BACPAC

Back Pain Consortium

BACPAC Data Portal User Guide

Version 2.0, March 2023

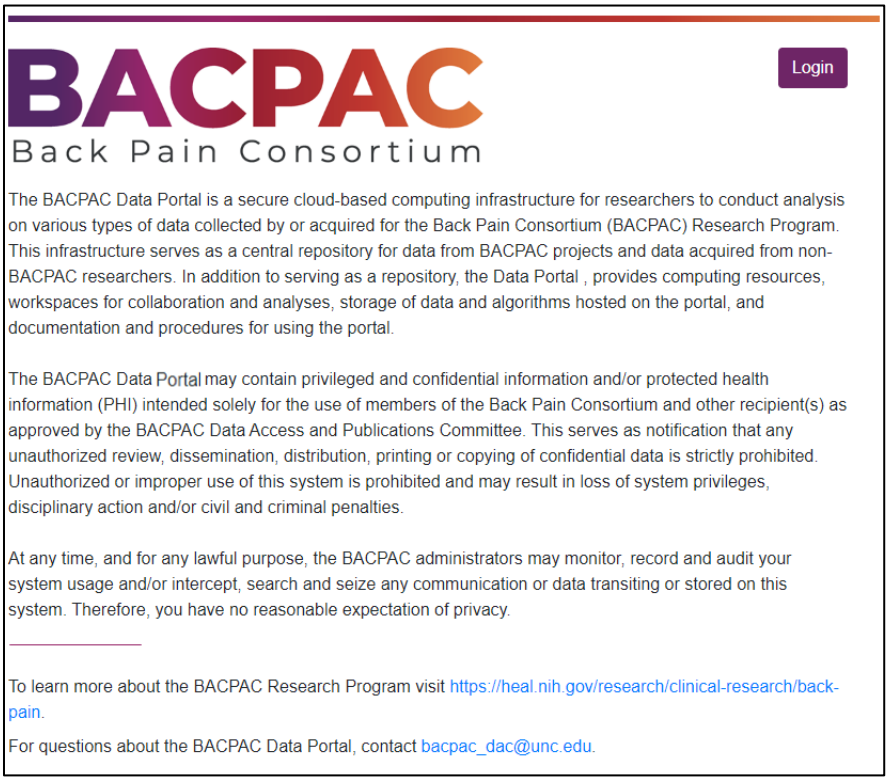
Prepared by the BACPAC Data Integration, Algorithm Development, and Operations Management Center (DAC)

Table of Contents

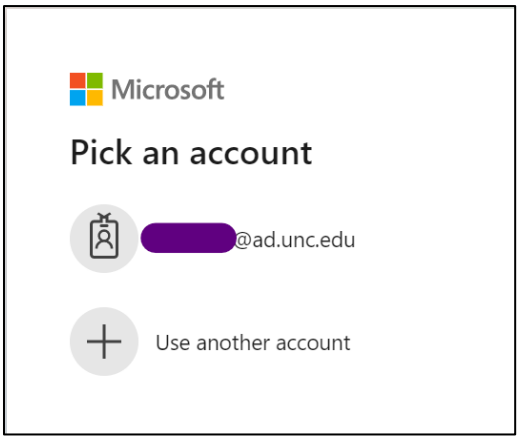
1. BACPAC Data Portal Overview

To access the BACPAC Data Portal, navigate to: **https://www.bacpacresearch.org/.**

The following page will be displayed. Follow the steps below to log into the BACPAC Data Portal.



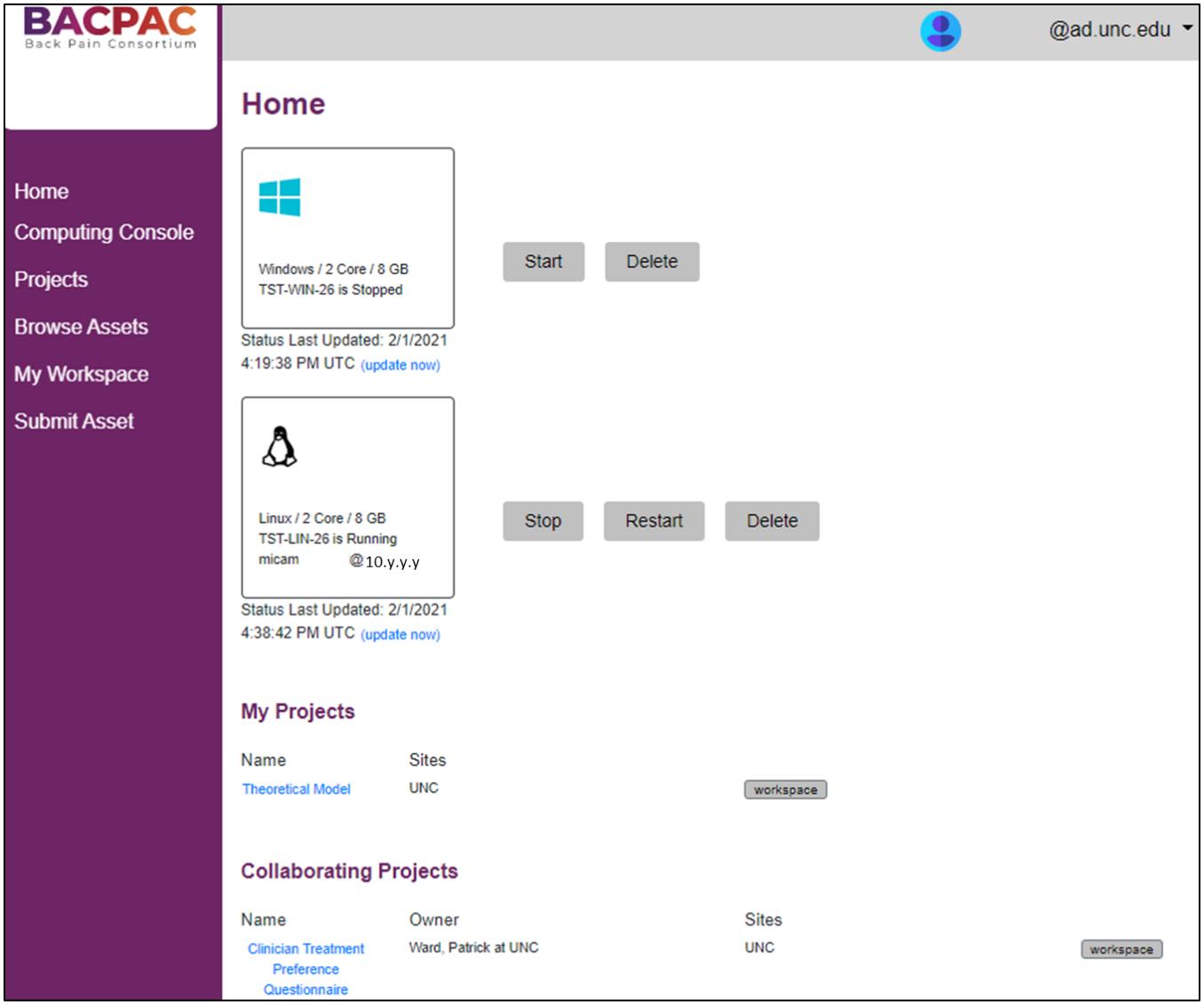
a. Click the login button.
b. On the Microsoft login screen (see image below), choose the account you use to login to the BACPAC Research Consortium MS Teams site:



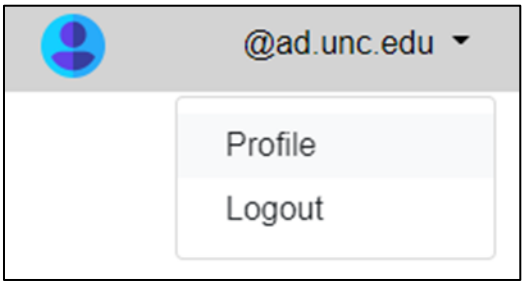
c. You will be re-directed to your organization's login page. Once you complete the login process, the Home page of the BACPAC Data Portal will display.

1.1. Home

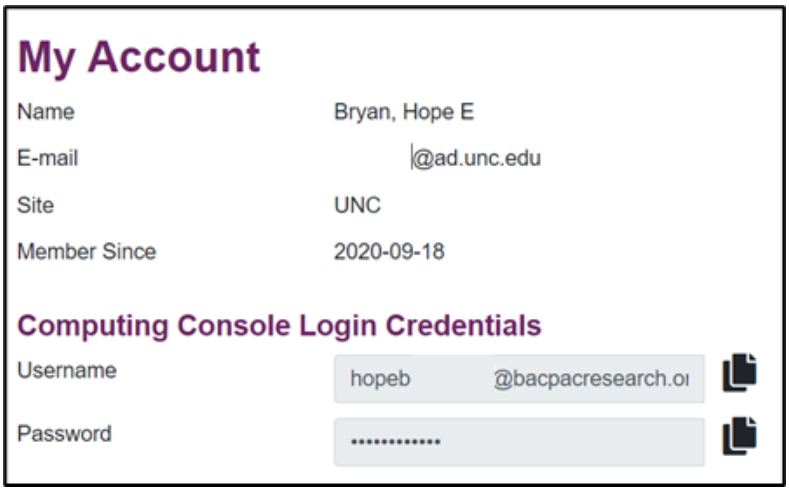
The Home page presents a summary of resources that you have access to. We will examine each of these resources using options on the left-hand menu.



a. To access information about your account, click your name in the top right-hand corner of the screen.

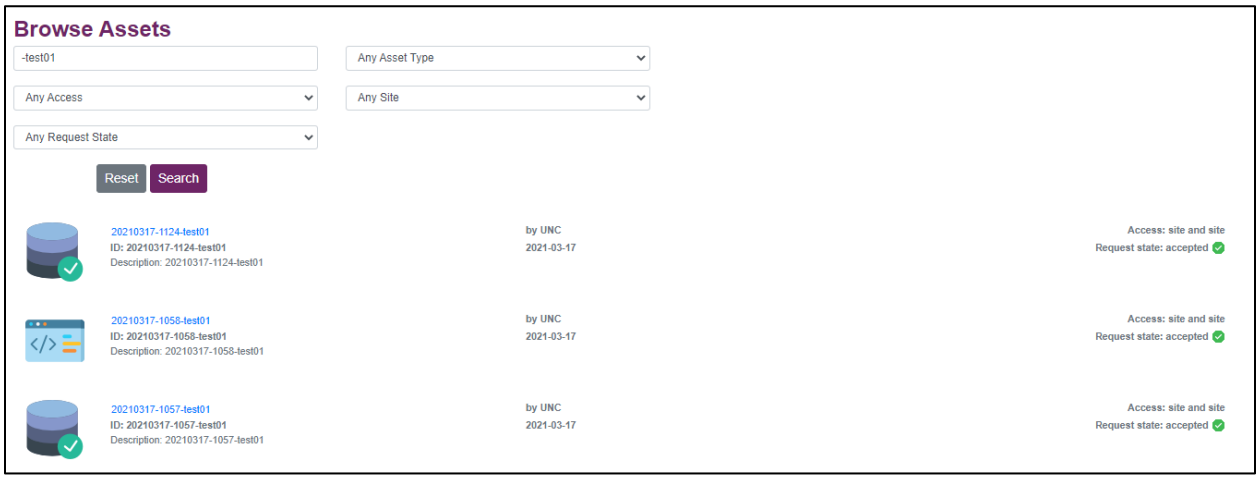


b. Select "Profile" to display a screen similar to this:



1.2. Browse Assets

The Browse Assets page lists datasets and program code which are available for use in the portal. All datasets are read-only. These datasets are sometimes referred to as "canonical data". These datasets cannot be modified by users in the portal.



This page shows identifying information about the asset (Note: the icon indicates whether the asset is data or code), the dataset name and description, the site which owns the asset and the access you have.

Your access to the asset is shown. It can be:

- None – meaning you do not have access to the asset and cannot use the asset for analysis.
- Site – meaning you have access to use the asset because it was produced by your research site.
- Submission – meaning you have access to use the asset because you submitted the asset to the BACPAC Data Portal for your research site.
- Project – meaning you have access to use the asset because you were granted access via the request process.

The asset request state or status is also shown to the right. It can be:

- Awaiting Submission – A new "Large File" asset submission was initiated, but the submission has not been marked as "Complete" by the submitter.
- Pending Approval – A new asset has been submitted to the BACPAC Data Portal and is under DAC review.
- Accepted – The asset and any new versions of the asset have been fully reviewed and approved by the DAC for use on the BACPAC Data Portal. Accepted assets are available to approved users from within their virtual machines.
- Awaiting Version Submission – The submission of a new version for an existing asset was initiated via the "Large File" submission mechanism, but the new version submission has not been marked as "Complete" by the submitter. Previously approved

versions of the asset are still available for use by approved users from within their virtual machines.

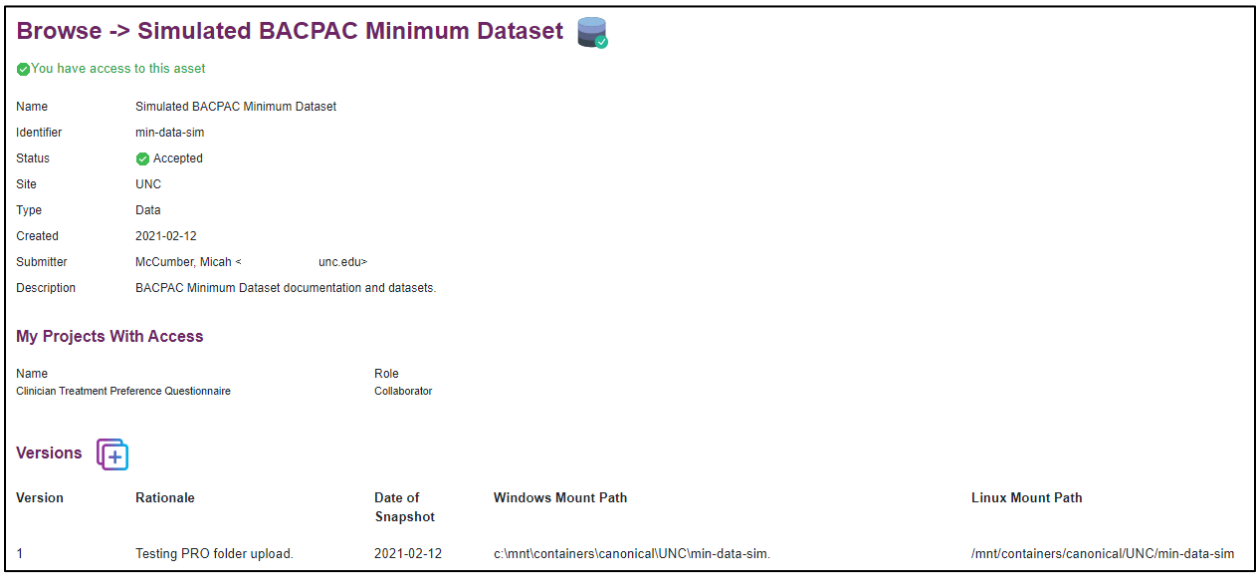
- Pending Version Approval – A new version of an existing asset has been submitted to the BACPAC Data Portal and is under DAC review. Previously approved versions of the asset are still available for use by approved users from within their virtual machines.

