

WHI Virtual Data Enclave (VDE) User Guide

Revised: 3/11/2024

Overview

The WHI Virtual Data Enclave (VDE) is a remote computing platform that you access over the internet for the secure analysis of restricted data. Due to the sensitive nature of the data, you cannot access to the internet, send email, copy files or print from with your VDE session. You cannot transfer any files into or out of the VDE. Any external files (e.g. code or data) that you wish to use in the VDE must be provided to the WHI-CCC to be transferred into your personal VDE folder. However, files that you wish to take out of the VDE will be transferred to you after being reviewed by WHI-CCC staff for restricted identifiable data. The statistical packages available in the VDE are SAS, STATA and R.

The WHI-CCC has limited staff for the support of the VDE. We have provided this document to help you work independently in the VDE environment. It is expected that users will have the basic skills with statistical software and knowledge of the WHI datasets to be mostly self-sufficient.

The VDE server is located behind a secure firewall on the WHI network. Signing in is a two-step process:

1. Log in to the WHI network using a virtual private network (VPN) connection.
2. Log in to the VDE server using a Remote Desktop Connection program.

WHI will provide you with a single username and password for both.

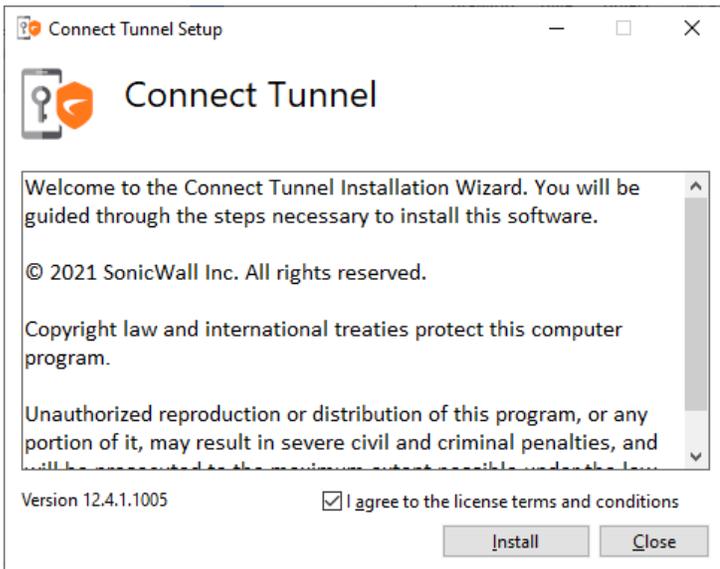
The VPN connection requires a one-time software installation on your computer. All Windows computers have a Remote Desktop Connection program built in. If you are using a Mac, a free third party program is available. Please see <https://docs.microsoft.com/en-us/windows-server/remote/remote-desktop-services/clients/remote-desktop-ios>

Getting Help

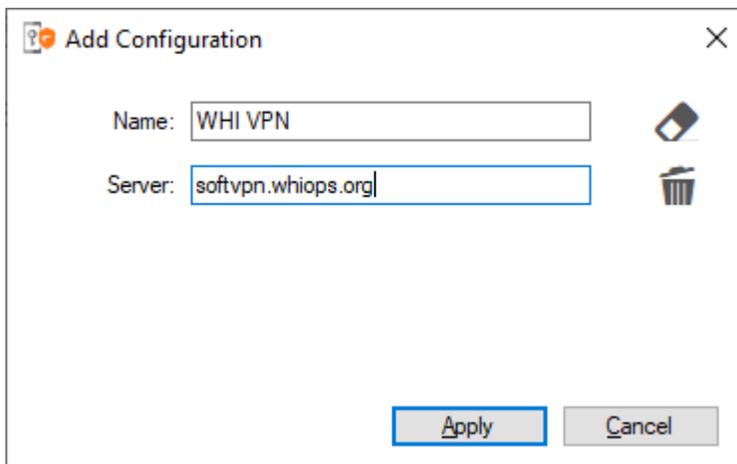
VPN installation, logging in, and the VDE environment: Ananya Roy aroy@whi.org; ananya@fredhutch.org
WHI data and analysis questions: WHI Helpdesk helpdesk@whi.org

Installing the VPN Software

1. Download the installation file that matches your computer:
 - [64 bit PC](#) (most common)
 - [32 bit PC](#)
 - [Mac](#)
2. PC: Go to the folder where you downloaded the .zip file. Double click the file to unzip it. Navigate to the folder containing the unzipped files.
Mac: Navigate (or locate) the downloaded the .dmg file.
3. PC: Double-click the “runme.bat” file.
Mac: Click or double-click the .dmg file.
4. Follow the prompts to install.

