INSTITUTE FOR
RESEARCH ON
INNOVATION & SCIENCE

IRIS VDE User Guide

Created by:
The IRIS Research Team
irisdatarequests@umich.edu

Last Updated:
January 18, 2018

Table of Contents

Introduction

Welcome to the community of IRIS researchers! We at IRIS look forward to working with you as you delve into the IRIS dataset via the virtual secure data enclave (hereafter, VDE or Enclave[1]) to meet your research goals.

The VDE is a virtual machine launched from the researcher's own desktop but operating on a remote server, allowing IRIS researchers to access the restricted use data without being physically present at the IRIS offices. In order to help prevent unauthorized disclosure, researchers are not allowed to transfer files directly to or from their personal computer when working within the Enclave. Researchers are encouraged to do as much of their work as they are able within the Enclave, and to that end, the VDE is regularly updated with statistical packages, programs, and other tools according to user needs and requests. IRIS recognizes that researchers need to access their analysis outside of the Enclave on occasion, and this document outlines the indirect transfer methods that are employed to achieve this while minimizing the risk of accidental or deliberate exposure of confidential information.

VDE users are asked to follow best practices in restricted-data use as outlined in the IRIS VDE Acceptable Use Policy. Thank you for helping us maintain the security and confidentiality of our data by following these rules. If you have any questions regarding practices or procedures for working with IRIS data in the VDE, please do not hesitate to contact us at irisdatarequests@umich.edu.

Sincerely,
The IRIS Research Team

[1] In some documentation you may also see the VDE referred to as the VDI, or Virtual Data Infrastructure. IRIS Enclave, VDE and VDI are used interchangeably but we will be using Enclave and VDE throughout this user guide.

Working with the IRIS Research Team

IRIS staff offices are located in the Perry Building on the University of Michigan campus at 330 Packard Street in Ann Arbor, Michigan. IRIS business hours are 8:00 a.m. to 5:00 p.m. Eastern Standard Time and office closures for holidays (or, on the rare occasion, inclement weather or other unplanned events) follow those of the University of Michigan.

IRIS has a Research team dedicated to supporting the community of researchers who use IRIS data. The Research Team includes Natsuko Nicholls, Research Manager; Beth Uberseder, Research Support Specialist; as well as IRIS research assistants.

Natsuko Nicholls
Research Manager
hayashin@umich.edu
(734) 763-0447
2362 Perry

Beth Uberseder
Research Support Specialist
ubersbe@umich.edu
(734) 763-5905
2342 Perry

Communications regarding the VDE will typically come from either Natsuko or Beth. As a VDE account holder you will also be added to the list-serv of current VDE users. This list is primarily used to alert all active VDE account holders of policies and scheduled maintenance to the server.

Please use the IRIS help ticket system to communicate with the IRIS Research Team if you have any questions or issues related to the VDE or IRIS data to ensure your request is seen and addressed in a timely manner. See below for details.

IRIS help ticket system

The IRIS Research Team maintains a help ticket system to track and resolve incoming researcher requests. Rather than emailing a team member directly, please use the e-mail irisdatarequests@umich.edu for any of the following:

- Issues logging in or setting up your VDE account
- Request for data import to the VDE
- Request for data export (disclosure review) from the VDE
- Questions or issues regarding VDE login
- Questions or issues regarding VDE software
- Requests for updates to the VDE
- Creation of new user accounts or shared group folders
- Questions regarding IRIS data
- Requests for access to new data files

When you send an email to irisdatarequests@umich.edu, a help ticket is automatically generated that all members of the IRIS Researcher Team will receive. As a help ticket is reviewed and addressed, the status will be changed and you will receive automated email notifications according to the designated status, as follows:

Status: Request

This is an automated email generated by the help ticketing system to alert the user that a request was successfully submitted. A status of Request does not indicate that a member of the IRIS Research Team has seen the request.

Status: In Process

If communication with the user is necessary for resolution of a request, the user will receive a response from the IRIS Research Team with a status of “In Process” assigned.

Status: Completed

Once a ticket has been addressed and resolved, the user will receive an email with the “Completed” status assigned.

Service level agreement

We at IRIS always seek to provide prompt and helpful service in response to your research requests. Our official service level agreement is as follows:

- A help ticket will be reviewed and assigned to a Research team member within 1-3 business days.
- The assigned IRIS Research team member will respond within 5-7 business days from the date of the original user request.
- Exceptions:
    - **Disclosure review requests:** See the policy on disclosure review for details on the service level agreement for all export requests. These requests may take 15 business days, so please plan accordingly.
    - **Requests for updates to the VDE:** If approved, these may only occur during the monthly maintenance window. VDE updates occur on the fourth Wednesday of every month unless otherwise noted.

A note on urgent requests:

As a general rule, help desk tickets are serviced during IRIS business hours, Monday through Friday from 8:00 a.m. to 5:00 p.m. Due to the often unpredictable nature of research, we understand that there may be occasions when you have a request that is urgent in nature. Please be sure to specify the time frame in which you need a request addressed if it is time sensitive, but be aware that IRIS does not guarantee that a researcher's deadline will be met in such cases. The service level agreement stated above applies to all cases whether a deadline is stated by the researcher or not, so please plan accordingly.

Creating your VDE account

As a new VDE user, there are several steps you need to take to gain access to the secure data enclave, including completing mandatory training, activating DUO Mobile for two-factor authentication, and setting up a remote connection on your local computer for the IRIS VDE. Details of each step are described in greater detail in the paragraphs that follow.

Step 1: Sign IRIS policy documents and complete trainings

After completing the application process and being authorized for a VDE account from IRIS, you will receive an email from the IRIS Research Team outlining the remaining steps that need to be completed prior to gaining access to your login credentials. All new users are required to:

1. Read and sign the IRIS VDE Acceptable Use Policy
2. View a 5-minute video on VDE acceptable use guidelines
3. Take a brief knowledge quiz
4. Download the DUO Mobile app

This mandatory training process helps to ensure that all of our researchers have a clear understanding of the terms of use and ways in which data security and confidentiality is maintained.

Step 2: Activate your DUO security account

To login to the IRIS VDE, you will need more than a username and password. IRIS requires two-factor authentication and uses a service called DUO[2] for this. To activate your DUO account for use with the IRIS VDE, you must first download the DUO Mobile app, which is available for Android, iPhone, and Windows Phone. The most efficient and effective way to utilize DUO is through this app, which can utilize cellular data or Wi-Fi to communicate.

[2] https://duo.com/

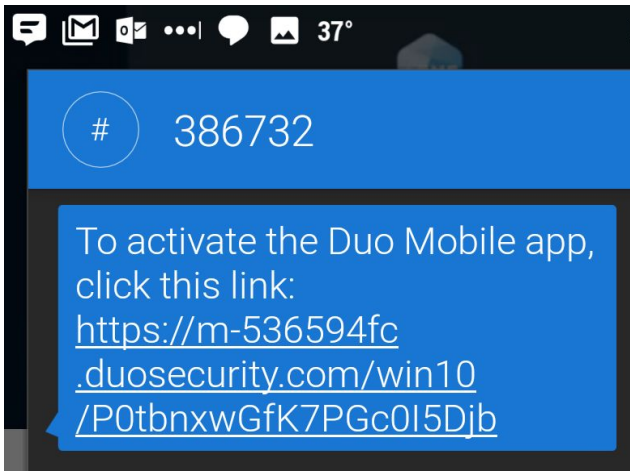
What if I don't have a smartphone?

If you do not possess a device capable of running the Duo Mobile app, an IRIS representative can configure the system to call a designated phone number (cell or landline) to authorize the login. This is a workable alternative to a mobile device, but will require that you have access to a designated phone number (cell or landline) whenever and wherever you need to login to the IRIS VDE.

DUO instructions are as follows:

1. **Download the DUO Mobile app to your smartphone:**
   Apple iPhone: https://itunes.apple.com/us/app/duo-mobile/id422663827?mt=8
   Android phone: https://play.google.com/store/apps/details?id=com.duosecurity.duomobile&hl=en
   Windows phone: https://www.microsoft.com/en-us/store/p/duo-mobile/9nblggh08m1g

2. **Provide IRIS with your mobile phone number[3]:**
   E-mail your mobile phone number to the IRIS Research Team. You will receive a text message that will look similar to the one shown below. **You must complete STEP 1 before we send you the activation code or it will not work, so please send your phone number after downloading the DUO Mobile app.**



3. **Activate DUO Mobile:**
   Tap the link to open it with the Duo Mobile app. This will activate your app for use with the IRIS VDE.

   When you receive this activation code please click it to complete the DUO set-up as it expires 24 hours after it is initiated.

If you need an activation code to be resent, please don't hesitate to contact us via irisdatarequests@umich.edu.

[3] Your phone number will not be shared with outside parties.

Step 3: Login to the IRIS VDE

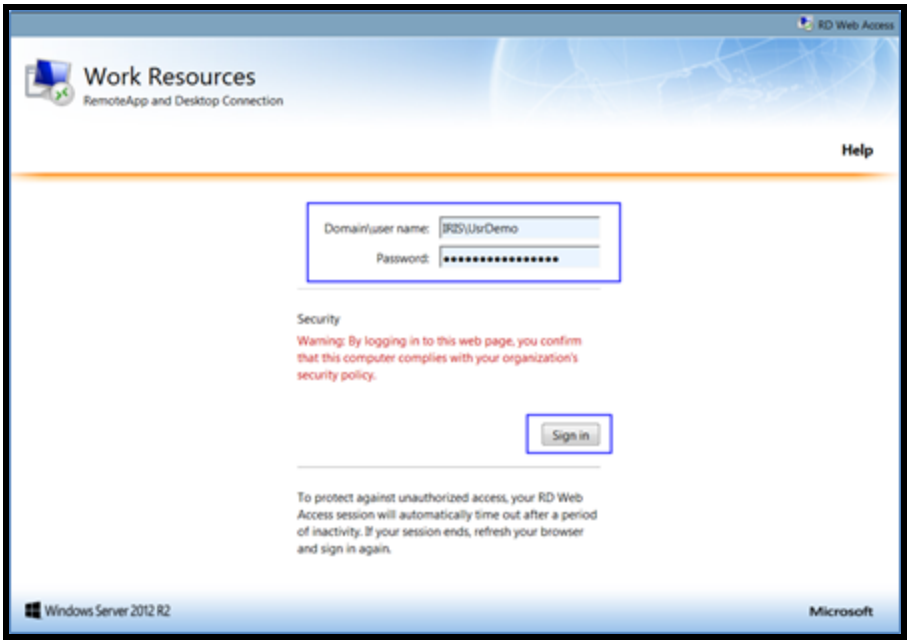
Upon completion of the training steps and activation of DUO Mobile, you will receive an additional email from the IRIS Research Team with your VDE login credentials, the location of your personal home folder and any group/shared folders you may have access to, and the name of your database schema. All users are given their own personal folder upon account creation but group or shared folder access may be granted at any time.

When you receive the welcome email, please take a moment to make sure you can login, and let the IRIS Research Team know if you have any issues or questions with the process. Steps with screenshots on how to create your remote connection are listed below.

