

OPEN ACCESS

Neutron Science TeraGrid Gateway

To cite this article: Vickie Lynch *et al* 2010 *J. Phys.: Conf. Ser.* **251** 012097

View the [article online](#) for updates and enhancements.

Related content

- [The SNS/HFIR Web Portal System – How Can it Help Me?](#)

Stephen D Miller, Al Geist, Kenneth W Herwig et al.

- [Doing Your Science While You're in Orbit](#)

Mark L Green, Stephen D Miller, Sudharshan S Vazhkudai et al.

- [The SNS/HFIR Web Portal System for SANS](#)

Stuart I Campbell, Stephen D Miller, Jean-Christophe Bilheux et al.



240th ECS Meeting

Digital Meeting, Oct 10-14, 2021

We are going fully digital!

Attendees register for free!

REGISTER NOW



Neutron Science TeraGrid Gateway

Vickie Lynch, Meili Chen, John Cobb, Jim Kohl, Steve Miller, David Speirs, and Sudharshan Vazhkudai

Oak Ridge National Laboratory, Oak Ridge, TN

lynchve@ornl.gov

Abstract. The unique contributions of the Neutron Science TeraGrid Gateway (NSTG) are the connection of national user facility instrument data sources to the integrated cyberinfrastructure of the National Science Foundation TeraGrid and the development of a neutron science gateway that allows neutron scientists to use TeraGrid resources to analyze their data, including comparison of experiment with simulation. The NSTG is working in close collaboration with the Spallation Neutron Source (SNS) at Oak Ridge as their principal facility partner. The SNS is a next-generation neutron source. It has completed construction at a cost of \$1.4 billion and is ramping up operations. The SNS will provide an order of magnitude greater flux than any previous facility in the world and will be available to all of the nation's scientists, independent of funding source, on a peer-reviewed merit basis. With this new capability, the neutron science community is facing orders of magnitude larger data sets and is at a critical point for data analysis and simulation. There is a recognized need for new ways to manage and analyze data to optimize both beam time and scientific output. The TeraGrid is providing new capabilities in the gateway for simulations using McStas and a fitting service on distributed TeraGrid resources to improved turnaround. NSTG staff are also exploring replicating experimental data in archival storage. As part of the SNS partnership, the NSTG provides access to gateway support, cyberinfrastructure outreach, community development, and user support for the neutron science community. This community includes not only SNS staff and users but extends to all the major worldwide neutron scattering centers.

1. Introduction

A gateway can bring advanced cyberinfrastructure capabilities to an entire science community. The gateway can provide an intuitive, effective, efficient organized architecture to provide an information interface to facility users while at the same time making available a large data archive (metadata and disk), high-end computing resources and, powerful analysis capabilities. The services and resources are coordinated, general purpose, and reliable. A science gateway enables scientific communities of users with a common scientific goal to use high performance computing, have a common interface, and leverage community investments. Some gateways are web portals with a web-based user interface in front and a service oriented architecture in the back. The gateway is not just a webpage. Rather it is a unified architecture for the planning and deployment of applications, services, and tools of interest to the community.

2. Gateway Support

The TeraGrid [1] is a network of high performance computers and resources at eleven different supercomputer centers supported by the US National Science Foundation. The partner facilities are connected via a dedicated wide area 10 Gigabit per second with over 1.5 petaflops of peak computing

performance across 136,740 CPU-cores. In addition non-traditional resources are also available on the TeraGrid including three specialized high-performance visualization clusters, storage services including 2.9 PetaBytes (PB) of spinning disk and 60 PB of tape archive as well as a Condor pool for capacity computing with an additional 99,000 available cpu-cores. TeraGrid resources continue to grow with additional computing, visualization, storage, capacity computing, and experimental architectures expected in the near future. Today there are over 4,000 TeraGrid user accounts for over twenty resources at the eleven facilities. Traditional high performance computing access requires accounts be created on every computing resource for every SNS proposal user. This could imply creating thousands of accounts in a few years of SNS operation.

