Introduction to Minerva Restricted Cluster (Minerva-Res)

Minerva Scientific Computing Environment

https://labs.icahn.mssm.edu/minervalab

Yiyuan Liu, PhD
The Minerva HPC Team

July 17, 2026

Icahn
School of
Medicine at
Mount
Sinai

Outline

- What is Minerva-Res?
- Getting Started: Project Allocation and User Account
- How to Login and Use Minerva-Res
- Data Migration During Transition Period
- Data Transfer After Transition Period
- Network Access
- Other Services: Open OnDemand, Data Archive
- Security Compliance

What is Minerva-Res?

A secure and dedicated HPC cluster designed for controlled data for Mount Sinai researchers.

- It is operated using NIST SP 800-171 Rev. 3 as the baseline security framework.
- There is no direct connection between Minerva and Minerva-Res.

Approved for handling:

1. Data from NIH controlled-access repositories
2. Centers for Medicare & Medicaid Services (CMS) Research Identifiable Files (RIF) and Limited Data Sets (LDS) data paired with a Minerva HPC Data Management Plan (DMP) Security Assessment Questionnaire (SAQ)

For complete details, see the Minerva-Res documentation
https://labs.icahn.mssm.edu/minervalab/documentation/minerva_restricted_cluster/

Minerva-Res Resources

System Specs

- **2×** login nodes
- **6.9 PB** encrypted GPFS storage at /arionencrypt
- **1.5 TiB** memory per compute node for a set of nodes (expandable)
- **GPUs** to be added
- **TSM** archival storage system on login nodes

These services are NOT available — please plan your workflows accordingly:

- × Database server
- × Web server
- × Posit Connect
- × Globus
- × On-the-fly scripts (Jupyter, RStudio IDE, Ollama)

Minerva-Res: Project Allocation

1. Allocation Request

- An active **Data Use Agreement (DUA) managed by the institution GCO** subject to NIST SP 800-171 is required.
- PIs submit **Minerva Restricted Cluster: Project Resource Allocation Request Form**.

2. Policies Agreement

- PIs review & agree to **Annual Minerva Restricted Supercomputer Policies and Procedures**.

  → *After logging in, navigate to **Minerva-Res Forms**, and select **Minerva Restricted Supercomputer Policies and Procedures**.*

Changes & Deactivation

- **Modifications:** Use your return code to log back into the same REDCap to make updates.
- **Deactivation (PI Responsibility):** occurs when the DUA expires, the project completes, or the institutional affiliation terminates.

Minerva-Res: User Account

PREREQUISITE | A project allocation has been created

1 User Request

Users review & agree to the **Annual Minerva Restricted Supercomputer Policies and Procedures**

2 PI Approval

PIs approve the user request to access Minerva-Res

NOTE

External accounts are not currently supported due to background screening security requirements.

For permission/access to a specific project directory, please submit an **HPC ticket**.

Logging In

Once your Minerva-Res access is granted, log in as follows:

Connection endpoints

Campus network or VPN

- **minerva-res.hpc.mssm.edu** — Round-robin redirect **• (recommended)**
- **ci06e01 / ci06e02.hpc.mssm.edu** — Specific login node (reconnect a screen session)

1 Connect

SSH to a login node using an endpoint above.

2 Password

Enter your Sinai School Password only.

3 MFA

Complete the Multi-factor Authentication (MFA) on your mobile device.

```
SHELL

ssh yourUserID@minerva-res.hpc.mssm.edu

(yourUserID@minerva.hpc.mssm.edu) Password: <Enter your Sinai School Password Only>

Enter Your Microsoft verification code<609017>
```

Minerva-Res Storage

- Storage is organized into the `$HOME` directory and one or more `Projects` directories.

Storage Type	Path	Quota
Home Directory	`/arionencrypt/users/<userid>`	30 GiB, 1M files
Project Data	`/arionencrypt/projects/<projectid>`	Per project allocation
CMS Project Data	`/arionencrypt/cmsdata/<projectid>`	Per project allocation

Check Current Usage & Quota

`du -sh /arionencrypt/users/<userid>` (for Home directory)

`df -h /arionencrypt/projects/<projectid>` (for project directory)

`df -h /arionencrypt/cmsdata/<projectid>` (for CMS data)

Software Environment

Module System

Software Availability

Software is available through the read-only Lmod module system at `/hpc/packages`.

