



IRIS VDE User Guide

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

Last Updated:
January 3, 2019

Table of Contents

Working with the IRIS Research Support Team	3
Our staff and hours	3
IRIS help ticket system	3
Service level agreement	4
VDE account basics	5
Database schemas and views	5
Saving your work in the VDE	5
Logging off from a VDE session	6
VDE access using Windows	7
VDE access using Mac	10
VDE access using Linux	12
Working with IRIS UMETRICS data	14
About IRIS UMETRICS data	14
Acceptable Use Policy	14
Updating software in the VDE	14
Importing files into the VDE	15
FTP connection steps	15
Transferring Files with Linux	16
Exporting files from the VDE	17
Preparing data outputs for disclosure review	17
Researcher Checklist for Disclosure Review Request	19
Publishing your research with IRIS UMETRICS data	20
Citations for IRIS UMETRICS data	20
Submitting publications to IRIS	20
Appendix A: IRIS VDE Acceptable Use Policy	21
Appendix B: IRIS Data Import Request Memo (template)	25
Appendix C: IRIS Disclosure Review Request Memo (template)	28
Appendix D: Instructions for using Git in the VDE	31
Appendix E: Instructions for using HeidiSQL in the VDE	33



INSTITUTE FOR RESEARCH ON INNOVATION & SCIENCE

Welcome to the community of IRIS researchers!

We at IRIS look forward to working with you as you delve into the IRIS UMETRICS dataset via the virtual secure data enclave (hereafter, VDE or Enclave¹) to meet your research goals.

A few reminders as you begin -

The VDE is a virtual machine launched from your 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.

We ask all VDE users 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 Support 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 Support Team

Our staff and hours

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



We are located in the IRIS offices at: 330 Packard Street, Ann Arbor, Michigan, 48104

Our business hours are: Monday through Friday from 8:00 a.m. to 5:00 p.m. (EST).

Exceptions to these hours include closures for holidays (or, on the rare occasion, inclement weather or other unplanned events), which follow those of the University of Michigan.

IRIS help ticket system

For any VDE or research related questions or requests, use the e-mail

irisdatarequests@umich.edu to generate a help ticket. As a help ticket is reviewed and addressed by the IRIS Research Support Team, the status will be changed and you will receive email notifications accordingly.

Status notifications:

A “**Request**” status notification is an automated email generated by the help ticketing system to alert the user that a request was successfully submitted, but does not indicate that a member of the IRIS Research Support Team has seen the request. The “**In Process**” notification includes a response from the IRIS team, and a notification of “**Completed**” status is assigned once a ticket has been addressed and resolved.

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 within 1-3 business days and the assigned IRIS team member will respond within 5-7 business days from the date of the original user request.
- Exceptions:
 - **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.
 - **Disclosure review requests:** (see below)

Disclosure Review Service Policy

- The IRIS Research Support 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. IRIS 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.
- Multiple disclosure review requests for simple revisions of the same file count as separate requests and therefore count against the maximum of 10 requests per research group per year.

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.

VDE account basics

After completing the application process and being authorized for a VDE account, you will receive an email from IRIS with a brief online training module. 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. The IRIS Research Support Team will work closely with you for the onboarding steps to set up your new account.

- **Step 1:** Sign IRIS policy documents and complete online training module
- **Step 2:** Activate your DUO security account
- **Step 3:** Login to the IRIS VDE (see instructions with screenshots on pages 7-13)

Database schemas and views

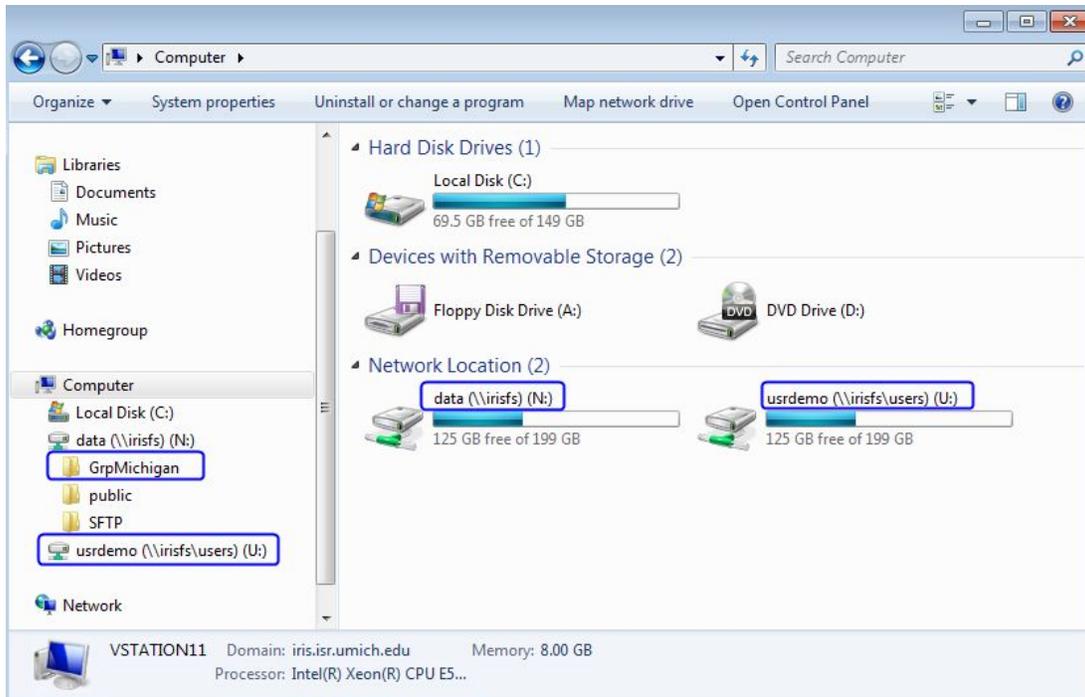
To give access to research data, IRIS grants VDE users access to a SQL server database where IRIS contains and manages data in tables and views. You will have access to requested data files in views in your schema and are able to create procedures, tables and views. For any questions related to the access rights and means of data access, please email irisdatarequests@umich.edu.