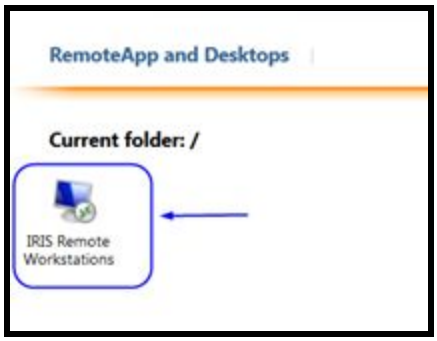
Steps for access using Windows

1. Visit the following URL: https://irisvdi.isr.umich.edu/

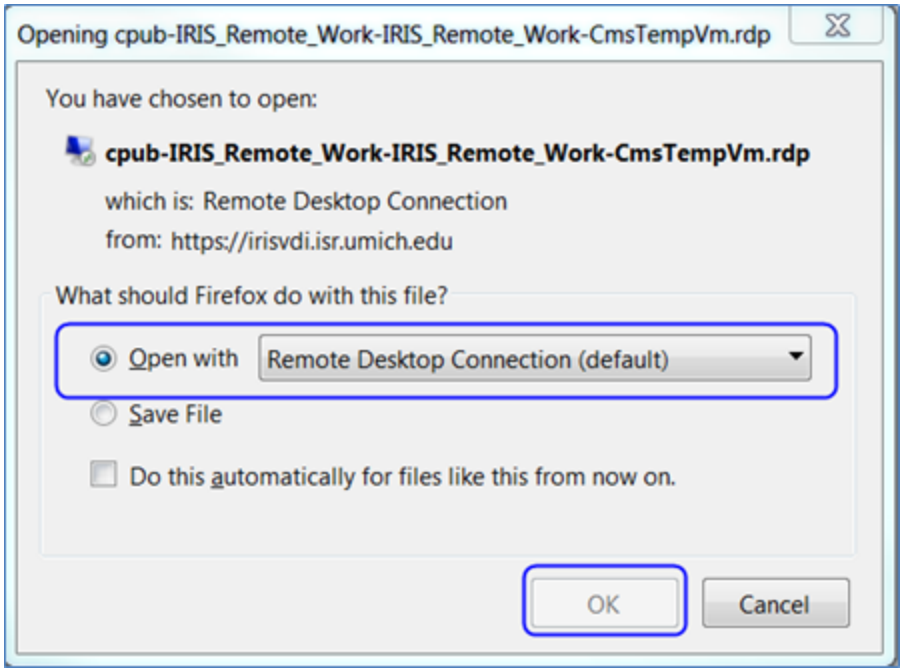
2. Enter your username and password provided by IRIS. Make sure to add “IRIS\” when typing your user name.



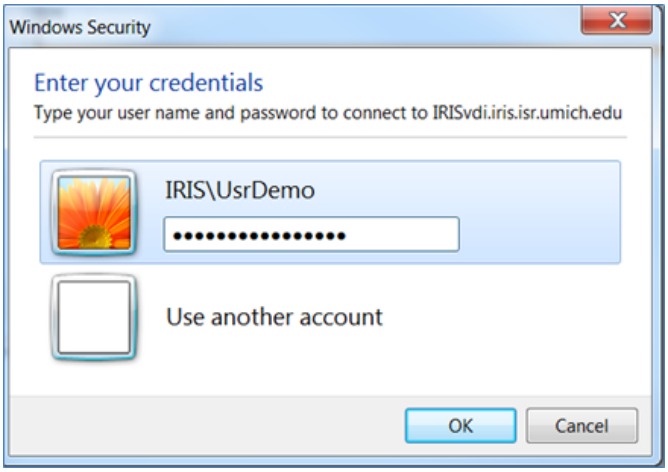
3. Click on the Remote Desktop Connection (RDC) icon "IRIS Remote Workstations" in the upper left of the window.



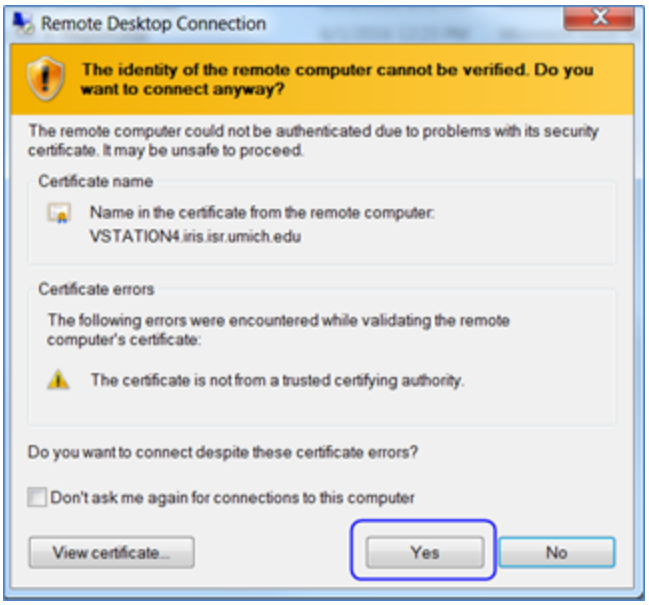
4. Once you login you will be prompted to download a customized RDC. You will be asked to open or save file. Open with "Remote Desktop Connection" and click on "OK".

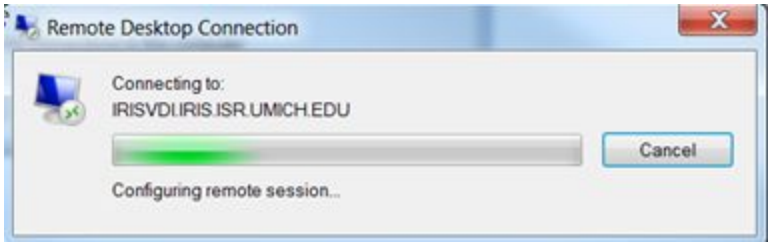


5. Once you are asked to enter your credentials, use the same username and password that you just used in Step 2. Again, make sure to add "IRIS\" when typing your user name.

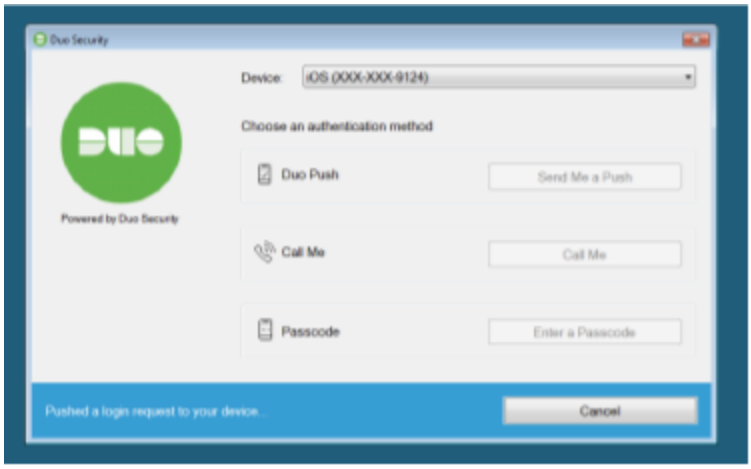


6. Click on “Yes” to Remote Desktop Connection.





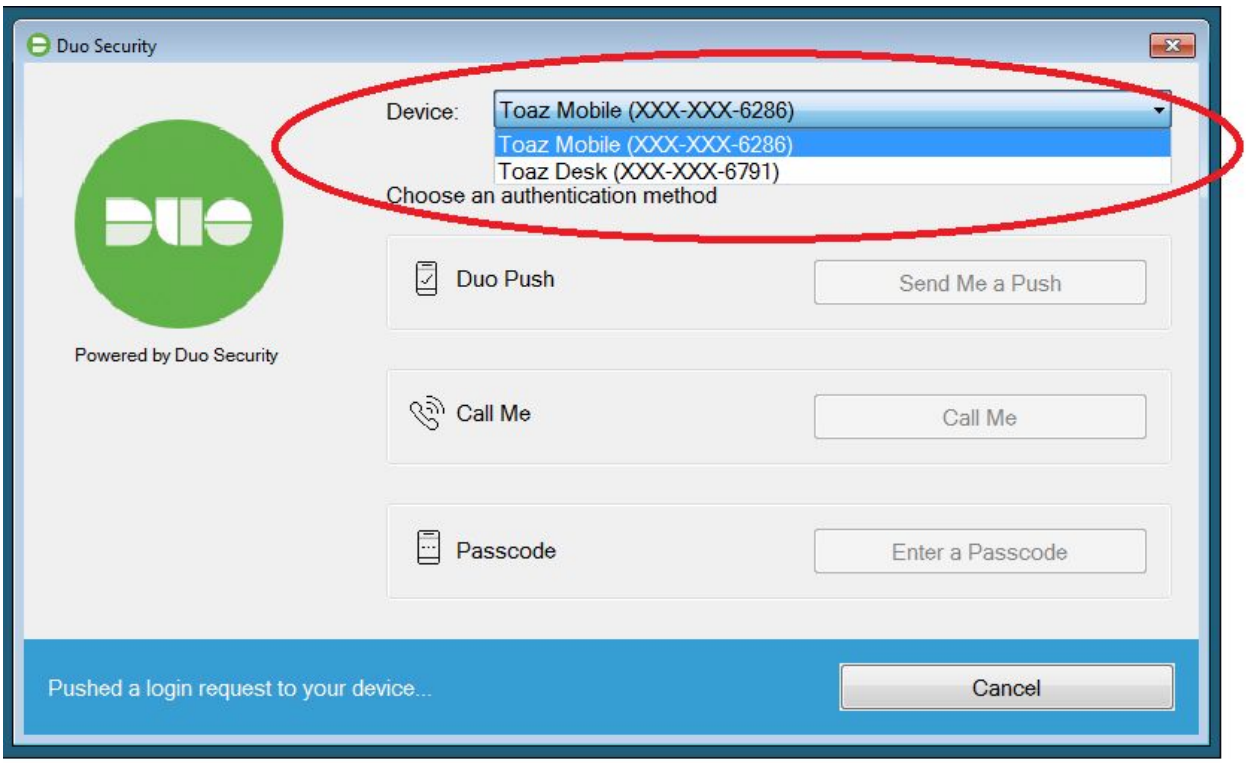
7. Once your username and password are accepted, you will see your virtual workstation starting up. Shortly you will be presented with screens that look like this…