Loading Applications

Use the module command to load an available application(s): `module load <software_name>`

Documentation & Support

For complete details, see the Minerva software documentation:
https://labs.icahn.mssm.edu/minervalab/documentation/software-environment-lmod/

Requesting Software

To request additional software, send in an HPC ticket: hpchelp@hpc.mssm.edu

Running Jobs (LSF Scheduler)

LSF Scheduler

Workloads are still managed by the IBM Spectrum LSF scheduler. All jobs must be submitted to the scheduler.

For complete details, please visit https://labs.icahn.mssm.edu/minervalab/documentation/lsf-job-scheduler/.

Job Queues

Queue	Description	Max Walltime
Express Queue	Rapid turnaround job	12h
Premium Queue	Normal submission queue	144h

Interactive Jobs

Suitable for real-time debugging, testing, and active development.

Command:

```
bsub -q express -W 1:00 -P
project_id -Is /bin/bash
```

Note: SSH X11 forwarding (-XF) is disabled.

Data Migration During Transition Period

ACTION REQUIRED

Who must act?

Researchers currently working on NIH controlled-access genomic data on the Minerva Cluster. This includes PIs and team members working under an active Data Use Agreement (DUA) managed by the institution GCO, subject to NIST SP 800-171.

- Specifically, PIs who signed the REDCap form with GCO: **"Agreement to use Minerva for federally designated controlled, unclassified genomic data sets."**

Data Migration During Transition Period

Minerva to Minerva-Res Migration window: 07/15 - 08/26 (6 weeks)

Minerva arion file system and home directory is available in **read-only mode** for data migration. Please migrate all controlled data and supporting scripts needed.

Phase 1: Before August 26th (Migration & Cleanup)

1. Users may migrate data from the Minerva project `/sc/arion/projects` and home directories `/hpc/users/<userid>` to the Minerva-Res storage under `/arionencrypt`.
2. Users must remove all NIH controlled-access data subject to NIST 800-171 completely from the standard Minerva cluster.
3. **PIs must formally attest that all controlled data has been fully removed from Minerva.**

Phase 2: After August 26th (Compliance & Restrictions)

- **Compliance Requirement:** No NIH controlled-access data is allowed on Minerva.
- **TSM Archived Data:** Please do not retrieve Minerva-Res archived data directly to Minerva.

Minerva-Res Data Transfer After Transition Period

Core Security Policies

Strict Restrictions

Standard SFTP/SSH and general internet access are **disabled by default**.

Zero Leakage Policy

Do **not** copy/paste, download, screenshot, or screen record data to unapproved environments.

Encryption Required

All approved data transfers must use cryptographic tools during transit.

Data Derivatives**

Any data derived from controlled-access datasets must be treated with the exact same level of protection (unless explicitly approved for release by the PI as non-controlled data)

** Data Derivative: Data derived from controlled-access datasets obtained from NIH-designated data repositories. Examples of derived data include imputed datasets and single nucleotide polymorphisms, or any data explicitly designated as Data Derivatives by NIH.
See NIH page at
https://grants.nih.gov/policy-and-compliance/policy-topics/sharing-policies/accessing-data/certification-agreement and
https://grants.nih.gov/news-events/nih-extramural-nexus-news/2025/05/protecting-human-genomic-data-when-developing-generative-artificial-intelligence-tools-and-applications

Minerva-Res Data Transfer After Transition Period

Transfer Summary — Data Classification & Transfer Procedures

Data Classification	Transfer Route	Required Action & Approvals
Non-Controlled Data (Standard & CMS)	**Data In (Ingress)**	**HPC approval**.
	Data Out (Egress)	**PI approval** required.*
Controlled Data (Standard & CMS)	**Data In (Ingress)**	• **CASE 1:** Complete the **External Connectivity Form** for direct data download from a remote server. • **CASE 2:** Otherwise, submit an **HPC ticket** for guidance.
	Data Out (Egress)	**STRICTLY PROHIBITED.** Requires DUA amendment & institutional review. Contact: **hpchelp@hpc.mssm.edu**

*** PI Responsibility**

PIs are responsible for the classification of the data and for approving that the data to be released is non-controlled and complies with the applicable Data Use Agreement (DUA) and institutional policies.

Non-Controlled Data Transferred Out of Minerva-Res

Non-controlled data may be transferred from Minerva-Res to Minerva Arion with PI approval.

1. Place your files in the designated staging directory:
   `/arionencrypt/projects/<project_res>/from-arionencrypt-to-arion`
2. Submit a **data transfer request**.
3. The PI will be notified to review and approve the request.
4. Upon PI approval, HPC staff will transfer the data to the approved Arion staging area.
5. Move the transferred files from the destination staging directory into your working directory.

