# The HORUS Project White Paper

**Version 1.0, July 25, 2024**

Muhammad Akhdhor[1], Shawn McKee[1], Bob Stovall[2], Pierrette Renee Dagg[2], Andrew Keen[3], Michael Thompson[4], Rob Thompson[4], Matthew Lessins[4], Michael Parks[5], Wendy Dronen[1], Shereen Ismail[2]

[1] Physics Department, University of Michigan, Ann Arbor, USA
[2] Merit Network, Ann Arbor, USA
[3] Institute for Cyber-Enabled Research, Michigan State University, East Lansing, USA
[4] Computing and Information Technology Department, Wayne State University, Detroit, USA
[5] Network Planning, Michigan State University, East Lansing, USA

Contact E-mail: smckee@umich.edu

Logo: The logo is a combination of the Egyptian word Horus and the theme of helping researchers. The triangle represents an Egyptian pyramid while also representing the joining of the three entities behind HORUS. Having the letter R as the largest letter at the peak of the triangle symbolizes the most important part of the HORUS project - the researchers. Finally, the yellow and blue shapes reflect a simplified image of the Egyptian god Horus. Horus was said to be the Egyptian god of the sky whose right eye was the sun. Note Horus was the son of Osiris.

# Abstract.

The HORUS project combines existing large-scale scientific data storage from the previously funded NSF OSiRIS project with a broad range of computational resources to enable accelerated scientific discovery for universities, colleges and community colleges in Michigan and the surrounding regions. The project focuses on providing easy access to diverse computing and storage resources, both of which are required for scientific research and analysis. This will be especially beneficial for researchers at smaller institutions, enabling them to more effectively contribute to society by expediting their research. In addition, through its connection to OSG and the PATh project, it will serve as an on-ramp to even larger scale resources across the nation when that is needed.

The project establishes a set of compute-servers to complement the existing OSIRIS storage system attached to the MERIT statewide research and education network in Michigan. The system consists of 2 GPU nodes (each with two A100 GPUs), 6 Large Memory compute nodes (each with two AMD Epyc 7F72 3.2 GHz 24 core processors, 1TB ram), 6

compute nodes (each with 2 AMD Epyc 7H12 2.6GHz 64 core processors, 512G ram) and 7 hybrid-memory-compute systems (ThinkSystem SR665, each with two AMD EPYC 9654 96C 2.4Ghz core processors, 1.5T ram). These resources and associated services provide computational services to a suite of universities including University of Michigan, Wayne State University, Oakland University, Eastern Michigan University, and Oakland Community College.

# 1. Introduction

Funded by the National Science Foundation, HORUS (Helping Our Researchers Upgrade their Science) is a collaboration of the University of Michigan, Michigan State University, Wayne State University, and Merit Network. It builds on the work done under the OSiRIS project[1], also funded by the National Science Foundation, to provide large-scale scientific data storage. It takes advantage of the high-speed research network created for OSiRIS, as well as Merit's state-of-the-art fiber-optic network.

By combining OSiRIS storage with HORUS processing power, more of Michigan's scientific minds can realize the potential of their ideas. It is free to use for members of public higher education institutions, those participating in NSF funded science projects and others involved in research and education.

# 2. The HORUS Project Vision

HORUS (Helping Our Researchers Upgrade their Science) is a collaboration of scientists, computer engineers and technicians, network and storage researchers and information science professionals from University of Michigan/ITComm (UM), Michigan State University/iCER (MSU), Wayne State University (WSU), and Merit Networks. HORUS is the high-performance computing fast lane for Michigan universities and community colleges. Researchers and students can tap into computing power typically only available to the state's top research universities.

HORUS is working to:

- Provide computing power to under-resourced researchers
- Accelerate scientific discovery at universities, colleges, and community colleges in Michigan and surrounding regions
- Give easy access to the diverse computing and storage resources needed for scientific research and analysis
- Remove barriers to the open science grid and serve as an on-ramp to national resources
- Introduce students to high-performance computing, data science, and artificial intelligence
- Support collaboration among community colleges with certificate programs in data science
- Encourage collaboration between research universities and the broader research community
- Promote new course development

# 3. HORUS Project Components

The HORUS team has extensive experience operating cyberinfrastructure for researchers. Besides the work on OSiRIS, team members have designed and operated infrastructures such

as the ATLAS Great Lakes Tier-2 for high-energy physics[2], the ICER infrastructure at Michigan State University[3], and Wayne State University's high performance computing center[4].
The HORUS team made use of available open source software, typically deployed in grid sites and data centers. The system was developed with researchers in mind:

- Easy to submit work and track jobs
- Fair sharing of resources
- Dynamic partitioning of resources to suit varying job sets
- Resource accounting and system metrics

## 3.1. Architecture

The HORUS project was designed as a computing extension to the previously funded OSiRIS project. New computing resources would be acquired and added to the same racks as the existing OSiRIS storage. These new resources would be connected to the same high-performance networks and would leverage, to the extent possible, the existing OSiRIS services and applications. New HORUS users would automatically have storage areas allocated in OSiRIS and a new batch scheduler, based upon SLURM, would provide fair-share access to the set of HORUS computing resources. The following sections detail the various components involved.

## 3.2. Computing Hardware

HORUS provides three distinct types of computational nodes, co-located with existing OSiRIS storage infrastructure, well-connected to an existing 100 Gbps research network and leveraging a set of open source software to make the resource broadly available, including to researchers outside the region via the OSG/PATh sharing described above.

