Louisiana EPSCoR
EXPERIMENTAL PROGRAM TO STIMULATE COMPETITIVE RESEARCH VOL.9 NO.9 JULY 2013

Data Mining Sessions Reap Wealth of Knowledge

Collaborative research creates very large and rich data pools, and storing and sharing this data can be the key to reaching research goals and leveraging innovative break-throughs.

To facilitate the storage and access of the enhanced data developed by Louisiana researchers, leading international scientists and practitioners in the field of semantic sciences convened in Baton Rouge to share their expertise.

The Louisiana Alliance for Simulation-Guided Materials (LA-SiGMA) team partnered with the LSU Center for Computation & Technology, the Environmental Molecular Sciences Laboratory of the Pacific Northwest National Laboratory, and LSU High Performance Computing to host a two-day workshop on Semantic Physical Sciences on June 7-8, 2013, at LSU.

LA-SiGMA is building the next generation of experimentally validated formalisms, algorithms, and codes for multiscale materials simulations with the goal of implementing them on supercomputers and educating our future highly skilled workforce of materials scientists and engineers.

The overarching goals of the workshop were to gain the collaborative knowledge needed to develop a stronger, long-term data management plan for the LA-SiGMA research consortium, and to provide further leadership for campuses as they enhance their data planning.

Participants were part of several lectures, collaborative discussions and hands-on training sessions regarding:

- Data and knowledge management and representation.
- eResearch tools and the use of data mining for the investigation of physical science data.

Dr. Nico Adams brought his wealth of knowledge about materials and manufacturing informatics from Australia's National Laboratory, Commonwealth Scientific and Industrial Research Organization.

Dr. Adams provided an introduction to linked data and ontologies, including Resource Description Framework (RDF) and Web Ontology Language (OWL) and their relationship to other semantic web technologies.

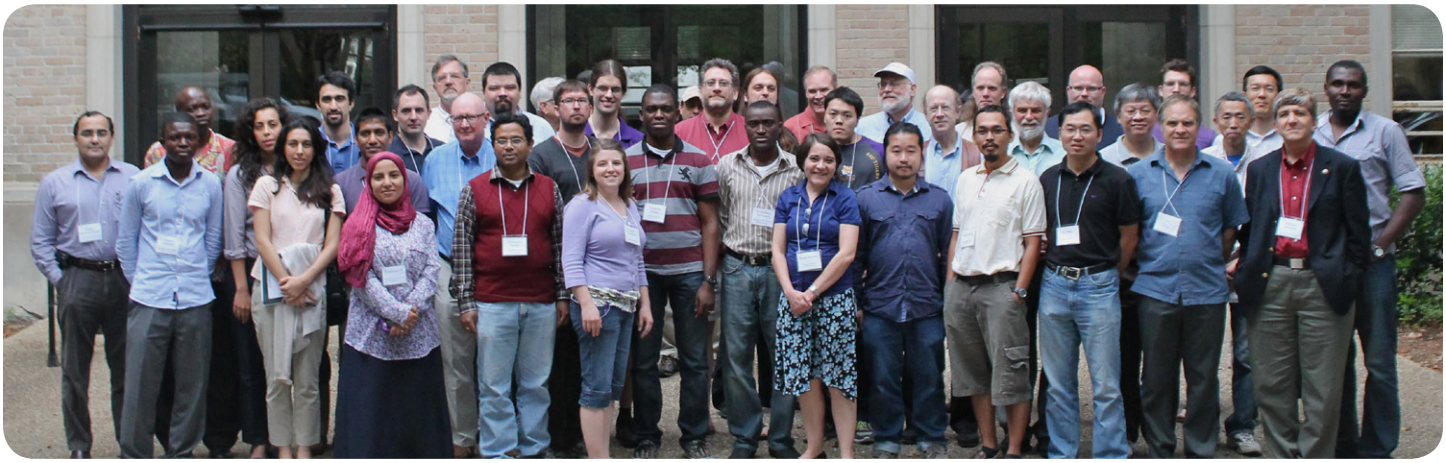
Dr. Adams specializes in web ontology, the study of how entities can be grouped, related in a heirarchy and subdivided according to similarities and differences. He led a hands-on practical session where participants constructed a small ontology to learn some of the features of OWL, how to reason with ontologies and how to use them to mark up data.

"The ability to extract information from large data sources, create semantically rich data and data ontologies has the potential to revolutionize all areas of inquiry currently underway at the Center for Computation & Technology."

- Dr. Mark Jarrell, Professor, Louisiana State University, Department of Physics & Astronomy

The second part of the session utilized a "hackathon" format for participants to use their own data to start constructing ontologies to describe it.

