Cloud Infrastructure for NextGen Water Resources Modeling

CIROH's R2OHC Cloud Platform

Arpita Patel[1], James Halgren[1], Benjamin Lee[1], Manjila Singh[1], Nia Minor[1], Harsha Vemula[1], Jeffrey C. Carver[2], Purushotham Bangalore[2], Steven Burian

[1]Alabama Water Institute, Tuscaloosa, AL, [2]Department of Computer Science, The University of Alabama, Tuscaloosa, AL

Paper Number: H41Q-1425
Thursday, December 18, 2025 | 08:30 – 12:00

ABSTRACT

➢Adopting advanced hydrological modeling frameworks requires weeks of setup, specialized compute access, and often lacks reproducibility. **CIROH's R2OHC (Research to Operations Hybrid Cloud platform)** solves these challenges for 28 consortium institutions through AWS/GCP cloud credits, free 2i2c JupyterHub, and on-premises HPC resources.

➢Key innovations include containerized workflows **(Docker/Singularity), Infrastructure as Code** and pre-configured environments—reducing setup time from weeks to 30 minutes. Results demonstrate impact: **13,000+ NGIAB downloads, daily CONUS-wide NextGen executions via NRDS, and SQL access to 2.7 million NWM forecast** points. While compute is consortium-funded, core tools remain open source, benefiting the broader water science community.

R2OHC Platform

➢CIROH's **Research to Operations Hybrid Cloud** combines public cloud (**AWS, Google Cloud, Azure**), managed services (**2i2c-JupyterHub, NWM BigQuery API)**, and on-premises **HPC (Pantarhei, Wukong)** to give researchers flexibility in choosing the right resource for each task.

➢Cloud resources scale on demand**, 2i2c-JupyterHub** provides ready-to-use environments, and HPC clusters handle computationally intensive simulations — all while maintaining reproducibility across platforms.

NGIAB +NRDS Ecosystem

- **Containerization** — Docker and Singularity package entire NextGen framework with all dependencies
- **Infrastructure as Code** — Configurations enable repeatable, version-controlled deployments
- **Pre-configured Environments** — 2i2c-JupyterHub and NGIAB images ready to use immediately+
- **NextGen simulations output on AWS** — S3 bucket

WHY IT MATTERS

- **NGIAB/NRDS** improvement could help improve NOAA operations
- Open tools and collaboration benefit entire water science community – **CIROH-UA GitHub**
- **DocuHub and Portal** – community resources

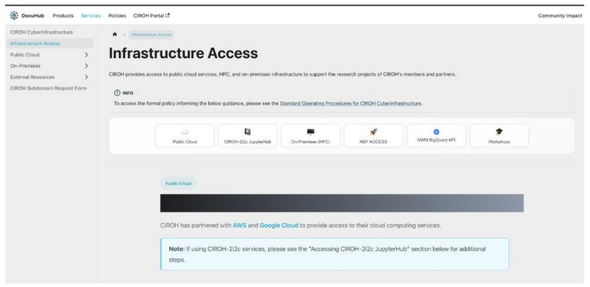
COMMUNITY & IMPACT

- **28** institutions | **405** users | **151** projects supported via R2OHC
- NGIAB: 13K+ Docker downloads | **NRDS**: 600 daily executions
- 10 peer-reviewed **publications** enabled by **CIROH cyberinfrastructure**
- 53 **public repos** | **47**+ TB cloud storage | 720 Slack members

GET STARTED

- **For NGIAB and NRDS -** https://ngiab.ciroh.org/
- **For Cyberinfrastructure access -** https://docs.ciroh.org/docs/services/access
- **DocuHub and Portal -** docs.ciroh.org, portal.ciroh.org
- **Opensource GitHub organization -** github.com/CIROH-UA
- **Contact** - ciroh-it-admin@ua.edu

PLATFORM ARCHITECTURE



Principal Investigators and researchers from 28 CIROH consortium institutions
Submit access requests via GitHub issue templates or forms at docs.ciroh.org
Choose which resources they need (cloud, JupyterHub, HPC)
Run hydrological models for their research projects

Scan to access NRDS outputs: continental-scale NextGen forecasts running daily via "CIROH's cloud infrastructure"

https://datastream.ciroh.org/index.html#v2.2

Through NSF ACCESS allocations, CIROH consortium members can obtain compute credits for Jetstream2

NSF Access
Jetstream2
Alloacations

Getting Started

Cloud Providers
AWS
GCP
Azure

CIROH's R2OHC platform leverages a multi-cloud strategy to maximize flexibility, scalability, and accessibility for the water research community. AWS serves as the primary operational backbone, providing robust workflow orchestration, event-driven compute, and scalable storage for high-volume daily model executions. Google Cloud Platform democratizes data access through managed JupyterHub environments where researchers can launch notebooks on-demand without infrastructure overhead, and BigQuery enables serverless SQL queries against petabytes of National Water Model data without requiring downloads.