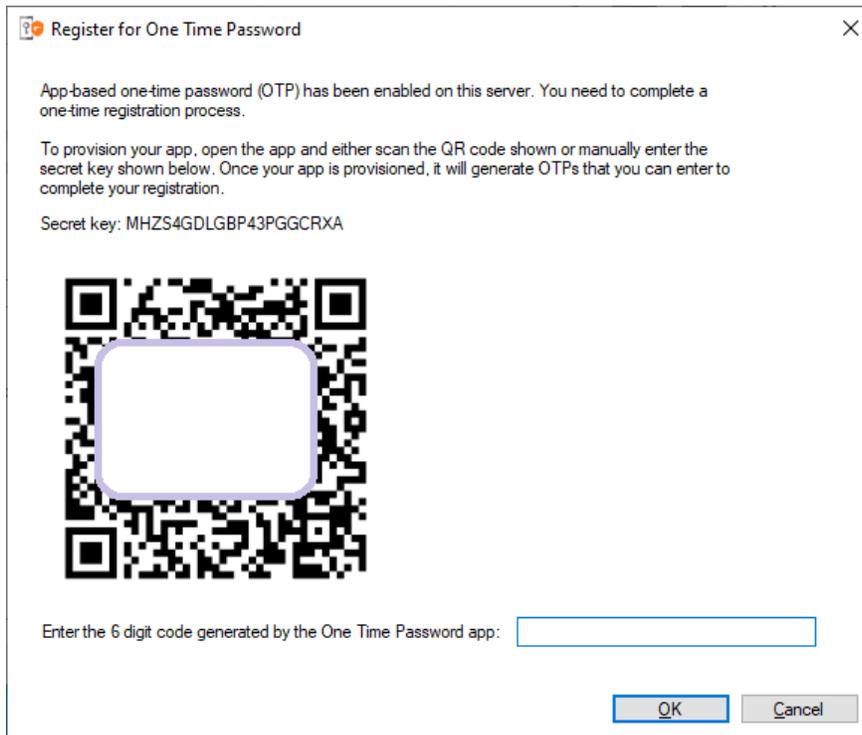


5. After the installation is finished, find and run the “SonicWall VPN Connection” icon on the desktop. You will be guided through a one-time configuration. Change the Hostname and Logon group to the values below:
Hostname: **softvpn.whiops.org**
Logon group: **WHI**



- Our VPN connection requires multi-factor authentication. Please install either the Microsoft Authenticator or Google Authenticator app on your smart phone or tablet. You will be required to type in a six digits code generated from the Authenticator app each time you sign into the VPN.

The first time you sign in you will need to register the QR code to your smartphone's mobile App like Google Authenticator and Microsoft Authenticator which both are available on Android and iPhones. Open your preferred mobile App, select Add new account and scan the QR code. Once registered, it will generate a six digits code for you. Type the code into the box to complete the sign in to the VPN.



- The VPN software is now installed and configured.
- To Sign off from the VPN, right click the "Disconnect" option at the lower right hand corner to disconnect/sign off from the VPN.

