

IRIS Data Access 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 STARMETRICS 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 50 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.

What steps are involved in the IRIS data access application process?

Researchers can download an application form and submit the form to IRIS. Within two weeks, an applicant will be notified of the review outcome and asked to review and sign the [IRIS Restricted Data Use Agreement](#). This [diagram](#) demonstrates the procedure.

Why is the de-identified IRIS dataset restricted use?

IRIS acknowledges that because of the granularity of the data, secondary reidentification is a significant 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.

Can researchers apply for IRIS data access without IRB project approval?

All applicants are required to attach the IRB determination letter from a sponsoring institution when applying for access to the IRIS data for research purposes. Our repository IRB approval by the University of Michigan requires that we obtain IRB documentations from research users.

What is the IRIS Enclave and Virtual Desktop Infrastructure (VDI)?

IRIS operates a virtual secure data enclave for researchers to access restricted use data. The VDI 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. VDI users are asked to follow best practices in restricted-data use as outlined in the IRIS [VDI Acceptable Use Policy](#).

How can VDI users export research output from the VDI environment?

VDI users should follow IRIS instructions (link) for preparing aggregate data files that will undergo IRIS disclosure review. This review is designed to eliminate the possibility of reidentifying an individual or organization from the output dataset. All VDI users must then submit a request to the IRIS Enclave representative to export research output from the VDI environment. Once approved, files will be securely transferred from IRIS Dropbox incoming to outgoing folders 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 VDI.

What type of workspace is provided to approved VDI users?

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

Can approved researchers access to VDI through a Mac computer?

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