Any datasets to which you have access will be included in drives mapped on the virtual machines you create in the portal.

Note the search boxes at the top of the page. You can filter the display of assets on the page by asset type, your access status, the [research] site who owns the asset, asset status/request state, or by an ad-hoc search string.

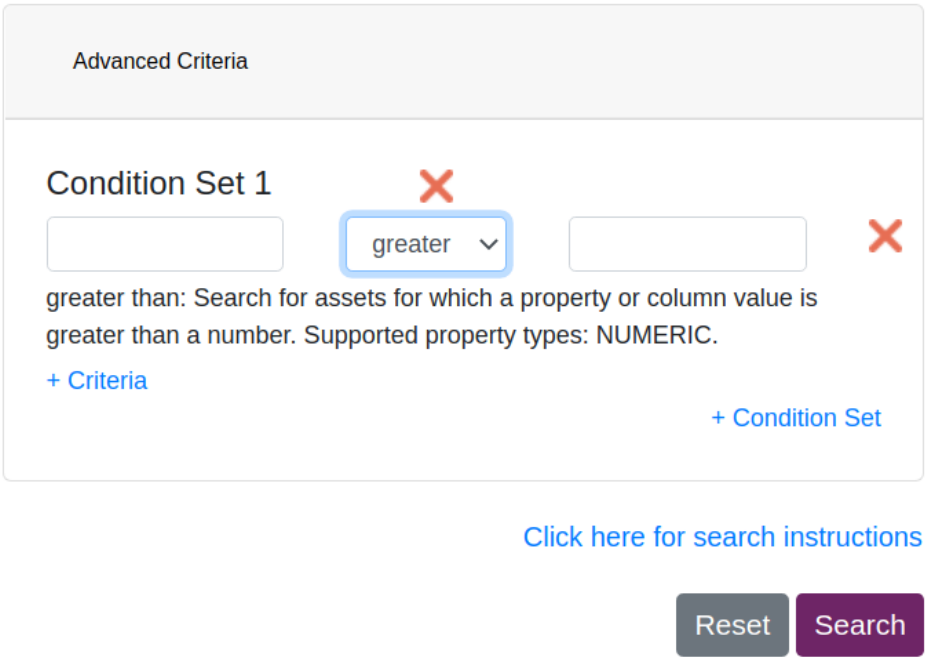
You can request access to data through the data access and publication request process in the BACPAC Research Consortium General channel on Microsoft Teams. At the top of the General channel, the tab called "BACPAC Data Access and Publication Request Submission" includes instructions on how to submit a request and a link to download the request form.

From the "Browse Assets" page, users can click on a specific asset to view more details about the asset, including initial date created, the user who submitted the initial asset, the asset description, associated projects, and the asset version history.



Advanced Search

Advanced Search criteria can be specified by clicking the “Advanced Criteria” Button. This allows users to search “meta-data” associated with each data set. Although this meta-data varies from set to set, typically searchable are things like the columns in the data set and their unique values and ranges.

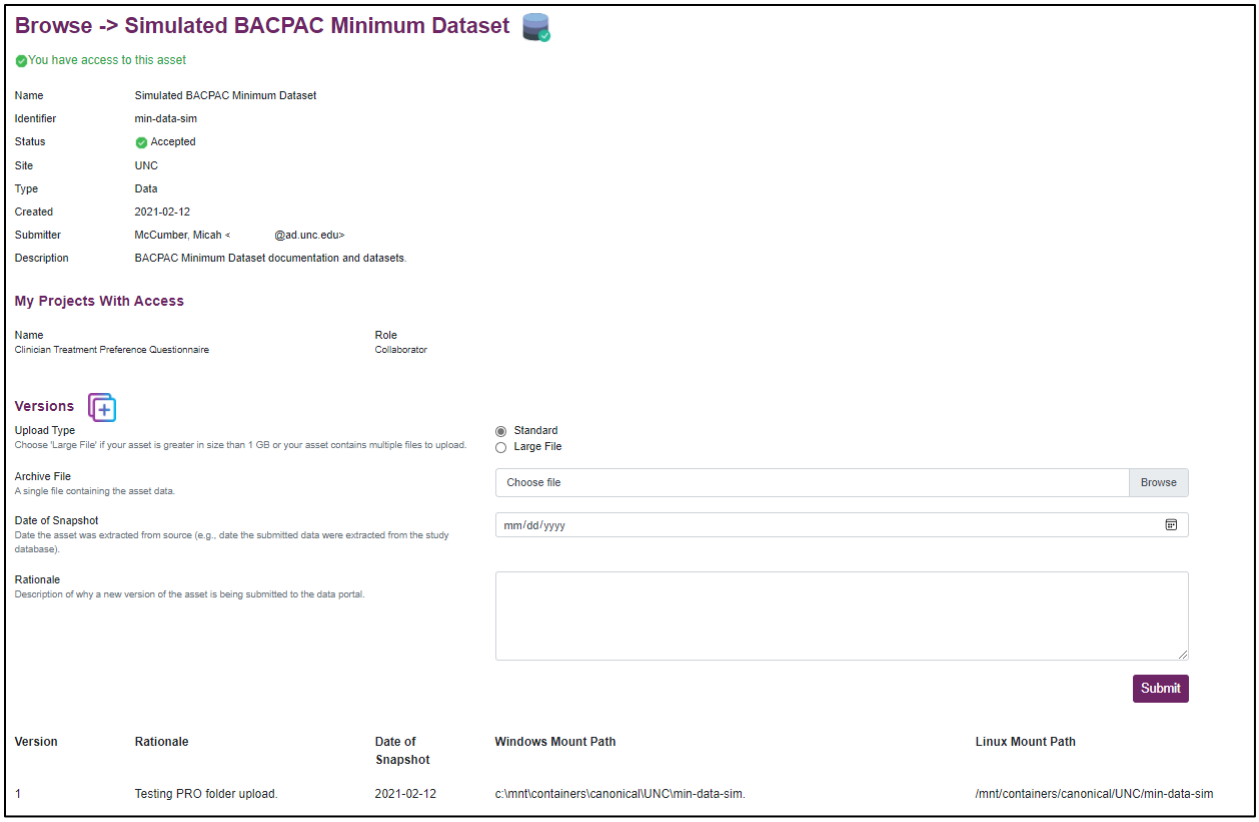


Click the “+ Condition Set” link to add more conditions. All conditions must be true for a data set to match.

Selecting a predicate will bring up documentation as to the behavior of the predicate.

Submitting a New Version of an Approved Asset

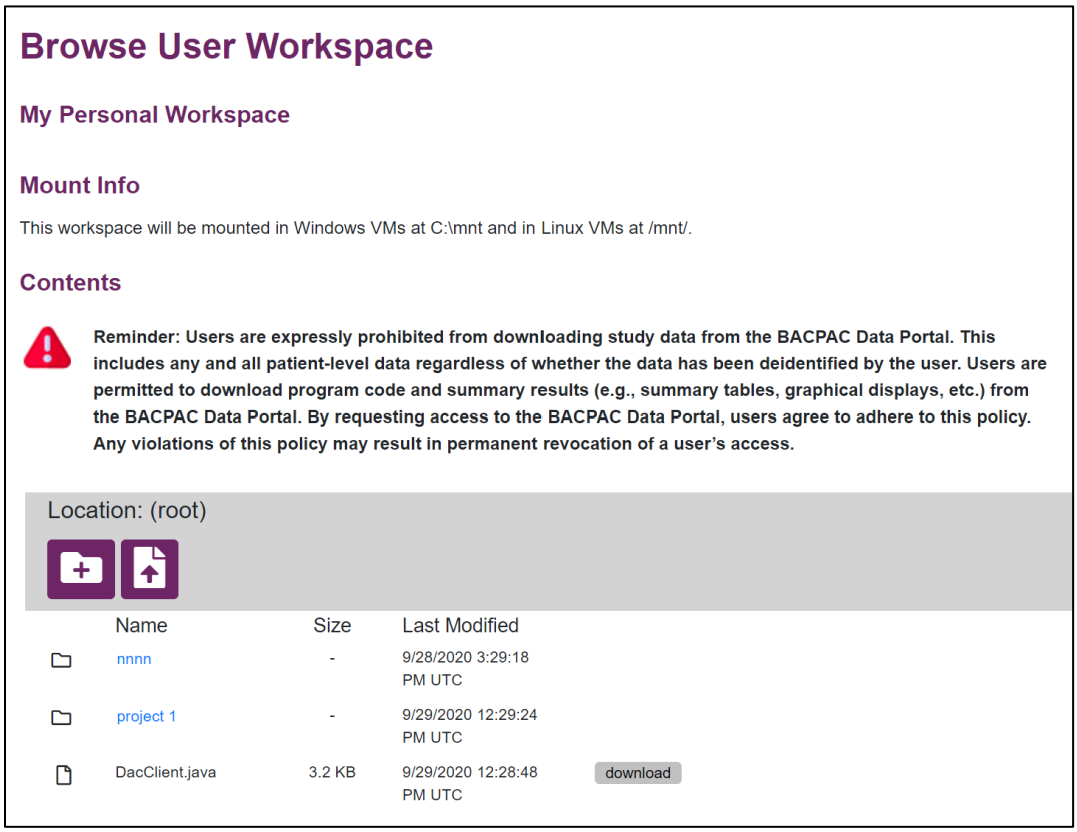
New versions of an approved asset are submitted via the detailed asset page, NOT from the "Submit Asset" page. To submit a new version, the submitter will click the next to "Versions" and complete the displayed form.



The form and process are similar to submitting a new asset (section 1.6), but the asset name and description are not required as those were assigned at the time of the initial submission. Standard uploads are submitted by completing the fields in the form shown above. The Large File uploads of a new version of an asset are initiated by completing this form. Once the form is complete, click the "Submit" button. For "Large File" submissions, additional instructions will appear at the top of the screen with a unique link to submit your file(s) using Microsoft Azure Storage Explorer (see section 2).

1.3. My Workspace

The BACPAC Data Portal allocates a workspace for each user.



The User Workspace comprises read / write file space accessible only to the user.

The workspace can be used for analysis output, logs, code, documents, or other files a user requires.

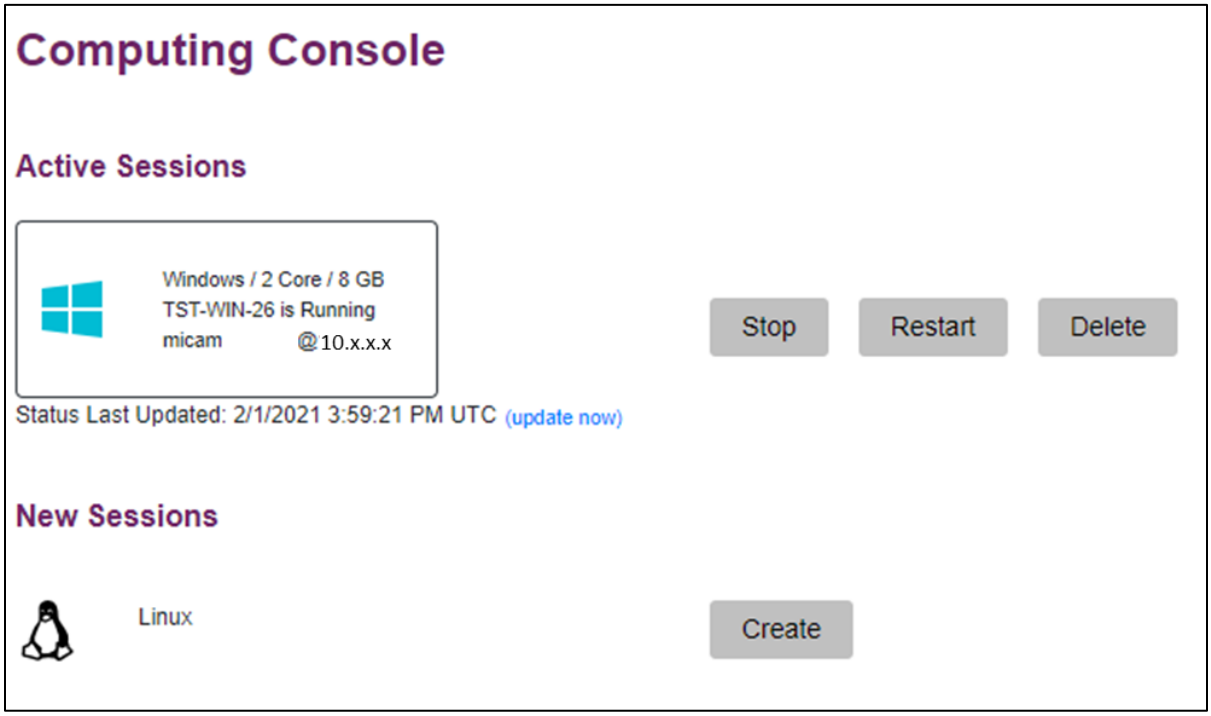
However, **it is forbidden to download study data from the BACPAC Data Portal**. This warning is prominently displayed on the User Workspace page. While the portal cannot prevent the download of study data, all file downloads are logged. The logs are scanned and the portal administrators are notified of any suspicious or prohibited activity.

From the user workspace page, you can create new directories and upload or download files.

- Click the  icon to add a new folder.
- Click the  icon to upload a file.
- Click the download icon to download a file. The download icon is displayed only for files, not for directories.
- To go into a directory to display its files, click the directory name.
- To navigate up the directory structure, click the "Location" links: Location: (root) / nnnn

1.4. Computing Console

The "Computing Console" page allows you to create, manage and access virtual machines (VMs) for data analysis.