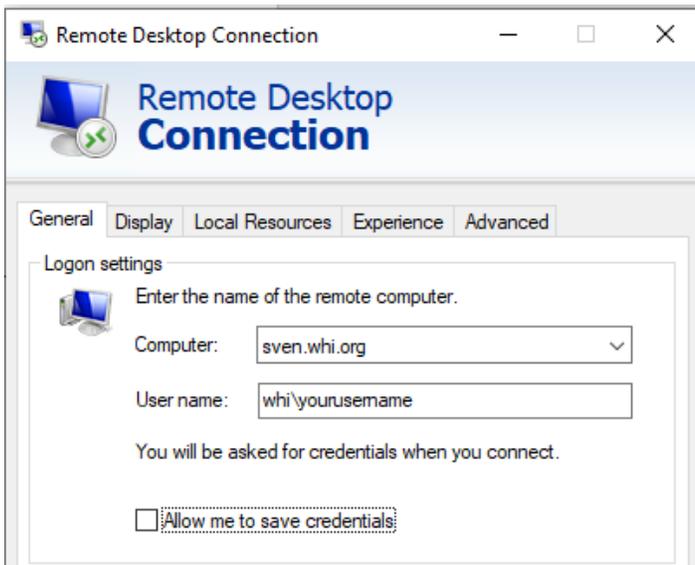


Accessing the VDE

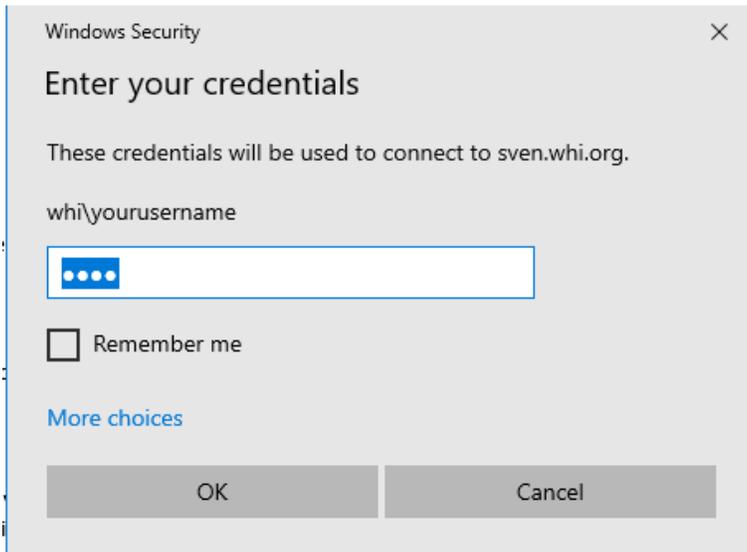
- Run the "Connect Tunnel" icon on the desktop and log in using your WHI account. *If at any time you are prompted to update the VDE software, run the update.*
- After the VPN is connected, open the remote connection program.
PC: Start menu/All programs/Remote Desktop Connection.
Mac: Find and run the remote desktop program.
- Enter the computer name: **sven.whi.org**



4. In the next dialog box click “Use another account”.



5. Enter your username prefixed with “whi\” as shown below. Enter your password in the next field and click Ok. Ignore whatever default username you might see above the username field.



6. Your VDE remote desktop should now appear. You can maximize it on your computer’s desktop.

Your VDE Desktop

When you connect to the VDE, you will see a Windows Desktop similar to a typical Windows PC (even if you are using a Mac). There are shortcuts on the desktop to your user folder, the WHI data folders and the statistical software. There are also shortcuts to Microsoft Office applications and the TextPad editor.

The VDE desktop shortcuts:



Statistical Software

SAS

Version: 9.4.

The following modules are installed:

- BASE Base SAS
- STAT SAS/STAT
- GRAPH SAS/GRAPH
- FSP SAS/FSP
- OR SAS/OR
- IML SAS/IML
- OROPT SAS/OR Optimization
- ORPRS SAS/OR Project and Resource Scheduling
- ORIVS SAS/OR Interactive Visualization and Simulation
- ORLSO SAS/OR Local Search Optimization
- ORACLEBNDL SAS/ACCESS ORACLE
- PCFILE SAS/ACCESS PC Files
- MYSQL SAS/ACCESS Intface MYSQL
- STATSTUDIO SAS/IML Studio
- WKSPSRVLOCAL SAS Workspace Server for Local Access
- WKSPSRVENT SAS Workspace Server for Enterprise Access

Before running any programs, you will need to set up the SAS libraries required by your project using LIBNAME statements.

Syntax: LIBNAME libref <path>;

a. LIBNAME = desired alias (1-8 characters, must start with an alpha character);

b. path = Local path to folder (e.g., "u:\folder\subfolder"). The path must be enclosed in quotes.

Just a reminder that the contents of any library can be viewed using either the Explorer Window accessed from the toolbar, or using the 'PROC DATASETS Library=libref;run;' statement.

The path names for the various data available in the VDE are provided in the Data Section.

Stata

Version: StataSE 15.

SAS datasets cannot be read directly into Stata. If you use Stata you must first read the datasets into SAS, and then export them to Stata datasets (.dta).

To avoid repeated writing of path names of data files in your do-file, you can set a local directory.

For example:

```
local pathname "u:/foldername"  
use "`pathname'/datafilename.dta", clear
```

Alternatively, to change the working directory, use:

```
cd "u:/foldername"  
use "datafilename.dta", clear
```

The path names for the various data available in the VDE are provided in the Data Section.