Researchers can access hydrologic data through the NWM BigQuery API for programmatic queries without downloading massive datasets, 2.7 million forecast points covering catchments and flowlines across all continental US watersheds, and NRDS public archives on AWS S3 providing historical and near-real-time NextGen simulation outputs in standardized formats.

Data Access
NWM BigQuery API
2.7M forecast points
NRDS

Products
NGIAB
NextGen Research Data Stream
Tethys
TEEHR

CIROH's core product suite includes NGIAB (NextGen In A Box), a containerized Docker distribution making NextGen modeling accessible on laptops to cloud; the NextGen Research DataStream (NRDS), an automated continental-scale forecasting pipeline producing daily simulations at datastream.ciroh.org; Tethys Platform for transforming hydrologic models into interactive web applications; and TEEHR for model evaluation

Managed Services
JupyterHub
BigQuery API
2i2c

People
28+ institutions
400+ users
Workshops

HPC Infrastructure
Jetstream2
Wukong
Pantarhei

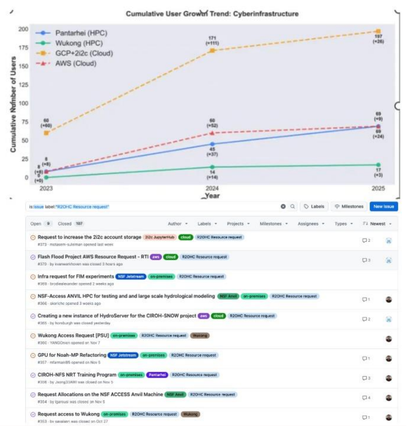
High-performance computing resources include Jetstream2, an NSF-funded national cloud resource for on-demand research computing; Wukong, the University of Alabama's HPC cluster with high-memory nodes for large-scale simulations; and Pantarhei, CIROH's dedicated on-premises system at UA optimized for hydrologic workloads with fast storage and GPU capabilities.

CIROH provides fully managed cloud services including JupyterHub environments pre-configured with hydrology packages for interactive computing, the BigQuery API enabling serverless SQL queries on petabytes of NWM data without infrastructure management,

NRDS TEEHR

**NextGen In A Box (NGIAB) is a community-accessible, containerized version of NOAA's NextGen Water Resources Modeling Framework...
No complex installation required.
Scan to learn more.**

CIROH CONSORTIUM



Learn more at DocuHub
CIROH's central documentation portal

R2OHC Platform provides researchers three pathways to computational resources:

- Public Cloud (AWS and Google Cloud at $500/month per project, plus NSF ACCESS allocations) for scalable on-demand computing;
- Managed Services (2i2c JupyterHub and BigQuery API) offering pre-configured environments and SQL access to 2.7 million NWM forecast points
- On-Premises HPC (Pantarhei at UA, Wukong at Penn State) for sustained high-performance computing at no cost.

This hybrid approach supports 151+ projects across 28 consortium institutions.

FUTURE DIRECTIONS

- Short-term priorities include expanding compute and storage across cloud and on-pre**mise HPC, integrating additional BMI-compliant models into NGIAB**, enabling multiple concurrent datastreams in NRDS, and merging DocuHub and Portal for improved usability.
- R2OHC aims to establish a **fully integrated research-to-operations pipeline** — empowering community-driven enhancements to national water prediction and making advanced water modeling accessible to all.

ACKNOWLEDGEMENTS

This Platform utilized Cloud and On-premises resources managed by **CIROH Cyberinfrastructure,** supported by the Cooperative Institute for Research to Operations in Hydrology (CIROH) with funding under award NA22NWS4320003 from the NOAA Cooperative Institute Program. The authors appreciate support from the CIROH Cyberinfrastructure team. Learn more: https://docs.ciroh.org/docs/services/intro