In the VDE's Public folder are "SQL Code snippets" where you will find some basic code that can help users interface directly with the SQL database using python, R, SAS, or Stata. You might find this quicker than exporting files. These code snippets will need to be modified to point to the specific schema to which you have access.

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” (example shown below).



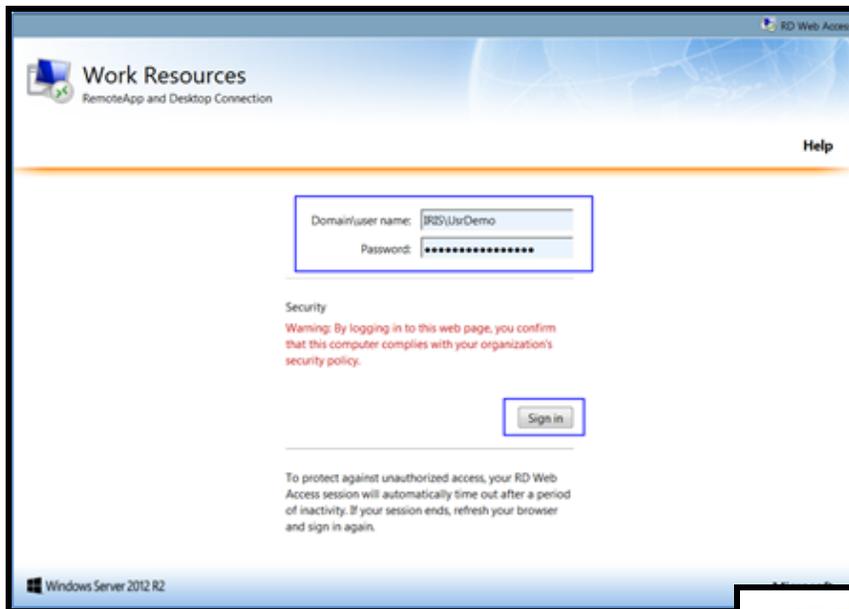
All users are set up with a personal home folder (on the U: drive) upon account creation that will have the same naming as the user login. In addition, research teams will generally have a shared group folder located on the N: drive with a name that begins with “Grp” (e.g., “GrpMichigan”). If you need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

Logging off from a VDE session

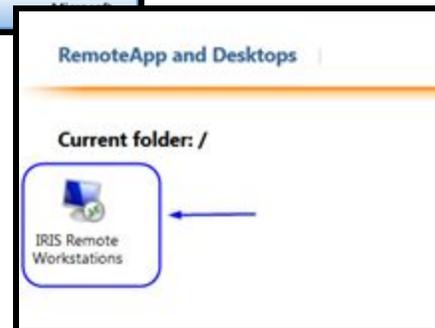
Once you complete your work in VDE, make sure to log off from the remote workstation. If you just close the window, the session remains open until it is automatically terminated after 7 days of idling time. Currently we have 15 virtual workstations (as of January 2019) for all account holders, so closing your session when you are finished allows others access to the VDE.

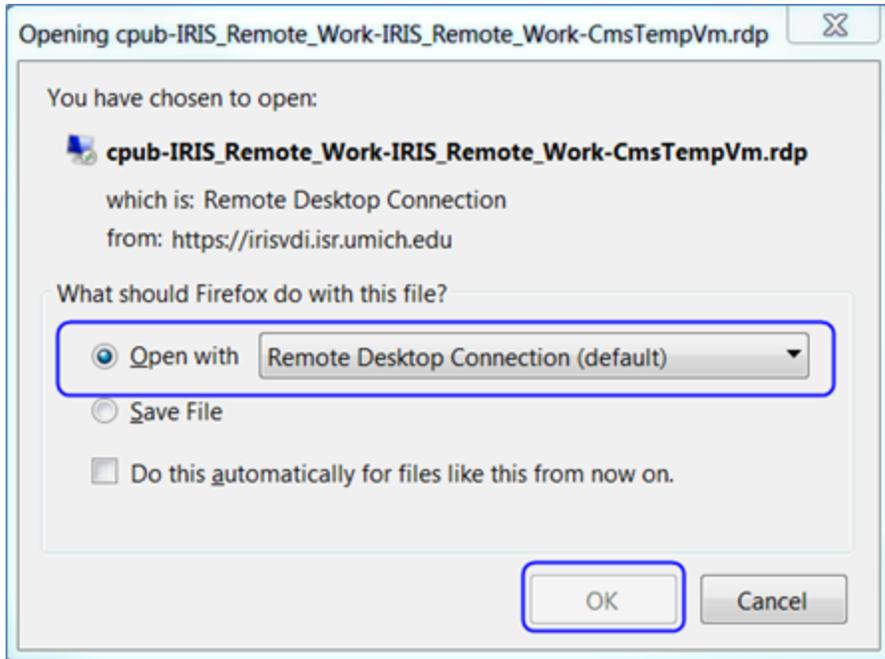
VDE access using Windows

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 the file. Open with "Remote Desktop Connection" and click "OK".

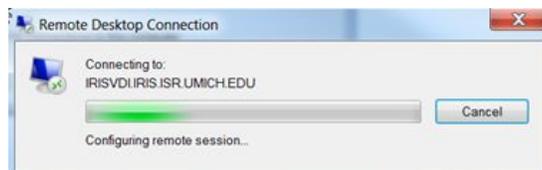
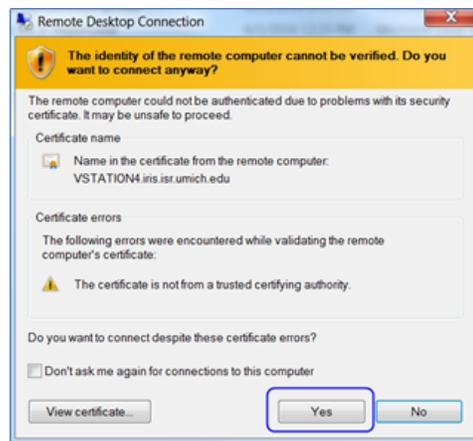