A solution to this problem is a science gateway. Instead of multiple accounts, a community user account with a community certificate is used for access to high performance computing. A “community user” is a dedicated account identity that stands for the often routine resources usage requests that would come from a specific scientific community. Examples could be web-services providing genomic searches, a general chemical modeling and simulation capability, or a general, repeatable workflow required as part of a facilities support of its experimental users. With a gateway, only one community account is needed on each distributed resource. The users do not have a login shell for the community account but rather use the community account as a role based access mechanism. The gateway developers take care of account management, accounting, certificate management using a community certificate, application code deployment on parallel processors, and user support.

Neutron Science TeraGrid Gateway [3] access is available via the Neutron Scattering Portal [2]. It is one of over 31 existing TeraGrid science gateways. It is charting new territory for wide cyberinfrastructure services such as federated user management, callback end-user identification with automatic service auditing across federated identity space gateway community accounts, grid interoperability, and participation in attribute-based authorization. The gateway user has privileges to run a limited set of neutron science application codes via the community account on the TeraGrid computers. The use of the Jimmy Neutron community account is a success of NSTG because it enables researchers to do their work by lowering the cost of entry to do high performance computing. These jobs are launched using Globus [4] commands from the portal. Input from the portal to the TeraGrid and output from the TeraGrid to the portal is done automatically using the GridFTP [5] servers deployed over a dedicated 10 Gigabit per second TeraGrid network. NeXus [6] which is based on hdf5 [7] is the standard format used in the portal. There is also a visualization capability in the portal including many features designed by the ISAW team [8].

3. Simulations

NSTG is coupling neutron science and cyberinfrastructure with McStas [9] simulations (Figure 1) that use the TeraGrid’s parallel computing resources. These simulations have the ability to simulate a neutron scattering experiment including the entire instrument and the sample material inside the instrument. They require accurate neutron scattering simulations of realistic experimental arrangements including all items in the beam scattering path (containers, cryostats, neutron optics ...) as well as modeling the chosen sample under study. It is necessary to accurately simulate neutron scattering off of all these materials. Theory driven simulation allows comparison of the number of neutrons that hit the detector with the experiment with few or no free parameters to empirically fit.

This approach is new for the neutron community and opens up new vistas for experiment planning and

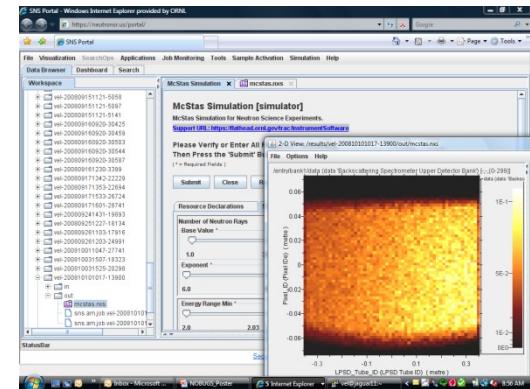


Figure 1: View of the portal showing simulation submission and visualization based on ISAW [8] of resulting NeXus [6] file.

experimental analysis. Experiments can be computationally designed and evaluated to test feasibility and optimization at the conceptual stage. It will also allow experimenters to “practice” in order to be better prepared to respond flexibly when the actual experiment is underway during the valuable allocated beam time, such as exploring observational optimizations beforehand. Simulations also allow more detailed analysis of experimental results by testing hypotheses about sample structure and dynamics to develop explanations of experimental results, for example by using a non-physical partial model to understand the significance of different sources of signal to the measured scattered neutron counts.

4. Fitting Service

Fast analysis of the data obtained from SNS instruments is critical for quick data comparison to resolution-convolved models and for making automated on-the-fly experimental adjustment/control possible. Collaborating with SNS instrument scientists, the Neutron Science TeraGrid Gateway team at ORNL has been designing and developing software tools, which use parallel computing techniques to perform fast data fitting (Figure 2) on the TeraGrid resources. The fitting service does a fit of theoretical models to the data files from the experiments. The user decides which model to use, which fitting algorithm to use, which parameters to vary, and gives initial guesses to the fitting parameters. An adaptive nonlinear least squares algorithm is implemented in parallel on the TeraGrid and shows linear scaling at least up to 32 processors. It minimizes a nonlinear sum of squares using an analytic or numerical Jacobian matrix. The fitting service uses the NL2SOL [10] or Dakota [11] software packages.

