# Invited Applications: Office of Nuclear Physics (NP) American Science Cloud (AmSC) Data Providers Program (DaPP)

## Overview

As a response to Administration and Congressional priorities, working in coordination with the Office of Advanced Scientific Computing Research (ASCR), NP invites labs to submit ambitious pilot applications to advance the NP scientific mission through the development of artificial intelligence (AI) ready datasets and AI models that can take advantage of them. NP is soliciting two-to-three-year proposals for awards up to $5,000,000 for the performance period to provide multiple AI-ready NP datasets and documented expert knowledge that will feed domain-specific AI models. The goal is to use the power of AI to accelerate progress on addressing fundamental questions on the nature of nuclear matter posed in the 2023 Long Range Plan for Nuclear Science: A New Era of Discovery.

Successful applicants should expect to actively participate in, and be guided by, the AI Consortium established by ASCR in a separate opportunity[1] and will work collaboratively with the awardees of ASCR's AmSC opportunity[2] to help define the computing, data, and networking infrastructure needs. Applicants are strongly encouraged to review ASCR's Consortium and AmSC National opportunities and develop NP-specific applications that integrate with supporting ASCR initiatives. The ASCR-led Request For Information is expected to further define AI activities[3] and future engagement with industry.

NP invites multi-laboratory applications to develop domain-specific data standards and meta-data for multi-modal data and expert knowledge documentation. The technical expertise required for these data standards is expected to help the AI Consortium develop SC-wide AI-ready data. Multi-modal is defined here to be datasets from distinct experimental collaborations, facilities, or across Office of Science programs. Data from different generations of an upgraded experiment, facility, or from multiple NP approaches to a single technical topic will be treated with a lower priority.

Applications should establish the data to be used, the AI-driven approach, and how the project would advance and accelerate NP science beyond current capabilities. Applications should describe the plan to develop and socialize data standards with communities of interest, curate and serve the datasets including documentation, make clear the new research questions this will be used to address, and any AI R&D necessary to realize these scientific benefits.

NP expects that principal investigators of successful pilots under this program will share progress and lessons learned to avoid common pitfalls and establish AI-ready data best practices for the broader NP community.

---

[1] Anticipated ASCR-led Opportunity "AI Project Consortium".
[2] Anticipated ASCR-led Opportunity "AI Project: The American Science Cloud (AMSC)".
[3] Anticipated Request for Information (RFI) on Partnerships for Transformational Artificial Intelligence Models

Applications must establish outcomes and deliverables for each year that lead to well-defined ultimate objectives at the conclusion of the pilot period of performance. Applications are expected to be collaborative and balanced, with all participating institutions meaningfully contributing clearly delineated roles directly tied to the project deliverables.

Priority for the NP AmSC DaPP is to support data-focused activities rather than equipment or computational hardware.

Applications should identify areas of collaboration with SC partners, who may be externally funded, and potential partnership with the ASCR-led AI Consortia and AmSC.

## Example Science Cases

Two key science questions within the NP mission are targeted by this Lab call:

*How do the rich patterns observed in the structure and reactions of nuclei emerge from the interactions of neutrons and protons?*
Answering this question requires data collected and disseminated from a diverse suite of scientific instrumentation at NP-supported national user facilities and cross-cutting activities in nuclear chemistry, theoretical physics, engineering, data science, and computer science. Efforts in this domain should partner with AmSC to reset current limits on automated data collection and data management. Successful scope will make the data and metadata from complex nuclear physics experiments AI-ready for pre-training and fine-tuning of AI models. The resulting outcomes, in addition to accelerating our understanding of nuclear structure and nuclear astrophysics, will also provide curated data broadly available to the SC community through NP's PuRe nuclear data resource.

*How do quarks and gluons make up protons, neutrons, and ultimately, atomic nuclei?*
NP supports research in hot and cold quantum chromodynamics (QCD). In this science domain, datasets collected by scientific instrumentation at RHIC, CEBAF, and CERN, up to the exabyte scale in size, generate high-dimensional data requiring complex, multi-level analyses to recreate events. Much of these data are unique, as the experiments are difficult or even impossible to repeat. The dominant model at scales probing the structure of an individual proton or neutron is lattice quantum chromodynamics (LQCD), a numerical approach for solving QCD on discrete space-time grids. LQCD is essential for studying strong nuclear forces and computationally intensive. Successful scope will generate AI-ready data for the AmSC across experiments to maximize scientific insights across energy regimes and probes.