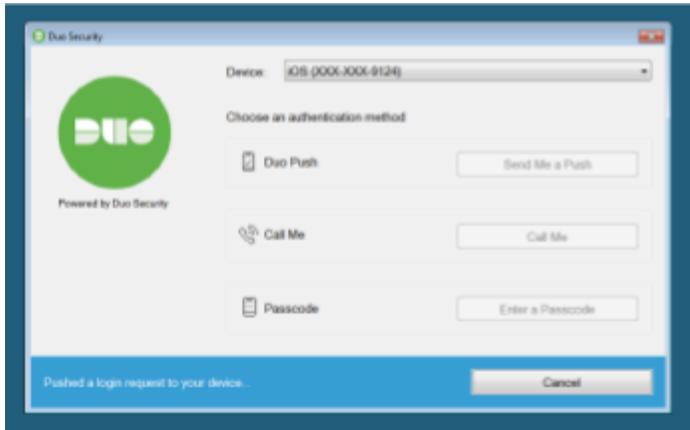
5. Once you are asked to enter your credentials, use the same username and password that you just used in Step 2. Again, make sure to add “IRIS” when typing your user name.



6. Click on “Yes”.



- 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...

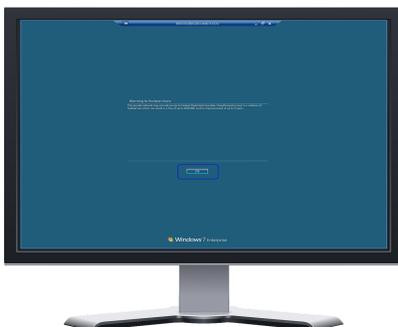
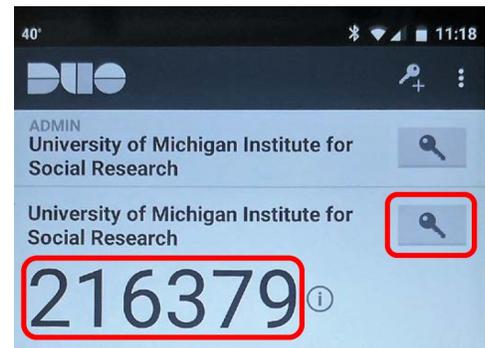


Computer screen



Smartphone screen

At the top of this window there is a drop-down box labeled “Device.” 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. 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 cellular or WiFi connectivity, 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 DUO app on your phone, and clicking the key next to “University of Michigan Institute for Social Research.” This code will generate whether you have any network connectivity or not.



- 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.

VDE access using Mac

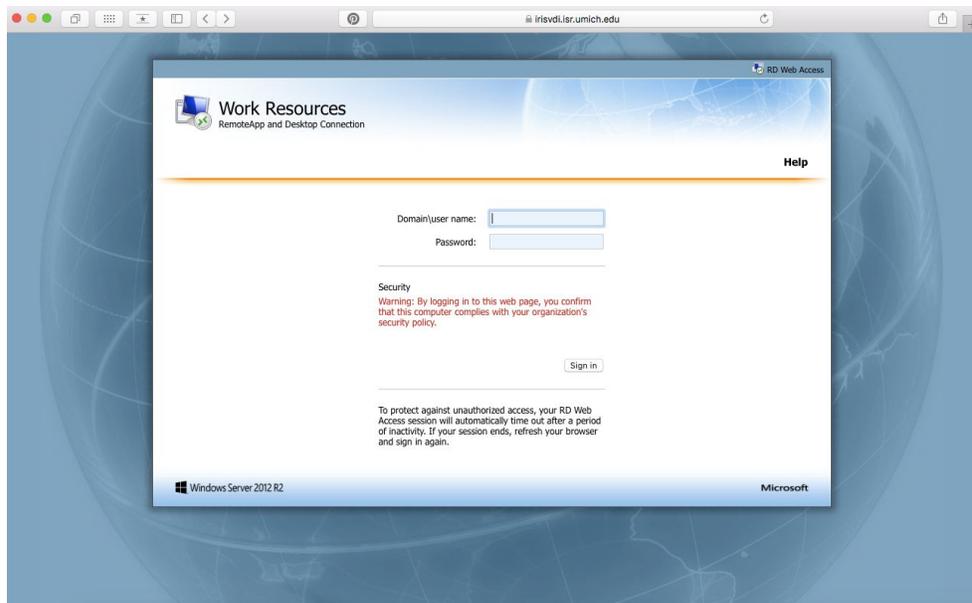


Prior to following the steps below, make sure you have Microsoft Remote Desktop version 8.0.43 installed. Microsoft Remote Desktop does not come pre-installed on a Mac, though most users will have it as part of their Office installation. Otherwise you can download it through the App Store.

1. Visit the following URL:

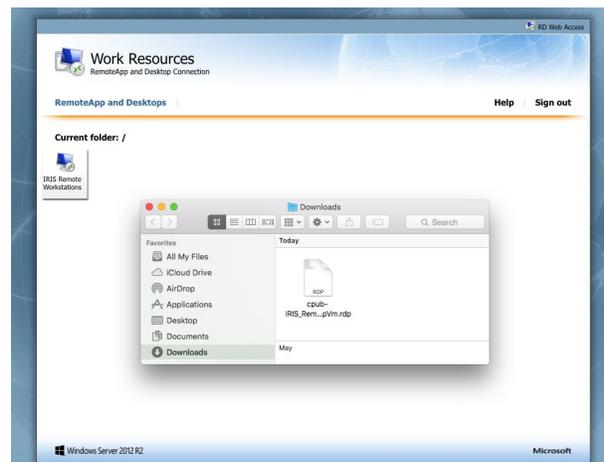
<https://irisvdi.isr.umich.edu/RDWeb/Pages/en-US/login.aspx?ReturnUrl=/RDWeb/Pages/en-US/Default.aspx>