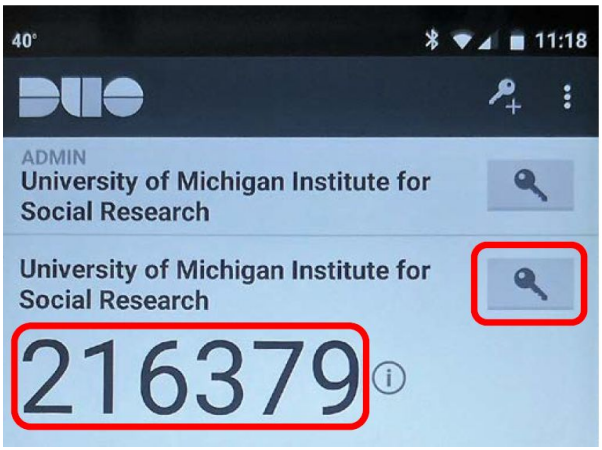
Computer screen

Smartphone screen

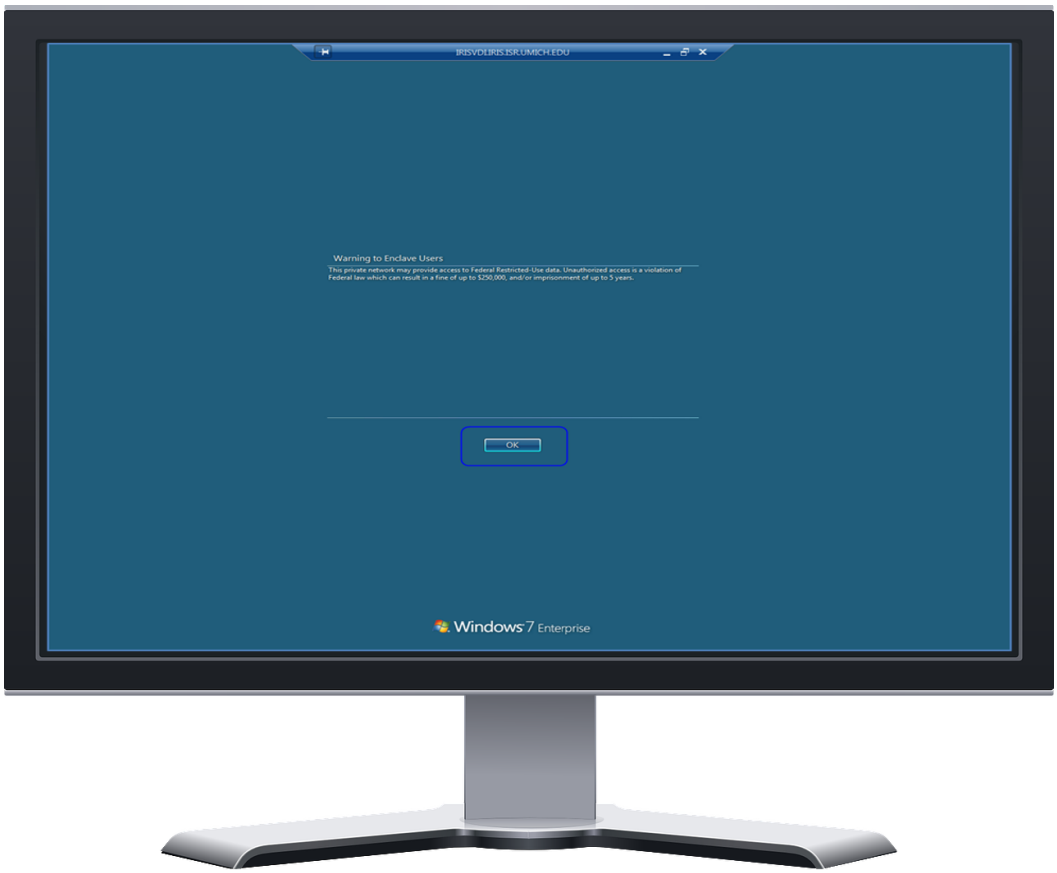
By default DUO will send a push notification to devices that have the DUO Mobile app installed and activated. This is the easiest and most reliable way to authenticate. At the top of this window there is a drop-down box. If you have chosen to have more than one device associated with your IRIS Duo login (i.e. your smartphone and a desk phone), you can switch the active device here.



This window also gives you multiple options for authenticating with your device. We recommend **DUO Push** (default) as it is the easiest, and most reliable. If you need to receive a phone call, you can click **Call Me** and DUO will call you and ask you to push a button on your phone's dialpad to authenticate. If you do not have connectivity (cellular or WiFi), you can click the **Passcode** button. This will bring up a text box for you to enter a code. You can get this code by opening the DuoMobile app on your phone, and click the key next to University of Michigan Institute for Social Research. This code will generate whether you have any network connectivity or not.



8. Click on “OK” to Warning to Enclave Users. After several seconds, you will be directed to one of our IRIS Remote Workstations. This may take longer if you are connecting to VDE for the first time.

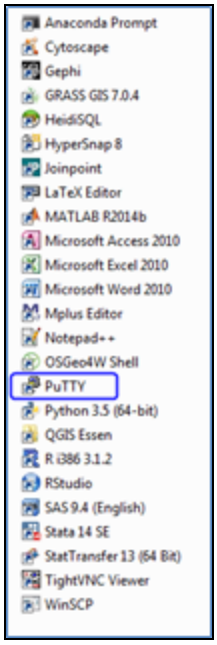


Steps for access using Mac

As long as you are using Microsoft Remote Desktop version 8.0.43 on a Mac, you can connect to the VDE. Microsoft Remote Desktop does not come pre-installed on a Mac, though most users will have it as part of their Office installation. Otherwise you can download it through the App Store. Some Mac users have experienced that a saved RDP link stops working. If you experience difficulty accessing VDE from a Mac machine, try to refresh the RDP link. Instead of using https://irisvdi.isr.umich.edu/, try the following full link: https://irisvdi.isr.umich.edu/RDWeb/Pages/en-US/login.aspx?ReturnUrl=/RDWeb/Pages/en-US/Default.aspx

Steps for access using Linux

If you prefer to work in the Linux environment, you must request a Linux user account to be added by e-mailing irisdatarequests@umich.edu. IRIS will provide an account to access Linux, at which point you should follow these steps:

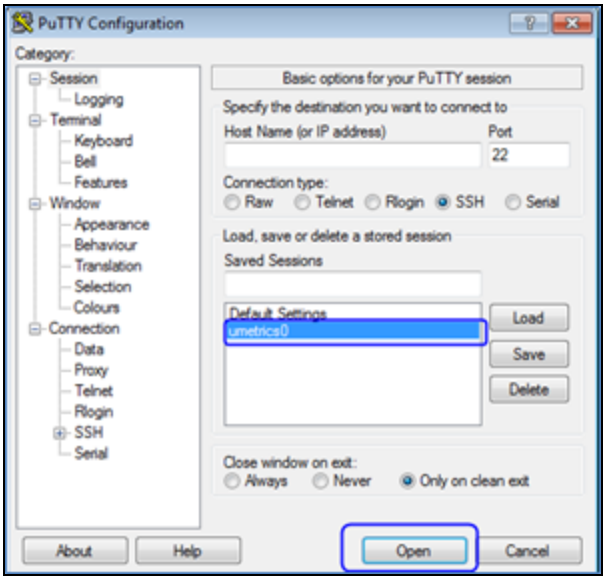


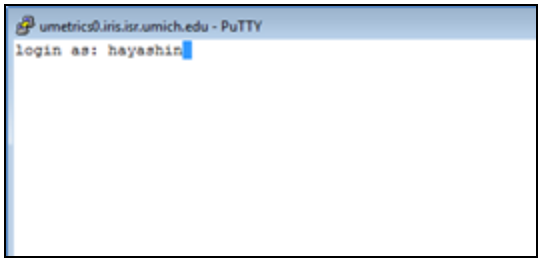
1. Make a connection to an IRIS VDE workstation—the same instruction applies to all VDE account holders.

2. Open the Main Applications folder on the desktop, and double click on the PuTTY icon.

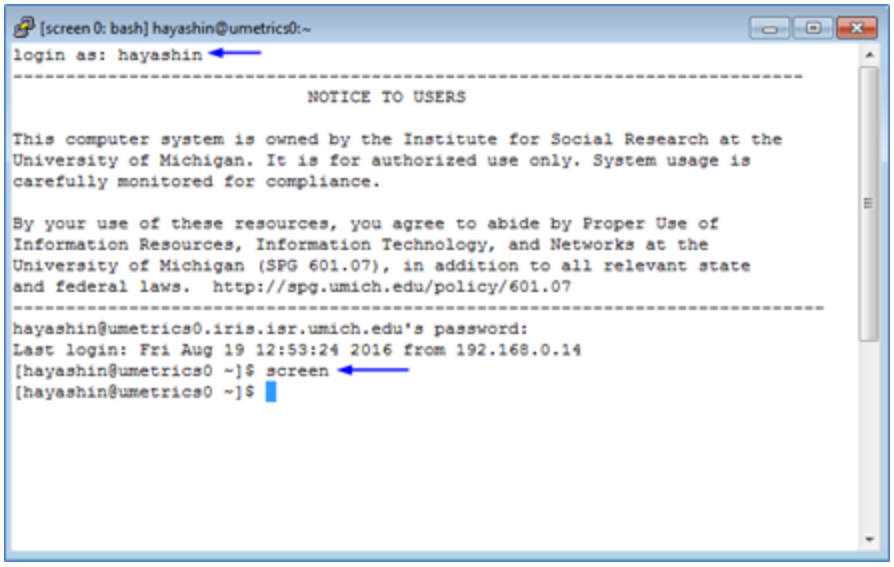
3. You should see a putty session screen and the login dialogue. Double click on the umetrics0 session.

4. Use the same credentials used for IRIS VDE account (username and password) at the "login as:" prompt and press the Enter/Return key. (Note: "IRIS\" is NOT required. Only type your user name after IRIS\.)



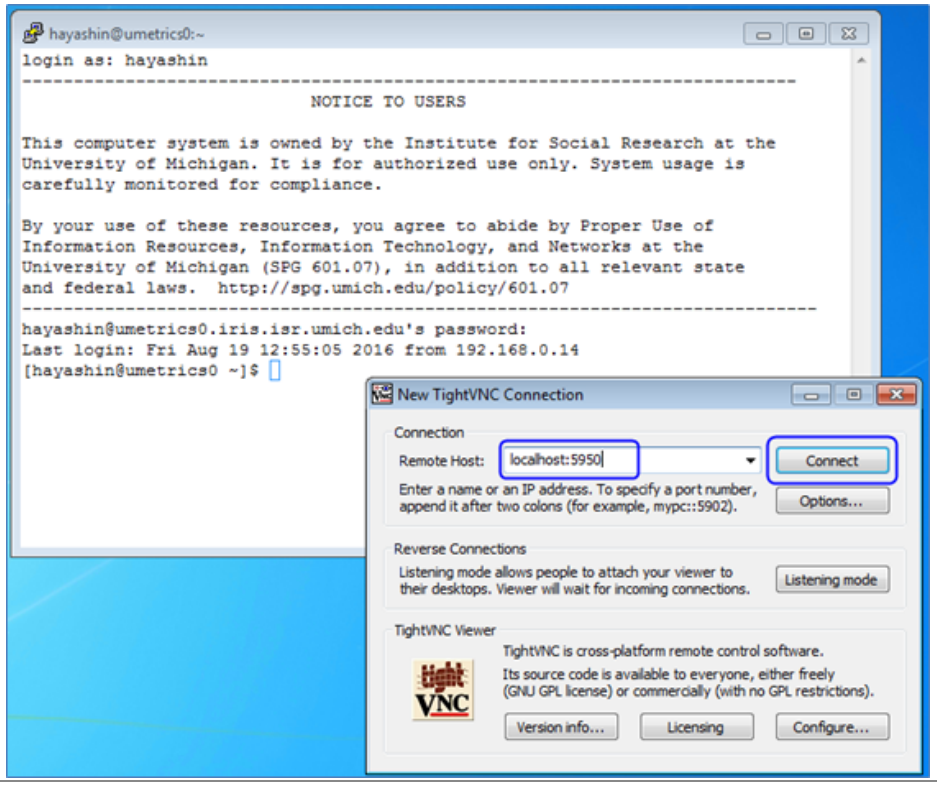


5. Once you see the notice screen and a password prompt, enter user IRIS VDE password at the "... password: " prompt and press the Enter/Return key. You should see the bash shell prompt "[userid@umetrics0 ~]$" and can now type in Linux commands. For example: [userid@umetrics0 ~]$ screen will let you have multiple screens (bash shells) in a single putty session.