5. Data Replication

One need of the SNS is to create replicated copies of its data archive both for resiliency of availability and advanced data placement to support close co-ordination with other computational tasks such as simulations and remote visualization. We have created a re-usable workflow to transfer data from SNS record copies to other, different locations. The first goal is to replicate the data to the HPSS storage system at the Leadership Computing Facility. This workflow includes moving the data, maintaining and updating archive status while observing proper data access controls. Follow-on tasks include replicating to other storage targets and technologies such as the OIC at ORNL, HPSS at Indiana, global file systems and logistical networking. Replication plans must address data policy and data ownership. The same person must own the data as it crosses into different user authentication zones such as the TeraGrid. Also privacy of the data must be maintained using data encryption.

6. Conclusions

A discipline specific Science Gateway can help facilities scale to a large number of users, give facilities access to high performance computing such as the TeraGrid, and enable a scientific community to use community software through a common interface. Thus researchers are more productive with a gateway because they use the same tools, use a common data format, and share data easily.

Closer integration of the SNS and the TeraGrid is the goal of this work. In the future, we plan for the NSTG to flow streaming data from SNS for data reduction workflows on the TeraGrid and move

data from the SNS data repository for further processing and analysis on the TeraGrid. Also universities with SNS proposals will use the high bandwidth interconnect of the TeraGrid to support data movement between SNS and their institution. Real-time simulations of experiments can use TeraGrid resources to predict experimental results as they are running or between experiments.

7. Acknowledgements:

We are thankful to the McStas developers, Emmanuel Farhi, Kim Lefmann, and Peter Willendrup, for their on-going collaboration with us and Mark Hagen from SNS for his help. We also would like to acknowledge our SNS-SCG collaborators including Shelly Ren, Jens Schwidder, and Jim Trater. Finally we need to acknowledge NSTG staff including Greg Pike and Dave Giles.

This material is based upon work supported by the National Science Foundation's Office of Cyberinfrastructure TeraGrid project (OCI-0525789 and OCI-0504074) as well as TeraGrid resource allocations under the following NSF programs: Partnerships for Advanced Computational Infrastructure, Distributed Terascale Facility (DTF) and Terascale Extensions: Enhancements to the Extensible Terascale Facility under current TeraGrid project allocation, TG-DMR060000T.

Prepared by Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6285, managed by UT-Battelle, LLC for the U.S. Department of Energy under contract number DE-AC05-00OR22725.

Notice: This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting this article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

References

- [1] Catlett C , et al 2007 "Teragrid: Analysis of organization, system architecture, and middleware enabling new types of applications" *Advances in Parallel Computing* (Amsterdam: IOS Press)
- [2] Website for Spallation Neutron Source information <http://neutronsr.us>
- [3] Cobb J W, Geist A, Kohl J A, Miller S D, Peterson P F, Pike G G, Reuter M A, Swain T, Vazhkudai S S, and Vijayakumar N N 2007 *J. of Concurrency and Computation: Practice and Experience* **19** 809-826
- [4] Foster I, "Globus Toolkit Version 4: Software for Service-Oriented" 2006 *IFIP International Conference on Network and Parallel Computing*, (New York: Springer-Verlag LNCS) 3779, 2-13
- [5] Allcock W, Bresnahan J, Kettimuthu R, Link M, Dumitrescu C, Raicu I, and Foster I "The Globus Striped GridFTP Framework and Server" 2005 *Proc. SC05*
- [6] Website for NeXus information <http://www.nexusformat.org>
- [7] Website for hdf5 information <http://www.hdfgroup.org/>
- [8] Mikkelsen D J, Mikkelsen R L, Worlton T G, Chatterjee A, Hammonds J P, Peterson P F, and Schultz A J, "Coordinated, Interactive Data Visualization for Neutron Scattering Data" 2002 *Proc. NOBUGS*
- [9] Willendrup P, Farhi E, and Lefmann K 2004 *Physica B* **350** 735
- [10] Dennis J E, Gay D M, and Welsch R E, 1981 *ACM Transactions on Mathematical Software* **7** 369-383
- [11] Eldred, M S, Adams B M, Haskell K, Bohnhoff W J, Eddy J P, Gay D M, Griffin J D, Hart W E, Hough P D, Kolda T G, Martinez-Canales M L, Swiler L P, Watson J P, and Williams P J, 2007 "DAKOTA: A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis. Version 4.1 Users Manual" Sandia Technical Report SAND 6337