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

Middleware and Grid Interagency Coordination (MAGIC) Meeting Minutes[1]

October 7, 2020, 12-2 pm ET

Virtual

Participants

Gabrielle Allen (UI)	Gabriel Martinez (DHS/HQ)
Lisa Arafune (CASC)	Linden Mercer (NRL)
Coleen Arviso (Navajo Tech)	Lauren Michael (UW-Madison)
Jason Arviso (Navajo Tech)	Melissa Midzor (NIST)
Keith Beattie (LBL)	David Morris (NSF)
Wes Bethel (LBL)	Alex Moura (RNP)
Eric Burger (OSTP)	Edward Moynihan (UI)
Richard Carlson (DOE/SC)	Elizabeth Niswander (UI)
Dhruva Chakravorty (TAM)	Donald Petravic (NCSA)
Damian Clark (Alabama A&M)	Manish Parashar (OSTP)
Michael Corn (UCSD)	Steve Petruzza (Utah)
Jessica Diep (DOL)	Lavanya Ramakrishnan (LBL)
Martin Doczkat (FCC)	Paul Ransom (NTIA)
Sharon Broude Geva (UMich)	Jared Ribble (NavajoTech)
Dan Fay (Microsoft)	Anne Richeson (CenturyLink)
Sharon Broude Geva (Michigan)	Kamie Roberts (NCO)
Monisha Ghosh (FCC)	Stefan Robila (NSF/CISE/OAC)
Nicholas Goldsmith (NSF)	Birali Runesha (UChicago)
Nakia Grayson (NIST)	Jennifer Schopf (UI)
Raj Grewal (NTIA)	Suhas Somnath (ORNL)
Linda Hayden (SCG)	Alan Sill (TTU and OGF)
Rob Hughes (NGC)	Todd Shechter (UW-Madison)
Ana Hunsinger (Internet2)	Jason Schuette (DARPA)
Margaret Johnson (NCSA)	Ed Synakowski (UWYO)
Michael Killean (UWYO)	Kevin Thompson (NSF)
Joyce Lee (NCO)	Wendy Whitcup (USC/ISI)
Chris Lowe (USDA/ARS)	Jonathan Williams (NSF)
Brian Lyles (LBNL)	Frank Wuerthwein (UCSD)
Miron Livny (UW-Madison)	Bevin A. Zauderer (NSF)
David Martin (ANL)	

[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.

Proceedings

This meeting was chaired by Richard Carlson (DOE/SC) and Stefan Robila (NSF).

Workforce Development Speaker Series: Diversity & Inclusion

Guest Speakers:

- Jason Arviso, Vice President of Operations for the Navajo Technical University
- Damian Clarke, Chief Technology Officer, Alabama A&M University
- Kevin Thompson, Program Director, Office of Advanced Cyberinfrastructure OAC, National Science Foundation

Kevin Thompson (NSF); Campus Cyberinfrastructure (CC*) Overview

Background: 2011 ACCI Task force on Campus Bridging initially created to increase capacity to universities - connectivity to support emerging science (see Slide 1 for link)

- Concept : networking as fundamental layer and underpinning of CI, driven by scientific R&E needs
- Most awards to campus networking upgrades (10/100 Gbps),
- Goes beyond external connectivity to national R&E fabric (i.e., I2) – include re-architecture of campus border;
- prioritizes science flows through a concept that ESnet made tangible in terms of new architecture and different path for scientific data in and out of network/ campus border (see diagram)
- Invested in other forms of CI, in addition to networking

Program-wide criteria

- Emphasizes strong campus level partnerships between researchers, users, educators and campus IT leadership
- Provide new capability provided by science, engineering and education
- Campus CI plan – supplemental document of what proposing with CC*; how it fits in larger context of strategic division of CI for that campus

CC*20 -507 solicitation – good year

- Area #1 – Campus Network upgrades (original)
- Area #2- Campus network upgrade (bit newer) –Addressing lesser resourced institutions in groups
- Area #3 – networking integration and applied innovation
- Area #4- Campus Computing
- Area #5- CyberTeam- Regionally coordinated team of CI practitioners applying their expertise to campus-wide scientific activity across multiple institutions. 2020 solicitation (Details in Slide 5)
    - more anchored facilitate campus, cluster cloud and computing resources for use across science projects and collaborations
    - Address multi-institutional science-driven needs and leverage and assisting CI on how those services may be bridged.
    - Planned engagements, tangible activities – encouraged to be multi-institutional
    - Fund up to 40 time equivalents throughout the 3 years
    - Want to see leadership institutions involved to institutionalize these positions to be long-lasting