A confluence of a number of trends described by Dr. Adams are set to deeply change the practice of science, in particular:

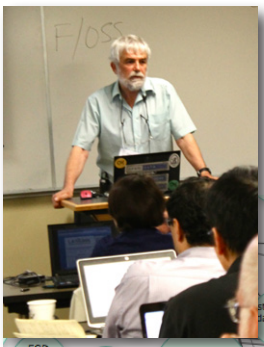


(Above) Data Management Workshop participants and speakers at LSU.

continued on page 2

- The web has established itself as the central, scalable platform for the production, exchange and dissemination of information and knowledge units.
- The beginning development of a data commons.
- Increasing political pressure and will to make open access data a reality.

"The combination of these trends is beginning to provide unique opportunities to integrate, query and derive new knowledge from disparate data sources. Linked data and ontologies will play a crucial role in this," said Adams.

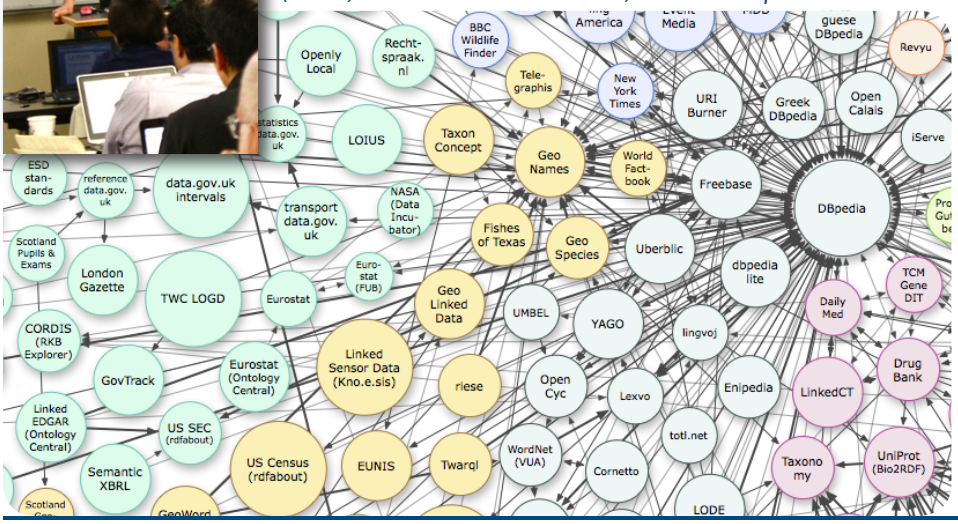


LA-SiGMA is in the process of contracting with the Texas Advanced Computing Center (TACC) in Austin, Texas to provide project data management and storage for the project. Mr. Christopher Jordan, TACC's data management lead provided an introduction of the center's capabilities to workshop participants.

"TACC is taking a leading role in providing national-scale data cyberinfrastructure in support of open science, with tens of petabytes of online storage capacity and hundreds of petabytes of offline archive," said Jordan.

"Our combined resources and expertise will facilitate coordinated data management across the project and will vastly expand the available resources for all project members, especially those at smaller institutions."

(Left) Molecular informatics researcher Dr. Peter Murray-Rust.
(Below) Zoom of one of his visuals, the Linked Open Data cloud.



WORKSHOP SPEAKERS

Dr. Nico Adams
Research Scientist, Materials Science and Engineering
Commonwealth Scientific and Industrial Research Organization (CSIRO), Canberra, Australia

Mr. Dave Cowley
Senior Research Scientist, WR Wiley Environmental Molecular Sciences Laboratory
Pacific Northwest National Laboratory (PNNL), Richland, Washington

MD. Bert de Jong
Lead, NWChem High Performance Software Development Group
PNNL, Richland, Washington

Dr. Marcus Hanwell
Technical Leader, Open Chemistry Scientific Computing Group
Kitware, Inc., Clifton Park, New York

Mr. Christopher Jordan
Manager, Data Management & Collections Group
Texas Advanced Computing Center, University of Texas, Austin, Texas

Dr. Peter Murray-Rust
Reader in Molecular Informatics and Senior Research Fellow
University of Cambridge, Cambridge, England

Dr. Neil Ostlund
President and CEO
Hypercube, Inc., Gainesville, Florida

Dr. William Shelton
Associate Director, WR Wiley Environmental Molecular Sciences Laboratory
PNNL, Richland, Washington

Louisiana EPSCoR
P.O. Box 3677
Baton Rouge, LA 70821-3677
#4475