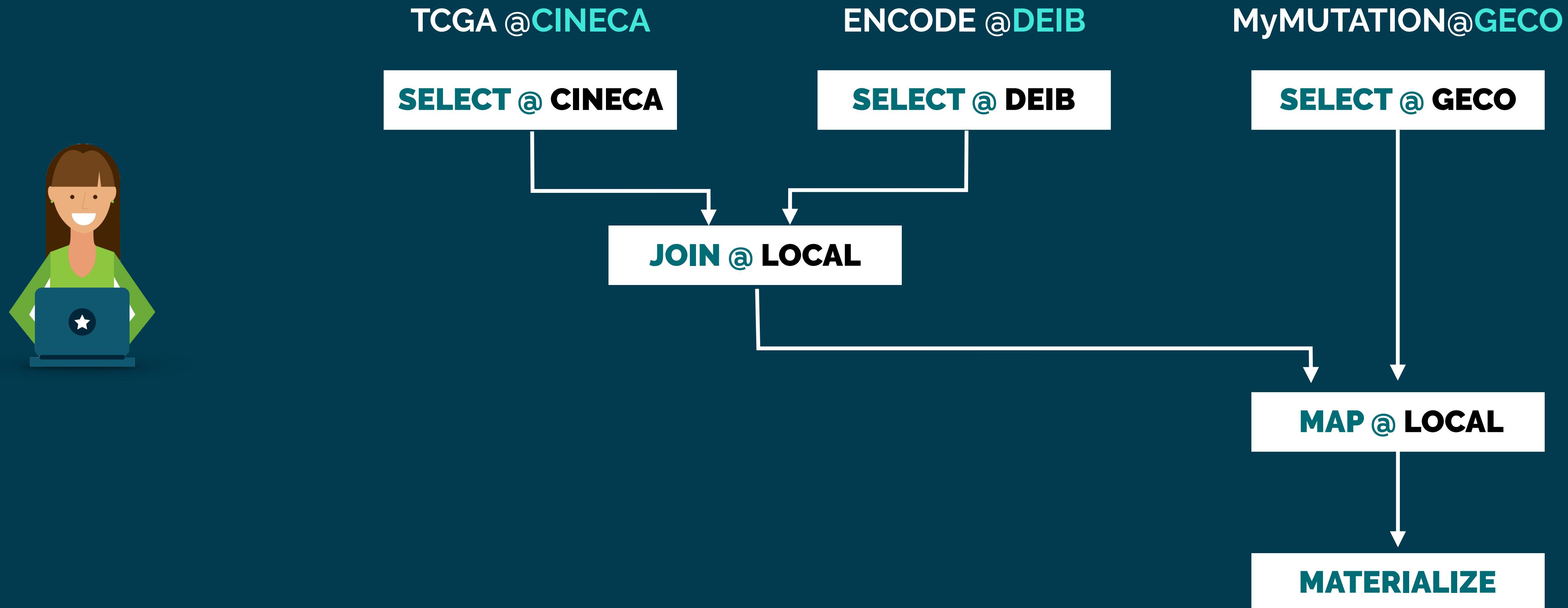
Federated Datasets



The screenshot displays the GMQL interface with several functional areas:

- Datasets:** A sidebar on the left containing a tree view of datasets. A red box highlights the "Federated" node, which is expanded to show sub-datasets: CINECA.HG19_TCGA_rnaseqv2_gene, DEIB.Example_Dataset_1, DEIB.HG19_ENCODE_NARROW_2019_01, and GeCo.MUTATION.
- Query editor:** The main workspace on the right. It includes:
 - A "Select query" dropdown menu.
 - A "Query name" input field set to "queryname".
 - An "Output format" radio button group between "Tab delimited" (selected) and "GTF".
 - Buttons for "Compile", "Execute", and a refresh icon.
- Metadata browser:** A panel at the bottom left showing a query: `DATA_SET_VAR = SELECT(at:CINECA) CINECA .HG19_TCGA_rnaseqv2_gene;`. It has buttons for "New condition", "Test", and "Download".
- Sample metadata:** A panel below the Metadata browser.
- Schema:** A panel on the far right.