Virtual machines will be in one of several states:

- Not yet created
- Creating
- Running
- Stopping
- Stopped

To create a new VM, click the "Create" button. **The creation process may take several minutes.** Refresh the page; when it is created its status will be "Running".

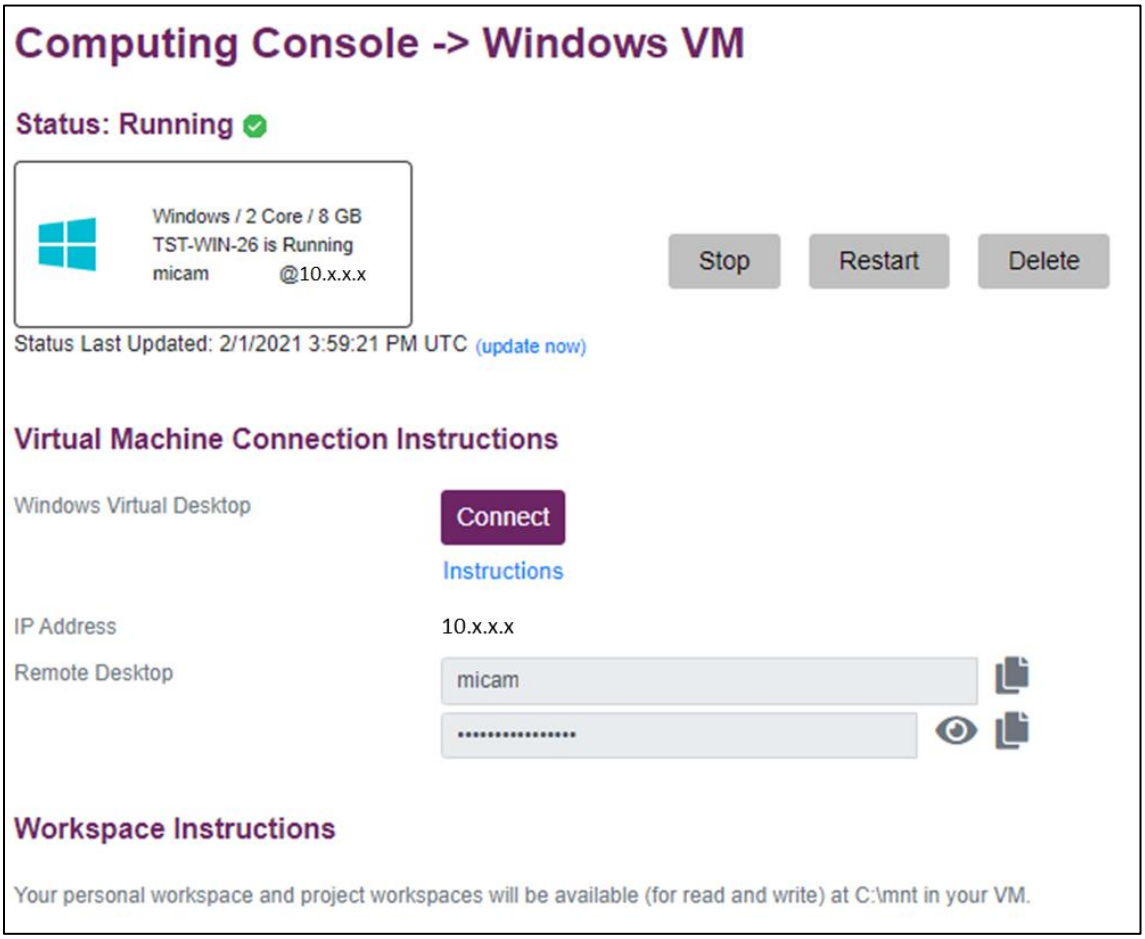
Active (running) sessions can be stopped, restarted or deleted.

- Stopping a VM keeps the VM instance but stops it from consuming resources.
- Restarting the VM starts a stopped VM and prepares it for your use.
- Deleting a VM stops the running VM and removes it from your account. To use it again you will have to create it again.

In the example screenshot above, a Windows VM has been created and is running and ready to be used for analysis. A Linux VM is not yet created.

It is good practice to **stop a VM once analysis programs are finished and you have logged out of the session**. This saves resources and reduces cost.

The VMs you create are accessible to you only. Once created, they exist until you delete them. Click the box describing the Windows VM. A detailed view of the running Windows VM is shown.



The "Workspace Instructions" at the bottom of the detailed VM page provides the mount location of your personal and project workspaces to which you have access. In addition to these, you will have read access to any canonical data you have been granted.

- You have access to any data from your research site.
- You also have access to any canonical data used for projects on which you collaborate.

To run data analyses, connect to the VM, open a remote desktop or command line session and you will have access to the software installed on the VM.

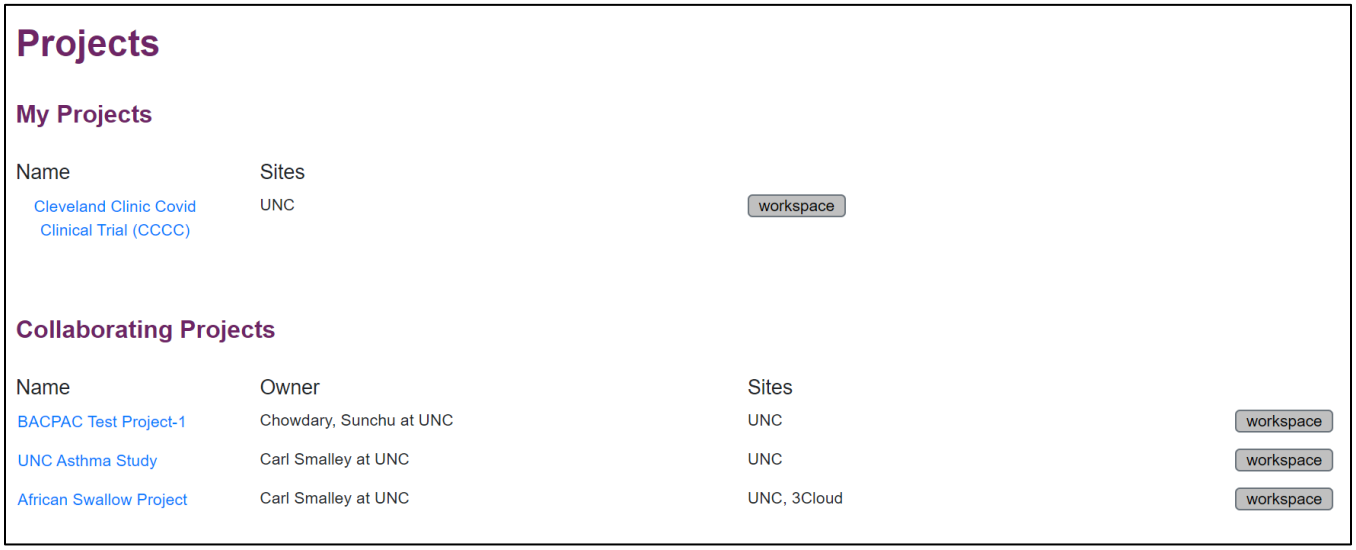
- If you have problems accessing a running VM, stop the VM and restart it.
- If that does not solve the problem, delete the VM and recreate it.

Step-by-step instructions for logging into Windows and Linux VMs are provided in section 2.

Instructions for using the Linux and Windows VMs, including accessing data, using provided software, and writing and sharing output results and code are provided in section 4.

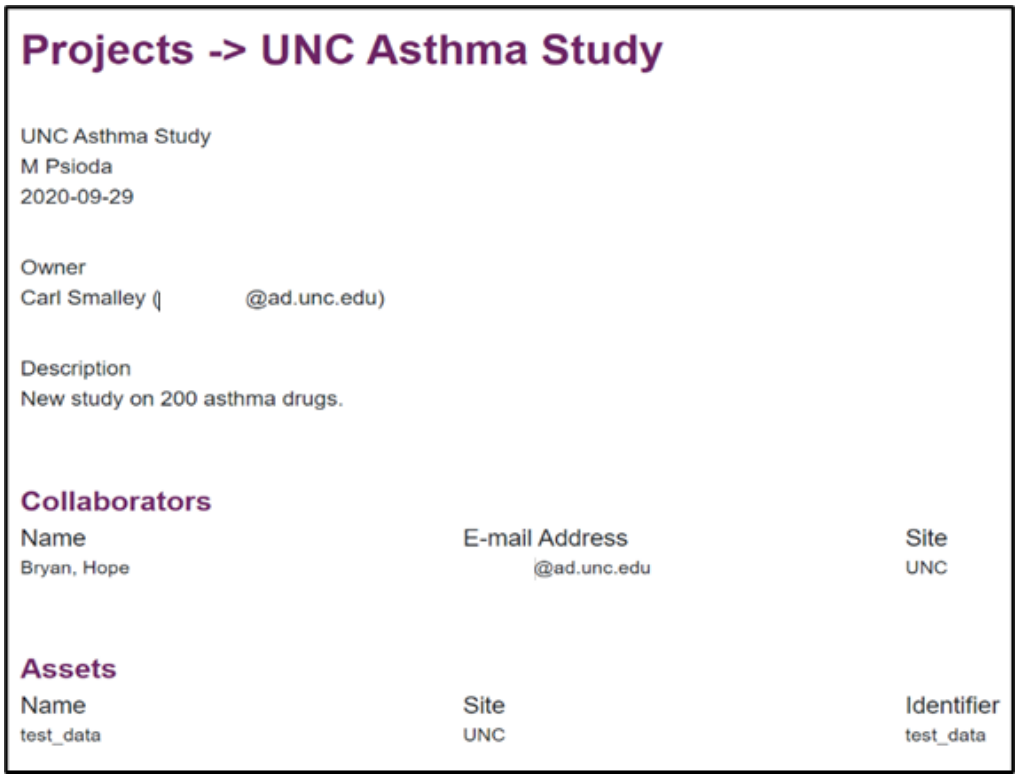
1.5. Projects

The Projects page lists all the projects you are working on, both your own and those on which you are collaborating with other researchers.

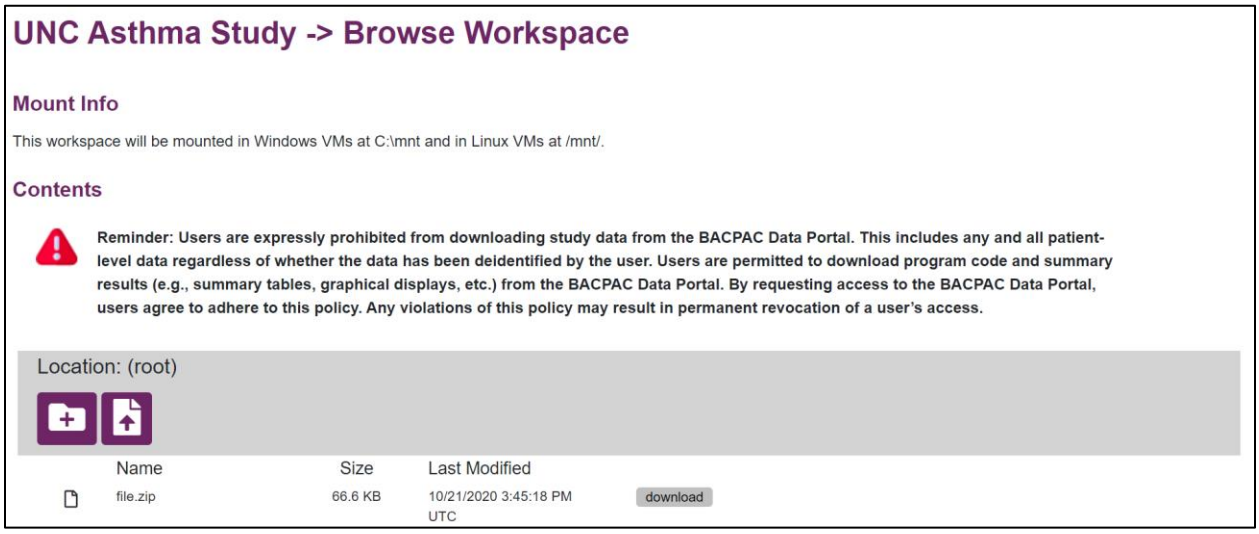


Each project has an associated workspace where files related to the project can be written.

Click on "Project Name" to display the project details.



Click on the "Workspace" button associated with a project to display the project details.



From the Project Workspace page, you can view, upload and download files associated with the project. Unlike User Workspaces, which are restricted to the single user, Project Workspaces can be used by all researchers collaborating on a project. This is the mechanism by which you share results, code and other project artifacts.

Note the warning on this page: **You are prohibited from downloading study data from the BACPAC Data Portal.** While it is possible to output an entire canonical dataset to a project or user workspace where it can be downloaded, that violates the portal user agreement. Such downloads will be recorded in the system logs so any violation will be detected.