## 3.3. OSiRIS building blocks

The OSiRIS project assembled a distributed software-defined storage infrastructure across the primary research Universities in the state of Michigan, providing 13.5 Petabytes of raw storage accessible via Ceph[5]. The project also provided InCommon Federated identity access and provisioning workflows, allowing its users login via their institutional identity. More details are available in the OSiRIS whitepaper[6].

## 3.4. Networking Description

The HORUS network environment is built on top of what the OSiRIS project developed and leverages the excellent regional network provided by MERIT. The Merit Backbone Network is shown below in Figure 1 and illustrates the comprehensive footprint MERIT has deployed as well as the high-performance interconnects to other regional, national and international networks.

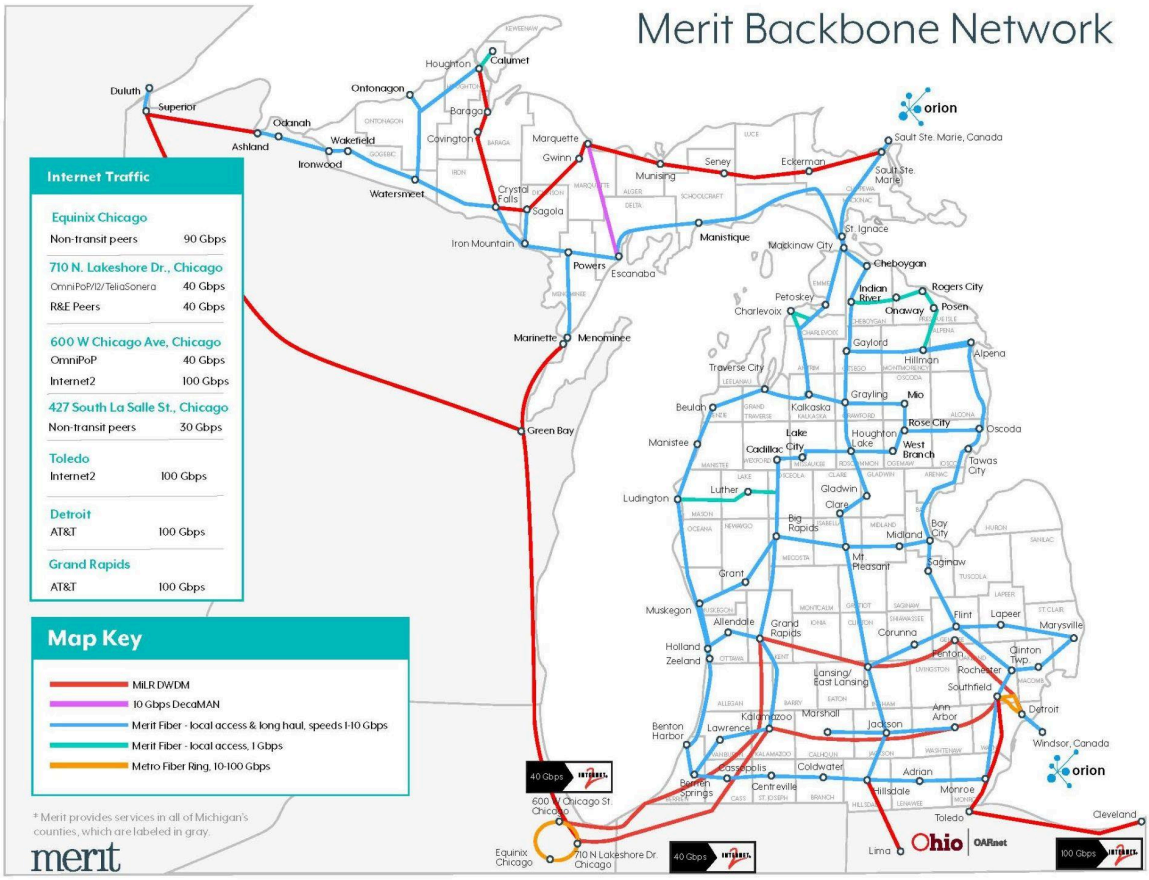


**Figure 1:** The Merit Backbone Network as of summer 2024 showing the link capacity and coverage across the region.

A high-speed research network was created as part of the OSiRIS project (see Figure 2 below). This network, combined with Merit's regional network, will support access to HORUS compute resources by educational institutions throughout Michigan and beyond. This network is highly resilient and features multiple 100 Gbps links internally and 100 Gbps connectivity to Merit and the wide-area network.

Merit's core strength is enabling researchers, educators, and learners to transfer critical data across its high capacity optical network in a safe, secure environment. Merit's network is designed with resiliency within its core network to our campuses and out to the national and global research communities. With Merit's next-generation network, HORUS creates a high-capacity compute research platform across three campuses and makes it available to researchers throughout Michigan and the region.