Federated Query



Federated Query



SELECT TCGA AT CINECA

TCGA = **SELECT**(tumor_tag == "BRCA"; **at:CINECA**) **CINECA.TCGA**;

SELECT ENCODE AT DEIB

ENCODE = **SELECT**(cell_line == "H1-hESC"; **at:DEIB**) **DEIB.ENCODE**;

JOIN DS1 AND DS2 AT DEIB

JOINED = **JOIN**(dist < 0; output: left_distinct; **at:LOCAL**) **TCGA ENCODE**;

SELECT MUTATION AT GeCo

MYMUTATION = **SELECT**(**at:LOCAL**) **MUTATION**;

Map MUTATION AT GeCo

RES = **MAP**(count_name: mut_count; **at:LOCAL**) **JOINED MYMUTATION**;

Materialize at GeCo

MATERIALIZE RES INTO ResGenes;

Federated Query



```
# SELECT TCGA AT CINECA
TCGA = SELECT(tumor_tag == "BRCA"; at:CINECA) CINECA.TCGA;

# SELECT ENCODE AT DEIB
ENCODE = SELECT(cell_line == "H1-hESC"; at:DEIB) DEIB.ENCODE;

# JOIN DS1 AND DS2 AT DEIB
JOINED = JOIN(dist < 0; output: left_distinct; at:LOCAL) TCGA ENCODE;

# SELECT MUTATION AT GeCo
MYMUTATION = SELECT at:LOCAL MUTATION;

# Map MUTATION AT GeCo
RES = MAP(count_name: mut_count; at:LOCAL) JOINED MYMUTATION;

# Materialize at GeCo
MATERIALIZE RES INTO ResGenes;
```

geco.deib.polimi.it/gmql-rest/ GMQL

GMQL GMQL-REST Demo Video Documentation Example Queries GeCo Hello Guest Logout

Datasets

0.0%

- > Private
- < Public
 - > Example_Dataset_1
 - > Example_Dataset_2
 - > HG19_ANNOTATION_GENCODE
 - > HG19_ANNOTATION_REFSEQ
 - > MUTATION
- > Federated