R

Version: 3.6.0 (2019-04-26) – “Planting of a Tree” with both 32 and 64 bit versions available.

In R, there are numerous ways to assign names to data and working folders using the ‘getwd’, ‘setwd’ and ‘file.path’ functions. The path names for various WHI data folders are provided in the Data Section.

Your VDE User Folder

There is a shortcut to your folder on the VDE desktop called “VDE User Folder”. The user folder is your workspace in the VDE. It is a private, secure folder to which you have full permissions (i.e. read, write, modify, create, delete). This folder (and any subfolders you create) is the location for your code, and the files that you create while working in the VDE (i.e. programming code, data files and output files). You may create as many subfolders as you like. Any outside files that you provided to the CCC for your use in the VDE will be found in your user folder.

Your user folder is mapped to the U: drive.

Data

Overview

There are shortcut(s) on your desktop to the various data folder(s) for which you have access. The folder on your desktop called “Investigator Data” is the folder containing the datasets provided by WHI for investigators’ use. All data folders and the files they contain are read-only.

The following data folders are available for access pending approval from the VDE Gatekeeper:

-  CMS
-  Geocodes
-  Investigators_datasets
-  LILAC
-  Rand_enroll

Each data folder is specified in the following sections. All VDE datasets are linked using the WHI Common ID.

Medicare (CMS)

The VDE currently provides Medicare data through 2017 to approved users. The Medicare datasets are in SAS format. These are read-only files and cannot be copied to other folders. All VDE users with permission to use the Medicare data have access to the Medicare Enrollment Data file. Access to other Medicare data is granted only to those files requested on the VDE intake form and approved by the CCC.

The path to the Medicare datasets folder is: M:\CMS\CMS_2017\sasdata

Medicare SAS Datasets:

- CarrierB (header)
- CarrierB_lineitems
- DME (header)
- DME_lineitems
- Home_health (header)
- Home_health_lineitems
- Hospice (header)
- Hospice_lineitems
- Inpatient (header)
- Inpatient_lineitems
- medicare_enrollment_2017
- Medpar
- Outpatient (header)
- Outpatient_lineitems
- PDE
- SNF (header)
- SNF_lineitems

Please read the document “Readme.txt” found in the folder “M:\CMS\CMS_2017\documentation”. This folder also contains listings of the contents and descriptive statistics of the variables in each data file and a document describing how the Medicare Enrollment data file was created and its contents. Also in this folder is the file named 'Medicare2017_SASformats.sas', which includes all the SAS formats used in the Medicare data files. Use an %INCLUDE statement in SAS to read the formats into your SAS session.

Geocodes

Geocodes based on participant addresses are in the “Geocodes” folder. The geocode datasets are in SAS format. Documentation on the geocoding process was provided by the Ancillary Study that produces the geocode data. The documents are:

- EEAWHI_address_documentation_final_1.pdf
- EEAWHI_address_documentation_final_2.pdf.

The path to the Geocodes folder is: G:\Geocodes

Investigator Datasets

The WHI Investigator Datasets which are currently available for download on the WHI website are available in the VDE. They are located in the folder “Investigator_datasets”. The datasets are tab-delimited ASCII files. Each dataset also has a corresponding data dictionary (PDF) and a SAS code file for loading the data. Before using these data, we strongly advise that you read *WHI_data_prep_and_use.doc* located in the same folder.

The path to the Investigator_datasets folder is: I:\Investigator_datasets

LILAC

The LILAC datasets include cancer treatment information and long-term cancer outcomes for WHI participants diagnosed with one of five selected cancers: breast, endometrial, ovarian, lung, and colorectal cancers.

Please read the data guide *WHI_LILAC_cohort_data_preparation_Sep2019.pdf* found in the folder “L:\LILAC” for additional information on data file setup, data conventions, and specific data set information.

The path to the LILAC folder is: L:\LILAC

Enrollment dates

If you are using Medicare data, you will need the participant enrollment dates in “ppt_rand_enroll_dates.dat” found in this folder.

The path to the enrollment dates folder is: E:\Rand_enroll

Transferring Files Out of the VDE

Files that you wish to take out of the VDE must be placed in the “Transfer” subfolder under your user folder. The “Transfer” folder is pre-created for you. File contents will be reviewed by CCC staff for compliance with confidentiality restrictions and conformance with your study proposal. File content that is considered restricted will be redacted prior to the release of the file. You will be notified by the CCC with details on the file transfer process.