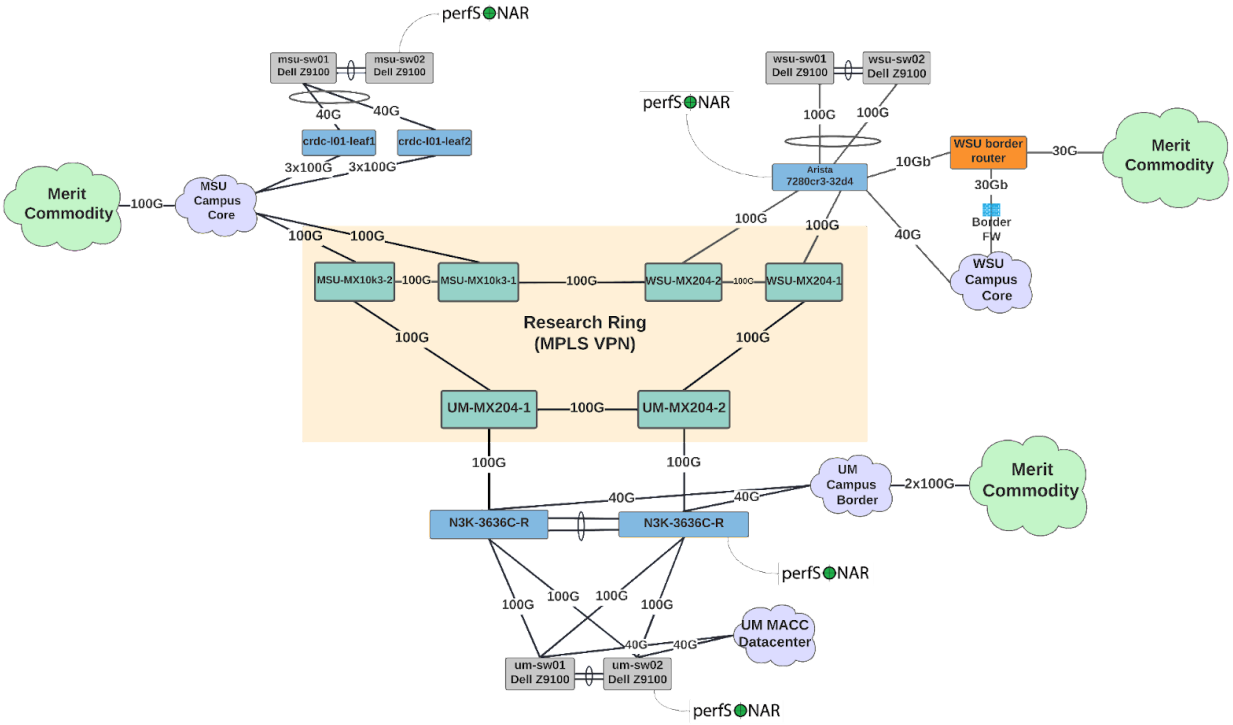


**Figure 2:** The logical diagram of the OSiRIS research network which interconnects the HORUS computing infrastructure at Michigan State University, the University of Michigan and Wayne State University, providing a resilient 100 Gbps research-focused network platform. Note that HORUS equipment is directly connected to the "gray" boxes at top and bottom of the diagram (MSU, UM or WSU, e.g., um-sw01 or wsu-sw02, etc).

## 3.5. Services, Authentication, Authorization and Accounting

HORUS uses software to adhere to its guiding principles of fairly sharing resources among users and maximizing the use of resources. The HORUS software architecture builds on a number of open source tools and applications, grouped by their particular role, many of which were inherited from the prior OSiRIS project.

### Authentication and Authorization

- **InCommon**, a set of community-designed identity and access management services
- **CoManage**, identity lifecycle management
- **Grouper** for creating and managing roles, groups, and permissions
- **CILogon**, for logging on to cyberinfrastructure

### Resource allocation and management

- **HTCondor-CE**, a meta-scheduler used as a "door" to a set of resources.
- **SLURM**, used to schedule jobs and interface with Open OnDemand (OOD).
- **NVIDIA MIG**, used to subdivide a GPU (cores and memory) into up to seven smaller instances, allowing more jobs to share a GPU.
- **Ceph**, with quotas to manage storage use in OSiRIS.

**Monitoring and Accounting**

- **Opensearch**, gathers data from syslogs and other sources for aggregation, visualization, analytics, and correlation.
- **CheckMK**, intelligent server and host monitoring system capable of validating service states and tracking resource usage.
- **perfSONAR[7]**, used to test and monitor network behavior across infrastructure.
- **AlmaLinux9**, for accounting and auditing to augment usage and security information.

**User Tools and Software**

- **OOD**, Open OnDemand which provides a user web interface to HORUS resources.
- **EasyBuild**, to allow users to build the software they need
- **EESSI** (European Environment for Scientific Software Installations), to provide access to prebuilt applications in common use.

Deployments of many of these tools (CoManage, Grouper, Elasticsearch, CheckMK, perfSONAR) already were in place for OSiRIS and have been reconfigured to accommodate HORUS. HTCondor is the HORUS batch scheduler, used with OSG/PATh to configure the appropriate connection to users' hosted Compute Element (CE) based upon HTCondor-CE. The HTCondor services, as well as all other required HORUS services, will rely on the virtualization platform already in place for OSiRIS. This virtualization platform includes four powerful virtualization hosts at each of our sites (UM, MSU, WSU) running libvirt. In addition, we have access to SLATE infrastructure, which provides the ability to orchestrate containers via Kubernetes if particular tools or jobs would benefit from that.

# 4. Science Domain Engagements

The HORUS project has a number of science domain engagements with different histories as we describe in the following sections.