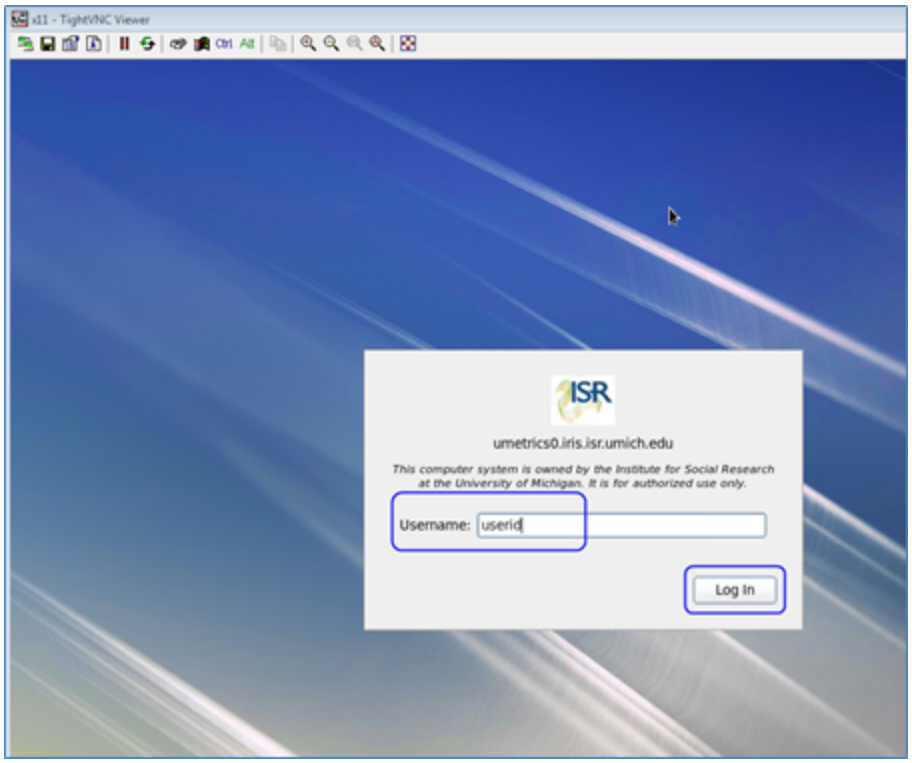
Additional steps for a graphical Linux environment

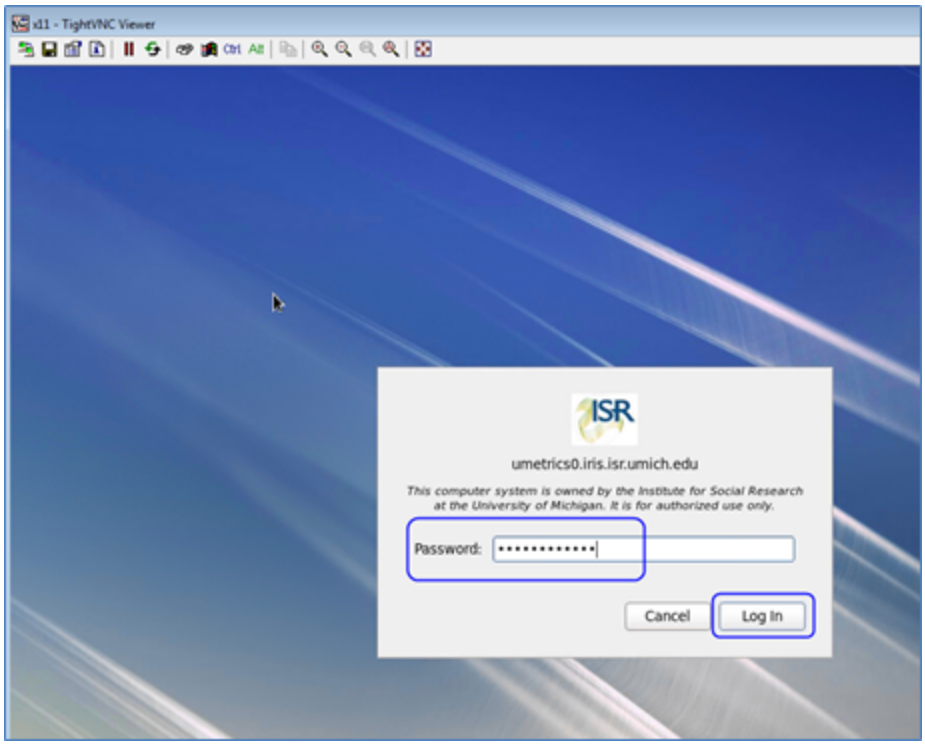
1. Remember to make a PuTTY connect first as described above. Note that the vncviewer uses the PuTTY channel to transfer information to the Linux server; the PuTTY must be running in the background with vncviewer.

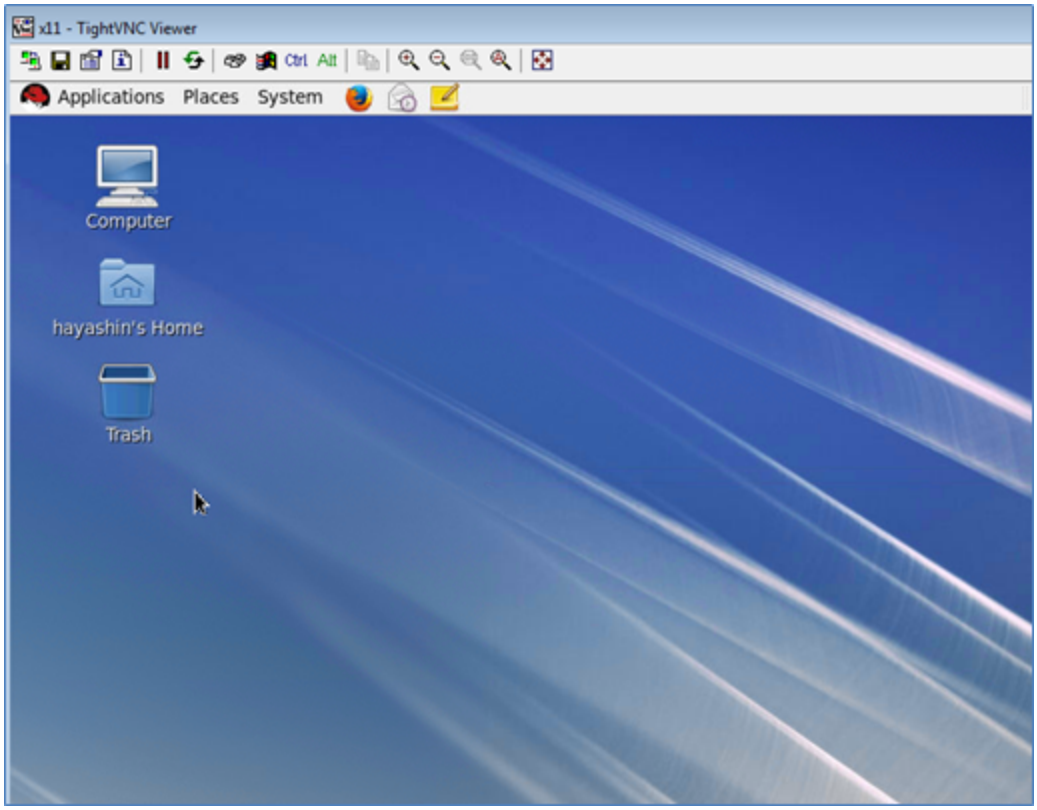


2. Open the “Main Applications” folder on the desktop, and double click on the TightVNC viewer icon. You should see the “New TightVNC Connection” form. The TightVNC viewer program is pre-configured.

3. Double click on the connect button. You should see the Linux graphical login prompt. Enter your IRIS VDE credentials and you should see the Linux Graphical Environment.




x11 - TightVNC Viewer
Ctrl
Alt
ISR
umetrics0.iris.isr.umich.edu
This computer system is owned by the Institute for Social Research at the University of Michigan. It is for authorized use only.
Password:
Cancel
Log In


x11 - TightVNC Viewer
Ctrl
Alt
Applications
Places
System
Computer
hayashin's Home
Trash

Introduction to IRIS data

About IRIS data

The IRIS website is a good resource for information about the IRIS data, including:

- The history of IRIS and the origins of the IRIS/UMETRICS data: http://iris.isr.umich.edu/about/

- Information about IRIS Research Data including annual data releases: http://iris.isr.umich.edu/research-data/

- Publicly available data documentation is available for each data release. For example, for the 2017 data release, summary documentation and a data dictionary are available here: http://iris.isr.umich.edu/research-data/2017datarelease/

- Research Data FAQ: http://iris.isr.umich.edu/research-data-faq/

- Papers published using IRIS data: http://iris.isr.umich.edu/research-data/publications/

Future data releases

IRIS plans to continue annual data releases as membership grows and member institutions continue to submit additional data. Temporal coverage of the data submitted by member institutions varies from one or two recent years to more than a decade. For the most updated information on data coverage please refer to the data documentation, or ask the IRIS Research Team. For reference, the list of current member institutions is available on the IRIS website here: http://iris.isr.umich.edu/membership/members/ but not all of these are currently represented in the research data release.

Working with IRIS data in the VDE

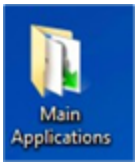
Database schemas and views

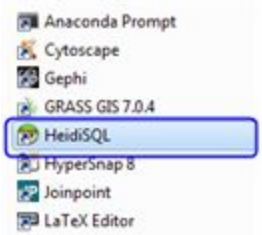
To give access to research data, IRIS grants VDE users access to a SQL server database where IRIS contains and manages data in tables and views. Upon VDE account creation, the IRIS Research Team will set up a unique database schema for you (and/or your research group). You will have access to requested data files in views in your schema. You are also able to create procedures, as well as, create tables and views in your schema.

For any questions related to the access rights and means of data access, please email irisdatarequests@umich.edu.