Add Delete Download UCSC

Query editor

Select query

1

Query name: queryname

Output format: Tab delimited (radio button selected) GTF

Show jobs

Compile Execute

Metadata browser

DATA_SET_VAR = `SELECT() Example_Dataset_1;`

Copy

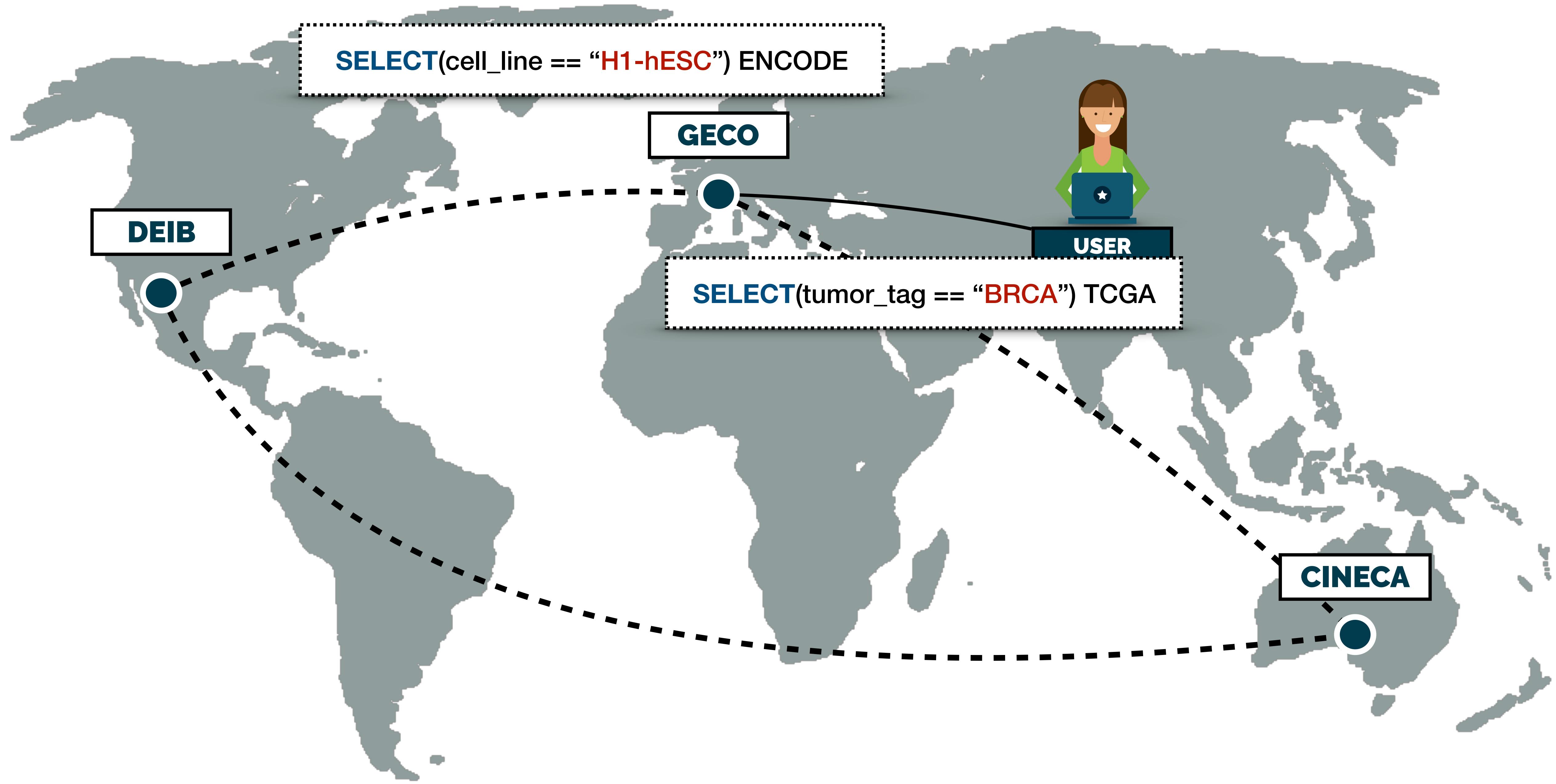
New condition Test Download

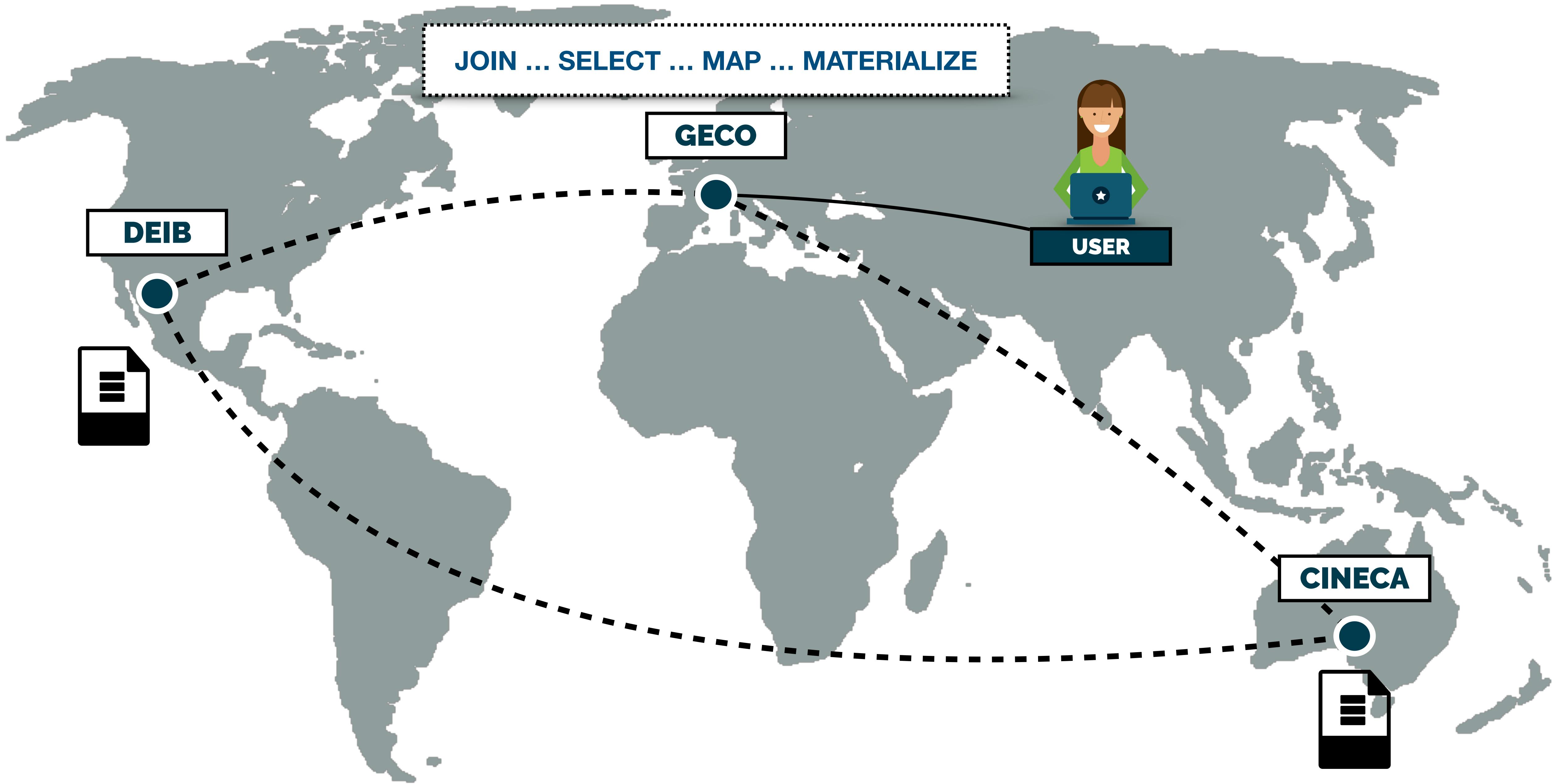
Sample metadata

Schema

Schema type: gtf

Field name	Field type	Heat map
seqname	STRING	
source	STRING	
feature	STRING	
start	LONG	
end	LONG	
score	DOUBLE	





geco.deib.polimi.it/gmql-rest/

GMQL

GMQL GMQL-REST Demo Video Documentation Example Queries GeCo Hello Guest Logout

Datasets

- Private
- + Public
- + Federated

[Add](#) [Delete](#) [Download](#) [UCSC](#)

Query editor

```

1 # SELECT TCGA AT CINECA
2 TCGA = SELECT(tumor_tag == "BRCA"; at:CINECA) CINECA.HG19_TCGA_rnaseqv2_gene;
3
4 # SELECT ENCODE AT DEIB
5 ENCODE = SELECT(cell_line == "H1-hESC"; at:DEIB) DEIB.HG19_ENCODE_NARROW_2019_01;
6
7 # JOIN DS1 AND DS2 AT DEIB
8 JOINED = JOIN(dist < 0; output: left_distinct; at:LOCAL) TCGA ENCODE;
9
10 # SELECT MUTATION AT GeCo
11 MYMUTATION = SELECT(at:LOCAL) MUTATION;
12
13 # Map MUTATION AT GeCo
14 RES= MAP(count_name: mut_count; at:LOCAL) JOINED MYMUTATION;
15
16 # Materialize at GeCo
17 MATERIALIZE RES INTO ResGenes;
18

```

Query name

Output format Tab delimited GTF

[Show jobs](#) [RUNNING](#) [Stop](#) [Compile](#) [Execute](#)

Metadata browser

```
DATA_SET_VAR = SELECT()
queryname_20190913_130954_ResGenes;
```

[Copy](#)

[New condition](#) [Test](#) [Download](#)

Sample metadata

Schema



DEIB - POLITECNICO DI MILANO

Federated GMQL

OVERVIEW

The background features a dark blue-grey grid pattern. Overlaid on this are several teal-colored diagonal stripes of varying lengths and orientations, creating a sense of depth and motion.