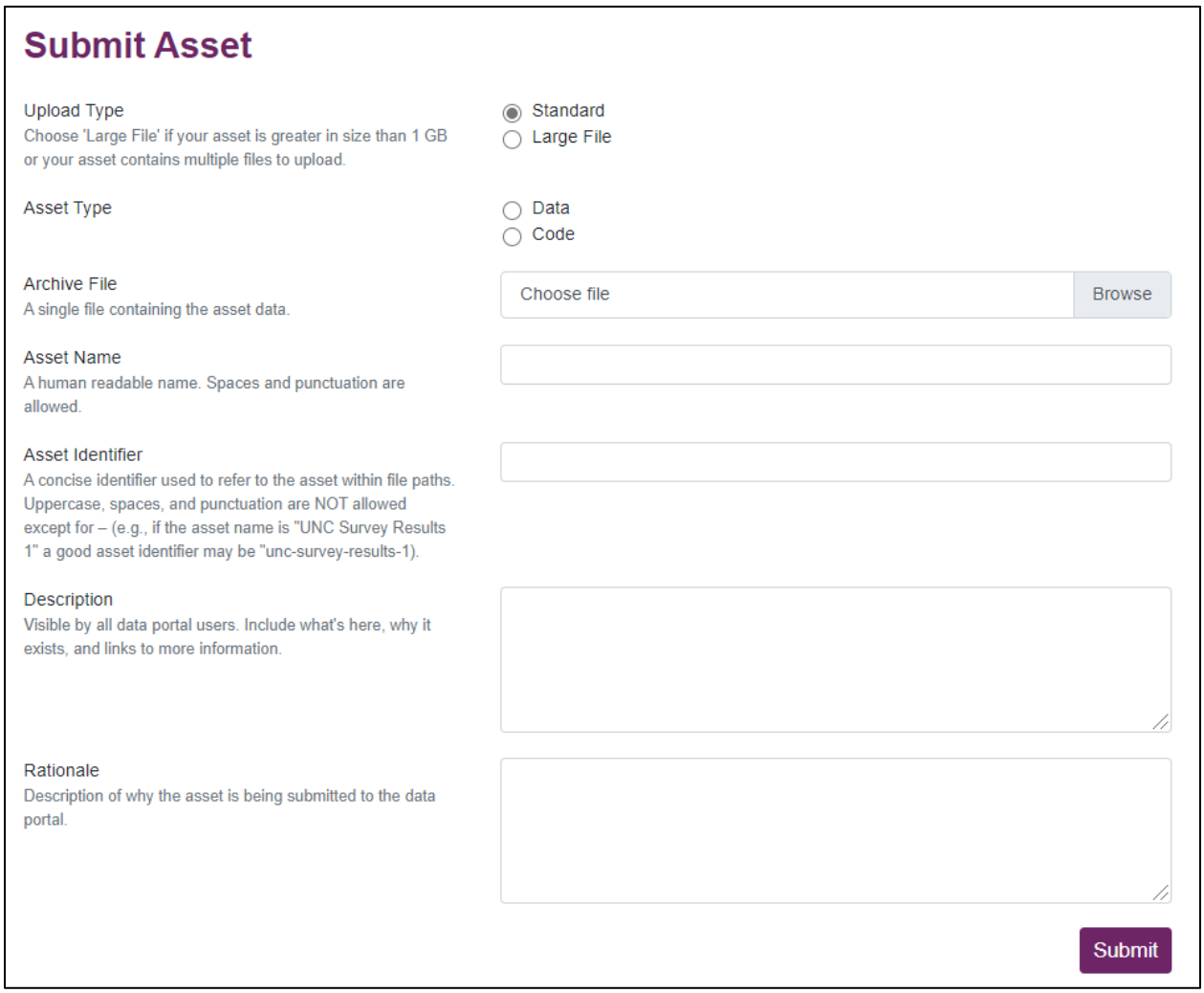
1.6. Submit Asset (*New Assets Only*)

Some users have the ability to submit assets to the BACPAC Data Portal. Assets comprise study (canonical) data and code useful in BACPAC study data analysis.

To submit a **NEW** asset, complete the form displayed below. The fields have short explanations that describe the information needed.

Standard uploads are limited to small, single files (e.g., an R program or single datafile with no documentation). Most asset submissions will be "Large File" uploads, which include all submissions with a single file greater than 1 GB or submissions with two or more files.

Once the form is complete, click the "Submit" button. For "Large File" submissions, additional instructions will appear at the top of the screen with a unique link to submit your file(s) using Microsoft Azure Storage Explorer (see section 2).



Once an asset is submitted and approved it will appear on the "Browse Assets" page as "Accepted". New versions of an approved asset can be submitted from the detailed asset page accessed via the "Browse Assets" page.

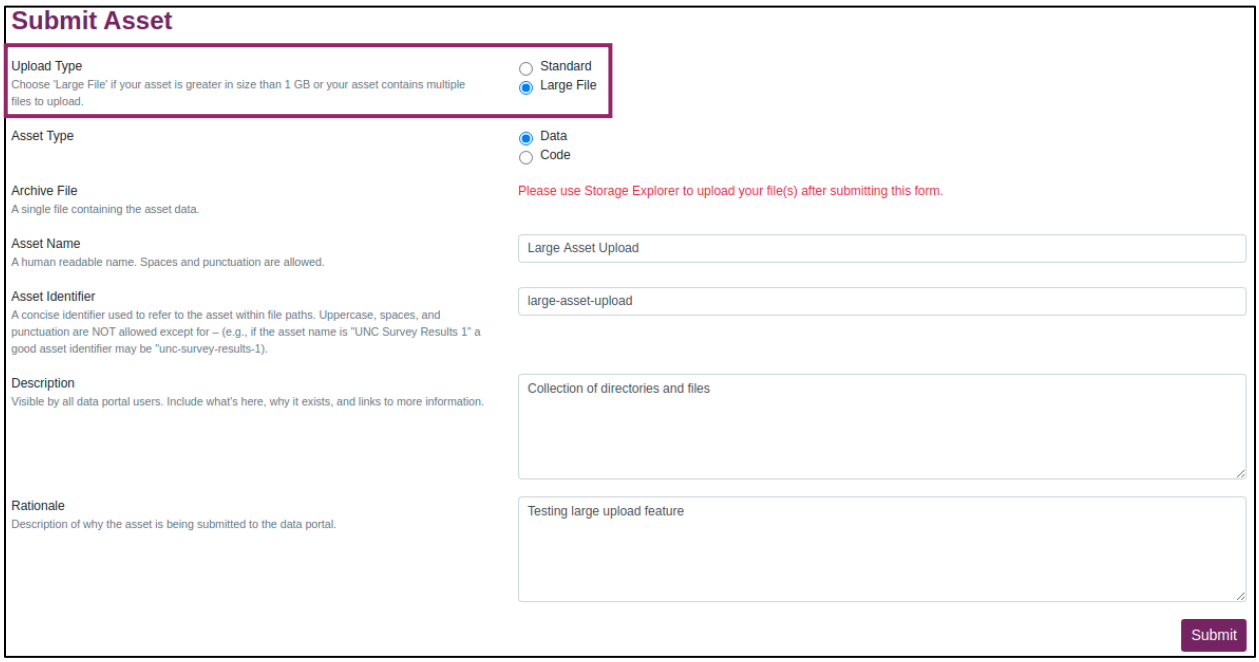
2. Submitting Assets using Large File Upload

In order to upload large files (e.g., any file which are larger than 1GB or any collection of files), BACPAC users should utilize the "Large File" upload feature of the BACPAC Data Portal.

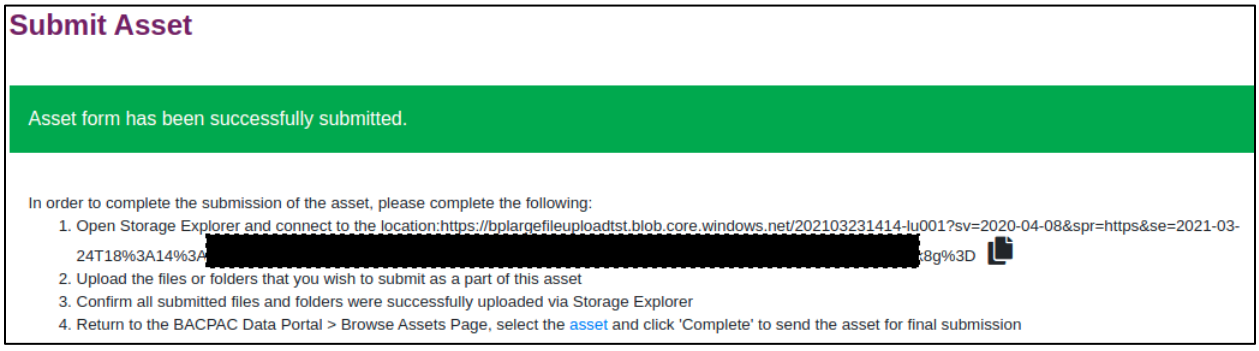
New assets are submitted from the "Submit Asset" page (see section 1.6) and new versions of an existing asset are submitted from the "Browse Assets" > detailed asset page (see section 2). From both of these pages, users can select "Large File" as the upload type.

***Note**: Prior to utilizing the large asset upload feature, please download "Azure Storage Explorer" and contact the BACPAC DAC [bacpac_dac@unc.edu] to provide your IP address.*

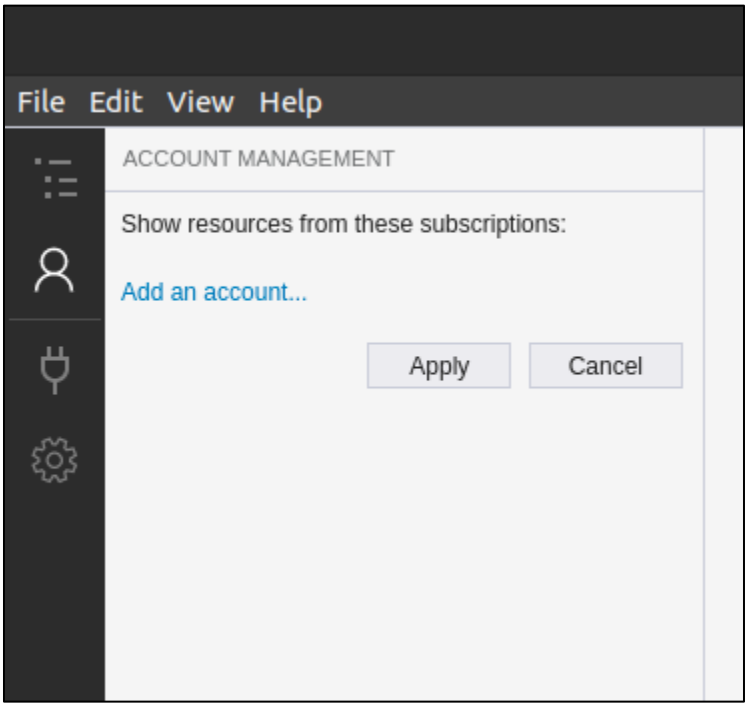
a. Select the "**Large File**" radio button and fill out the remainder of the submission form.



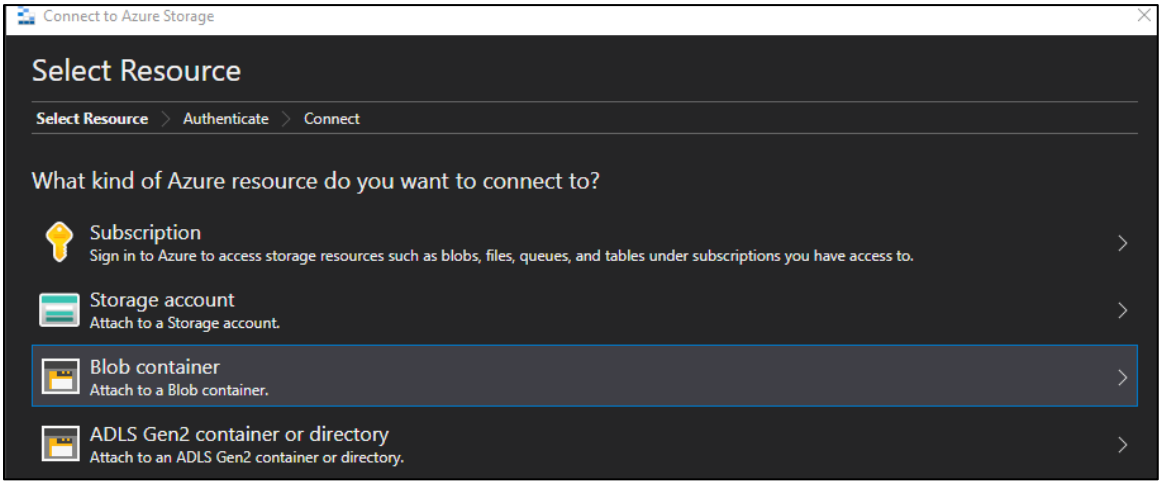
b. After clicking on "**Submit**", you will be presented with a link (i.e., SAS URI) at the top of the page, copy the entire SAS URI (e.g., from "https://..." onward) by clicking , the copy icon.



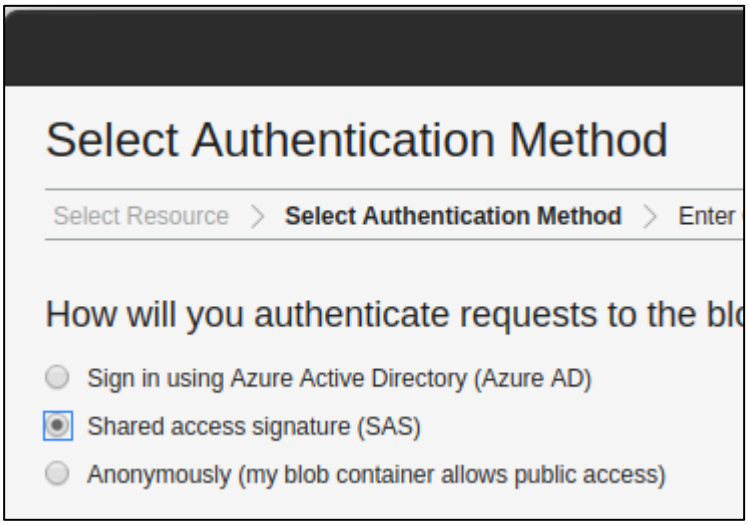
c. Open Azure Storage Explorer, click  , the "Open Connect Dialog" icon, on the left-hand side.



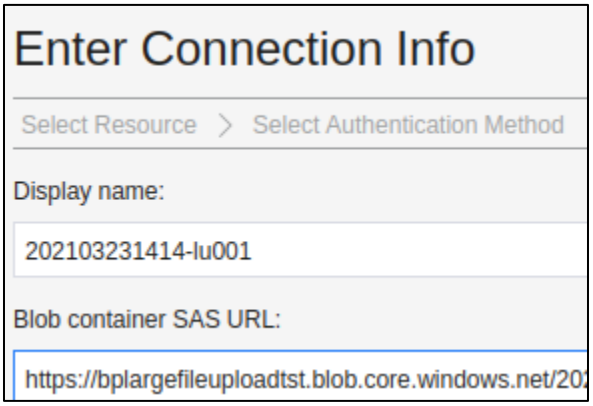
d. In the resulting pop-up select "Blob container" as shown below.



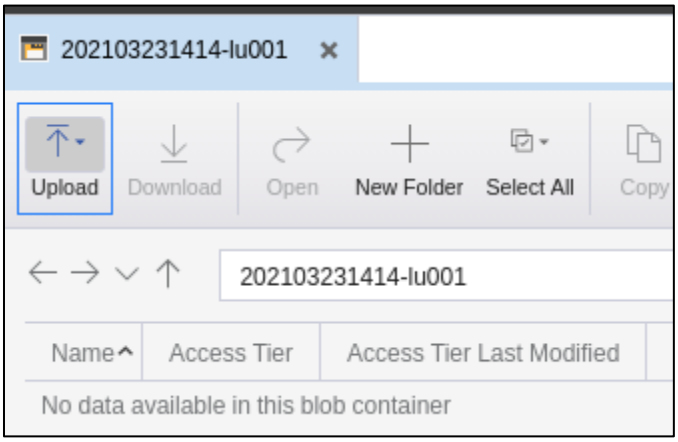
e. In the resulting pop-up (“Select Authentication Method”) select the “Shared access signature (SAS)” radio button.