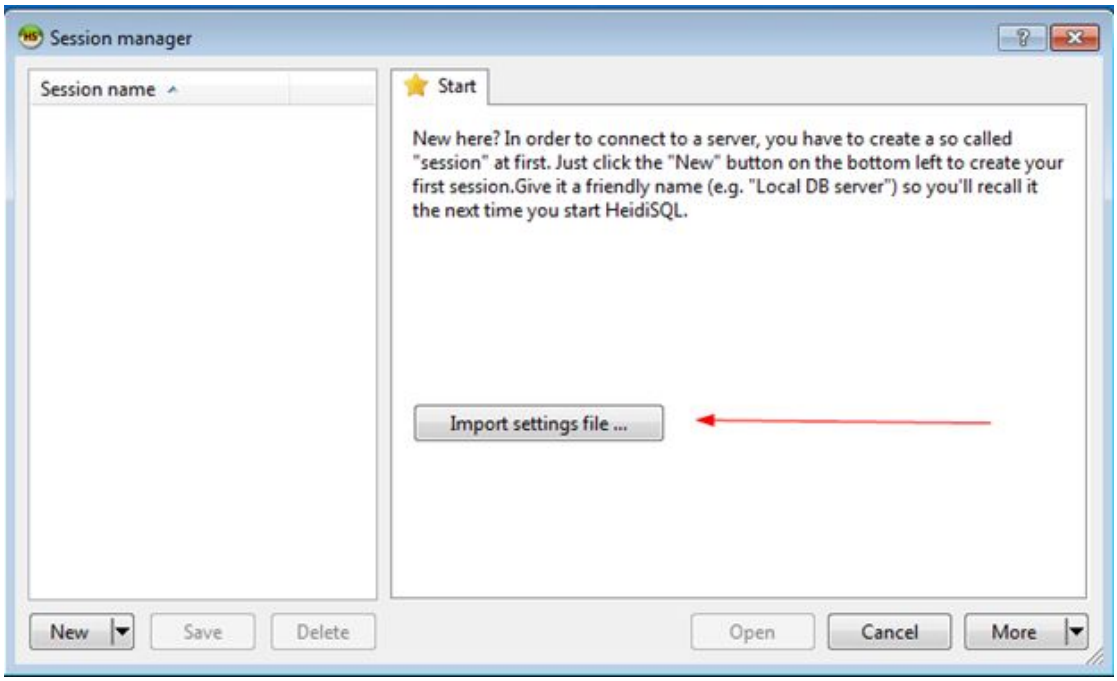
Setting up access to IRIS data using HeidiSQL

HeidiSQL is a GUI interface installed in the VDE that allows you to connect to the IRIS SQL database and access data files (tables/views), run queries, and export data as csv files. You can create and drop your own tables in your schema and manage the data as you need. Below is a step by step guide on how to connect to the database using HeidiSQL.

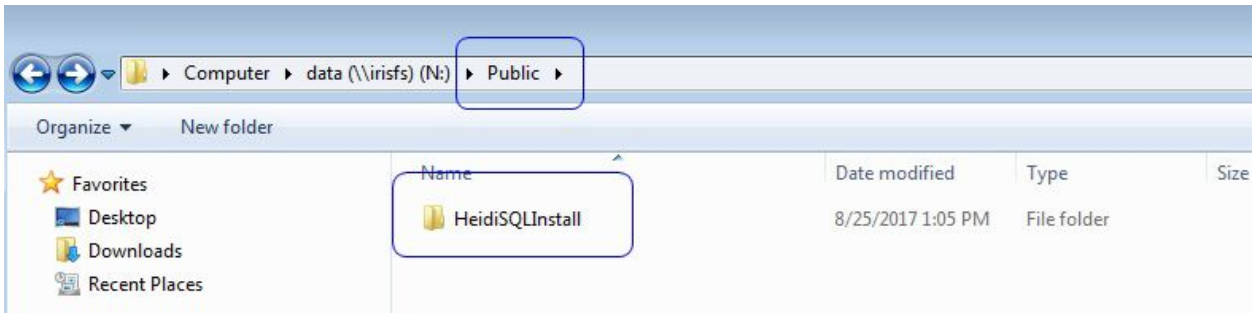
1. Click on the “Main Applications” folder on your desktop. 

2. Double click on HeidiSQL to start the program. 

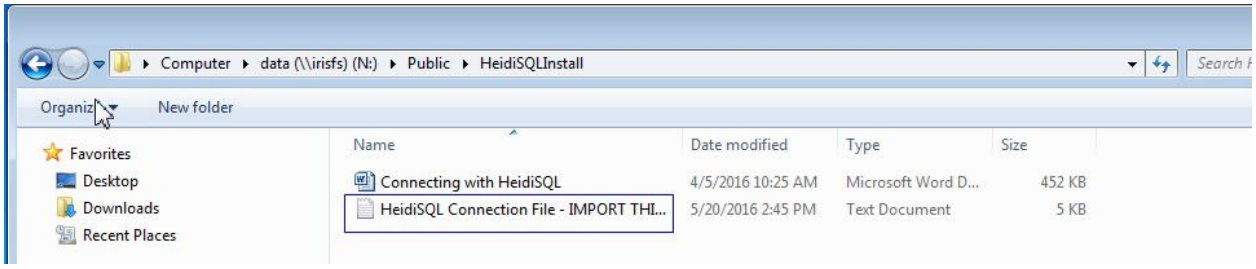
3. The first time you open HeidiSQL in the VDE you will need to configure the connection to the database. The first window you see when you open HeidiSQL is the “Session manager” window. Select the button named “Import settings file …”

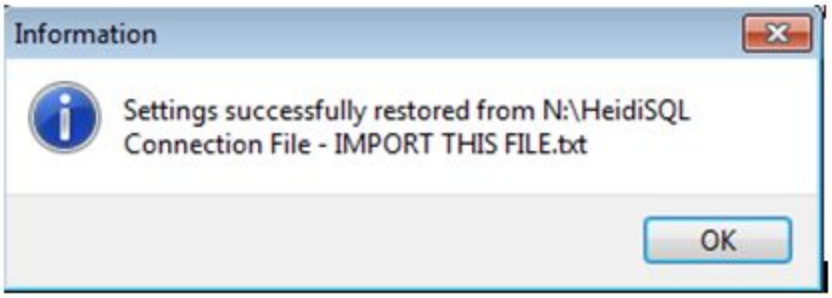


4. Next, navigate to the settings file found in the following network drive location: (\\irisfs)(N:) There is a public folder in which you will find a “HeidiSQLInstall” sub-folder.

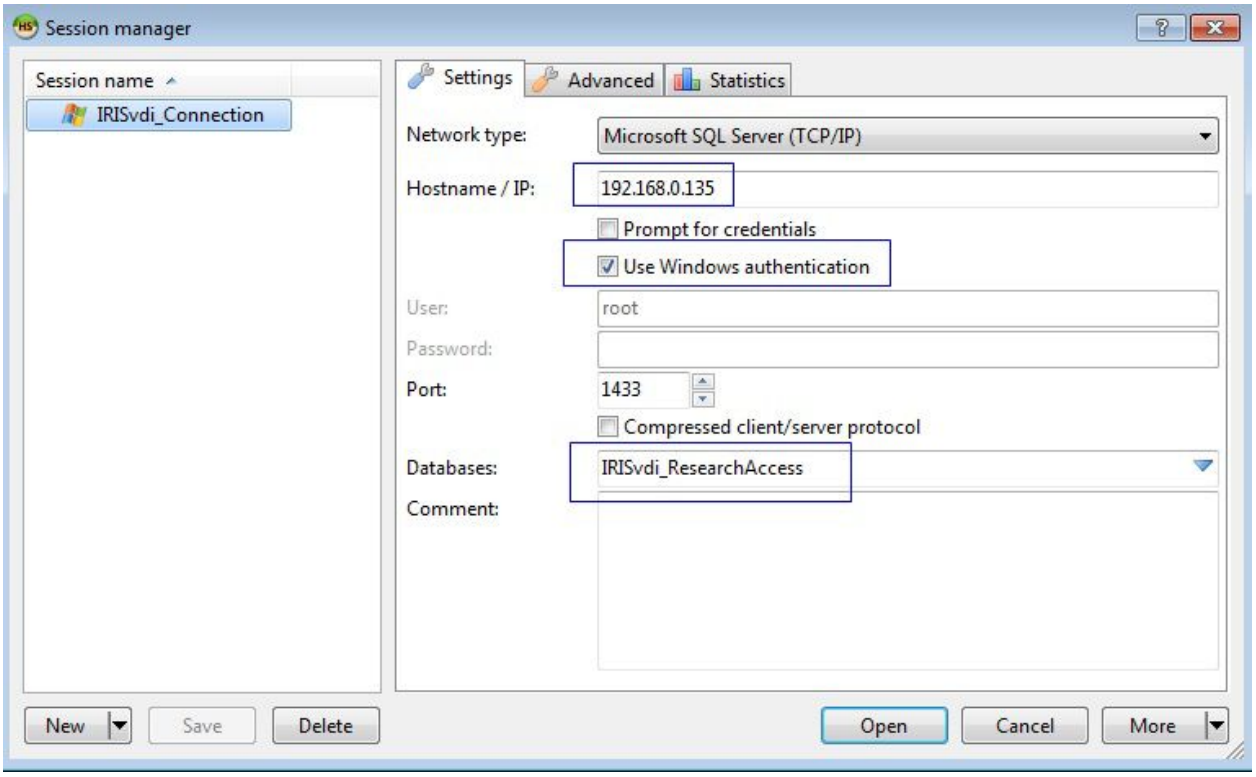


5. Double click on the Heidi connection settings file. If everything works correctly you should see a “Success” message.



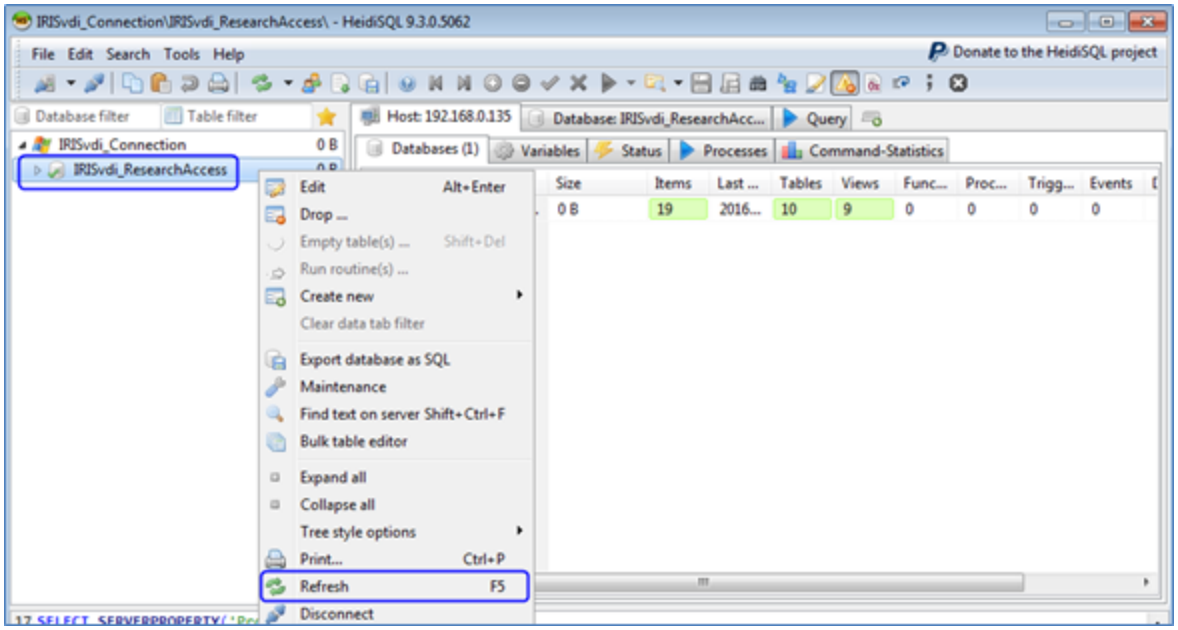


6. Click on the “Session Name” and you will see the connection details. Select “IRISvdi_Connection”, “IRISvdi_ResearchAccess”, and leave everything else as is.