DEIB - POLITECNICO DI MILANO

Federated GMQL

SETUP

Installing a GMQL Instance

- **Using Docker**

Follow the guide at : <https://github.com/DEIB-GECO/GMQL-Docker>

- **Full Installation**

Follow the guide at [https://github.com/DEIB-GECO/GMQL-WEB/
wiki/installation](https://github.com/DEIB-GECO/GMQL-WEB/wiki/installation)

Name Server

- Open Source Project
- Enables querying hundreds of datasets and thousands of samples
- Runs on a cloud-computing system based on Apache Spark
- Publicly available through a user-friendly web interface
- Well maintained public repository which integrates genomic datasets from ENCODE, TCGA/GDC and Roadmap Epigenomics

| Joining the Federation

- Register your instance on the Name Server
- Login
- Add the required configuration to your **repository.xml** file

Non sicuro — genomic.elet.polimi.it/nameserver/static//index.html#!/signup

Name Server

Register a new Instance

Username (instancename)

Email

Description

Password

Repeat Password

GMQL API URL

[Login](#) [Register](#)

© GeCo - Politecnico di Milano - View it on [GitHub](#).

Non sicuro — genomic.elet.polimi.it/nameserver/static//index.html#!/signup

Name Server

Name Server

Register a new Instance

Username (instancename)	<input type="text" value="MyInstance"/>
Email	<input type="text" value="myemail@myinstitute.com"/>
Description	<input type="text" value="My Instance"/>
Password	<input type="password" value="....."/>
Repeat Password	<input type="password" value="....."/>
GMQL API URL	<input type="text" value="http://myserver.com/gmql-rest/"/> http://myserver.com/gmql-rest/

[Login](#) [Register](#)

© GeCo - Politecnico di Milano - View it on [GitHub](#).