## 4.1 Inherited Science Collaborations from OSiRIS

The precursor to HORUS had a number of science domains that utilized OSiRIS's storage to make their data shareable and accessible. The domains below also were interested in getting access to computing resources that could be used to process, filter or transform their data. The HORUS project inherited them as collaborators during its startup.

Two groups at Oakland University in Rochester, Michigan are leveraging OSiRIS storage for their research. The Battistuzzi research lab focuses on long-term evolutionary patterns of microbial life, and the OU Genomics group aims to use and promote next-generation

sequencing and bioinformatics technologies for research and education at the OU Biological Sciences Department. More Information

Brainlife.io: An online platform to accelerate scientific discovery by automated data management, large-scale analyses, and visualization. Brainlife plans to switch over to OSiRIS as their primary archival storage system before the end of this year. No single computing resource has enough storage capacity to store all of their datasets, nor is it reliable enough so that users can access the data when they need them. Brainlife will rely on OSiRIS to store archived datasets and transfer data between computing resources. More Information

Building on existing collaboration between MSU and the Grand Rapids based Van Andel Institute, OSiRIS has installed NVMe-based Ceph OSD nodes and an NFS gateway at the institute to enable direct access to bioinformatics research data. OSIRIS at VAI will enable VAI bioinformaticians to work with MSU researchers to better understand Parkinson's disease and cancer. OSiRIS facilitates data access for VAI researchers to leverage the computational resources at MSU Institute for Cyber Enabled Research. More Information

U.S. Naval Research Lab is collaborating with researchers at U-M to share their high-resolution ocean models with a global community. This unclassified data was stored on US Navy computers that were not easily accessible to many researchers. OSiRIS enables scientists worldwide to leverage these models. More Information

The ATLAS Event Service is designed to leverage object stores like OSiRIS for fine grained physics event data which can be retrieved and computed in small chunks and leverage transient compute resources. More Information

Homed at U-M Institute for Social Research, the NIH 5 year project 'Effect of the Placental Epigenome on Stunting in a Longitudinal African Cohort' uses OSiRIS to store and share data to a wider community. More Information

At WSU the Microscopy, Imaging & Cytometry Resources (MICR) core

leverages OSiRIS to enable wider access to imaging data.

The JETSCAPE collaboration at WSU is an NSF funded multi-institutional effort to design event simulators for ultra-relativistic heavy-ion collisions. OSiRIS provides a universal storage platform for collaborative access.

Global Nightlights at U-M is the only complete archive of NOAA nighttime imagery from 2 different satellite programs: DMSP (1993-2016) and VIIRS (2012-ongoing). By keeping portions of this archive on OSiRIS we enable wider usage of the datasets by researchers outside the institution.

## 4.2 HORUS Project Planned Science Collaborations

The following projects have driven the development of HORUS and point to the types of use cases others may apply. A summary table also defines the computing details and requirements for each of these projects.

**Material Simulation — University of Michigan**: Explores material properties by running material simulation applications using so-called integral files (typically 1-2 TB each) to do computation. With recent GPU optimization, the bottleneck for certain calculations is now completely related to the I/O efficiency, especially when one considers disorder (where hundreds of 1-2 TB files are used in each simulation). For HORUS, it will be interesting to see how these codes perform with quick read access to multiple such files in OSiRIS storage. This discipline is well matched to HORUS GPU use in parallel with the OSiRIS storage.

**Pritzker Neuropsychiatric Disorders Research Consortium**: The Pritzker Consortium aims to 1) discover neurobiological and genetic causes of major depression, bipolar disorder, and schizophrenia 2) identify biomarkers for better diagnosis and novel targets for their treatment. The Consortium involves researchers from several institutions including Cornell, Stanford, UC Irvine, HudsonAlpha Institute for Biotechnology, and the University of Michigan. This area of work has significant computational and data challenges.

**Van Andel Institute**: Van Andel Institute hosts a busy genomics facility that is used heavily both internally and by neighboring regional collaborators in West Michigan (such as Michigan State University). Last year nearly $1 million of mainly genomics services was provided to external partners (mainly MSU). This effort represents dozens to hundreds of TBs of unique data to be analyzed and transformed, which is a challenge to move and even harder to analyze. The analyses required are each a special challenge, requiring a mix of central VAI computation mixed with those of the individual customers (researchers).

**Cryo-EM** — University of Michigan: Single particle cryo-electron microscopy allows scientists to determine the atomic structure of biological macromolecules. Given that the 3D structure of a protein determines its function, understanding the structure of proteins allows structural biologists to understand basic biology, evolution, human health, and disease. At UM, computing

and storage is a huge bottleneck for cryo-EM researchers on campus. Currently, each lab that uses cryo-EM instruments needs to accommodate their own storage and computing, which means each lab has local workstations. A centralized compute-resource and data storage would be a boon for the campus. Existing solutions are not optimal for a large number of users. On campus, more than 30 laboratories are using the UM Cryo-EM resource.

**Evolutionary Genomics** — Oakland University: This discipline explores deep evolutionary histories (e.g., early life evolution) which require large-scale simulations and analyses of empirical data. These kinds of analyses use Maximum Likelihood and Bayesian approaches that require extensive computational time, even in a parallelized environment. Additionally, the reconstruction of ancient phylogenetic events are highly dependent on assumptions and priors given at the start of the analyses. Extensive hypothesis-testing to determine the robustness of results to these assumptions is required. Additional centralized computational resources and data storage are key to continue this line of research at Oakland University and to grow its reach over time (e.g., ancestral state reconstructions).

