



The government seeks individual input; attendees/participants may provide individual advice only.

Middleware and Grid Interagency Coordination (MAGIC) Meeting Minutes¹

October 2, 2019, 12-2 pm ET
NCO, 490 L'Enfant Plaza, Ste. 8001
Washington, D.C. 20024

Participants (*In-Person Participants)

Lisa Arafune (CASC)	Joyce Lee (NCO)*
Richard Carlson (DOE/SC)	Don Petravick (NCSA)
Vipin Chaudhary (NSF)	Nathan Tallent (PNNL)
Sharon Broude Geva (UMich)	Birali Runesha (UChicago)
Brian Lyles (ORNL)	Sean Wilkinson (ORNL)

Proceedings

This meeting was chaired by Richard Carlson (DOE/SC) and Vipin Chaudhary (NSF). September 2019 meeting minutes were approved.

Data Life Cycle Summary & Report

MAGIC will summarize the data life cycle speaker series in a brief 4 page report. An announcement for volunteers is forthcoming.

MAGIC Annual Planning Meeting (APM)

No objections to draft MAGIC APM slides, which will be presented at the Large Scale Networking APM on October 8, 2019.

FY21 proposal: Individual topics instead of multi-month series; lay out 6-8 topics and assign to someone to identify potential speakers. Updates from today's meeting in red font.

Data integrity for scientific endeavor within data life cycle context (data provenance, security)

- Data integrity, etc. within data life cycle (working on encrypted data (secure nodes); sites setting up data enclaves associated with HPC centers)
- Wide range of topic areas: from working on encrypted data (secure nodes); sites setting up data enclaves associated with HPC centers
- Potential speakers: NSF Cybersecurity Center of Excellence (Von Welch or Jim Basney)
 - Zachary Ives, Dept Chair of Computer Information Science (UPenn)– Data integration

Data confidentiality updates

¹ Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program.

- New data confidentiality projects? Note NIST proposals requesting input: Data confidentiality – identifying and protecting against breaches and how to recover from breaches
- Potential speakers: [Wu Chen \(NIST\)](#)

Implications of new AI on science work

- DOE running series on AI for science (need summary) – [check October meeting. DOE AI Town hall series](#)
- AI for optimizing HPC – work on IO systems
- Annotating data where annotation environment is difficult; multi-domain (e.g., earth science and medicine); viewed as infeasible by investigators (Don Petravick)
- [ML as new approach to statistical discovery; paths forward \(Brian Lyles\) – Conner Coley; Natalie Eyke, Klaves and Jenson \(MIT\) Autonomous Discovery in the Chemical Sciences – hypothesis vs. data driven approach \(kfjensen@mit.edu\)](#)
- [AI IWG Speaker](#)
- Potential speakers: [Geoffrey Fox \(Indiana U\)- streaming data and ML workshop summary](#)

ROI and cost efficiency for academic and lab based computing

- Potential Speakers
 - Craig Stewart (Indiana University) – calculating ROI quantitatively and financially and in human terms.
 - CASC members? CASC discussing topic of optimizing cost efficiency in delivering computing
 - [Working on deliverable for ROI; cloud ROI \(Craig Stewart, Rick McMullen and \[TBD\] re: PEARC papers. Sharon Geva & Alan Sill \(CASC working group\)](#)
 - [Cloud ROI \(Craig Stewart, Rick McMullen\); Daniel Reed \(NSF-funded project\); Kate Keahey \(ANL\)](#)

Direct integration of energy sources and computing facilities (A. Sill)

- Multi-disciplinary : integrating energy production and computing more closely. Many startups locating data centers remotely, near sources of renewable energy
- Training is requiring larger and larger machines – using HPC resources do conduct training. Training is power intensive
- Potential Speakers: Andrew Chen ([UChicago](#))- zero carbon cloud; working on feasibility of stranded power to power small data centers

Re-examine networking infrastructure underlying middleware

- Attempts to put more middleware in network layer (authentication, encryption, security)
- Major network providers could discuss what doing to support high bandwidth distributed computing
- Putting storage into network (Miron Livny)
- National Research Platform- latest developments (Kevin Thompson)

MLAI, Data science, virtualization and containerization approaches -context of workflow (Dhruva Chakavorty)

- Latest occurrences in these areas. 20 min overview from someone working in all 3 areas or workshop?

SuperComputing 2019

DOE labs will discuss their efforts to federate computing resources. Topics to be covered include: InCommon, CoManage, Jupyter notebooks, configuring machines, nuts and bolts of distributed ecosystems.

Roundtable

NSF: Vipin Chaudhary

[CSSI](#) deadline on November 1, 2019

Meetings

November 19, SC19, Colorado Convention Center, Rm 711 (1:30 – 3:30 p.m. MT); Remote dial-in will be provided

December 9-12, Los Angeles, CA, Streaming Systems and Realtime Machine Learning (STREAM-ML) Workshop" of the IEEE Big Data Conference 2019. See <http://ipcc.soic.iu.edu/STREAM-ML>

Please note the following important deadlines.

Abstracts due October 31, 2019

Submissions due October 31, 2019

Paper assignment November 1, 2019

Review due November 10, 2019

Next Meeting: November 19, Colorado Convention Center, (3:30 – 5:30 p.m. EST)