

# IRIS Data FAQ

## **What kind of data does IRIS collect and manage?**

IRIS builds on, and houses, university financial and personnel administrative data provided by member institutions and other related data pertaining to scientific investments, including university data on federal awards, employee, vendor, and subaward transactions. The IRIS data infrastructure is built on UMETRICS, an initiative begun by the CIC universities in March 2013 to enhance the data that was produced for the Federal STAR METRICS program.

## **What kind of data does IRIS curate and share?**

IRIS receives microdata from member universities that contains individual and business names and partial birth date information (month and year) so that award, researcher, and vendor information can be linked to research activities and outputs such as journal articles, citations, patents, and new business startups. These data are then de-identified prior to being made available for researcher access. Access to Census-linked IRIS data is possible only through the [Federal Statistical Research Data Center System](#). That access is administered by the [U.S. Census Bureau](#).

## **Which universities and which years are represented in the IRIS data?**

IRIS has over 30 member universities and the membership is quickly growing. The temporal coverage of the data submitted by member institutions varies from one or two recent years to over a decade.

## **Who has access to the data housed in the IRIS repository?**

Our goal is to enable the research community to access data as easily as possible, subject to privacy and confidentiality restrictions. Thus, public elements will be made available. Researchers from all disciplines are encouraged to apply for access to a restricted, de-identified IRIS dataset.

## **Why is the de-identified IRIS dataset restricted use?**

IRIS acknowledges that because of the granularity of the data, secondary reidentification of people and institutions is a concern. We therefore restrict user access and enforce data use terms and conditions in order to meet the confidentiality and privacy protections we promise to our member universities.

## **What is the IRIS Enclave and Virtual Data Enclave (VDE)?**

IRIS operates a virtual secure data enclave for researchers to access restricted use data. The VDE provides a point of access to the requested data elements and is a virtual machine launched from the researcher's own desktop but operating on a remote server. VDE users are asked to follow best practices in restricted-data use as outlined in the IRIS VDE Acceptable Use Policy.

### **How can VDE users export research output from the VDE?**

Research results generated through the analysis of IRIS data can only be exported from the IRIS 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 the VDE to ensure that individuals and organizations cannot be reidentified from the aggregated datasets.

We ask researchers who wish to export analysis results and/or code to fill out a *Disclosure Review Request form* to initiate review. Requests may take up to 15 business days, depending on the type of output. IRIS encourages researchers to produce tabular data outputs (and charts and figures) for disclosure review. If the researcher is producing non-tabular data products for publication, the disclosure review time may be longer than 15 business days. Once approved, the requested output files will be securely transferred to an outgoing folder for researchers to collect with a remote server using SFTP (SSH File Transfer Protocol).

### **What kind of software, packages, and libraries are installed on IRIS virtual machines?**

All virtual machines are loaded with Windows 7 and the following software, packages, and libraries: Microsoft Office, Stata 14, SAS 9.4, R / RStudio, MATLAB, LaTeX, HeidiSQL, MS SQL Server Management Studio 2014, Gephi, Cytoscape, QGIS, GRASS GIS, Adobe Acrobat Pro,, Notepad++, Python, Anaconda, Jointpoint, PuTTY, WinSCP, and TightVNC Viewer. Researchers can contact IRIS for any questions concerning existing software or to request the installation of additional applications in the VDE.

### **What type of workspace is provided to approved VDE users?**

Each user will be given a home folder in the IRIS network drive so that one can save work generated in the VDE. Upon request, IRIS can also provide a shared space for a research team.

### **Can approved researchers access to VDE through a Mac computer?**

Yes. Mac users are able to access the IRIS VDE as long as they have the most recent version of Microsoft Remote Desktop application (version 8.0.33 as of July 2016).