**Network Telescope — Merit**: Network telescopes, or darknets, consist of networking instrumentation that receive and record unsolicited internet traffic destined to large swaths of unused IP space, covering about 500,000 contiguous IPs. Internet traffic destined to this “dark IP space” is inherently suspicious since no real user services are hosted in the IP space monitored by the darknet. Thus, darknets provide a unique vantage point to computer networking and cybersecurity researchers for studying internet-wide scanning activities, distributed denial of service attacks, misconfigurations, and even network outages. Merit operates one of only two large network telescopes in the country and maintains an archive of darknet data dating back to 2005. Currently, Merit’s darknet collects more than 100 GB of compressed PCAP data daily, and the archive exceeds 345TB. Processing the telescope data is challenging because of the large amounts of CPU, memory, and storage required. Tasks include: 1) longitudinal studies, 2) performing annotations with external, third-party data sources (e.g., Censys.io, historical DNS data, etc.), and 3) employing ML/AI/statistical techniques for data clustering, predictive analytics, inference, anomaly detection, and other applications.

**Population Health Informatics — Oakland University**: This discipline involves multiple research projects using natural language processing and machine learning for biomedical and healthcare analytics. Heterogeneous datasets include massive amounts of unstructured textual data as well as alphanumeric tabular data and image data. Natural language processing is GPU-intensive; currently, the relatively few number of GPU nodes result in extended wait times for jobs in the queue to start running.

**Oakland Genomics Oleksyk Lab — Oakland University**: Genomic analysis on a population scale often operates with extensive volumes of data and requires multi-step heterogenous (cpu/IO intensive) analysis. For example, a basic population genomics dataset of 100 samples amounts to an estimated 9 TB of data and 60 billions of short unassembled reads. During the initial steps of this analysis (sequence alignment and variant calling) the total dataset temporarily triples in volume. In addition to the approaches to optimize the later downstream analysis, some of the approaches in the field (i.e. population structure using dense sequencing

data analysis) involve model-based Bayesian clustering and similar computationally intensive data. The requirements for storage, IO, and CPU increases and expands with every additional step used for comparative analyses of multiple populations that involve thousands of samples. Additional data storage and storage-lenient computational resources would allow Oakland University and our laboratory to stay current with the demands of big data analysis, enable use of the latest computational instruments, and further promote research and collaboration with other groups in the field.

**Autophagic Body Modeling — Eastern Michigan University**: Autophagic body modeling is an important tool for improving our basic understanding of the process of autophagy, a conserved intracellular recycling pathway important for human health. We measure the size and number of autophagic bodies formed in yeast with altered levels of key autophagy proteins. Accurate estimation of the original size and number of autophagic bodies from random TEM sections requires the use of a simulation to compare with the actual data. We are developing improved versions of this simulation that use CompuCell3D (CC3D) to model the bodies as more realistic non-spherical shapes. Currently, a single CC3D run takes ~1 hour on a laptop, but our automated workflow will require thousands of runs to generate enough simulated data for statistical accuracy.

**JETSCAPE — Wayne State University**: The JETSCAPE collaboration (jetscape.org) is a consortium of about 50 researchers from 14 institutions in the US, Canada, Japan, and Germany, including physicists, computer scientists, and statisticians. The collaboration develops simulation frameworks for the simulation of very high energy nuclear collisions. These collisions carried out at the Brookhaven National Lab. and at CERN produce nuclear-sized exploding droplets of matter that reach temperatures over 3 trillion degrees Celsius. The software is made available to the worldwide community via Github. However, these simulations are extremely compute-intensive. From time to time, the collaboration will compete for and receive computing allocations to carry out large-scale simulations and compare with experimental data. However, the typical allocations fall far short of the required time; hence, the collaboration has divided the simulation into several stages which can be simulated separately. The results of the earlier stages are then stored in large storage facilities such as OSiRIS where they can be retrieved to carry out the later stages of the simulations. HORUS will provide needed resources for the collaboration as well as a mechanism for broader access to the models and datasets.

**Heavy-Element Quantum Chemistry — Oakland University**: Successful heavy-element chemistry depends on the availability of powerful tools of theoretical chemistry to predict the non-trivial behavior of the elements at the bottom of the periodic table. The development of reliable infrastructure (basis sets, force fields, density functionals) for predictive relativistic quantum-chemical modeling of complex chemical systems — involving actinides, heavy main-group, and superheavy elements — relies on a readily accessible computing facility with hundreds of CPUs, including large-memory nodes, and massive and fast scratch space. The continuation of this line of research at Oakland University is predicated on sustainable maintenance and growth beyond the current high-performance computing cluster in the wake of an influx of new STEM faculty with growing demands for expansive parallel computing.

**Certificate Program in Data Science — Oakland Community College**: The Data Science Certificate is designed to provide a strong foundation for data investigation, including data wrangling, cleaning, security, regression and classification, prediction, and data communication. It will be a 35-hour program including Python, R programming for data science, machine learning, big data, and NoSQL. OCC needs easy-to-use compute, storage, and software resources to provide a broad introduction to the exploding field of data science.