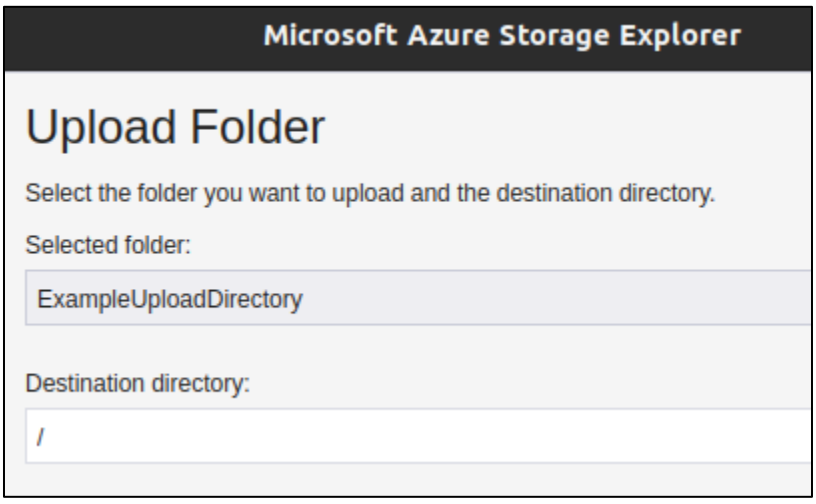
f. Paste the SAS URI copied from the BACPAC Data Portal page in step b in the "Blob container SAS URL" box in the resulting pop up and click "**Next**".



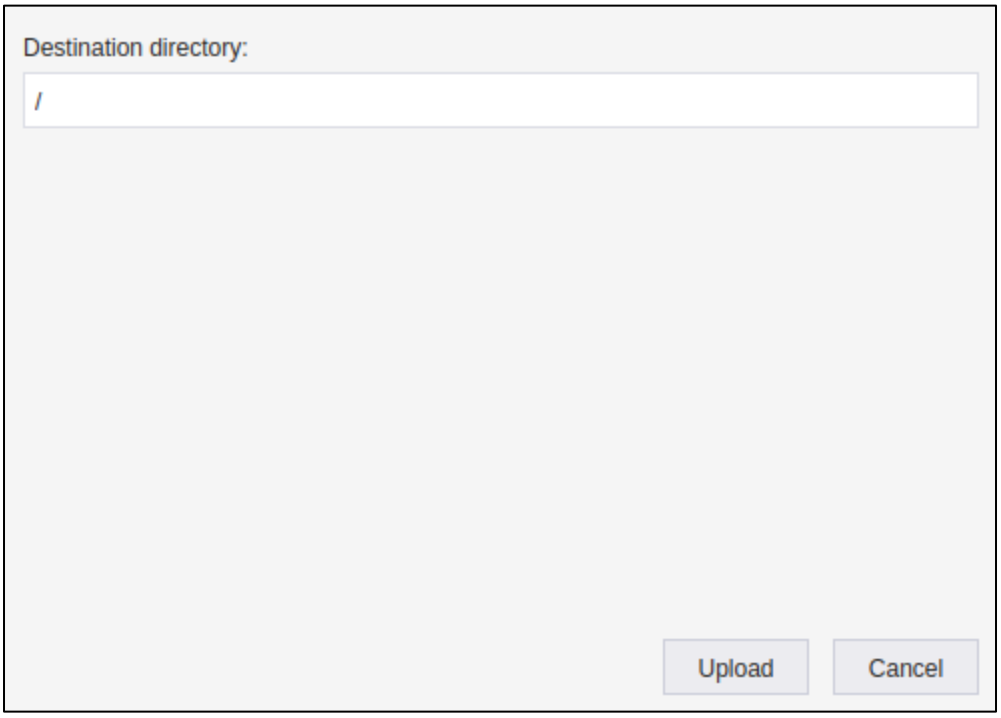
g. Click the "**Upload**" icon.



h. Select the folder or file that you would like to upload.



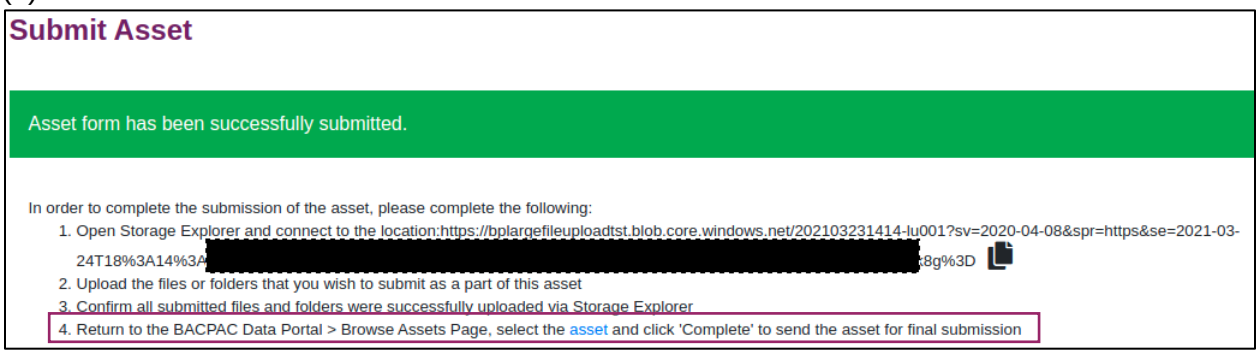
i. Select "**Upload**" at the bottom of the dialog box (**do NOT adjust any other fields after selecting the folder to upload**).



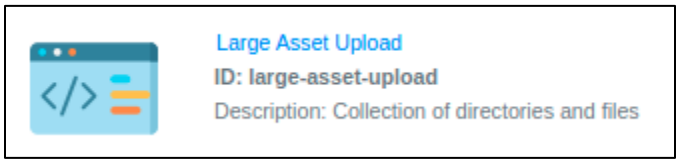
j. Once the transfer has completed you must return to the BACPAC Data Portal and mark your asset upload as "Complete". *Note: This step is necessary in order to notify BACPAC staff that your asset is ready for review and approval; you may continue to upload related files or folders for this asset until you have marked the upload as completed*

To complete the asset upload, you must navigate to the detailed asset page for your asset. This can be done by (1) clicking the "asset" hyperlink in step 4 of the Large File upload instructions or (2) returning to the BACPAC Data Portal "Browse Asset" page, searching for your asset, and clicking on your asset to get to the detailed page.

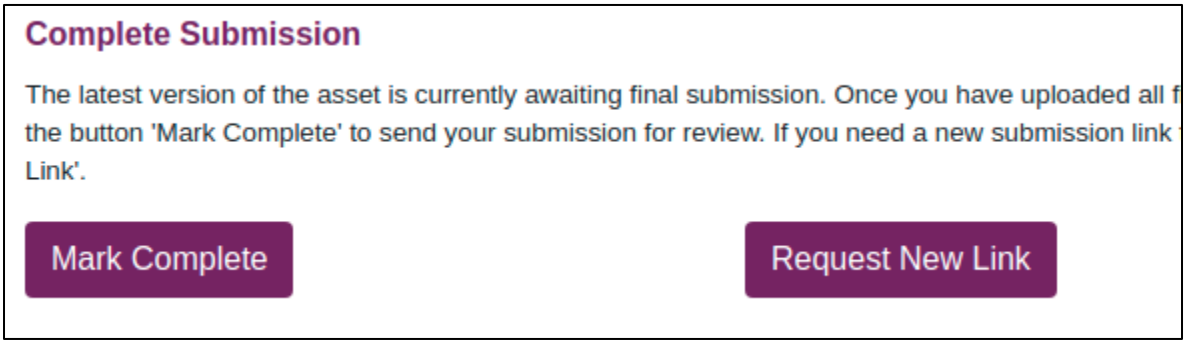
(1)



(2)



k. Click on "**Mark Complete**":



*Note: In the event that your SAS URI expires prior to your upload completing, you may click on the "**Request New Link**" button in order to generate a new SAS URI which will allow you to continue your upload*.

3. Creating and Logging into Windows and Linux VMs

It is helpful to understand user accounts in the BACPAC Data Portal. Each user has two accounts; the first to log in to the BACPAC Data Portal and the second to access and log in to the virtual machines (VMs).

1. The account used to log in to the portal is the account to which the invitation to access the BACPAC Data Portal was sent. Most often, this will be the account you use to access MS Teams.

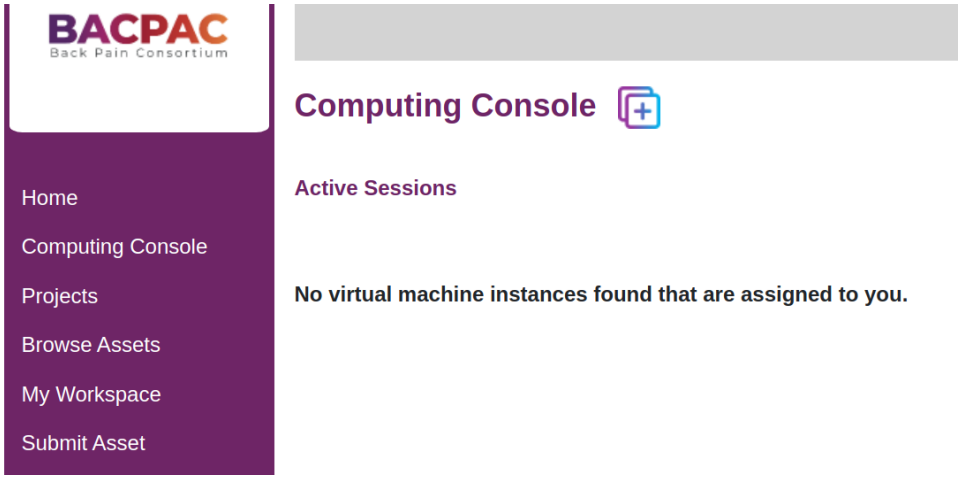
2. The account to access and log in to the VMs is provided when you create a VM within the BACPAC Data Portal. This second account is only used when accessing VMs from the portal. The account has the format:

   user first name + last initial + 5 digits @ *bacpacresearch.org*

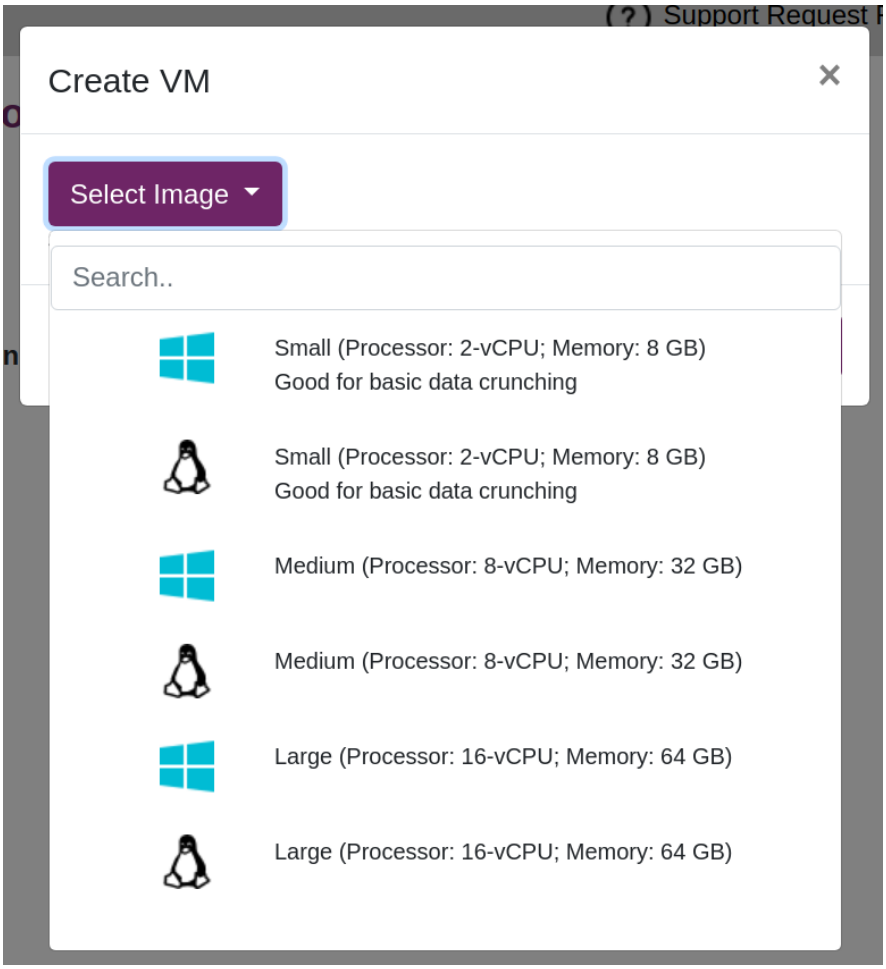
You will login to the VM using your assigned user name, which is the first part of the *bacpacresearch.org* account (user first name + last initial + 5 digits).

3.1. Creating Virtual Machines

Users have the option of creating Linux or Windows virtual machines in three sizes and may run up to 3 machines simultaneously. Initially, the user will have no machines when they visit their Computing Console page:



To create a virtual machine, click the "+" icon next to the text Computing Console. This will open a menu from which you can select your virtual machine type and size:



After selecting a machine and confirming, you will be directed to a page reporting the machine's status. After a few minutes it should (upon refresh) show something like this (for a Linux virtual machine).