7. Now just click “Open” and you will enter into your connection session with the database.

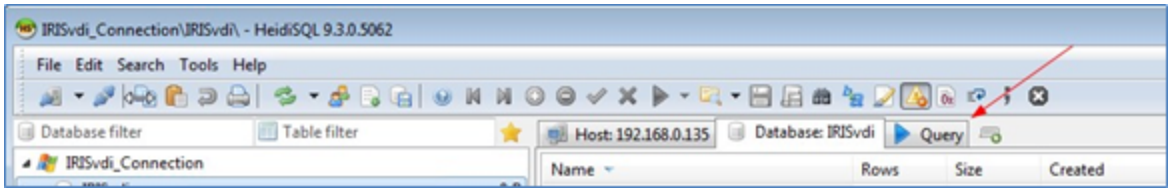
8. THIS NEXT STEP IS VERY IMPORTANT. You need to refresh the table list so your schema is visible. If you do not, the queries you write will not work. Simply right-click on “IRISvdi_ResearchAccess” under “IRISvdi_Connection” and click refresh from the list.



9. You will now see all the tables/views you have access to and schema(s), e.g., “UMich”. You are now ready to write queries and work with the data.

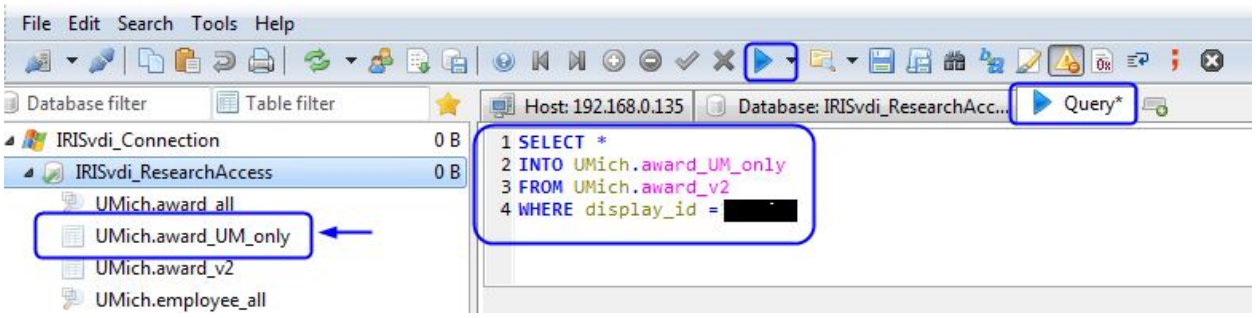
Writing simple queries

1. To write a query just click on the “Query tab” in HeidiSQL:



2. Try writing a simple query to retrieve some data for your analysis. Click on the blue arrow “Run Query” button after you have written your query. The result will appear below the query. For instance:

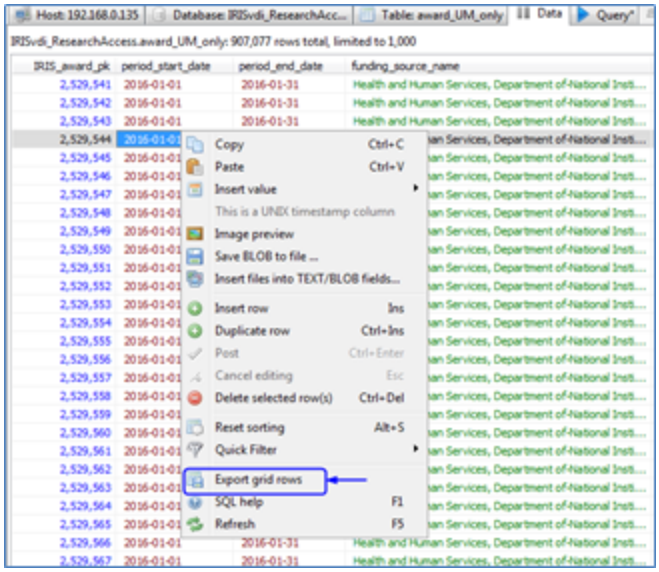
3. If within your own schema, you can create a new table as well. Click on the blue arrow “Run Query” button after you have written your query. For instance:



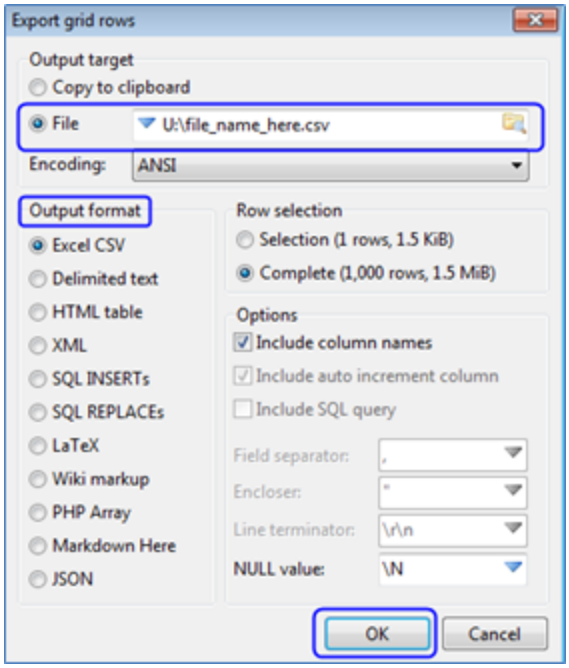
Exporting files from HeidiSQL

You may want to export data as a csv so that you can use it in other applications, like SAS, Stata, or R.

1. Once you run the query just right-click anyplace in the resulting data grid and click “Export grid rows”.

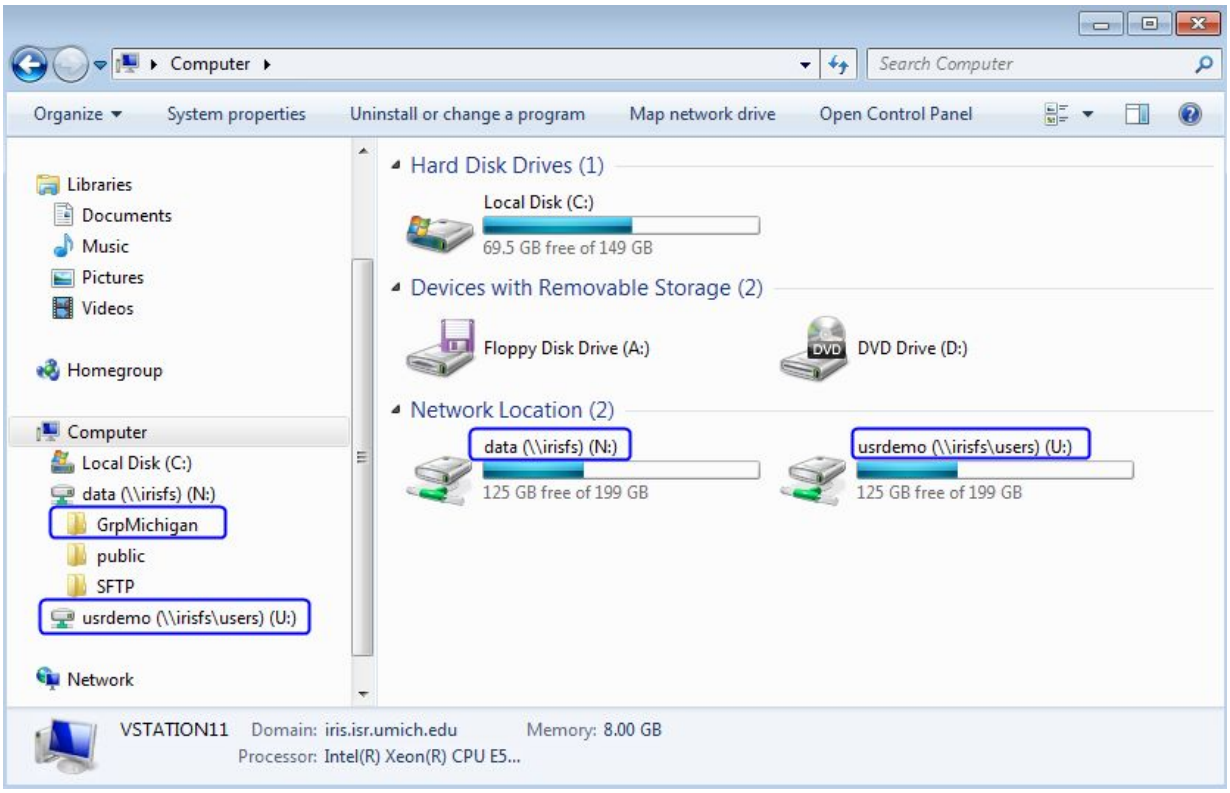


2. Select “File” and then type in a file name. Select the format that works best for you (e.g. Excel CSV) and then click OK. If OK is greyed out, it means that you need to click on an Output Format. Even though it defaults to Excel CSV, you still need to click on it. **We recommend saving your data files in your network drive, not in My Documents or Desktop.** They will be safer there and less likely to be irretrievable in case your connection is unexpectedly lost.



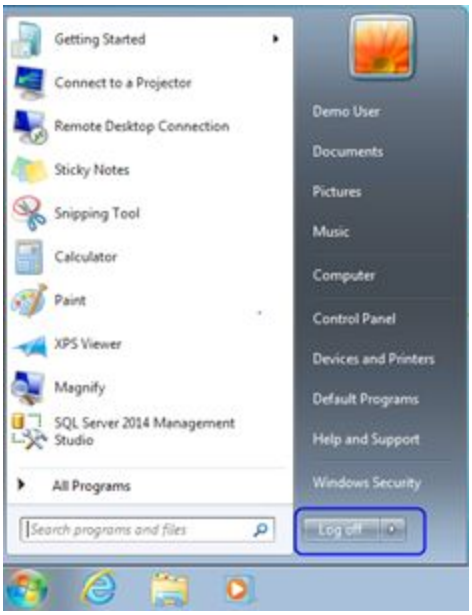
Saving your work in the VDE

When working with the IRIS data within the VDE, you need to make sure to save your work within the IRIS file server, either within your home folder (U:\ ) or a shared group folder (N:\ ); **otherwise, all work will be lost after a current VDE connection session is terminated.** To locate these folders open Windows Explorer and view the items located under “Network Location” (shown below).