# HORUS Science Driver Compute Needs

The following table describes the resource requirements that drive the HORUS computing needs. There are a diverse set of requirements that HORUS was designed to support and many of the science use-cases are complementary in terms of resource use (e.g., some require GPUs but little CPU, some require CPU and large memory, some require lots of CPU).

| Science Description & Challenges | Resource Types | Codes | Core Range | Runtime Range | Memory |
|---|---|---|---|---|---|
| **Materials Simulation — University of Michigan** : Explores materials and their behaviors by running detailed simulations requiring large input datasets and optimized use of GPU and CPU resources | GPU, CPU, I/O streaming , large inputs | Green's function codes (green.physics.lsa.umich.edu) | One iteration uses a full GPU and associated CPU core | One iteration is one hour; need 10-100 iterations to complete the work | Memory: 20-40 GB on GPU & replicated on CPU |
| **Pritzker Neuropsychiatric Disorders Research Consortium**: Works to discover neurobiological and genetic causes of major depression, bipolar disorder, and schizophrenia, and to identify biomarkers for better diagnosis | GPU, CPU, large inputs | Neuroinfo, Imaris, Amira, Metamorph, Huygens image analysis, Halo, Volocity, Zen and Arivis, Image Pro Plus, ImageJ, Ilastik,Tarastitcher, STAR aligner, Portcullis, Matlab, Python, R | Four cores using 32G | Varies from minutes to several weeks; majority of jobs are in range of several hours to a couple of days | GPU: 16G/GPU CPU 8G/core |

| | | | | | |
|---|---|---|---|---|---|
| **Van Andel Research Center**: Conducts diverse set of genomics research between the center and external collaborators; challenges include multi-TB size data and complex joint computation required to analyze the data | CPU, large inputs, large outputs | Illumina tools, Aligners, tertiary analysis tools, bio-statistical tools | 40 to 400 based upon sample sizes | Hours to days | Maximum of 256GB / server (4GB/core) |
| **Cryo-EM — University of Michigan**: Single particle cryo-electron microscopy allows scientists to determine atomic structure of biological macromolecules; currently, computing and storage is a bottleneck for cryo-EM researchers on campus. Over 30 labs that use our cryo-EM instruments need to accommodate their own storage and computing | Each project needs CPU, GPU and large input data | RELION, cisTEM | RELION: GPU jobs will use four GPUs plus any associated CPUs on the GPU node (typically 12-24 CPUs); cisTEM: 100 CPUs | RELION: 0-24 hours; cisTEM: 0-3 hours | CPU: 2-5 GB CPU RAM / core GPU: 12-24 GB / GPU |
| **Evolutionary Genomics — Oakland University**: Explores deep evolutionary histories (e.g., early life evolution) which require large-scale simulations and analyses of empirical data using Bayesian and Maximum Likelihood techniques; additional centralized computational resources and data storage are key to growing this line of research at Oakland University | CPU, large input data | RAxML, BEAST, MrBayes, IQ-Tree, MCMCTree | Single CPU per job | Runtimes of 1-30 days | 2GB / core |

| | | | | | |
|---|---|---|---|---|---|
| **Network Telescope — Merit Network**: Cybersecurity research using "darknets" record unsolicited internet traffic destined to large swaths of unused IP space, generating large amounts of data and the need for continual analysis | GPU/CPUs, storage, memory, I/O streaming data | Custom-built Go software, PyTorch, R/Python | One core per hour of data gathered | 20-30 minutes per PCAP for Darknet event extraction | 128 GB / server |
| **Population Health Informatics — Oakland University**: Multiple research projects using natural language processing and machine learning for biomedical and healthcare analytics; GPU-intensive and requiring large inputs | CPU, GPU, high IOPs, large input datasets | Python, MATLAB, R | Typical job requires a CUDA supported GPU with >= 16G RAM | Job runtimes may vary from several minutes to a few (1-2) days | >=16G RAM per GPU and 16-64G CPU RAM per node |
| **Oakland Genomics Oleksyk Lab — Oakland University**: Conducts genomic analysis on a population scale requiring extensive volumes of data and multi-step heterogenous (cpu/IO intensive) analysis | High IOPs, CPU, GPU | BWA, GATK, BEAGLE, fastPHASE, IMPUTEv2,bcftools, MACH, ShapeIT, CHROMOPAITER, fineSTRUCTURE and others | Single CPU per job but typical work needs 200 jobs | Varies: tasks from tens of minutes to 200-300 hours@128 cores for alignment and variant calling pipeline for 100 samples | Majority of the tasks 2GB/core, some tasks 4GB/core |

| | | | | | |
|---|---|---|---|---|---|
| **Autophagic Body Modeling — Eastern Michigan University**: Important tool for improving basic understanding of the process of autophagy; individual runs on a laptop take about one hour, but statistical accuracy requires thousands of runs | CPU | CompuCell3D, python | Single CPU per job but requires thousands of jobs for a study | Minimum runtime for a single simulation ~1 hour | 2-4 GB / core |
| **JETSCAPE — Wayne State University**: Develops frameworks for the simulation of very high energy nuclear collisions; extremely compute-intensive and generating large datasets, making it challenging to acquire access to enough CPU and storage for collaboration | CPU, large output datasets | Physics codes in C++, statistical analysis codes in Python; general scripting for large-scale runs in Parsl (Python) | Using a single thread on a single core, a single event would take 30 hours | 100-200M thread hours to calibrate the model; 17 unknown parameter s to set | Single event on a single thread will use a little over 2GB of memory |
| **Heavy-Element Quantum Chemistry — Oakland University**: Develops reliable infrastructure (basis sets, force fields, density functionals) for predictive relativistic quantum-chemical modeling of complex chemical systems involving actinides, heavy main-group, and superheavy elements requiring hundreds of CPUs, large-memory nodes, and large, fast scratch space | CPU, high streaming , large memory | Gaussian, NWChem, TURBOMOLE, Dirac, Columbus, EXP-T, in-house codes | 80-400 cores | Hours to weeks per job. | 192G / server but some tasks would benefit from 1TB servers |

| | | | | | |
|---|---|---|---|---|---|
| **Certificate Program in Data Science — Oakland Community College**: Introduction of HPC, machine learning, data management, working with large datasets | GPU/CPU storage, memory | Stata, R/Python, MATLAB | One to four cores, but requires hundreds of jobs for each class | Expect short runtimes for class projects | 128 GB / server |