*** PI Responsibility**

PIs are responsible for the classification of the data and for approving that the data to be released is non-controlled and complies with the applicable Data Use Agreement (DUA) and institutional policies.

Data Transfer for CMS-Related Projects

Key points

- **Same procedure —** staging, request submission, and approval are identical to the standard non-controlled data process.
- **Different directories —** dedicated Data In (Minerva Arion) and Data Out (arionencrypt) locations, separate from the standard project paths.
- Use the CMS-specific paths from the documentation — not the standard project staging paths.

Full instructions

For the complete procedure and exact CMS staging directory paths, see the Minerva-Res Data Transfer documentation:

https://labs.icahn.mssm.edu/minervalab/documentation/minerva_res_data_transfer/

Controlled Data Transferred into Minerva-Res

CASE 1:

Download Controlled Data from Approved NIH Repo

1. Complete the **External Connectivity Request Form**
2. All requests are subject to security review and approval based on research justification
3. Once approved, the requested IP will be added to the allow-list
4. Perform the approved data ingress activities. Access will be removed once the transfer is complete.

CASE 2:

Otherwise, please contact **hpchelp@hpc.mssm.edu** for data upload guidance.

Controlled Data Transferred Out of Minerva-Res

PROHIBITED

Export or transfer of controlled-access data from the Minerva-Res environment is not permitted — unless expressly authorized under the applicable Data Use Agreement (DUA) and approved through the institutional review process.

REQUEST A REVIEW

Email the HPC team to begin the institutional review process:

hpchelp@hpc.mssm.edu

External Network Access

Login and compute nodes have no external network connectivity. To request access:

- Complete the **External Connectivity Request Form**
- The HPC team will evaluate the research justification and security requirements
- Approved IP sources will be whitelisted for the designated time frame

`module load res_proxies` (to access whitelisted URLs)

Open OnDemand

Web Portal Access

A dedicated Open OnDemand portal provides browser-based access to applications: https://ondemand-res.hpc.mssm.edu/

Available Applications

Access standard interactive workloads and environments directly from your browser, including:

- **Desktop & Jupyter**
- **RStudio**
- **SAS & Stata**

Important Restriction

Data transfer (upload or download) is disabled in the Open OnDemand Restricted environment.

Documentation

For additional information and step-by-step guides, see the Open OnDemand documentation page:

https://labs.icahn.mssm.edu/minervalab/documentation/open-ondemand/

Long-Term Data Archive: TSM

Direct Node Access

TSM clients can be accessed directly on Minerva-Res login nodes.

Documentation & Support

For complete details, please visit the official documentation page:

Access TSM with Command Line Documentation

TSM Archived Data

Please do not retrieve Minerva-Res archived data directly to Minerva.

Security & Compliance

Minerva-Res is a lockdown environment designed to protect controlled-access data and maintain NIST SP 800-171 compliance.

Access & Network Controls

Minimum Necessary Principle: Users must only access data specifically required for their approved research.

Network Tunnels: Port forwarding is disabled. Bypassing security controls, creating unauthorized network tunnels, or using unapproved proxies is forbidden.

Monitoring & Logging

Session Monitoring: Sessions that remain idle will be automatically terminated after 1 hour.

Command Logging: All command activity is logged to central security platforms (Graylog/SIEM) for security auditing and system health monitoring.

Violations of compliance may result in immediate access revocation per institutional policy.

Acknowledgement

NIH Grant Support

Minerva is supported by the Clinical and Translational Science Awards (CTSA) Grant No. **UL1TR004419** from the National Center for Advancing Translational Sciences, National Institutes of Health. Citing this grant number in Minerva-enabled publications is required.

Contact & Support

Got a question? Need help?

Send an email to the Minerva ticketing system:

hpchelp@hpc.mssm.edu

Thank You!

NIH AI Guidance

Crucial Rule: When using AI models with controlled-access human genomic data, researchers are responsible for complying with all applicable institutional and funding agency policies.

Key Protections: The NIH Guide Notice

The Genomic Data Sharing (GDS) Policy sets strict boundaries on generative AI tools (see NIH Guide notice):

- **No Sharing with Public AI Tools:** The GDS Policy does not allow sharing controlled-access data or their data derivatives with unauthorized users (including public generative AI tools).
- **Model & Parameter Prohibitions:** Sharing generative AI models and their parameters trained on human genomic data obtained from NIH controlled-access repositories is strictly prohibited.
- **Future Frameworks:** The NIH intends to release future guidance outlining the responsible use and sharing of generative AI models based on controlled-access data.