All users are set up with a personal home folder (on the U: drive) upon account creation that will have the same naming as the user login. In addition, research teams will generally have a shared group folder located on the N: drive with a name that begins with “Grp” (e.g., “GrpMichigan”). If you need an additional shared folder set up, please contact IRIS.

Logging off from a VDE session



Once you complete your work in VDE, make sure to log off from a remote workstation. If you just close the window, the session remains open until it is automatically terminated after 7 days of idling time. Currently we have 12 virtual workstations (as of December 2017) for all VDE account holder, so closing your session when you are finished allows other users access to the VDE workstations.

VDE policies

IRIS VDE Acceptable Use Policy

VDE users should follow best practices in restricted-data use as outlined in the IRIS VDE Acceptable Use Policy (found on the IRIS website and in Appendix A of this guide). For details on common procedures related to these policies, please read below.

Updating software in the VDE

Researchers must refrain from modifying the virtual desktop interface environment (Windows or Linux) or the software installed within it. Adding user-configurable or unauthorized software is prohibited and existing software may not be used for anything other than its intended research purpose. If you wish to install additional software in the VDE, this request must be made via irisdatarequests@umich.edu and, if available, **the request will be completed during the next monthly maintenance window.**

Importing files into the VDE

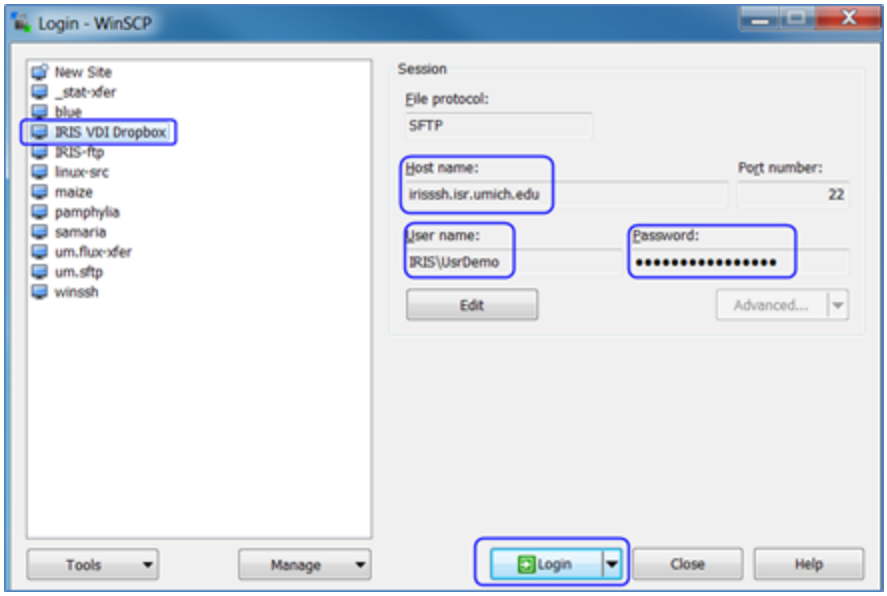
If a researcher needs to bring data files into the Enclave, the files may be uploaded via FTP to a designated temporary storage area for review and import by the IRIS Research Team. The researcher should include information about this data import request in the initial data access application form or submit a Data Import Request Memo to irisdatarequests@umich.edu.

Any data which could potentially re-identify IRIS data will be prohibited from importation. This qualifier is subject to review by IRIS staff. The IRIS Research Team will review the file(s) requested for importation and, if approved, install them in the researcher's personal Enclave user space as requested.

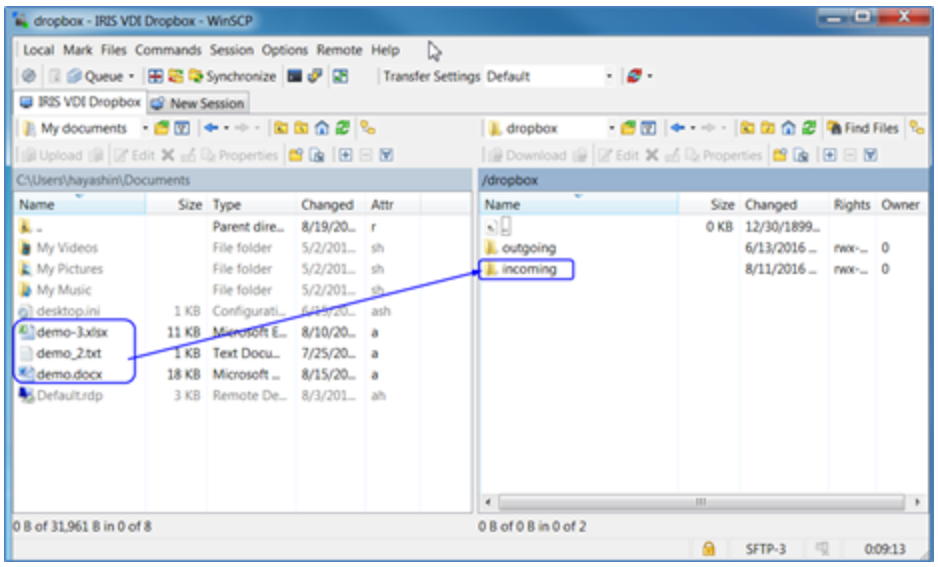
Steps for importing files to the VDE

If you wish to request IRIS to merge user provided data with IRIS data in the VDE environment, you can connect and upload files to the IRIS FTP server using tools built-in to Windows and OS X (e.g., FileZilla), or a dedicated FTP client like WinSCP. Below, we describe the procedure using WinSCP.

1. When using WinSCP, you will need to login to access IRIS FTP server (host name: irisssh.isr.umich.edu). Use the same credentials given for VDE access. Select IRIS VDI Dropbox as a destination folder to upload your own data files.



2. Once you have access to “IRIS VDI Dropbox” via IRIS FTP server (irisssh.isr.umich.edu), you will find the “incoming” folder. Select the file(s) that you wish to transfer to ftp (from a left panel) and drag and drop the file(s) into the destination folder (incoming) in the right panel.



3. E-mail IRIS at irisdatarequests@umich.edu with your Data Import Request Memo. IRIS will collect your data files to review them.

4. Upon approval, the IRIS representative will remove your files from the "incoming" folder and move them into the location you specified on personal home or shared folder in the U:\ drive. You will be notified via email when this transfer is complete.

Exporting files from the VDE

Research results generated through the analysis of IRIS data can **only** be exported from the VDE after a disclosure review by IRIS staff for compliance with the IRIS Data Use Agreement and IRIS VDE Acceptable Use Policy. All analysis output and programming code are subject to disclosure review before being exported from VDE to ensure that individuals and organizations cannot be re-identified from the aggregated datasets.

Researchers who wish to export their research results and/or programming code from VDE will need to follow the disclosure review procedures as described below. Researchers should contact irisdatarequests@umich.edu if there are any questions regarding disclosure review.

Disclosure Review Service Policy

- The IRIS Research Team is responsible for handling disclosure review requests within **15 business days** from the date the request is received, depending on the nature of the request. The IRIS Research Team will alert researchers promptly if the review will exceed the stated time frame.
- As part of the data access seat fee agreement, each research group is given 10 requests for disclosure review each calendar year.
- Research teams will be notified when they are approaching their maximum number of requests and when they have reached 10 requests.
- Multiple disclosure review requests for simple revisions of the same file will still count as separate requests and therefore count against the maximum of 10 requests per research group per year.

What Researchers may **usually** export

- Statistical summary information including frequency tabulations, magnitude tabulations, means, variances, correlation coefficients, and regression coefficients.
    - allowed only if it does not permit the identification of any individuals and organizations
    - aggregated observations should include at least 3 IRIS member institutions

What Researchers may **never** export

- Raw data or microdata files or analysis output containing individually identifying information
- Program/code that includes information on database connection (e.g., database name, connection credentials) or embeds identifiable information

Common reasons for decline
Problem: File includes organization names **Solution**: Replace names with preferred anonymized designations (e.g., University A or Vendor A) or with the numeric organization id (e.g., vendor ID, subawardee ID)
Problem: Disclosure request lacks description **Solution:** Send a description of the file(s) for which you are requesting disclosure review. Be sure to follow the Disclosure Review Request Memo template to provide accurate and complete details
Problem: Frequency count in cells is too small **Solution:** Suppress cells for which the values are smaller than 10
Problem: Risk of secondary disclosure **Solution:** Do not make disclosure requests for files that are very similar or contain overlapping information such that the risk of secondary disclosure is high

Preparing data outputs for disclosure review

If a researcher needs to export analysis results from the Enclave, the researcher should first review the information on what may/may not be exported as well as common examples of issues that will cause a disclosure review request to be declined. Preparing output files with these guidelines in mind and avoiding common pitfalls such as those described will help to speed the time it takes to process the disclosure review request.

Disclosure proofing is intended to prevent the identification of a particular individual or

organization. For example, in the employee data, this means that an employee earnings measure can only be released if the number of cases is high enough that no single person can be identified. Disclosure proofing also involves comparison of successive output files, so that a small change in the number of cases from one file to the next does not result in the identification of any one individual. Applying these tests can sometimes mean that certain data points must be excluded from your outputs. The disclosure review process influences potential outputs in several ways:

1. Individual Data

Information on individual employees can only be released if it is based on a sufficient number of cases to ensure that no single person can be identified in the data.

2. Vendor and Subaward Data

Information on business establishments can only be released when it is based on a large enough number of establishments and when the concentration rate is low. A concentration rate would be high if a small number of vendors accounted for a large percentage of spending in a particular category. Some by-county or by-NAICS code measures cannot be released because of these issues.

3. Secondary Disclosure and Implicit Samples

Effective disclosure proofing requires attention to the possibility that data in a single output, or data disclosed in more than one output can be used to infer information about individuals or organizations. For instance, if we were to approve the release of a file with data that differed from the data in a prior disclosure release by only one or two data points, comparing files would allow you to identify information about those data points, compromising privacy and confidentiality.

Likewise, the size and (for businesses) concentration of unreported categories must be taken into account. For example, if we break information about individuals into three categories and release data on two of them along with information about the total number of individuals on which the data are based, the size of the omitted category can be easily inferred.

Concerns about secondary disclosure and implicit samples can sometimes lead to suppression of information for categories that would pass disclosure when reported alone. Suppose, for instance, we wish to report information on an industry, 1, and on its four sub industries 1.a, 1.b, 1.c, and 1.d. If 1.d is too small (or too concentrated) to pass disclosure, its suppression from the report will also require suppression of 1.a, 1.b, or 1.c because those three data points along with total information about 1 would allow information about the suppressed category (1.d) to be inferred. This is why effective privacy protection requires that IRIS also scrutinize contextual information such as the number of cases associated with a particular report or figure and the match rates for individuals and organizations.