## 4.3 Additional Science Domains

After the beginning of the HORUS grant, we created a few webinars to advertise HORUS to the broader research community in our region. This resulted in engagements with two additional science domains described below. We are continuing to seek other science domains in need of the unique set of capabilities HORUS provides.

**Large Eddy Simulation with Correction - Michigan Tech University**: Large eddy simulation (LES) turbulence models have proven themselves useful tools in modeling the incompressible Navier-Stokes equations as well as coupled problems such as fluid-fluid interaction and magnetohydrodynamic flows. However, they often prove too computationally expensive in practical applications. Large eddy simulation with correction (LES-C) is a new class of turbulence models, built on the existing LES models, which has been shown both theoretically and in simulations to resolve flows more quickly. This is done by leveraging an easily-parallelizable defect-correction loop on the modeling error. However, current research is preliminary and focuses on idealized simulations. High performance computing resources are needed to test the models on a more real-world suite of test cases and verify their utility on larger parallel architectures.

**Rusakov Group Quantum Chemistry - Oakland University**: We are a quantum chemistry research group in the Department of Chemistry at Oakland University. We develop theoretical models for heavy-element compounds and apply them to a range of systems, from small molecules to complex pharmaceuticals, materials, and exotic species of the heaviest artificial elements. In particular, we focus on quantifying the effects of the relativistic motion of electrons on chemistry. Our approaches range from established quantum-chemical methods to novel techniques theoretical chemists gratefully adopt from the condensed-matter physics community.

## 4.4 OSG-PATh Collaboration

A very important component of HORUS is enabling access to its resources for those outside the scope of the grant proposers. The NSF CC* Regional grant that funded HORUS requires that we provide access to 20% of the resources for researchers anywhere in the United States. To meet this requirement, we engaged with the OSG-PATh projects[8,9]] to enable their users to

transparently access HORUS resources. This has been a great success and we have provided significant resources for these users since January 11, 2024. Figure 3 shows the OSG-PATh resource use of HORUS.

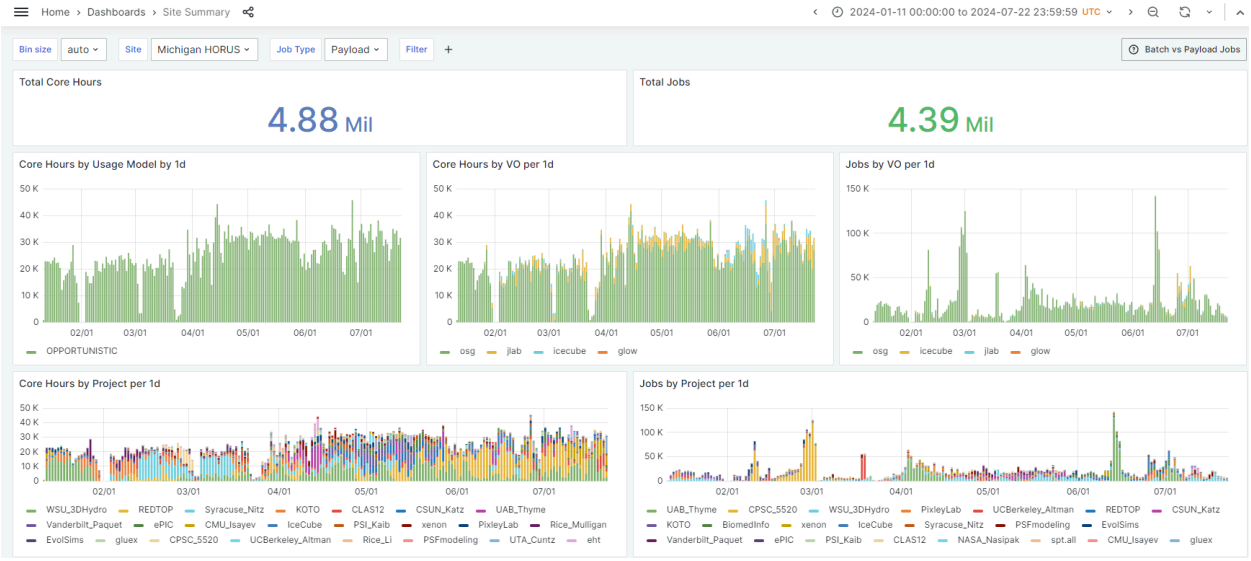


**Figure 3**: The Grafana monitoring page from the OSG-PATh project showing the amount of resources used (from January 11 to July 22, 2024) and also the types of projects and science domains using them.

