MAGIC Meeting Minutes
March 2, 2016

Attendees

Rich Carlson	DOE/SC
Devarshi	LBNL
Shantenu Ja	Rutgers U
Dan Katz	NSF
Ken Klingenstein	I2
Padma Krishnaswami	FCC
Peter Lyster	NCO
Rajiv Ramnath	NSF
Alan Sill	TTU
Jack Wells	ORNL
Dean Williams	LBNL

Action Items

Proceedings

The meeting was chaired by Rich Carlson, DOE and Dan Katz of the NSF. Rajiv Ramnath discussed the NSF programs for Software Infrastructure for Sustained Innovation (SI)2 and Dean Williams discussed the Earth System Grid Federation.

The MAGIC members noted that Dan katz will be leaving the NSF at the end of the month and expressed their thanks for the vision and leadership Dan has provided as CoChair of the MAGIC Team for several years. We wish him well in his new position at the University of Illinois, Champagne/Urbana. The MAGIC members also welcomed Rajiv Ramnath of the NSF and expressed their interest in having Rajiv CoChair the MAGIC Team when Dan moves on.

Software Infrastructure for Sustained Innovation (SI)2: Rajiv Ramnath, NSF

NSF takes a comprehensive view of cyberinfrastructure motivated by research priorities and the scientific process. The cyberinfrastructure (CI) must accommodate the recent revolution in the scientific workflow to many interfaces to shared services. It must also support advanced CI to accelerate discovery and innovation, coordinating data, networking and cybersecurity, high performance computing, software, and people.

CI software (including services) is essential for most science thie software must be sustained after it is developed. Many software programs are cross-cutting and within directorates other than CISE (e.g., biological informatics, geoinformatics, BIGDATA). ACI provides a sustainable software-enabled ecosystem for advancing science by:

- Supporting scientific software research and development
- Supporting foundational CI research and development,
- Influencing community policies and environment for sustainability of the ecosystem
- Helping develop a trained workforce

The flagship program is Software Infrastructure for Sustainable Innovation (SI)2 which provides:

- Elements support: $500K/3 years
- Frameworks: $1M/year for 3-5 yrars

- Institutes: $3-5M/year for 5-10 years

SI2 priorities include:

- Multidisciplinary and omni-disciplinary software building on other NSF programs
- Techniques, tools, and processes for rapid integration of software
- Embedded innovation and research
- Security, trustworthiness and reproducibility
- Sustainability
- Science-inspired education
- Comprehensive metrics

Broader software challenges include funding models, international funding, career paths for software-focused researchers, incentives, training, interdisciplinary work, portability, and dissemination.

Example projects include:

- A Next-Generation Open-Source Computational F;luid Dynamic Code for Polydisperse Multiphase Flows in Science and Engineering: Alberto Passalacqua, Iowa State Un.
- Adding Research Accounts to the ASSISTments' Platform: Neil Heffernan, Worcester Polytechnic Institute
- Distributed Workflow Management Research and Software in Support of Science: Ewa Deelman, USC and Miron Livny, Un of Wisconsin

Institutes include:

- Science Gateways Software Institute: A hub for a software ecosystem for developing science gateways
- Chemistry and Materials Research Institute

Future directions include:

- Example societal Grand Challenges
- Research needs: e.g., distributed, dynamic, multi-disciplinary collaborations
- Research processes need to be dynamic, e.g., toolboxes of composable computational research methods, workflows that adapt to data and humans-in-the-loop, security, reproducibility, and trustworthiness.
- Systems integration
- Human, computer interfaces (HCI)
- Social aspects
- Education
- Domain science

These CI software programs depend on, interact with, and contribute to significant other NSF programs.

For the complete briefing, please see:
https://www.nitrd.gov/nitrdgroups/index.php?title=MAGIC_Meetings_2016#March_2016_Meeting

Earth System Grid Federation (ESGF) : Dean Williams (ESGF Executive Committee Chair)

ESGF is the world's leading source for climate modeling data with over 4PB of data, 27,000 users, and over 2000 CMIP peer reviewed articles based on ESGF data. ESGF automates infrastructure for archiving and comparing simulation results, diagnostics, and validations. It leads the creation of a flexible, extensible infrastructure for future climate efforts. It exploits broad data sharing, and empowers integrating data streams for a virtual laboratory. ESGF currently expedites distributing over

20 PB of climate data per year worldwide (and increasing exponentially). It expedites combining model output with various types of observation and reanalysis. Data and algorithms must be tracked.

ESGF has led data archiving for the Coupled Model Intercomparison Project (CMIP) for studying general circulation model output. This spans over 62 models in over 20 countries. The World Climate Research Program (WCRP) serves as the primary coordinating body for this research. The CMIP runs are key components of the Intergovernmental Panel on Climate Change (IPCC) assessments.

ESGF is a coordinated multi-agency, international collaboration of institutions that develop, deploy, and maintain software needed for climate study. It provides a federated distributed data archival and retrieval system. The deployable software stack includes services for:

- Publishing
- Search
- Transfer
- Computation
- Resource monitoring and allocation
- Security
- User interfaces
- Exploration: remote analysis and visualization

The architecture implements modularity and abstraction as a composition of interacting software services. Each service can be evolved or replaced individually. The architecture supports distributed, dynamic, scalable, resilient, fault tolerant, and secure services. The infrastructure implements common and consistent standardization and peer-to-peer services. ESGF best practices enable transport of 10s of Petabytes of data per year currently. The immediate goal is 4 Gbps (1PB/month) with a stretch goal of 16 Gbps (1PB/week). Current ESGF work is focused on the CMIP6 data lifecycle (CMIP6 output and ESGF data publication).

ESGF also supports the Accelerated Climate Modeling for Energy (ACME) end-to-end workflow In 2016 ESGF is focused on Federated databases. In early 2018 ESGF plans to implement ecosystem data sharing. By 2022 ESGF plans to implement a virtual laboratory integrating satellite, in situ climatology, and diagnostics ecosystems.

ESGF lessons learned include:

- Constantly improve and adapt the architecture, e.g., software security, root certificate authorities policy
- Integrate already proven technologies and applications
- Promote participation and involvement
- Establish a governance model
- Avoid single points of failure
- Funding is always a struggle

For the complete briefing please see:
https://www.nitrd.gov/nitrdgroups/index.php?title=MAGIC_Meetings_2016#March_2016_Meeting

Discussion among the MAGIC members

MAGIC should discuss international privacy initiatives.

InCommon became a part of EduGAIN, going from 600 Id providers to 3000.

Stream 2016 will be held in March.

Next MAGIC Meeting

April 6, 2016, 2:00-4:00 Eastern, NSF Room TBD