Actionable Takeaways

Do Not Upload — Do not upload human genomic datasets to public chatbots or external APIs.

Secure Local Environments — Utilize secure local environments (such as the centrally managed, local `minerva-llm` module) to ensure data never leaves the secure boundaries of the Mount Sinai HPC cluster. See **minerva-llm module Documentation**.

Using minerva-llm Module on Minerva-Res

Security Mandate: The **minerva-llm** module allows researchers to run approved AI models locally within the Minerva environment, ensuring that controlled-access data never leaves our secure boundaries.

1. Environment & Proxy

Load the specific restricted proxies to establish secure outbound access for setup.

```
# 1. Load restricted proxy
module load res_proxies/1
# 2. Load Miniforge
module load miniforge3/26.1.1-3
```

2. Conda Isolation

Create environment & block external unapproved user packages for strict isolation.

```
# Create & activate env / Prevent user-site load
conda create -n llm-res-env python=3.12 -y && conda activate llm-res-env
export PYTHONNOUSERSITE=1
# Install DL core
conda install pytorch transformers accelerate sentencepiece protobuf wrapt -y
```

3. Execute Module

Load centrally managed model configurations and execute scripts locally.

```
# Load LLM module & verify
module load minerva-llm/1.0 && echo $MINERVA_LLM_MODELS
# List approved models
minerva-llm-list
```

Using minerva-llm Module on Minerva-Res

4. Python Implementation Example

Create an execution script (e.g., example.py) to run local model inference:

```
# example.py
import os
from transformers import AutoTokenizer, AutoModelForCausalLM

# Access models folder via environment variable
model = os.path.join(os.environ["MINERVA_LLM_MODELS"],
"TinyLlama-1.1B-Chat-v1.0")
print(f"Loading model: {model}")
tokenizer = AutoTokenizer.from_pretrained(model)
llm = AutoModelForCausalLM.from_pretrained(model)
```

Important Considerations

Login Node Thread Limits

If running on login nodes, restrict CPU threads to avoid resource limits (libgomp errors) by setting the environment variables before execution:

```
# Restrict thread counts
export OMP_NUM_THREADS=1
export MKL_NUM_THREADS=1
export OPENBLAS_NUM_THREADS=1
export NUMEXPR_NUM_THREADS=1

# Now execute your inference script
python example.py
```

Non-Controlled CMS Project Data

DATA IN (Ingress) — Arion ➔ Minerva-Res

1. Stage Files: Place files in your pre-created staging folder:
`/sc/arion/cmsdatastaging/from-arion-to-arionencrypt/<your_userid>`

2. Submit Request: Submit a **data transfer request**.

3. HPC Execution: The HPC team runs the secure transfer. Files arrive at:
`/arionencrypt/cmsdata/<yourproject_res>/from-arion-to-arionencrypt/`

4. Verify & Move: You must verify data integrity and move files immediately out of staging to your working project directory.

Documentation

For detailed information, please refer to the documentation page:
https://labs.icahn.mssm.edu/minervalab/documentation/minerva_res_data_transfer/

Non-Controlled CMS Project Data

DATA OUT (Egress) — Minerva-Res ➔ Arion

Requires PI approval to verify non-controlled status.

1. Stage Files: Place files in:

`/arionencrypt/cmsdata/<project>/from-arionencrypt-to-arion/`

2. Submit Request: Submit a **data transfer request**.

3. Transfer: Upon PI approval, HPC staff transfers files to:

`/sc/arion/cmsdatastaging/from-arionencrypt-to-arion/<userid>`

Documentation

For detailed information, please refer to the documentation page:

https://labs.icahn.mssm.edu/minervalab/documentation/minerva_res_data_transfer/

Controlled CMS Data

DATA IN & OUT (Ingress / Egress) — Controlled CMS Data

DATA IN (Ingress):

- Submit an HPC ticket specifying **Controlled CMS Data Ingress**.
- The HPC team will verify your CMS DUA permissions and coordinate the transfer directly.

DATA OUT (Egress):

- **Strictly Prohibited** unless explicitly authorized under the CMS DUA and approved by a formal institutional review.
- To initiate a review, submit an HPC ticket.

Documentation

For detailed information, please refer to the documentation page:
https://labs.icahn.mssm.edu/minervalab/documentation/minerva_res_data_transfer/