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

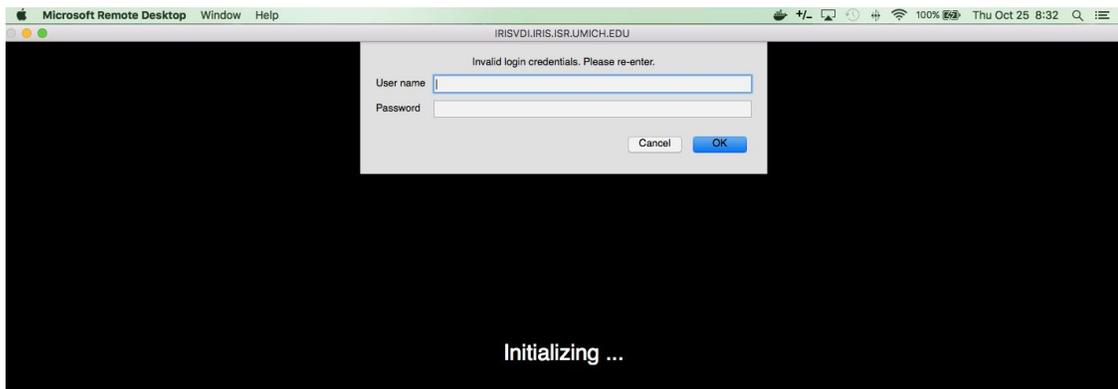


3. Click on the “IRIS Remote Workstations” icon in the upper left of the window and then click to open the remote desktop connection.

4. Once you are asked to enter your credentials, use the same username

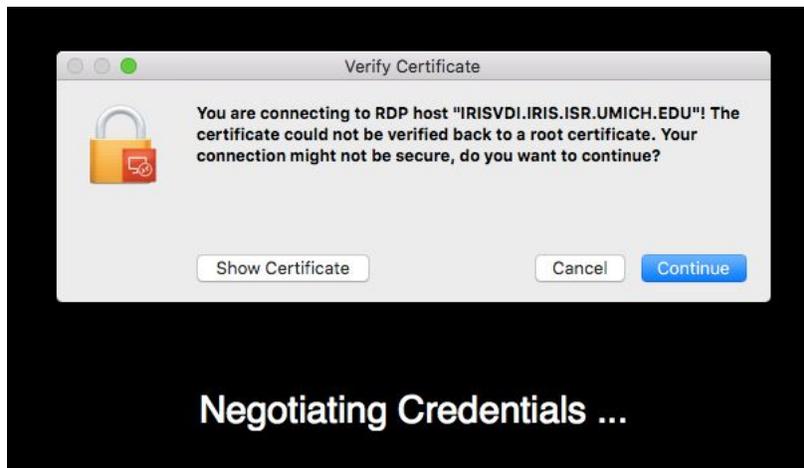


and password that you just used in Step 2. Again, make sure to add “IRISV” in front of your user name.



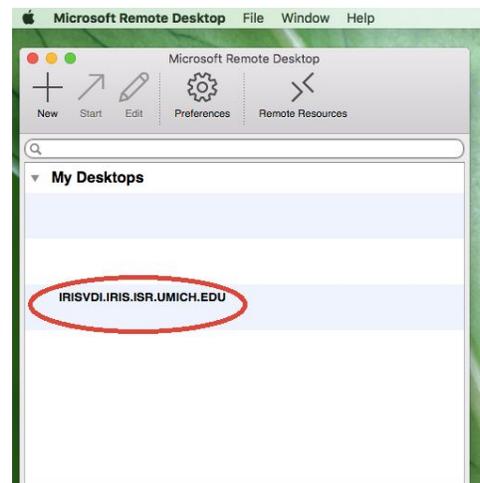
5. Click “Continue.”

6. Once your username and password are accepted, you will see your virtual workstation starting up and the DUO two-factor authentication prompt. (See page 9 for more on DUO.)



7. Click on “OK” to the “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.

8. Once this initial connection is made, you will be able to access the IRIS VDE via your Microsoft Remote Desktop application going forward. **PLEASE NOTE:** 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 by repeating the steps above.



VDE access using Linux

If you prefer to work in the Linux environment, you must request a Linux user account by e-mailing irisdatarequests@umich.edu. IRIS will create an account to access Linux, at which

point you should follow these steps:

1. Make a connection to an IRIS VDE workstation as described earlier.

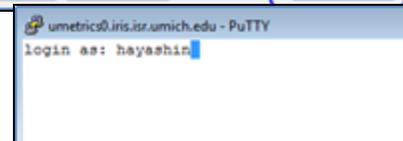
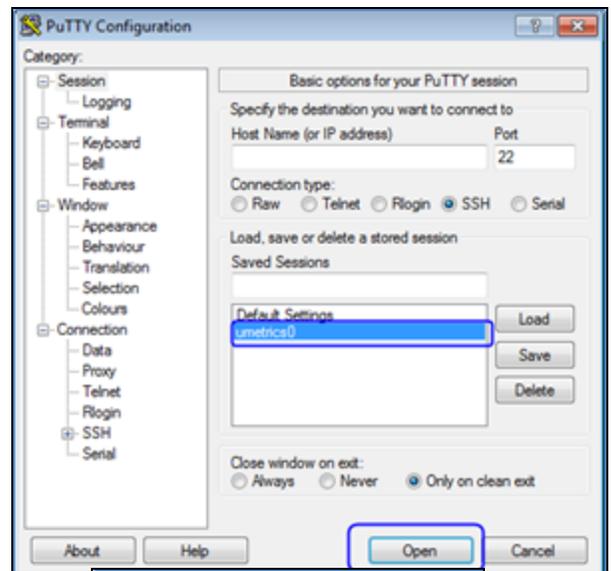
2. Open the Main Applications folder on the desktop, and double click on the PuTTY icon.