Summary numbers for CI Program 2012-2020

- 340 awards
- Categories; mostly networking
- Also cyber team/CI engineer 33 total (5 in 2020) –predecessor to cyber team category

CC* status

2020 – last year for Cyberteam. Note: Alan Sussman runs cyber training program and other related funding opportunities in NSF/OAC, but will note carry forward in CC* directly.

CC* Cyber Team Awards (2020, 2019, 2017) – most expensive category award in CC* (see slide 9)

- AIHEC (American Indian higher Education Consortium) – TTU in North Dakota, partnering with Dakota State University
- HBCU award in bottom of table – planning activity
- Fewer awards in 2019 and 2017

Quad Charts

- The Great Plains Regional CyberTeam (Slide 10): 2019 Award. Participating institutions: U South Dakota, South Dakota State
- SWEETER: South West Expertise in Expanding Training, Education and Research (Slide 11) PI: Dhruva Chakravorty (TAM) – impact beyond Texas and to Mexico and Arizona
- KyRC- Kentucky Research Computing Team. R&E drivers
- Rocky Mountain Advanced Computing Consortium built off long-standing regional computing consortium

CC* Experiences

Damian Clarke (Alabama A&M)

CC* requirements:

- Cyberinfrastructure Plan – required. So must understand current CI.
- Science drivers – How Science DMZ will help specific researchers. So must have proper and open dialogue with your campus researchers. No division that is dedicated for research at smaller campuses, so need to break down silos with researchers. Explain how grant would help support current and future research. Understand research that is performed and what type of networking infrastructure, compute and storage infrastructure, is needed.
- Prepare before grant drops in November.

Jason Arviso (Navajo Technical University)

Tribal college

- One of 37 tribal colleges nationwide. Each has challenges of addressing big data or lack of infra. Supporting big data or ability to have large network resources

NSF Tribal College of university Program –

- address infra capacity in New Mexico campus.
- Grow central STEM programs: Offer baccalaureate and some graduate degrees

Focus:

- Research because research capacity has developed.
- Addressing Lack science drivers promoting use of the requirement for large data sets to be transferred to and from locations

2 CC* grants: regional network design and small institutions.

- addressing increasing the WAN on Navajo nation to address data research and distance education challenges, and
- implementing Science DMXZ and advanced wireless testbed to deliver homework gap solutions for students and faculty at Navajo Tech
- Enabled us to reach out to 5 sister colleges in Southwest U.S.
- Tribal colleges' capacity to support and sustain that type of infrastructure is quite cumbersome. Relationship established with our mulligan AZ University helps us understand and build formidable capacity. Enables us to take upon the responsibility of maintaining and operating networks efficiently ourselves- this knowledge and ability to address these challenges has been the greatest reward. Interactions with lead institutions helped us to develop and grow

Primary goals of our grants: training, educating, facilitating collaboration to prepare tribal colleges for active participation in regional and national networking.

- This has helped improve access to national research and education research at tribal colleges.
- Helped design and implement campus networking for small institutions to support availability of participating in big data research.

Building more collaborations with tribal colleges

- Consortium's CC* award helped them host discussions across tribal colleges and the .US.
- Helps in understanding challenges and supporting each other; CC* instrumental in these internal discussions and collaboration

Knowledge of architecture and infrastructure to support big data.

- Will be hosting training on that instruction – to our SW institutions and Navajo Nation colleagues who are very interested. Hosted workshops in coordination with them

Homework Gap: how bring access to homes. COVID impacted tribal nation.

- CC* provided additional capacity; able to work with carriers and increased capacity across Navajo nation.
- Currently, coordinating to introduce resources to K-12 education. Hope to provide major loan-type network delivery into those communities to make available those R&E resources throughout Navajo Nation.