Sito web non sicuro

Name Server

Name Server

Login

Username: myInstance

Password:

[Register](#) [Login](#)

© GeCo - Politecnico di Milano - View it on [GitHub](#).

Non sicuro — genomic.elet.polimi.it/nameserver/static//index.html#!/home

Name Server

Name Server Home Datasets Instances Groups Logout

myInstance

Your **token** for communication with the NameServer is:

7cadeca22c38b9321f1029c267938e7342e74510

(eye icon)

You should have the following configuration in your repository.xml file:

```
<property name="GF_ENABLED">true</property>
<property name="GF_NAMESERVER_ADDRESS">http://genomic.elet.polimi.it/nameserver</property>
<property name="GF_INSTANCENAME">myInstance</property>
<property name="GF_TOKEN">7cadeca22c38b9321f1029c267938e7342e74510</property>
```

Your instance URL is the following:

http://myserver.com/gmql-rest/

Save

© GeCo - Politecnico di Milano - View it on [GitHub](#).

Non sicuro — genomic.elet.polimi.it/nameserver/static//index.html#!/home

Name Server

Name Server Home Datasets Instances Groups Logout

myInstance

Your **token** for communication with the NameServer is:

```
7cadeca22c38b9321f1029c267938e7342e74510
```

You should have the following configuration in your repository.xml file:

```
<property name="GF_ENABLED">true</property>
<property name="GF_NAMESERVER_ADDRESS">http://genomic.elet.polimi.it/nameserver</property>
<property name="GF_INSTANCENAME">myInstance</property>
<property name="GF_TOKEN">7cadeca22c38b9321f1029c267938e7342e74510</property>
```

Your instance URL is the following:

```
http://myserver.com/gmql-rest/
```

Save

© GeCo - Politecnico di Milano - View it on [GitHub](#).

| Federating a Dataset

- Login to the Name Server
- Go to the **Datasets** section
- Click on **Add a new dataset**

Non sicuro — genomic.elet.polimi.it/nameserver/static//index.html#!/datasets/create

Name Server

Name Server Home Datasets Instances Groups Logout

Add a new dataset

Name MYDATASET

Author ENCODE Project

Description ENCODE data mapped to HG19 human genome. Broad (or Regions) Peaks format is selected and archived/revoked data are avoided. The release date of this dataset is August 2017. <https://www.encodeproject.org/>

Privacy

Available instances:

- GMQL-ALL (Group)
- admin
- genomic
- canakoglu
- imn_gqml_instance
- andreagulino
- luca_portatile

Selected instances:

- myInstance
- GMQL-ALL remove

Repositories

Available instances:

- admin
- genomic
- canakoglu
- imn_gqml_instance
- andreagulino

Selected instances:

- myInstance

Non sicuro — genomic.elet.polimi.it/nameserver/static//index.html#!/datasets/create

Name Server

Name Server Home Datasets Instances Groups Logout

Add a new dataset

Name MYDATASET

Author ENCODE Project

Description ENCODE data mapped to HG19 human genome. Broad (or Regions) Peaks format is selected and archived/revoked data are avoided. The release date of this dataset is August 2017. <https://www.encodeproject.org/>

Privacy

Available instances:

- GMQL-ALL (Group)
- admin
- genomic
- canakoglu
- imn_gqml_instance
- andreagulino
- luca_portatile

Selected instances:

- myInstance
- GMQL-ALL remove

Repositories

Available instances:

- admin
- genomic
- canakoglu
- imn_gqml_instance
- andreagulino

Selected instances:

- myInstance

Non sicuro — genomic.elet.polimi.it/nameserver/static//index.html#!/datasets/create

Name Server

Name Server Home Datasets Instances Groups Logout

Add a new dataset

Name MYDATASET

Author ENCODE Project

Description ENCODE data mapped to HG19 human genome. Broad (or Regions) Peaks format is selected and archived/revoked data are avoided. The release date of this dataset is August 2017. <https://www.encodeproject.org/>

Privacy

Available instances:

- GMQL-ALL (Group)
- admin
- genomic
- canakoglu
- imn_gqml_instance
- andreagulino
- luca_portatile

Selected instances:

- myInstance
- GMQL-ALL [remove](#)

Repositories

Available instances:

- admin
- genomic
- canakoglu
- imn_gqml_instance
- andreagulino

Selected instances:

- myInstance