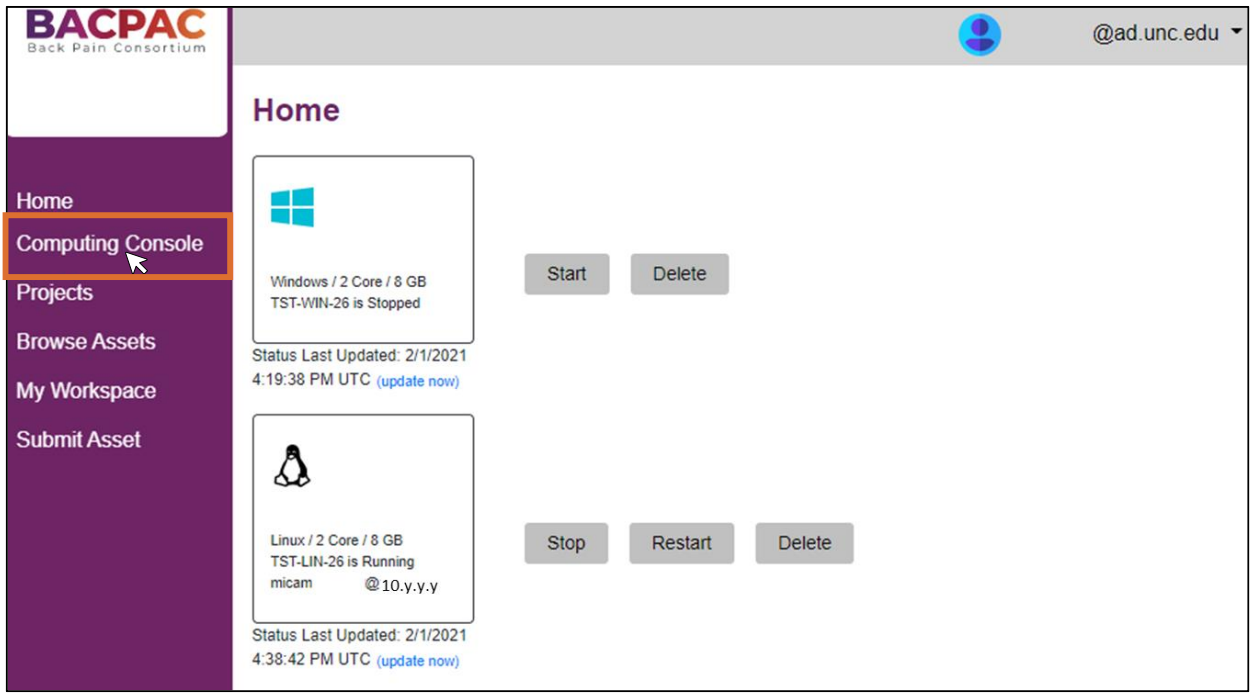
Virtual machines have a running cost. After you are finished with one, you can return to this page and press the “stop” button to shut the machine down until it is needed. The “delete” button will permanently delete the machine along with all the data on it (except that in your personal space, see Section 1.3).

3.2. Accessing a VM

All VMs must be accessed by first logging into the BACPAC Data Portal. Once in the portal, users will need to connect to a VM and enter their *bacpacresearch.org* credentials two times to successfully accessing their VM.

1. The first login is to access the Windows Virtual Desktop.
2. The second login is to access the user's specific VM.

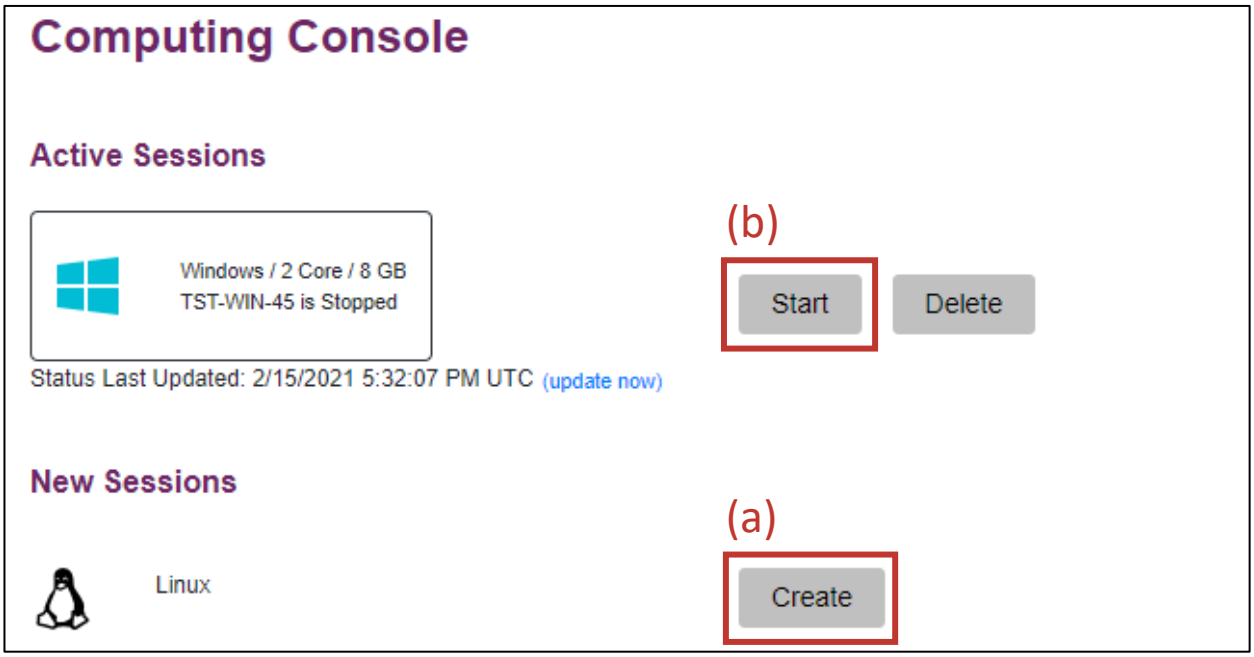
Step A. Navigate to the Computing Console page to connect to a VM.



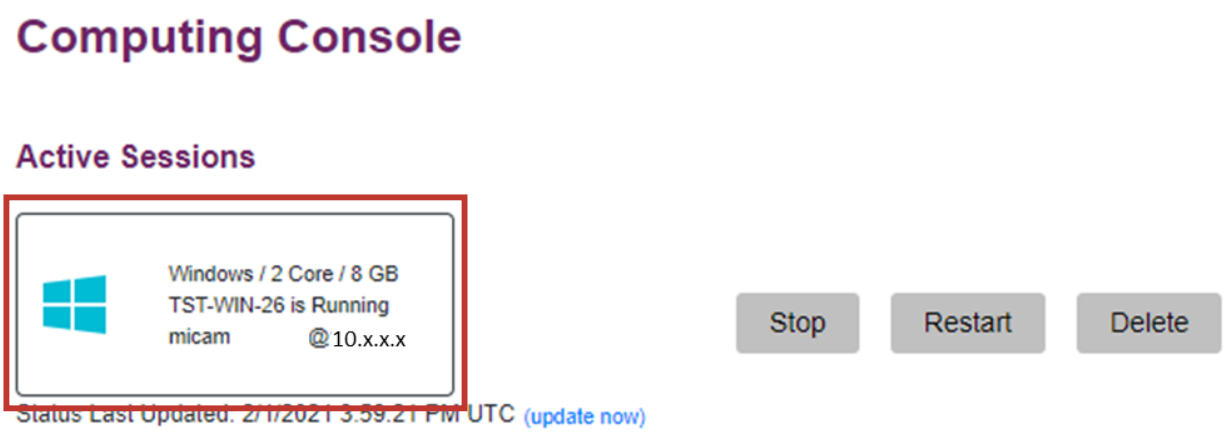
Once on the Computing Console page, you will need to confirm you have an active and running VM.

(a) If you do not currently have an active session for your VM, then you need to click "Create" to start a new session as shown for the Linux VM below.

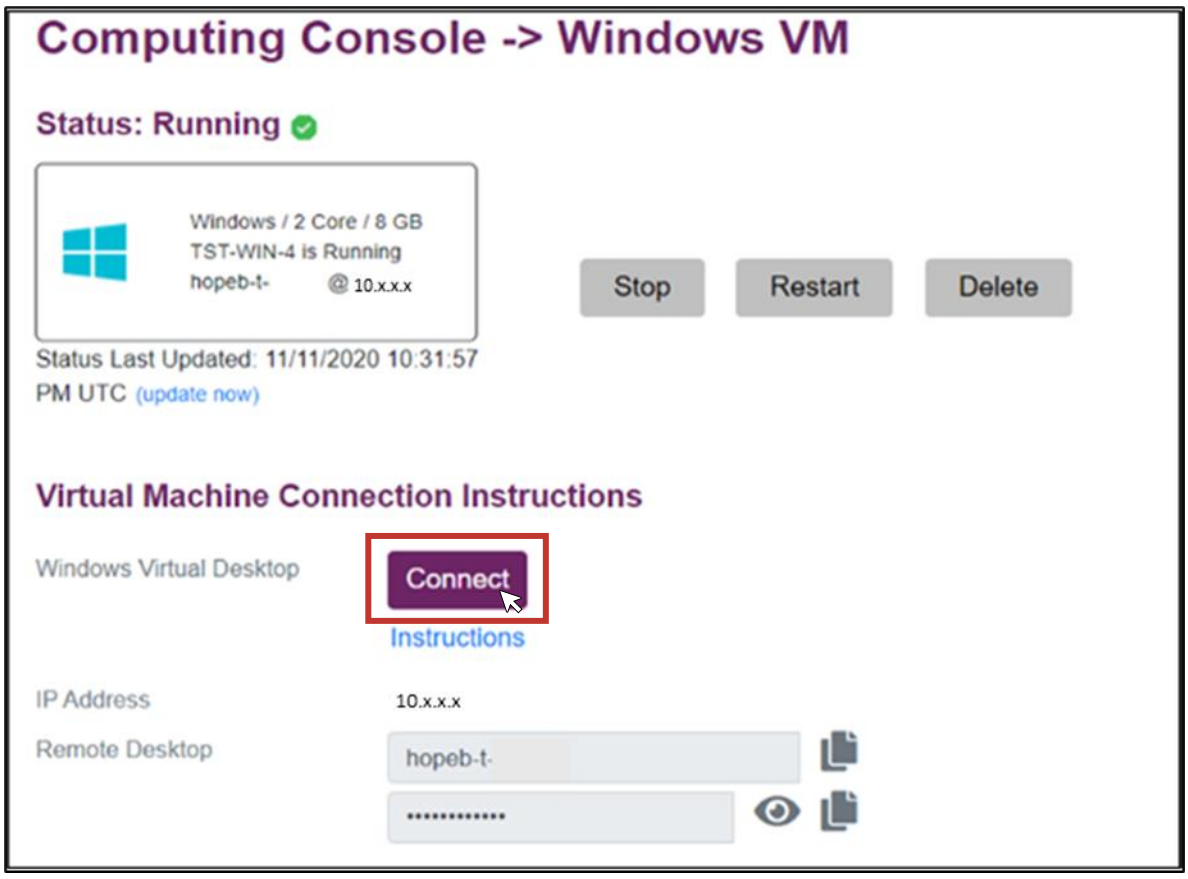
(b) If you have an active session for your VM that is stopped, then you need to click "Start" as shown for the Windows VM below.



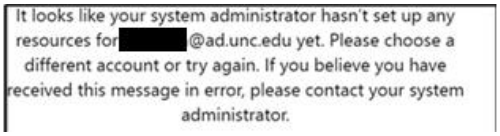
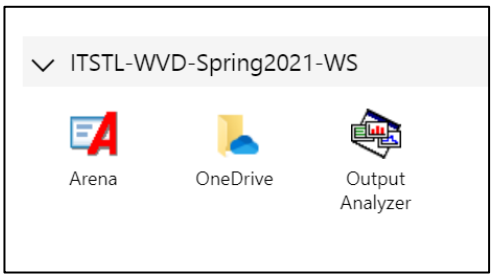
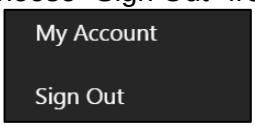
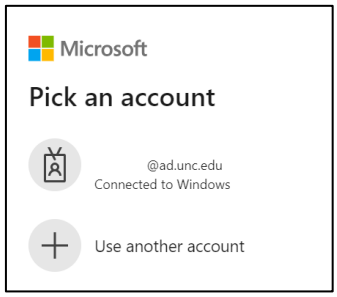
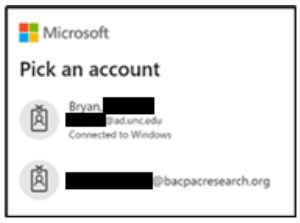
Step B. When you have an active VM that is running, then you will click the VM box to go to the detailed VM page as shown for the Windows VM below.



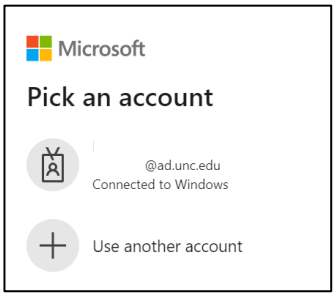
Step C. To access a VM click "Connect" from the Computing Console -> Windows VM or Computing Console -> Linux VM pages.



A new browser tab will open. In the new browser tab, you will see one of several screens. In the table below, find the screen you see and follow the associated instructions.

	Screen	Instructions
Option 1	 *OR* 	**O1. If you see the Option 1 screen, follow the steps below:** **O1.1. Step 1:** In the upper right corner of the screen, click your name and choose "Sign Out" from the menu.  **O1.2. Step 2:** Close the browser tab and then continue to the steps below. You do not have to choose an account if a list is shown. **O1.3. Step 3:** Return to the Computing Console -> VM page and click "Connect". O1.3.a. If you see a screen similar to the one below WITHOUT your account ending in *bacpacresearch.org*, follow the steps outlined for Option 2 (starting at O2.1).  O1.3.b. If you see a screen similar to the one below WITH your account ending in *bacpacresearch.org* displayed, then follow the steps outlined for Option 3 (starting at O3.1). 

Option 2



O2. If you see the Option 2 screen, follow the steps below:

O2.1. Step 1: If you do not see the account ending in *bacpacresearch.org*, then select "Use another account".

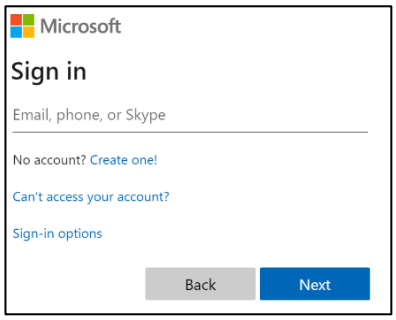
O2.2. Step 2: Enter your user name.

O2.2.a. In a separate browser tab, go to the BACPAC Data Portal page.

O2.2.b. In the top right-hand corner of the BACPAC Data Portal page, click your portal account then click "Profile" in the dropdown beside your user name.