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

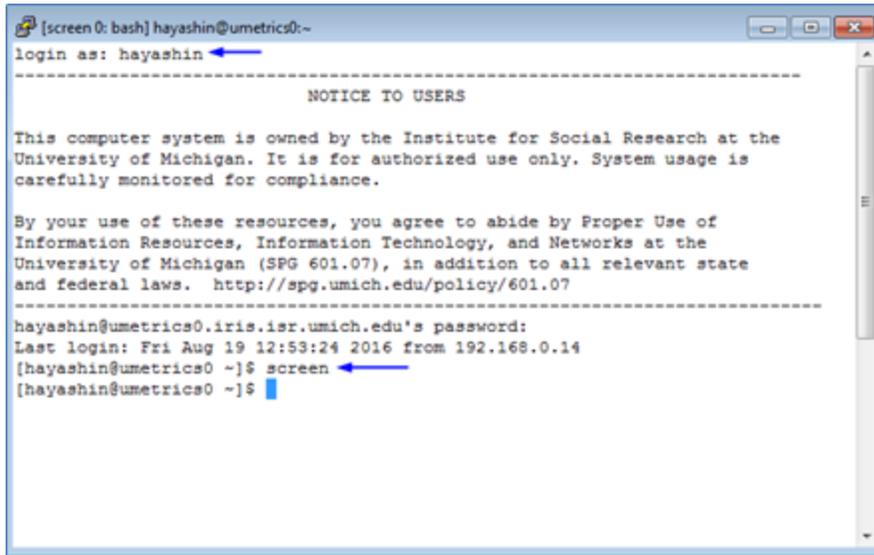
4. Use the same credentials used for IRIS VDE account (username and password) at the “login as:” prompt and press the

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

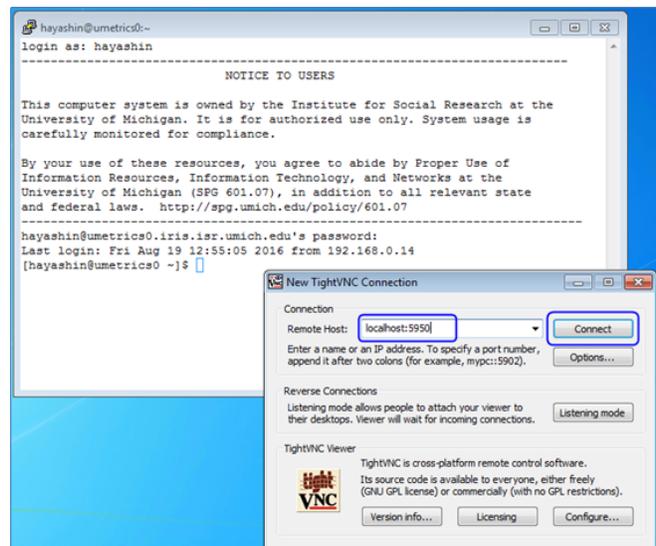


screens (bash shells) in a single putty session.



Additional steps for a graphical Linux environment

1. Make a PuTTY connection 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 to access the Linux Graphical Environment.



Working with IRIS UMETRICS data

About IRIS UMETRICS data

The IRIS website is a frequently updated resource for information about the IRIS UMETRICS data and annual research data releases. IRIS plans to continue annual data releases as membership grows and member institutions 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 Support Team.

Data Documentation

Documentation for the current IRIS UMETRICS data release and a data dictionary are available in the Public folder of the VDE. **This documentation is not to be shared outside of the VDE.** Public data documentation is available on the IRIS website.

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 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.**

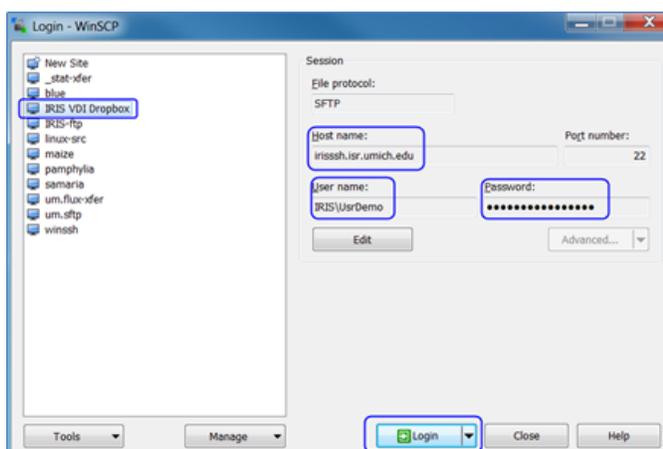
Importing files into the VDE

Any data which could potentially re-identify IRIS UMETRICS data will be prohibited from importation. This qualifier is subject to review by IRIS staff. If a researcher needs to bring data files into the Enclave, the files should be uploaded via FTP (see instructions below) to the designated temporary storage area (“dropbox”) for review by IRIS. The researcher should email irisdatarequests@umich.edu, including information about this import request using the “Data Import Request Memo” template available in Appendix B. IRIS will collect your files from the “incoming” Dropbox folder to review them. If approved, IRIS will move the files to the location you specify in the memo (e.g., personal or shared folder) and notify you when this transfer is complete.

FTP connection steps

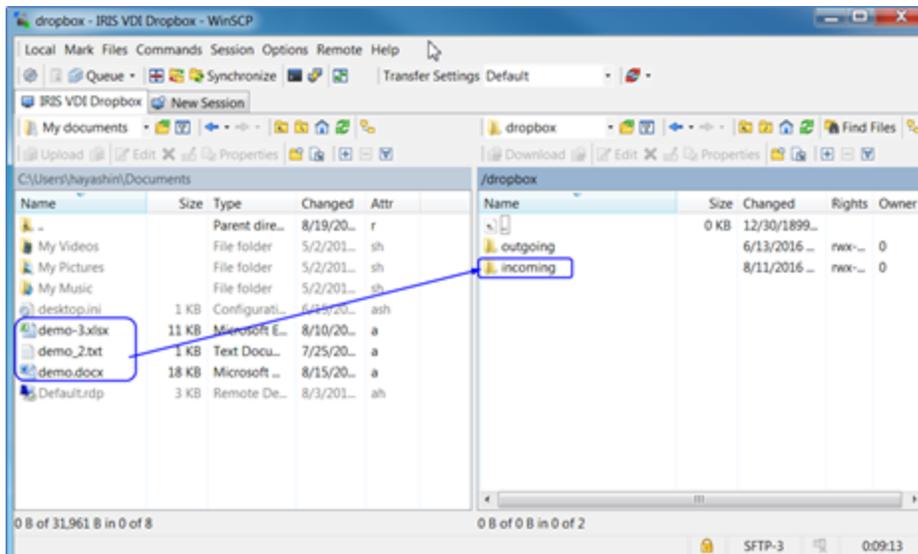
To import files to the VDE, you can connect and upload files to the IRIS FTP server using tools built in to Windows and OS X (e.g., FileZilla), or a dedicated FTP client like WinSCP, as described below. Imported files should be placed in the “incoming” Dropbox folder. You may follow similar connection steps for exporting files (after disclosure review) from the “outgoing” Dropbox folder in the VDE.