Dhruva Chakravorty (TAM); SWEETER: South West Expertise in Expanding Training, Education and Research –Co- PI Emily Hunt

Large regional collaboration extends from TX, AZ, New Mexico (including Texas, UTA, and R4 schools).

- Focus on training, research and building research collaborations; engaging folks to use CI resources.
- Extending work to community colleges. Received community building grant bringing research innovation to community colleges; will be making huge impact in TX.

- Cyber team members also funded for other CC* grants:
    - New Mexico State University funded for compute cluster.
    - West Texas A&M won award for cluster.
    - We have extended the scope of our cyber team and helping these institutions bring up their clusters and get researchers on board. Fantastic experience.

Adapting to new norm (COVID)- online research exchange.

Discussion

Open Science Grid works with CC* awardees- helps them share capacity of new clusters with externa users. Wonderful experience to work with campuses to whom we've had no visibility in the past – see what they do and what we can do to do to help. 2020 CC* awardees (approaching 20). Demonstrates what can be and needs to be done in terms of what happens on our campuses. What is impact of deploying these advance capability (networking and computing) on your ability to train students and staff that later contribute to your surrounding community? Is there value in installing this in-house instead of using remotely?

- Value in both. Some need to be installed in house while others need to be done remotely.
- Want to move from manually walking fields and recording data on paper to drone infrastructure doing spectrum and soil and plant analysis; need in house storage to store that data. But need some services like OSG. Looking at this tool to train in house researchers to use.
- On-prem resources are crucial– huge step forward in attracting talent pool. In smaller institutions, huge interest in training students. Also, the continuum is important – some should be on-prem., so should be in distributed resource. For some things, like data lake, need high availability resources that can scale internationally.
- The continuum is important, but not concerned about high end part of continuum. Putting significant national resources into the high end. Want to encourage everyone who participate and experience to send message on the importance of the low end of the continuum. And the need to place these capabilities in the campuses, in the tribal colleges. We can help connect these low end to learn more about working together and collaborating. Important that the message comes out strong from those who participate and experience the value of this in workforce development because if the students never see a computer, we will lose many opportunities to train.
- HBCU perspective: More campuses need to have the equipment and services for students and faculty to touch, learn and grow from a workforce development perspective.
- Texas: No state resources guiding compliance for colleges and small institution with Texas administrative code 202, particularly with clusters and HBCU environments. Estimates 50 community colleges noncompliant due to lack of knowledge. VPN licensing prohibitively expenses (E.g., Prairie View asked students not to use VPN- cannot access campus cluster, which Texas A&M is allowing access to its resources. Think about bottlenecks and challenges facing small institution; need trained people to address these issues.

MAGIC as a platform- but also need to self-organize to address issues "bottom-up", rather than waiting for someone to rescue us.

- Agreed. Not technology, rather a "person" obstacle. Need to sit down and try to work on something together. Not everyone can do a cluster, some could only do Jupyter notebook, but some of us can do a little more and share resources.
- Sharing/collaboration of resources would help tribal college which is also trying to provide resources to our government. E.g. Abandoned uranium mines – how can we clean it and eventually make an impact on community. Tribal colleges are always looking for collaboration opportunities.
- OSG would be happy to contact you with anything we can do to help. Helping campuses, but need to get organized at campus and small college level.
- Locally and regionally important HBCU and data science use cases are a good motivator.

Concerns:

- Cost inefficiency of commercial cloud computing;
- Potential loss of understanding of the fundamentals of advanced computing.
- Scrimping on on-prem resources – not beneficial in bulk computing from cost view point. Cloud providing lower threshold and lower cost access is not borne out by years of experience.
- With that drastic lowering of costs and physical sizes of advanced HPC. Astonishing amount of computing just a few years ago, due to advances in GPUs, etc.
- Point in the direction of workforce development with locally available resources Put all pieces together: local and regional interest; need for continuing DIY, hands on computing to learn the field; potentially drastic reduced cost to physical infrastructure, size needs to deployment. Nvidia etc. would probably easily satisfy the modeling needs for the uranium village. Working on attempting to improve access to low cost electricity from regionally available and renewable resources (maybe add to mix)

Cloud- To develop skills for HPC, running own resources and not just relying on magical preconfigured resource in cloud is an important workforce development item.