O2.2.c. The profile page lists the user name and password you will use for the VM login account. Copy the user name.

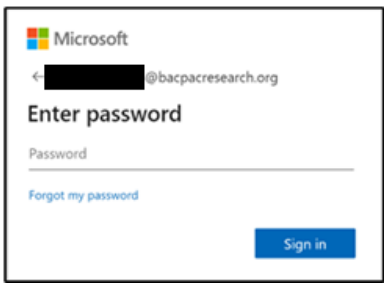
O2.2.d. Go back to the VM browser tab and paste the user name into the account sign in page and click "Next":



O2.3. Step 3: Enter your password.

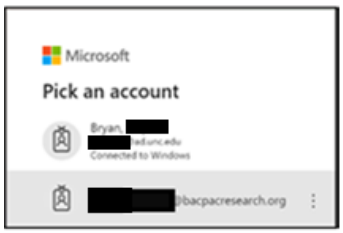
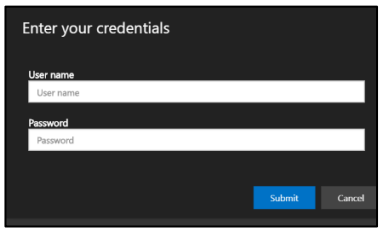
O2.3.a. Return to the portal profile page and copy the second field (i.e., the password).

O2.3.b. Paste the password into the sign-in screen.



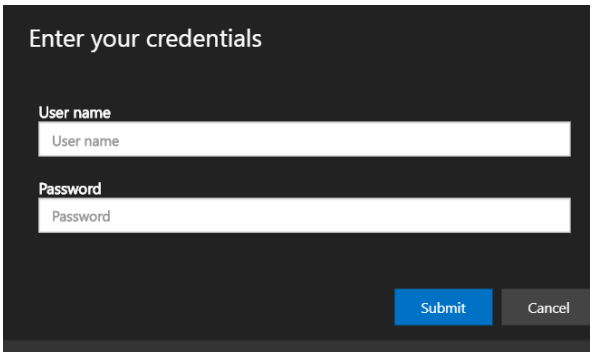
O2.4. Step 4: Click "Sign in".

O2.5. Step 5: Follow the steps outlined for Option 4 (section O4).

Option 3		**O3. If you see the Option 3 screen, follow the steps below**: **O3.1. Step 1:** If you see your account ending in *bacpacresearch.org*, select that account. **O3.2. Step 2:** Go to the Computing Console VM page where you clicked the "Connect" button, there are two fields labeled "Remote Desktop". **O3.3. Step 3:** The first field is your user name. Copy and paste the second field value into the Password field. **O3.4. Step 4:** Click "Sign in". **O3.5. Step 5:** Follow the steps outlined for Option 4 (section O4).
Option 4		**O4. If you see the Option 4 screen:** **Click on "Putty" for command line access to Linux Compute** **OR** **Click on "Remote Desktop for Windows or Linux GUI".** **Continue to Section 3.3. Windows Virtual Desktop (WVD).**
Option 5		**O5. If you see the Option 5 screen:** **Continue to Section 3.3. Windows Virtual Desktop (WVD).**

3.3. Windows Virtual Desktop (WVD)

This is the Windows Virtual Desktop (WVD) login screen which will allow you to access your Linux and Windows VMs.



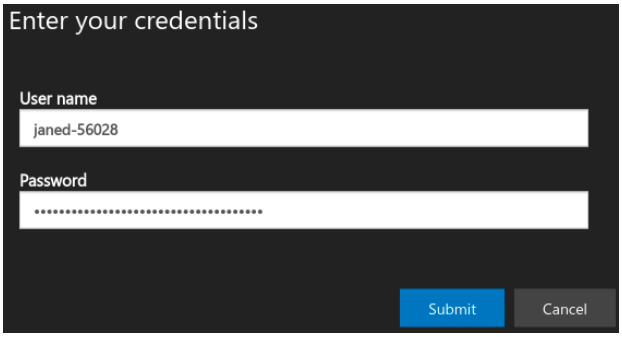
From this screen, do the following:

a. Click on the BACPAC Data Portal tab in the browser.

b. On the Computing Console VM page where you clicked the "Connect" button, there are two fields labeled "Remote Desktop". Copy and paste the first into the "User name" field in the dialogue shown above. "User name" will have the format FirstNameLastInitial-12345.

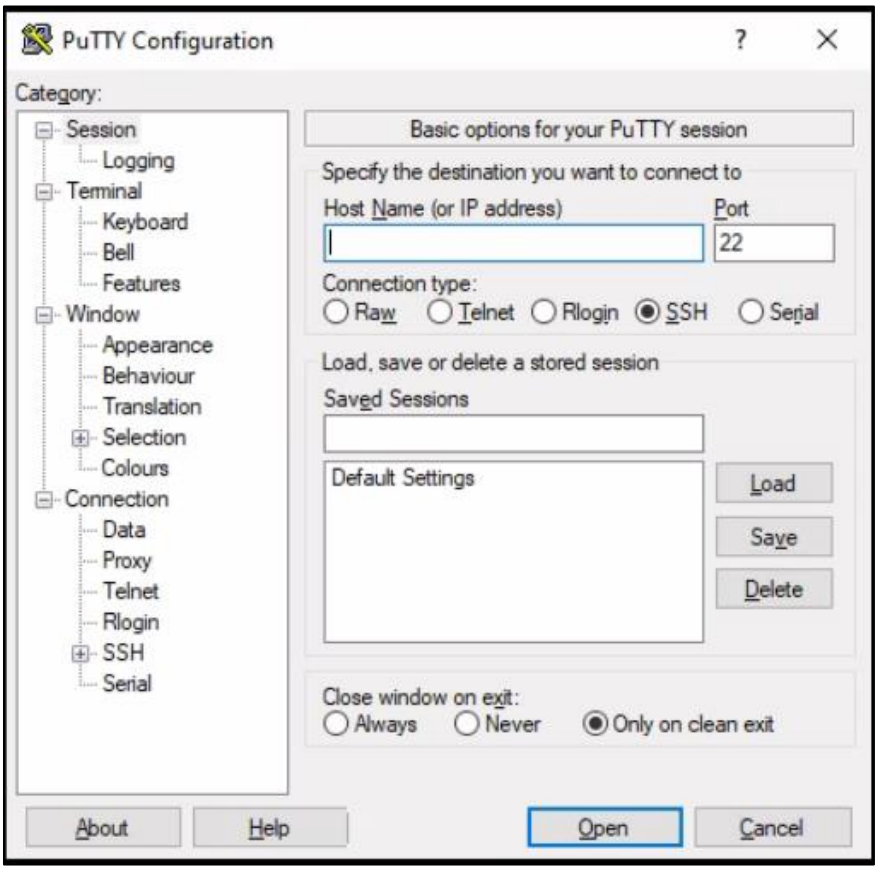
c. Return to the BACPAC Data Portal tab in the browser to copy and paste the second into the "Password" field in the dialogue box.

d. Click "Submit".



3.4. Putty (Linux VM)

If you have chosen to connect to a Linux VM using "Putty", the configuration screen will be shown:



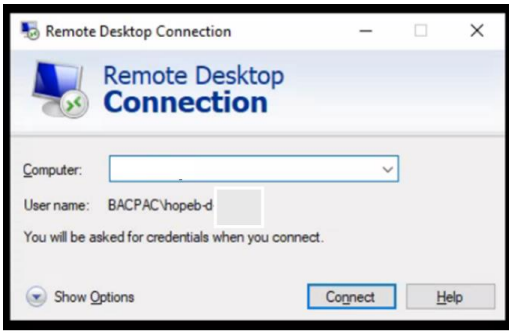
a. Click the browser tab to go back to the BACPAC Data Portal page displaying the Computing Console.

b. Locate the IP address for the Linux VM and enter it into the "Host Name or IP address" field. **Note: You cannot use copy / paste to enter the IP address.**

c. Click "Open" and the Linux VM computer console will display.

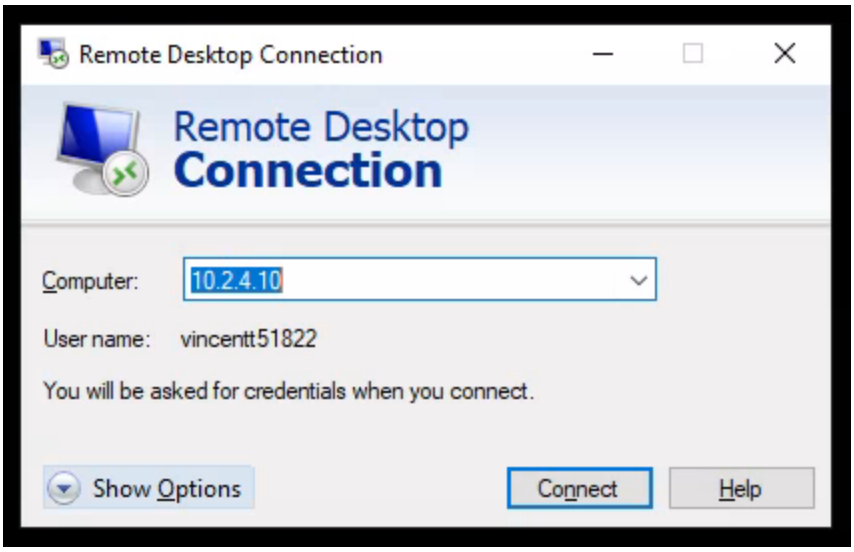
3.5. Remote Desktop (Windows VM)

If you have chosen to connect to a Windows VM using "Remote Desktop", you will see this screen:

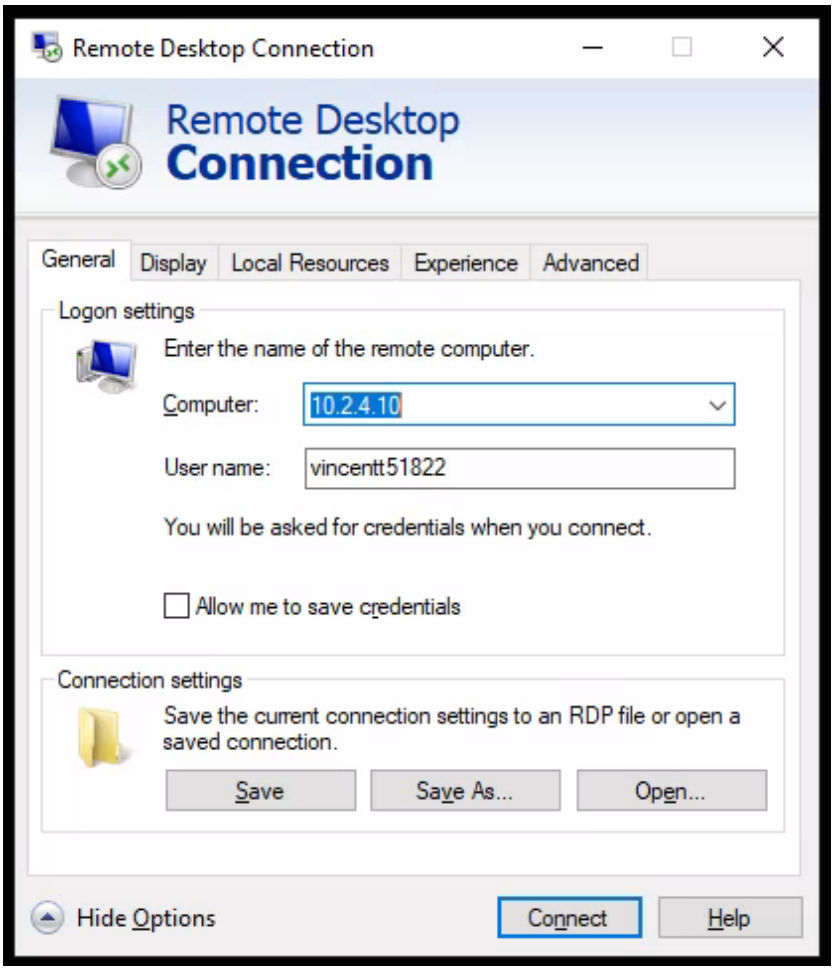


a. Click the browser tab to go back to the BACPAC Data Portal page displaying the Computing Console.

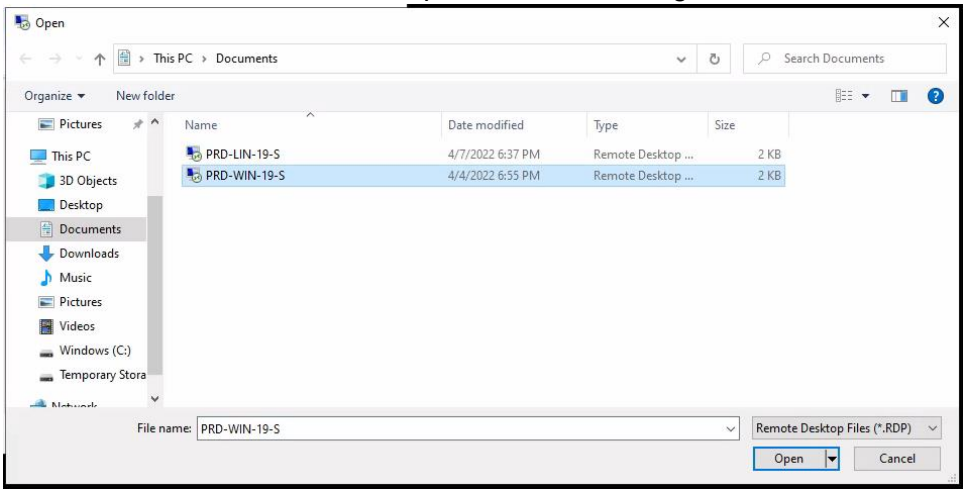
b. Click the “Show Options” button on the bottom left of the Remote Desktop Connection Window.