1. On 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 data files.



2. Once you have access to “IRIS VDI Dropbox” via the 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.



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 IRISs server and the Linux server like any other windows share.

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.

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

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 and common examples of issues that will cause a disclosure review request to be declined. 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.

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

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.

Copy the file(s) into a single folder named “Disclosure_Review_MMDDYY,” replacing the “MMDDYY” with the date of your 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. Please save it in the same folder as the files for which you are requesting review. 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 3: Email the IRIS Research Support Team via irisdatarequests@umich.edu.

Please note the location of the folder to be reviewed.

Step 4: Download approved file(s).

If approved, the requested output will be copied to your “outgoing” Dropbox folder for your retrieval. (See instructions for connecting via FTP on page 15.) **PLEASE NOTE: FTP folders should not be used as a permanent storage area. Files left over 3 days are subject to deletion.**

Publishing your research with IRIS UMETRICS 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”.² 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 UMETRICS data

Citations for each data release are made available for download in .BibTeX and .ris format on the IRIS website. For example, the citation for the 2017 Core Files is:

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

Researchers are encouraged to contact the IRIS Research Support Team for more information on 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.

² *IRIS Data Use Agreement*, [2017] page 2, line 4.h.

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

Instructions for placing a disclosure review request:

- Fill out memo as fully as possible.
 - Label a folder in the VDE with the file(s) requested for review using the naming convention DisclosureReview_MMDDYYYY. This folder should contain only the files you wish us to review, plus this memo.
 - Email irisdatarequests@umich.edu to place your review request and be sure to let us know the folder name to review.
-

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)

Please note that disclosure review requests may take up to 15 business days to review.

Appendix D: Instructions for using Git in the VDE

NOTE: for more general documentation on Git, see: <https://git-scm.com/doc>

Setting up an account

1. Email irisdatarequests@umich.edu if you would like to create an account with Git.
 - o The default setup will be a new userspace. If requested, we can also set up a project namespace. Please specify this in your request if desired.
2. We will email you with an account name and password for using Git within the VDE.
3. Within the VDE, you may use the Git GUI (user interface) or access Git via Git Bash (terminal window).



Creating and accessing your Git repository

1. Login to the VDE with your usual login and open Git Bash
2. Browse to your personal drive (the U:/ drive) i.e., in Git Bash use the command **cd u:**
3. To initiate a git repository in your personal drive folder, use one of these commands:

[for a user namespace]:

git clone ssh://usernamehere@irisgit:22/public/usernamehere.git

replacing “usernamehere” with your login name

[for a project namespace]:

git clone ssh://usernamehere@irisgit:22/public/projectnamehere.git

replacing “usernamehere” with your login name and “projectnamehere” with your project namespace.

4. At the password prompt, use the password provided by IRIS for your Git account. This will create your git repository folder, then you may add code files to this folder to use git within the VDE.

Please note that all Git repositories in the VDE are public within the VDE and accessible to any user in the VDE who has created a Git account.

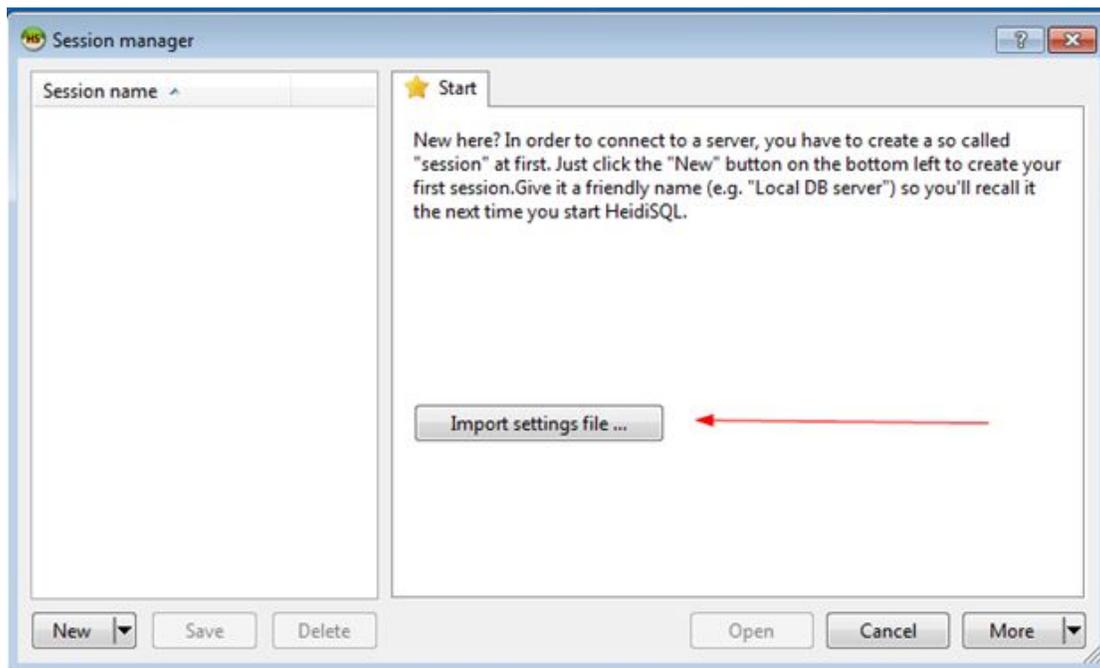
Appendix E: Instructions for using HeidiSQL in the VDE

HeidiSQL is a GUI interface installed in the VDE that allows you to connect to the IRIS SQL research 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.