The OSG-PATh Grafana monitoring page that generated the above plot is accessible at https://gracc.opensciencegrid.org/d/000000079/site-summary?orgId=1&var-site=Michigan%20HORUS&var-type=Payload&from=1704931200000&to=1721692799000 and can be customized as desired.

# 5. Outreach and Engagement

Outreach and engagement efforts for Project HORUS have included information sharing to beta users and spreading awareness of the computing resources to researchers in the region. Outreach efforts have included personal outreach, presentations, webinars, and email campaigns. A public-facing website, https://horus-ci.org[10] was created to provide a technical overview of the project, sample use cases and to house documentation. Personal outreach efforts included greater than 40 personal emails from project personnel to researchers and administrators at Michigan, Ohio, and Tribal colleges, several of which resulted in personal meetings. Project HORUS was also presented to Merit's board of directors, which is composed of CIOS from 12 of Michigan's 13 public universities. Project HORUS was presented individually to 12 public universities during annual updates with university CIOs and technical leadership teams. Presentations at public conferences included the Northern Tier Network Consortium annual meeting in July, 2023, and the Great Plains Network annual conference, both of which included outreach to Tribal communities in attendance.

Webinars named Project HOURS: Helping our Researchers Upgrade their Science, were hosted on December 13, 2022, October 25, 2023, and May 21, 2024. A total of 96 people from Michigan and Ohio institutions attended in total. Eighteen emails were developed to advertise the project and the webinar. These emails were sent to researchers and research administrators in Michigan. Mailing list sizes ranged from 11 to 949 recipients. Open rates on these emails ranged from 5% - 64%.

# 6. Next Steps and Future Plans

We are planning to work in a number of areas to improve and harden our HORUS infrastructure including:

- Continuing to tune and optimize our underlying storage to meet HORUS needs
- Monitoring and tweaking our batch system to ensure fair sharing of resources across the set of science users.
- Updating and evolving the user facing tools like OOD, Easy Build and EESSI (European Environment for Scientific Software Installations).
- Providing best-effort support for HORUS users through August 2026 and perhaps beyond.
- Maintaining the HORUS/OSiRIS hardware and software, including necessary security patches and updates.
- Transitioning the project into a campus operated service without dedicated external funding. All of our member institutions have committed to continuing HORUS if it is seen to be a useful service and, as of the end of the project in August 2024, we have secured support for continued HORUS operations from our member institutions on a best effort basis through August 2026.

We also intend to look for future relevant solicitations to take the next steps in better helping our researchers meet their scientific computing needs.

# 6. Summary and Conclusion

The HORUS project, as of July 2024, is well underway, delivering significant resources for OSG-PATh users and initial HORUS science domain users. Our intent is to operate HORUS (and the underlying OSiRIS storage infrastructure) in a best effort way for two years after the HORUS grant ends on August 31, 2024.

# 7. Acknowledgements

We want to thank the National Science Foundation for both the HORUS and prior OSiRIS grants which made all our work possible: HORUS **OAC-2232628** and OSiRIS **OAC-1541335**.

We want to acknowledge the significant support we have received from our hosting institutions including Michigan State University, the University of Michigan, Wayne State University and Merit Networks.

The HORUS Logo was designed and created by Michelle David, Communications Manager for the Institute for Cyber-Enabled Research (ICER) at Michigan State University.

# 8. References

[1] OSiRIS: a distributed Ceph deployment using software defined networking for multi-institutional research, Shawn McKee, Ezra Kissel, Benjeman Meekhof, Martin Swany, Charles Miller and Michael Gregorowicz, JJ. Phys.: Conf. Ser. 898 062045, DOI 10.1088/1742-6596/898/6/062045. Website https://osris.org/, accessed July 25, 2024.

[2] ATLAS Great Lakes Tier-2 Center, website https://www.aglt2.org/, accessed on July 25, 2024.

[3] ICER Institute for Cyber-Enabled Research at Michigan State University, website https://icer.msu.edu/citing-icer, accessed on July 25, 2024.

[4] Wayne State's High Performance Computing center, website https://tech.wayne.edu/hpc , accessed on July 25, 2024.

[5] Ceph: a scalable, high-performance distributed file system, Sage A. Weil, Scott A. Brandt, Ethan L. Miller, Darrell D. E. Long, and Carlos Maltzahn. 2006. In Proceedings of the 7th symposium on Operating systems design and implementation (OSDI '06). USENIX Association, USA, 307–320. https://dl.acm.org/doi/10.5555/1298455.1298485

[6] The OSiRIS Project Whitepaper, available on Zenodo https://zenodo.org/doi/10.5281/zenodo.12826973, DOI 10.5281/zenodo.12826973.

[7] Andreas Hanemann, Jeff W. Boote, Eric L. Boyd, Jérôme Durand, Loukik Kudarimoti, Roman Łapacz, D. Martin Swany, Szymon Trocha, and Jason Zurawski. 2005. PerfSONAR: a service oriented architecture for multi-domain network monitoring. In Proceedings of the Third international conference on Service-Oriented Computing (ICSOC'05). Springer-Verlag, Berlin, Heidelberg, 241–254. https://doi.org/10.1007/11596141_19

[8] OSG. (2015). Open Science Data Federation. OSG. https://doi.org/10.21231/0KVZ-VE57

[9] PATh Facility. (2022). https://doi.org/10.21231/k4r7-s230

[10] The HORUS Project Website, website https://horus-ci.org , accessed July 25, 2025.