- What can be done on the cloud that cannot be done on-prem? Build out the continuum.
- Not just about getting the cords, but the total cost of ownership (involves actual dollars) – also nontangible cost aspect is important. Even in large institution, need help getting on the cloud and on-prem clusters. Part of ownership (network access, fire walls, network access). May be unable to continue pro bono internally and externally due to financial issues (big universities). Need to look at everything – everything that is needed and what we've been relying on for funding. Then step back and solve it; what can be done in practical way.

Cyberinfrastructure (CI) resources:

- Large universities may be unable to continue pro bono internally and externally due to financial issues (e.g., furloughs, job loss for staff, faculty.) Much uncertainty on ground.
- Possible solution in lieu of pro bono work: Minority Servicing Cyberinfrastructure Consortium – doing something that cannot do alone via a separate entity - 501(C)3, external funding needed by also funded by institutions. Consortium would provide cyberinfrastructure resources and be

source of skillset and workforce development for the institution and members would be smaller institutions lacking skill set inside individual schools. Trying to push this solution long before COVID.

- Move away from mentality of asking for more to focusing on what we can improve with what we have
- Big opportunity to increase ROI: Track funds spent on bringing people to meetings and the total cost.
    - Question for CASC: What are the metrics for ROI? Measuring cost for an hour of computing is not the question of return. Rather, the question is how much science are we delivering?
        - CASC: Need to define return in financial and nonfinancial terms. CASC has developed metrics and qualitative measures of things that need to be addressed. Investment has to include investment in people as well as hardware.
        - What we have learned from our community engagement from short-form and long-form surveys.
            - Can compare ROI if not measuring.
            - Workshop trying to get everyone to right starting point – have tools to evaluate ROI as it applies to your institution. Need to define ROI in terms that is meaningful for your institution; but no conversation is occurring (perhaps because afraid to draw attention to their expenses to the university).

    - Need to think about CICE – would tie us closer to the software world. In the next 5 years, will use more tools that are industry standards (e.g. Terraform will become more common; advantage because where much technical computer science education is headed) So start align ourselves with forward-facing trends; then solve many workforce issues because move from handful of people trained on site to much larger pool. And alignment with industry would be significantly stronger. So, more forward-facing view of looking at things.; <u>New technologies to build forward</u>.
        - Changing cyber ecosystem – Can't build next 50 years of innovation on the last 40 years of technologies. But in terms of industry readiness, we do have technologies that we can build into (e.g., 5G, IPV6, whole range of broadband sets of technologies) that can scale to new needs and expanded connectivity for universities and travel colleges and universities. We're moving to exascale computing which is whole new sets of development. Moving towards big data science and ML. All of these represents new challenge in scale and sets of technologies. But have scalable technologies available from industry standards today, but need to let go of what doing for last 50 years (IPv4). Have new technologies to build forward.
        - Bigger picture- includes compute cost for cloud (which includes whole new paradigm where difficult to do things lie asset management because virtualized) vs. on-prem with nice network perimeter and know how to do asset management with fixed devices in a data center are back mounted, for example

- National tribal collaborative report addresses this issue.
- Thus, there are challenges and new technologies that are essentially new scale that have been exacerbated by COVID

Big picture important – and even bigger picture exists.
Need to ensure that right people are participating in these conversations.

- Concern: when we look for commercial cloud, we're getting the leftovers that were tailored for commercial use - need to try to fit research needs in it.
    - Need to be able to define our needs. If part of our plan, needs to be tailored to our research needs; open to changing research practices. There are unique research needs, if continue to use what was built for other things, won't work as well. Then will default to what we have always done.
- How fund these efforts. Can see the proposal mechanism breaking.
- Not all technologies (e.g., container) are confined to cloud; just a modern software practice limitation. Invest in software development and teaching it is important part of workforce development regardless of where deployed (e.g., building and deploying applications; crucial to adapt new architectures – cloud based or on prem.
- Building curriculum for research computing –

Recap:

