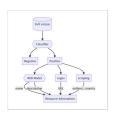protocols.io

Jul 08, 2022

⊕ Set Up Biodata Resource Inventory in Google Colab

Kenneth Schackart[1]

[1]University of Arizona

Kenneth Schackart: Global Biodata Coalition Consultant

In Development   ⊶ Share

This protocol is published without a DOI.

↗ Kenneth Schackart
   University of Arizona

ABSTRACT

This protocol will guide you on how to get everything in place to update the Biodata Resource Inventory.

This protocol describes how to setup Google Colab, connect your Google Drive, and clone the repository.

Some of the information here came from the following article:
https://medium.com/analytics-vidhya/how-to-use-google-colab-with-github-via-google-drive-68efb23a42d

PROTOCOL CITATION

KEYWORDS

GBC, Biodata resource inventory, Google Colab

LICENSE

CREATED

Jul 07, 2022

LAST MODIFIED

Jul 08, 2022

PROTOCOL INTEGER ID

66228

MATERIALS TEXT

An active Google Drive account.
An active GitHub account.

Prepare Google Drive

1   In your Google Drive home directory, create a new folder called GitHub.

   Inside the GitHub folder, create another folder called biodata_resource_inventory.

Connect Colab to Drive

2   Go to Google Colab.

   If you need to change the account you are using, close the pop up by clicking Cancel at the bottom right. Change your Google account by clicking the icon at the top right.

3   A menu will pop up. At the bottom right, select: "New Notebook"

4   In the cell, paste or type the following:

```
from google.colab import drive
drive.mount('/content/drive')
```

   Click the "run" button to the left of the cell or run with ctrl+enter

   You will be prompted to allow Google Colab to access your Drive, say yes.

   Do not try to change accounts at this stage, it will cause an error. If you need to change your account, go back to step 2.

5   Create a new code chunk by clicking "+ Code" under the top-left banner. In that cell paste or type the following, then run:

```
%cd /content/drive/MyDrive/GitHub/biodata_resource_inventory/
```

   This has changed your current directory to biodata_resource_inventory.

🐾 protocols.io   2

If you named the folders something else, you will need to modify the above command to reflect the actual folder names.

Create GitHub Access Token

6   Go to GitHub.

   Go to *Settings,* navigate to *Developer settings* and click on *Personal access tokens*.

7   Click *Generate new token* button on the top right corner of the page.

8   Click the *repo* checkbox under *Select scopes*.

9   Click the *Generate token* button at the bottom of the page.

10 You have now created a new access token. Copy the token to your clipboard. **Do not close this page yet**.

   ⚠ Do not share this access token.

Clone the Repository

11 Go back to your Colab Notebook, create a new Code cell, and enter the following, pasting your access token into the quotations, then run:

```
git_token = ''
```

12 Create a new code cell and paste or type the following, then run:

```
!git clone
https://{git_token}@github.com/globalbiodata/inventory_2022 .
```

   Note: During development, the main branch is mostly empty. To switch to the development branch run the following in a code cell:

```
# Switch to development branch if desired
!git checkout inventory_2022_dev
```

Congrats! You now have linked Google Colab to GitHub for access to their API, and you have cloned the repository into your Google Drive. You can now close the page with your access token.

Running the pipelines

13 You can now access all the files in the repository via traditional Google Drive.

14 You will find two notebooks in the repository:
- running_pipeline.ipynb: This allows reproduction of the original results
- updating_inventory.ipynb: This guides you to update the inventory

Each notebook provides instructions on their respective process, and includes already filled-in code chunks. So you should have to do little more work to run those pipelines.