Addressing questions #1 and #2 requires unleashing the full potential of DOE-SC accelerator-based user facilities. AI-management of thousands of accelerator parameters will substantially improve operational efficiency through precise and informed adjustment of accelerator components. Meeting the experimental needs of future machines to achieve unprecedented energies and intensities requires advances in accelerator performance, driven by AI. Successful scope will develop AI-ready data standards for the autonomous control of accelerators and scientific instruments, optimizing the control of complex systems to fast-track scientific discoveries. Establishing data standards that can be broadly applied across SC

facilities is a priority, and collaboration with other SC programs is encouraged.

## Proposal Expectations and Award Details

Each DOE national laboratory with Nuclear Physics as a recognized core capability can submit one application as team-lead; other DOE labs with active NP sponsored programs are eligible to participate as partners. There are no limitations on the number of applications a lab may participate in as a partner. Within 7 days of receiving this announcement, labs should submit to NP an expression of interest (EOI) to submit an application as primary. The EOI should include: 1) the identity the principal investigator (PI); 2) a list of laboratory, U.S. university, and/or industry partners; and a one-paragraph summary that identifies the effort aligned with the science and/or technical cases listed above and the approach the team will take in generating AI-ready data.

NP will review the expressions of interest and invite selected PIs to submit an application through PAMS. The application should have a narrative no more than 5 pages in length and should include a description of the science question(s) to be addressed, the datasets to be used, work needed to prepare the data for AI training, teaming and management arrangements, and the project milestones and deliverables. Labs are encouraged to coordinate to ensure that NP receives the most competitive and scientifically ambitious applications from the most appropriate teams.

Each application is expected to include a lead and at least one partner and may include any number of additional partners as appropriate, including relevant subject matter experts from DOE laboratories, U.S. universities, and/or private companies. The lead lab is restricted from retaining more than 51% of the total budget.

Applications from a single lab with no partners and/or scope focused on a single NP experiment or facility will be considered nonresponsive.

NP anticipates making ~$10M available to these pilot efforts. NP anticipates making up to 5 awards in FY 2025 ranging from $3,000,000 to $5,000,000 with a duration of two to three years.

Applications must be received by ***September 12, 2025,*** and should be submitted following an invitation from PAMS. The NP Lab PI who is invited to submit the application should follow the instructions in the PAMS invitation email. Expressions of interest and all questions should be sent to both Paul Mantica and Sharon Stephenson (Paul.Mantica@science.doe.gov, Sharon.Stephenson@science.gov). Applications will be reviewed by Office of Science staff. Decisions are expected to be announced in September 2025.

## Review Criteria

The following review criteria and food for thought questions will be used as part of the review, as well as programmatic factors:

1. Scientific and/or Technical Merit of the Project
    a. What new capabilities will this work enable?

   b. What is the likelihood of achieving valuable results including adoption of data standards?
   c. How would the proposed work advance the NP mission?
   d. Does the Data Management Plan meet the standards of the community?
2. Appropriateness of the Proposed Method or Approach
   a. How logical and feasible are the proposed methods and approaches?
   b. Are there alternative approaches that could enable the same science?
   c. Does the applicant take advantage of community software tools to efficiently accomplish the work?
   d. Are the objectives and milestones clear and achievable?
   e. Have existing and relevant data sets been appropriately identified and are they accessible to both the team and the approach?
   f. Does the application identify possible connections with the AI Consortium?
   g. Does the application identify possible interfaces and capabilities of the AmSC?
3. Competency of Applicant's Personnel and Adequacy of Proposed Resources
   a. Does the proposed work take advantage of unique facilities, capabilities, or expertise of the participating institutions?
   b. Does the team include the necessary skills and experience to carry out the proposed work?
   c. Is the scientific and technical work appropriately distributed between the lead and partner institutions?
   d. Are there synergies with ongoing or previously supported efforts in AI and ML?
4. Reasonableness and Appropriateness of the Proposed Budget
   a. Are the proposed budget and staffing levels appropriate for the proposed work?