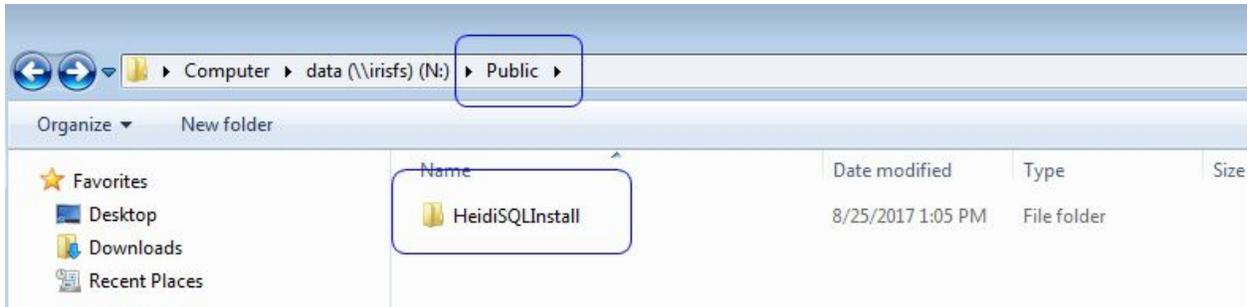
1. Click on the “Main Applications” folder on your desktop then double click on HeidiSQL to start the program.



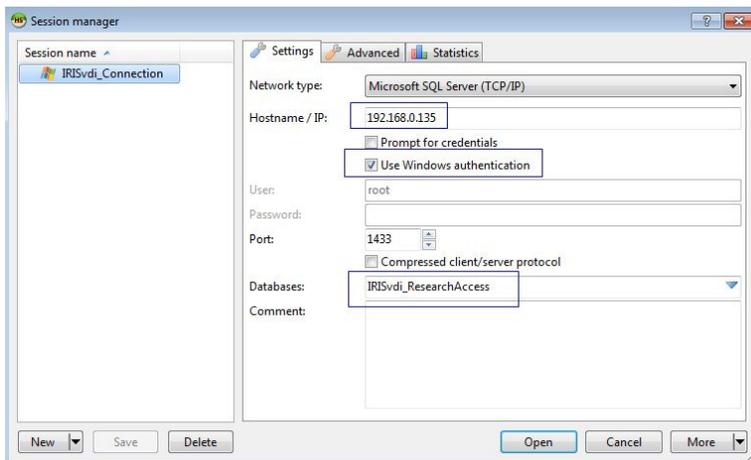
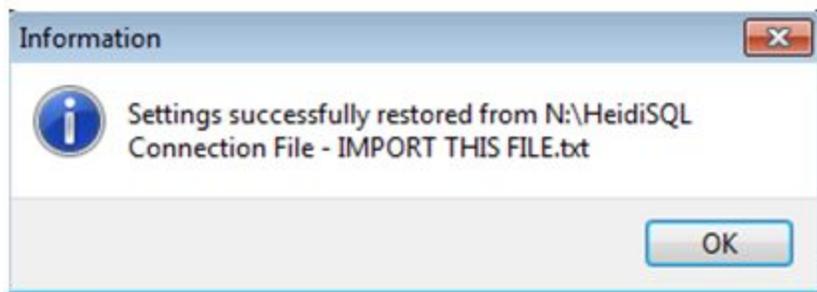
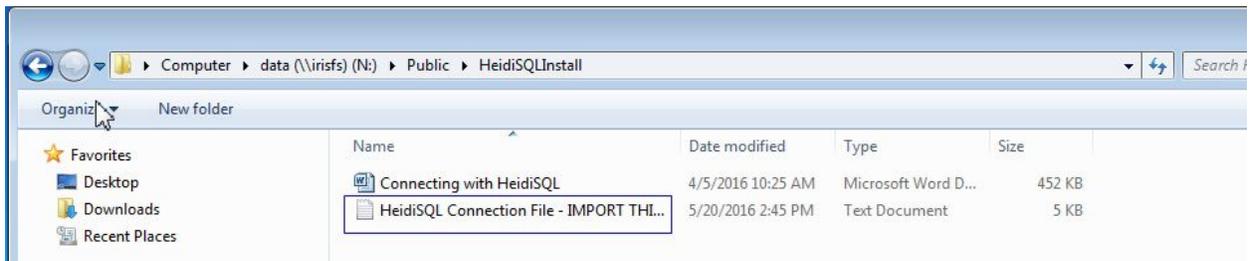
2. The first time you open HeidiSQL in the VDE you will need to configure the connection to the database. The first window you see when you open HeidiSQL is the “Session manager” window. Select the button named “Import settings file ...”



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

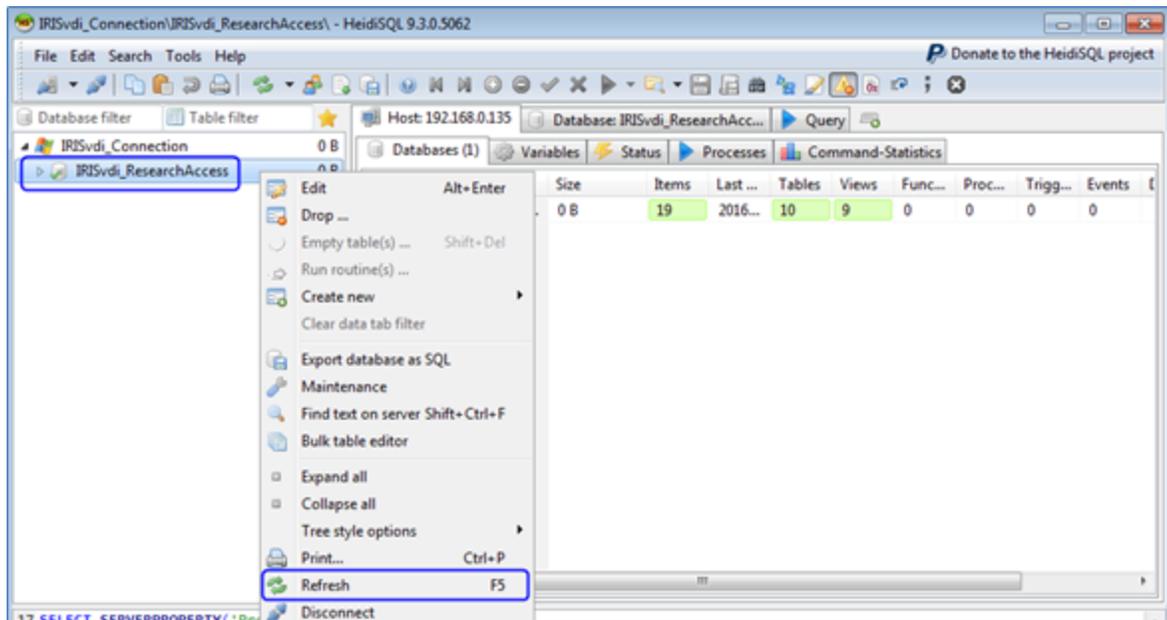


- Double click on the Heidi connection settings file. If everything works correctly you should see a “Success” message.