- 1st workforce session
    - Workforce development trends (e.g., forging partnerships) and program implementation
- 2nd session:
    - Began with NSF CC* program, recipient experiences and other ripple effects from participation (e.g., partnerships, opportunities).
    - Other themes: have more people participate in the discussion and more discussion as well as a way forward.
- What can MAGIC provide? Roundtable
    - While it was proposed as a series, it is more of an ongoing discussion (e.g., updates from consortiums/workshops, diversity across scientific domains)
    - Attendees involved in today's discussion represent organizations that do not traditionally participate in MAGIC which typically discusses technologies; their perspectives triggered a discussion on what is happening on the ground at these institutions.
    - Roundtables have worked in other public team, Joint Engineering Team's cyberinfrastructure security roundtable – people attend when able and provide updates
    - NITRD doesn't have a very active workforce program component area, but MAGIC has flexibility to provide a forum.
    - Subcommittee/ subgroup with deliverables and has additional meetings. Charter in a way that would be attractive to keeping these partnerships and collaborations
        - Look for common ground within MAGIC's scope. Idea discussed earlier – many technologies are common across modes of delivery (e.g. CIC – not care if on cloud or on prem). Next ingredient: rapid advance of technologies in these areas that's unique to advanced computing. Another motivator: skills development, the

component that would need to be retained. Still tools that Sharon mentioned that is unique to our scope of work.
        - Important to bring them in a bigger way. As they alone know what is hindering the use of these resources on campuses. Faculty of minority serving institution is interested, but need training to integrate into existing/developing courses. Include those with interest in improving teaching and research skills. No institutional motivations to encourage them
        - Gen-cyber route for cyber training: NSF sponsored NSA gen-cyber program which has had tremendous impact because it is a "deployment programming" – take existing skills; give the skills. Wish something akin to that in K-12, university, grad. Gen-Cyber promotes both student and teacher camps. Link: www.gen-cyber.com
    - Worth asking the following question, which puts us in the sweet spot for MAGIC: Suppose have all the computing needed, how would you organize your workforce development? What would you want to train people on? How would you incorporate classes? What societal problems would you solve? What resources would you need (including professional advance and assistance on a national, regional or consortium level)?
        - MAGIC is forum exchanging information for community and among agencies; identify topics needing further definition and sharpening and maybe to motivate agencies to consider offering funding
        - Agencies have permanent workforce development mission. Worth dedicated effort. Contribute to on an ongoing basis as research priorities will change.
        - Also, note larger institutions' motivation for regional coordination for small institutions (Area #2 of CC* program). So note, federal, larger and smaller institutions.

Foundational work needed:

- Sustainability issue: Minority consortium found that understanding of research computer infrastructure isn't there. Not understand existence of "enterprise computers – the infrastructure that facilitates research computing. Need to explain what is CI and why is this important to college presidents, provosts. etc. Then can discuss workforce development from the perspective of this discussion. Minority consortium planning on spring workshop to lay groundwork regarding importance of CI.
- Great if MAGIC representative can present at Minority consortium. MAGIC can reach out to Minority consortium.
- Workforce Development roundtable still useful for information exchange and participation by more institutions. We can see how sustainable it is.

Digital dichotomy of new technologies which are different and not backwards compatible with old technologies.

- Workforce development building forward into the future is directly impacted; need to be trained in new technologies. Key component of workforce development today and into the future. Incorporate into these discussions.

FY22 Potential MAGIC Tasking:
Present at Large Scale Networking Annual Planning Meeting each year. Proposed direction for FY22:

- Task 1: Workforce development (3-4 month series)
    - Forward looking solutions
    - Career paths (e.g., cyberinfrastructure)
    - Diversity and inclusion in HPC (includes science domain inclusion)
- Task 2: Role of Cloud in Scientific Discovery (4-6 mo. series).
    - Stakeholders' perspectives on research needs and how cloud benefits research (list of stakeholders)
- Task 3: Other Topics (single sessions)
    - National Research Cloud – direction and agency impact.
- MAGIC Workshops/Roundtables
    - Workforce Development
        - partnerships/collaborations (e.g., HBCU)
        - Science domains

Workshops may come up in the course of a speaker series. MAGIC will discuss at the time. Otherwise, do a summary report.

Announcements

- October 13: CASC workshop. Open to anyone dealing with ROI considerations for Research Computing and Data at their institution.
- October 14 -16: CASC member meeting

Next Meeting

November 4 (12 p.m. ET)