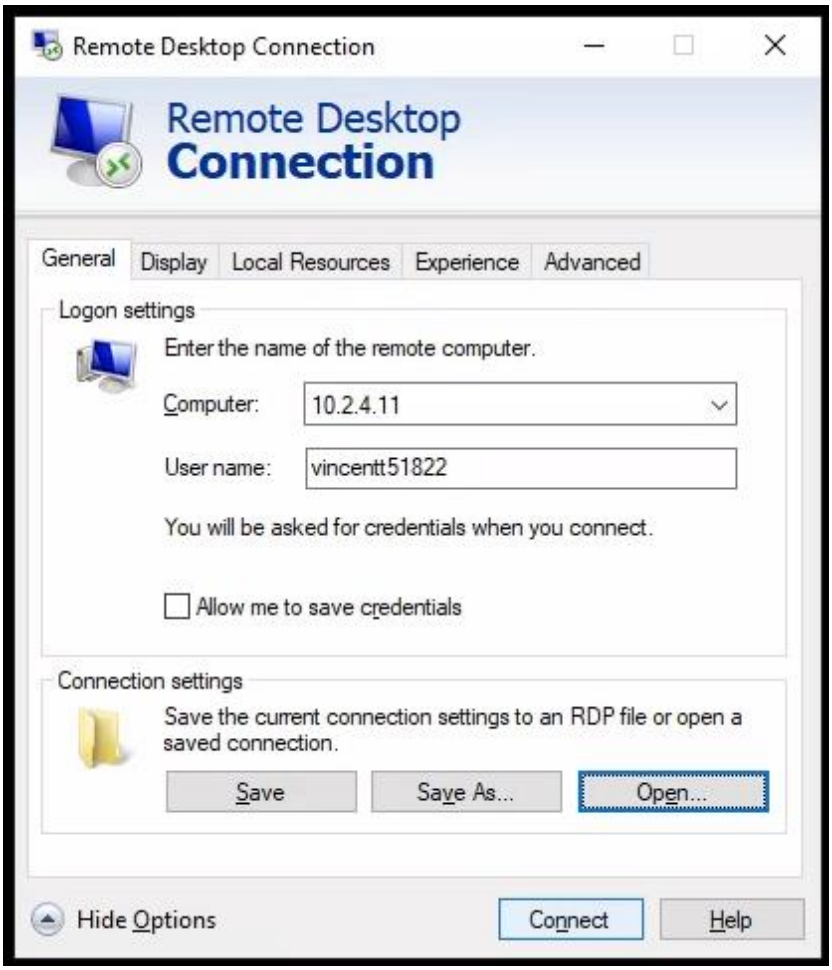
c. Then click “Open,” which will allow you to select a profile:



d. Select the profile corresponding to your VM OS and Size (here the **WIN**-19-**S**) means a small Windows VM. The click “Open”. This will configure the connection for that VM.



e. Click “Connect”



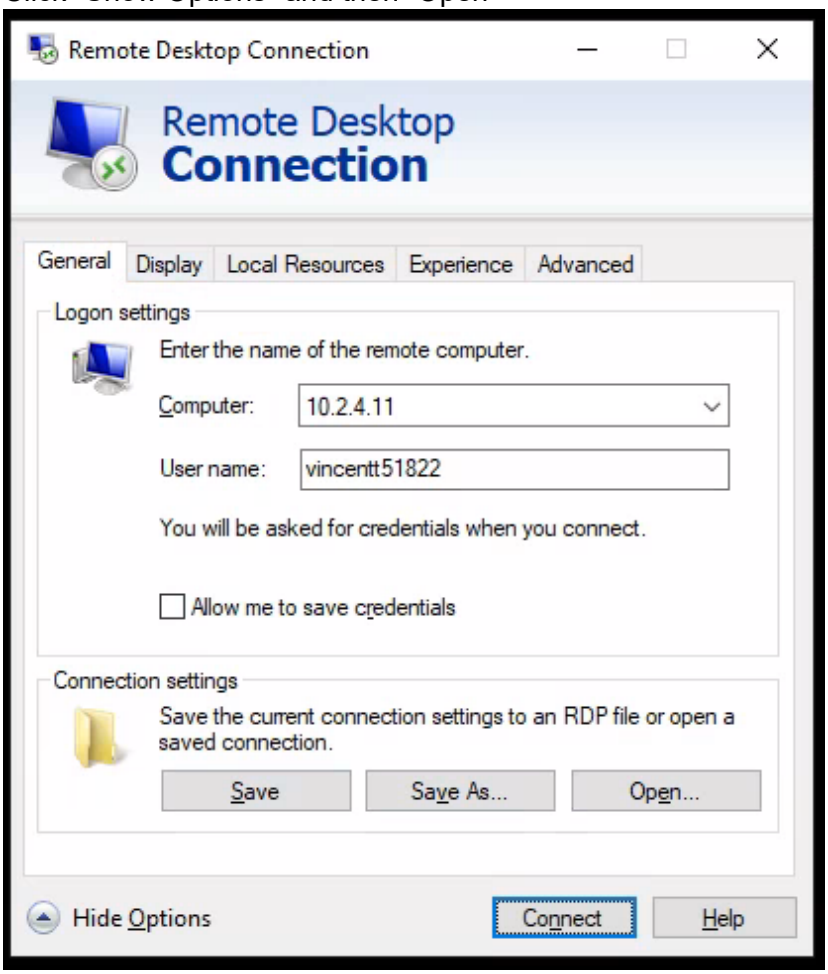
f. When you see this screen click "Yes" and the Windows VM desktop will display.

Note: For newly created VMs, it may take a few minutes for the desktop to load.

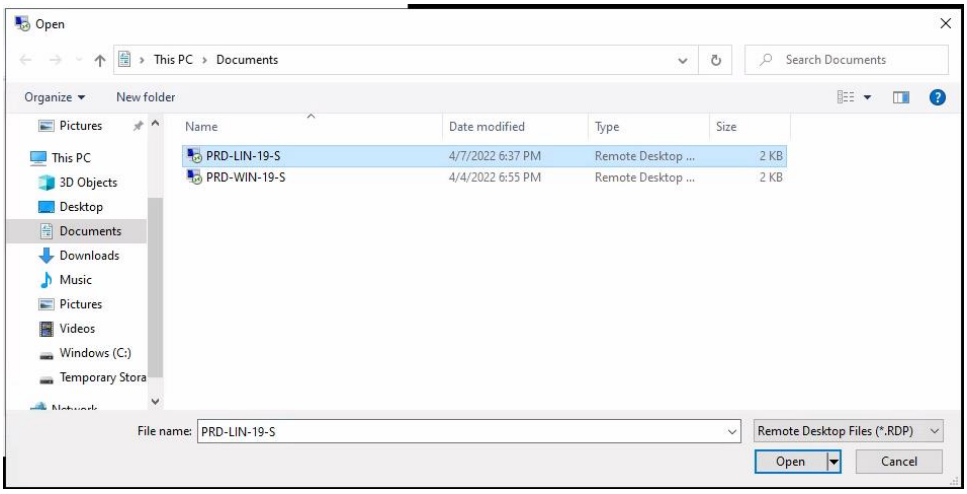
3.6. Remote Desktop (Linux GUI VM)

If you have chosen to connect to a Linux VM using "Remote Desktop", you will see this screen:

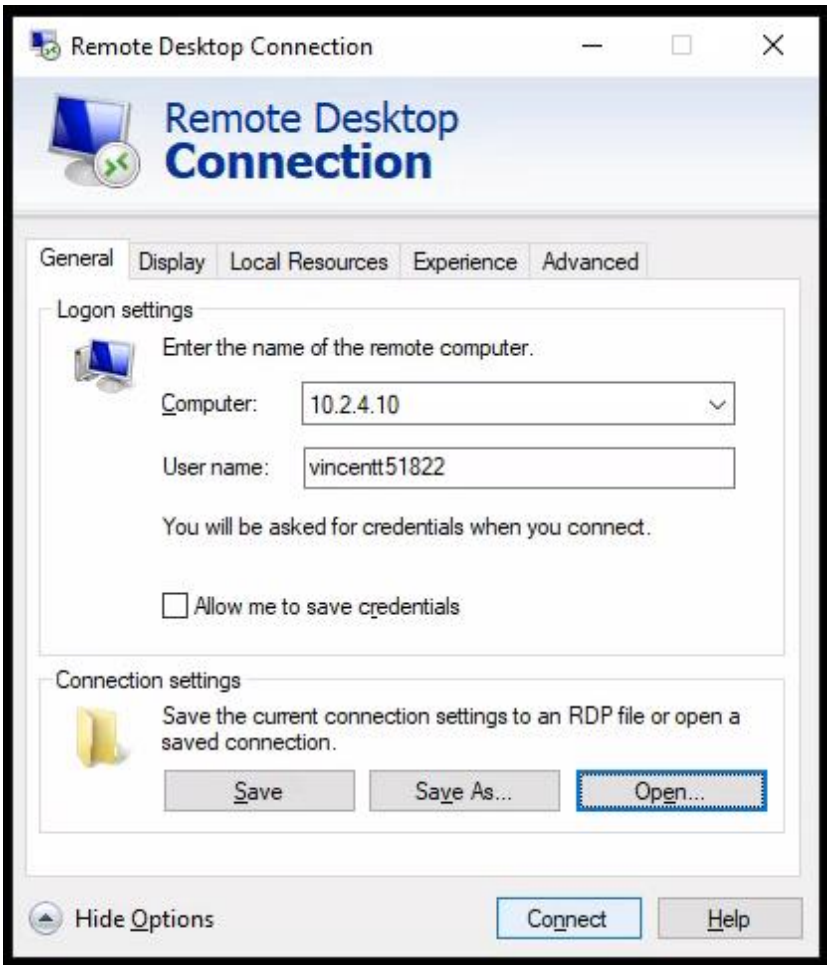
a. Click “Show Options” and then “Open”



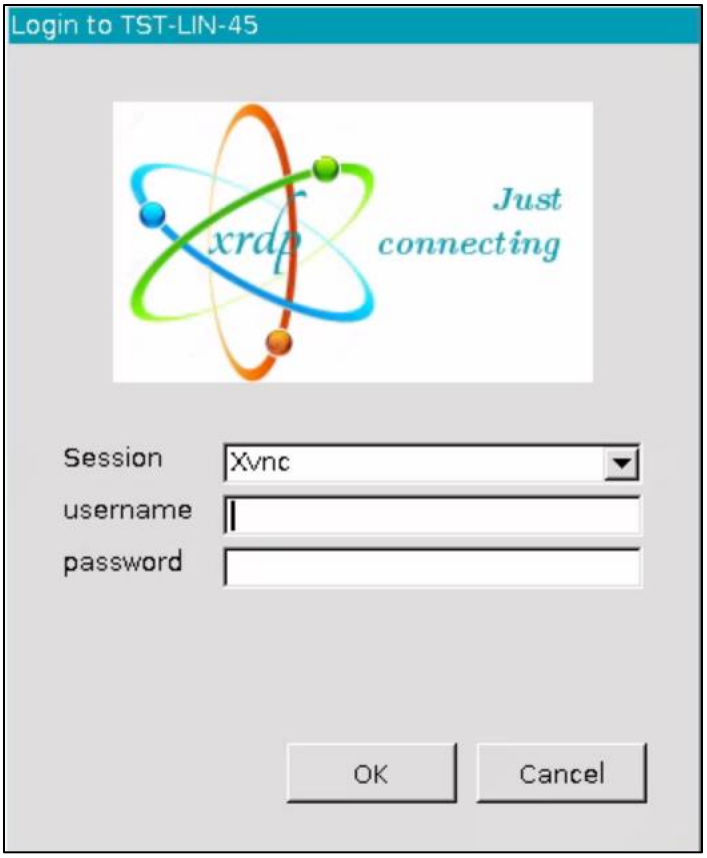
b. Select the profile appropriate to your VM. In this case “PRD-**LIN**-19-**S**" indicates a small linux VM. After selecting the appropriate VM profile, click “Open” to configure the connection.



c. Then click “Connect”

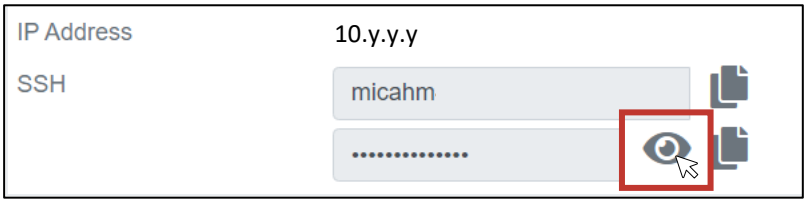
d. 

e. After you click "Connect", you will then see this screen:



f. Click the browser tab to go back to the BACPAC Portal page displaying the Computing Console.

g. Enter your user name found on the Compute Console page into the "username" field shown above.



h. View the password on the Compute Console page and manually type it in the Linux login screen shown above. **Note: you cannot use copy / paste for this field.**

i. Click "OK" and the Linux VM desktop will display.

3.7 Using the Native Client

On Windows a substantially better user experience is available if you use the Native AVD Client available at https://tinyurl.com/dataportalclient. Once you have installed and opened the client, click “Add” near the top right corner of the application, then click “Workspace.” Now use your Virtual Machine username and password. Then follow the instructions above.

4. Using the Windows and Linux VMs

4.1. General VM Usage Guidelines and Etiquette

Always stop VMs when not in use.

You may delete a VM when it is no longer needed or when a security update requires a new VM be utilized. When a VM is deleted, all work not saved in the personal or project storage locations will be lost.

4.2. Software

VM’s are pre-populated with statistical software including SAS, R, and Python. Other software is available by request. While most of the internet is blocked on the VMs, cran and pypi should be accessible so you can install R and Python libraries. It is also possible to upload software and libraries directly to the VM via your personal or project space.

4.3. Accessing Canonical Data

All data is located under the /mnt hierarchy on linux and C:/mnt on windows.

4.4. Output and Code Storage

Personal Storage Space

The BACPAC Data Portal allocates a workspace for each user. The User Workspace comprises read / write file space accessible only to the user. No other users have access to the User's personal workspace. The workspace can be used for analysis output, logs, code, documents, or other files a user requires.

It is forbidden to download study data from the BACPAC Data Portal. This warning is prominently displayed on the User Workspace page. While the BACPAC Data Portal cannot prevent the download of study data, all file downloads are logged. The logs are scanned, and the portal administrators are notified of any suspicious or prohibited activity.

Project Storage Space

Unlike User's Personal Workspaces, which are restricted to the single user, Project Workspaces can be used by all researchers collaborating on a project. This is the mechanism by which you share results, code and other project artifacts. A project is created based on approved BACPAC Data Access and Publications requests.

Note the warning on this page: You are prohibited from downloading study data from the BACPAC Data Portal. While it is possible to output an entire canonical dataset to a project or user workspace where it can be downloaded, that violates the portal user agreement. Such downloads will be recorded in the system logs so any violation will be detected.

5. FAQs and Troubleshooting Solutions