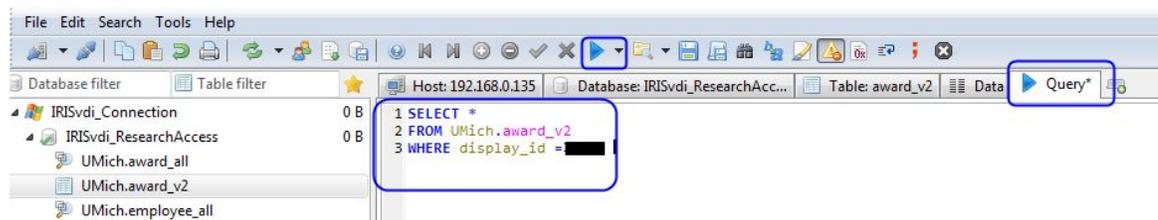


- Click on the “Session Name” and you will see the connection details. Select “IRISvdi_Connection”, “IRISvdi_ResearchAccess”, and leave everything else as is.

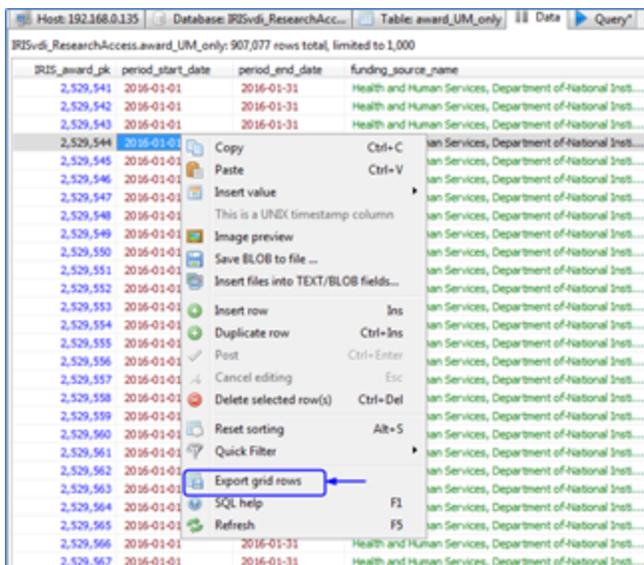
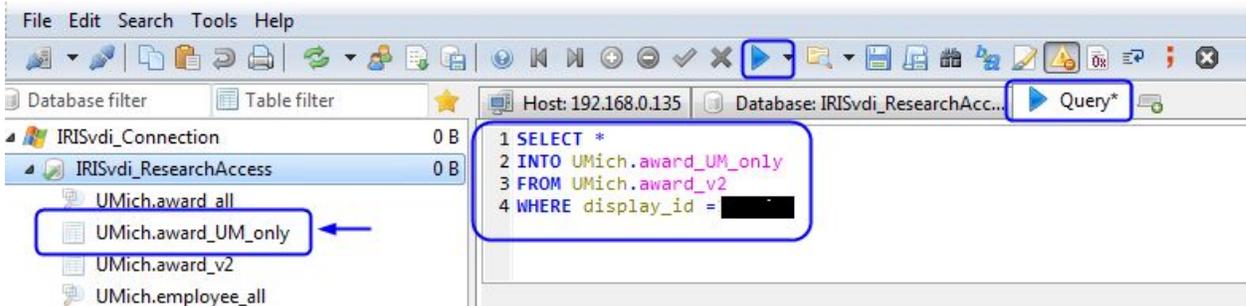
- Now just click “Open” and you will enter into your connection session with the database.
- THIS NEXT STEP IS VERY IMPORTANT. You need to refresh the table list so your schema is visible. If you do not, the queries you write will not work. Simply right-click on “IRISvdi_ResearchAccess” under “IRISvdi_Connection”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



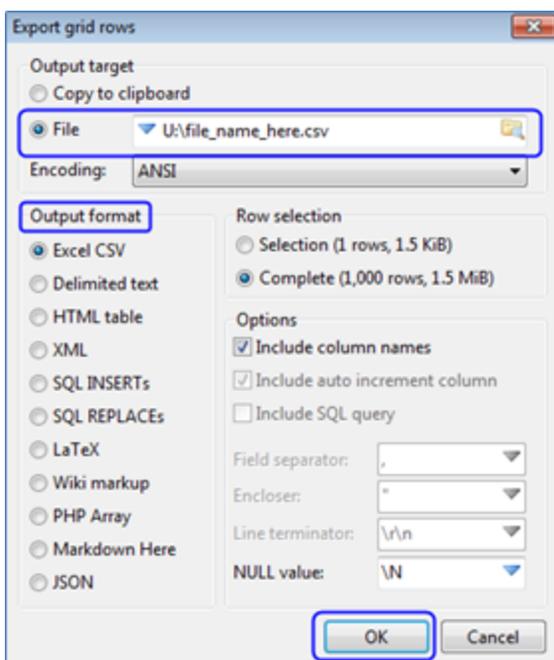
- You will now see all the tables/views you have access to and are now ready to write queries and work with the data.
- To write a query** just click on the “Query” tab. Click on the blue arrow to run your query. The result will appear below the query.



10. 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:



11. You may want to **export data as a csv** so that you can use it in other applications, like SAS, Stata, or R. Once you run the query just right-click in the resulting data grid and click “Export grid rows.”



12. 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.