Researcher Checklist for Disclosure Review Request

Step 1: Prepare files for disclosure review request.

Step 2: Write a Disclosure Review Request memo.
This document is essential as it provides information necessary to the reviewer as well as creating a reference for your own research group to use when creating future disclosure requests.

Step 3: Send memo to the IRIS Research Team via irisdatarequests@umich.edu.
An e-mail help ticket will be generated upon receipt of the request. A memo must always be included with a disclosure review request or it will be denied regardless of the content of the files requested. IRIS will only review complete requests.

Step 4: Download approved file(s).
If review results are approved by IRIS, the requested output will be copied to your FTP "outgoing" folder. A copy of the approved disclosure review request memo will be saved to your personal and/or project folder.

PLEASE NOTE:

FTP folders accessible to researchers on the IRIS VDE should not be used as a permanent storage area. Files left over 3 days are subject to deletion.

Transferring Files with Linux

When you login into an IRIS VDE Remote Workstation, Linux users should see L:\ drive and it should automatically map to your IRIS Linux home directory. Should this not occur you can map a drive letter in windows using the following network path:

\\umetrics0\userid
where “userid” should be replaced with your IRIS userid.

After your L: drive is mapped you should be able to drag and drop files between the IRISfs server and the Linux server like any other windows share.

Publishing your research with IRIS data

Any publications resulting from research using the IRIS data must "acknowledge the IRIS at the University of Michigan as the source of the Data and [researchers] will provide IRIS with an electronic copy of the publication".[4] Publications must include appropriate acknowledgement of the IRIS Project (e.g., "This work was enabled, in part, using resources provided by IRIS.") and the necessary dataset citations once your research is published and deposited into trusted repositories for public sharing and long-term storage and accessibility.

Citations for IRIS data

Citations for the 2017 data release (UMETRICS 2016Q3a) are listed below. Citations for subsequent data releases will be made available on the IRIS website.

- **UMETRICS 2016Q3a Core files[5]:**
  The Institute for Research on Innovation & Science (IRIS). UMETRICS 2016Q3a Dataset: Core Files. Ann Arbor, MI: IRIS [distributor], 2017-03-15, doi:10.21987/R70080
- **UMETRICS 2016Q3a Crosswalk files[6]:**
  The Institute for Research on Innovation & Science (IRIS). UMETRICS 2016Q3a Dataset: Crosswalk Files. Ann Arbor, MI: IRIS [distributor], 2017-03-15, doi:10.21987/R7B66D

Researchers are encouraged to contact the IRIS Research Team for more information for best practices in sharing data and metadata, versioning, and data citation.

Submitting publications to IRIS

Please submit citations and/or electronic copies of your publications to IRIS via the IRIS publications form.

[4] *IRIS Data Use Agreement*, [2017] page 2, line 4.h.
[5] See: http://iris.isr.umich.edu/research-data/2017datarelease-corefiles/
[6] See: http://iris.isr.umich.edu/research-data/2017datarelease-crosswalkfiles/

Appendix A: IRIS VDE Acceptable Use Policy

Institute for Research on Innovation and Science (IRIS)

VDE Acceptable Use Policy

Researchers who are authorized to access de-identified data from the IRIS repository for approved uses within the IRIS virtual data enclave (the "Enclave" or "VDE") must be aware of, and follow, all regulations and guidelines below.

Data Access Protocols

- To ensure data security, access to IRIS data is only possible within the VDE.
- Researchers must hold a VDE user account to access the data. The user account is issued for one year periods (renewable) as indicated in the IRIS Restricted Data Use Agreement (the "DUA") and may only be used for the purposes described in the approved research proposal (and in Appendix B: *Research Activity* in DUA).
- The terms of DUA, which permit usage of the VDE, expires after one year unless extended. Renewal will be considered upon the submission of an annual project report by researchers.
- The data cannot be used for any other purpose, disclosed to third parties, or used for commercial purposes.
- No individuals besides those with explicit project approval are allowed to view or work with the provided data.

Data Security and Protection Procedures

- Researchers agree to safeguard the information contained in the Enclave from unauthorized use, disclosure, inadvertent modification, destruction, or denial of service.
- Access to the data and analytical software/tools in the Enclave environment is a revocable privilege and is subject to security testing and monitoring, including user provided content and working materials stored in users' home folders within the VDE network drive.
- Researchers must refrain from modifying the Enclave environment (Windows or Linux) or the software installed within it. Adding user-configurable or unauthorized software is prohibited and existing software may not be used for anything other than its intended research purpose.

- Researchers agree not to take screenshots or other video or image grabs of any displayed data.
- Researchers are prohibited from sharing personal VDE user accounts and authenticators (passwords) or permitting the use of remote access capabilities to any unauthorized individual.
- Researchers are required to immediately notify the IRIS Enclave representative of any suspicious system activity or concerns over the use or disclosure of the data.

Penalties for DUA Violation

- Those who violate the terms of the DUA will be removed from their active research projects in the VDE; upon termination of the DUA, the researcher's data access via VDE will end.
- Any violation of the DUA may result in a permanent ban on data access.
- Any attempt to re-identify IRIS data will result in immediate and permanent revocation of all future access to the Enclave.

Data Import Protocols

- Researchers may arrange to merge their own data with IRIS data. This arrangement must be made in advance during the process of submission of their research proposal (and in Appendix A: *Requested Data Elements* in DUA) for internal review.
- Any data which could potentially re-identify IRIS data will be prohibited from importation. This qualifier is subject to review by an IRIS Enclave representative.

Data Export and Disclosure Review Protocols

- All results of analyses researchers wish to export from the Enclave must be reviewed for disclosure risk by the IRIS Enclave representative. These include: quantitative data as well as qualitative notes on discussion and presentation of regression results. All research results may only be published after review by the representative.

Post-Approval Modifications to Submitted Materials

- Researchers must notify the IRIS Enclave representative in writing of any proposed changes in research plans. Amendments to the project include adding new collaborators to a project using Enclave data.

Publication Acknowledgement and Data Citation

- Publications must include appropriate acknowledgement of the IRIS Project (e.g., "This work was enabled, in part, using resources provided by IRIS.") and the necessary dataset citations once your research is published and deposited into trusted repositories for public sharing and long-term storage and accessibility.
- Researchers are encouraged to contact IRIS for more information and / or review IRIS wiki for best practices in sharing data and metadata, versioning, and data citation.

By signing this user agreement, I am acknowledging that I accept and will abide by all the terms and conditions described above.

Signature Date

Printed Name

Appendix B: IRIS Data Import Request Memo (template)

Data Import Request Memo

A memo to request user provided data to be imported into the IRIS VDE should be emailed to irisdatarequests@umich.edu and contain the following information. You may download this form and edit as needed, otherwise please create a memo with elements numbered as follows:

1. PROJECT NAME

2. PROJECT PRINCIPAL INVESTIGATOR NAME

3. REQUEST SUBMITTED BY

4. GENERAL INFORMATION
    a. How many files are requested for import?
    b. What is the purpose of use (why is the file being requested)?
    c. Please state how the data are part of the research project as approved. (You may summarize or copy descriptions from your research proposal).

5. DESCRIPTION OF RESEARCH SAMPLES
    a. How was the file created?
    b. How will the file be merged or combined (with what data elements of IRIS data)?
    c. What is the sample size/representation (i.e., what are the cases and/or subsets of data represented)?
    d. Identify any implicit samples or potential secondary disclosure concerns.

6. FILE INFORMATION
   For SAS data set please paste “PROC CONTENTS”
   For STATA data set please paste “DESCRIBE”
   For other data sets please provide the following:
    a. File name (including file format, e.g., “output.rtf”)
    b. File size (KB)
    c. Number of records
    d. List of variables (record layout is sufficient)

7. DESTINATION LOCATION FOR FILES
   If approved for import, where would you like this file to be saved?
    a. Group folder (list folder name) or,
    b. Personal folder (list folder name)

8. PERMISSIONS FOR FILES
   Indicate all that apply:
    a. The data are public use and downloaded from the internet here (list URL)

b. The data are public use and procured other than from the internet (attach evidence of public use)
c. The data are purchased (attach receipt or documentation from institution that subscribes to data)
d. The data are proprietary (attach written permissions from data custodian or vendor providing consent to import to IRIS VDE and noting any restrictions on use)

9. OTHER INFORMATION OR COMMENTS (OPTIONAL)

Attach this memo to an e-mail and send it to irisdatarequests@umich.edu

---

For IRIS use only:
Date request received: - -
Date reviewed: - -
Approved [ ] Not Approved [ ] Changes made before import, if any:

_____

_____

Appendix C: IRIS Disclosure Review Request Memo (template)

Disclosure Review Request Memo

A memo to request disclosure review should be emailed to irisdatarequests@umich.edu and contain the following information. You may download this form and edit as needed, otherwise please create a memo with elements numbered as follows:

1. PROJECT NAME

2. PROJECT PRINCIPAL INVESTIGATOR NAME

3. REQUEST SUBMITTED BY

4. GENERAL INFORMATION
    a. What and where are the files requested for export?
    b. How many files are requested for export?
    c. What is the purpose of use (why is the file being requested)?
    d. Please state how the outputs are part of the research project as approved. (You may summarize or copy descriptions from your research proposal).

5. DESCRIPTION OF RESEARCH SAMPLES
   How was the file created? For each sample, please describe your selection criteria and how the research sample differs from the samples underlying or other samples you have used. Take as much space as you need for each; add samples as needed.
    a. SAMPLE 1
    b. SAMPLE 2
    c. SAMPLE 3

6. RELATIONSHIP BETWEEN SAMPLES
   Describe how your samples relate to each other and identify any implicit samples.

7. OUTPUT FILES
   For each research output file to be removed, please enter the following information:
    a. File name (including file format, e.g., “output.rtf”)
    b. Description of file (e.g., “tables relating to …”)
    c. Program that produced the file (e.g., “output.sas”)
    d. Research Sample Number that underlies the file
    e. The file name/location that contains supporting statistics
    f. The name of the program that generated the disclosure analysis file (e.g., “output_disc.sas”)
    g. Comments (any other information you wish to add)

8. VARIABLE DEFINITIONS
   Please include all variables in the research output you wish to remove and in supporting disclosure output.
   a. VARIABLE NAME:
   b. DEFINITION (include type, e.g., continuous, binary (0,1), other discrete):
   c. SOURCE:
   d. COMMENTS:

9. OTHER INFORMATION OR COMMENTS (OPTIONAL)

Attach this memo to an e-mail and send it to irisdatarequests@umich.edu
Please note that disclosure review requests may take up to 15 business days to review.

For IRIS use only:
Date request received: - -
Request ___ of ____
Date reviewed: - -
Approved [ ] Not Approved [ ]
Changes requested for resubmission, if needed:

_____

_____

_____

_____