National Cyberinfrastructure Coordination Service Conference

Rethinking NSF's Computational Ecosystem for 21st Century Science and Engineering

Conference Report
NSF Proposal: 1930161

July 25, 2019

Program Committee: Alan Blatecky, Dana Brunson, Thom Cheatham, Joel Gershenfeld, Angel Hedberg; Jim Bottum and Dan Reed were also on the program committee but were not able to attend.

RTI INTERNATIONAL | turning knowledge into practice

National Cyberinfrastructure Coordination Service Conference Final report
Rethinking NSF's Computational Ecosystem for 21st Century Science and Engineering
July 24, 2019

Overview

The overarching objective of the National Cyberinfrastructure Coordination Service Conference was to rethink the nature, composition, and execution of NSF-supported, national-scale coordinated CI services that are essential to 21st century science & engineering research and education. The overall conclusion of the participants was that advances in computation, software, and networks combine with new ways of thinking about data and social systems that together justify a shift to viewing the whole as an ecosystem, rather than as a set of separate activities or domains.

The conference was conducted June 27-28, 2019 at the Westin Hotel in Alexandria, Virginia, and a copy of the agenda can be found on page 9. Because of the complexity of the issues, the conference was limited to 35 invited participants to ensure adequate time for exploration and discussion. A list of attendees can be found on page 14. Attendees were selected based on their thought leadership in their respective domains and for diversity of areas of research and expertise.

The astounding growth in computation, data, software, IOT, and most recently in AI/machine learning, is significantly transforming business, research, education, and government. The ability to combine and integrate these technologies and capabilities to explore new solutions is essential for 21st Century Science. Cyberinfrastructure (CI) enables and accelerates these advances. Research challenges and science projects that were impossible to consider 5 years ago are now beginning to be addressed because of innovative and capable CI services.

The conference used 3 vectors to explore how emerging CI capabilities should be used to support 21st Century Science. These three vectors were:

1. Providing national-scale coordinated CI services
2. Developing the next generation of researchers and technical workforce
3. Integrating emerging CI capabilities and supporting a balanced portfolio of advanced CI capabilities.

To help expedite discussion and considerations, a series of questions (page 11) were generated for each vector. The first half of the Conference consisted of a series of presentations and lightning talks from leading scientists and CI experts to help set the stage for the discussion. In the second half of the Conference, each vector team was charged to explore how the CI Ecosystem should be adapted to address the evolving needs of researchers as well as respond to the series of questions raised for their vector.

The focus of the conference was to inform and better understand how to maximize the impact on scientific discovery by the ever-evolving coherent and coordinated national CI ecosystem. The conference and list of questions, by design, generated considerable discussions across the spectrum of CI issues and services. While some of the discussions were quite spirited, the diverse group of participants and facilitators helped ensure that the discussions were kept on track. While the conference did generate a set of findings (page 3) and recommendations (page 10), the most important output from the conference may be the set of six (6) observations about the rapidly evolving new order of the role of CI in scientific (page 2), and the questions on the nature of the CI services (page 8).

Conference Observations

Observation One – Transforming Science: The growth in CI capabilities is significantly transforming all aspects of our lives, and we are rapidly approaching a new inflection point on how research and education are being conducted. This growth includes:

- Unprecedented computational and data processing capabilities at all scales.
- Increasingly available CI capabilities at the network edge including 5G and IOT.
- Growing importance of campus CI as a key enabler of S&E research, such that NSF is providing an ever-shrinking piece of the national CI.
- Proliferation of commercial cloud services supporting new models of operation and allocation of resources (on-demand or elastic usage models, novel technologies and compute approaches (Google Tensor Processing), containers, are driving commercial applications and ever-changing business models.
- Continuing revolution of adoption and use of open source software, AI/ML across many areas of science including, integrated data, machine learning, and complex workflow management.
- Continuing rapid growth in the amount and importance of data, not just for replication, but for reuse in many ways.
- An extremely competitive market place for CI talent; it is hard to recruit and retain CI professionals.

The ability to combine and integrate these technologies and capabilities to address problems and explore new solutions promises, has been and will continue to be remarkable, with advances that were impossible to even consider 5 years ago. This inflection point will be marked by accelerated advances and will not only require new types of CI services and capabilities, but provision of these services will have to be significantly accelerated as well.

Observation two – It's an Ecosystem: the core of CI is no longer just about HPC or Data writ large. The evolving CI is an ecosystem that includes people, software, networks, data, and compute capabilities. All are necessary components of the ecosystem. All must be present and supported in a balanced way in order to be effective in enabling scientific discovery and education. Services and support must be bundled in unique ways to leverage and extend the CI Ecosystem to the entire research and education community; simple extension of existing methodologies and processes will not be adequate. Supporting computation, networks and software each involves distinctive approaches with varying time horizons and investment requirements. Support for social systems (including people) and data bring additionally distinct time horizons and investment requirements.

It is important to note that the CI Ecosystem isn't about a future state; it is already here. It urgent that new efforts be developed immediately to build a professional coordinated CI community to order to support science and engineering.

Observation three – New Domains Served: The CI Ecosystem underpins today almost all scientific domains and frequently emerges as the pace-setter for advances in these domains. State of the art CI capabilities are becoming fundamental for many disciplines and sciences that have not typically been part of the compute world, including the social sciences, the digital humanities, many field-based domains of the physical sciences, and citizen science. While many of these domains are referred to as the Long Tail of science, this term can be misleading as it implies that Long Tail of science is of lesser value or may have fewer requirements or resource needs than "short head" of science. The CI ecosystem has been and will continue to transform all areas of science and powers new remarkable approaches and discoveries, from small science to big science, from seasoned researchers to the next generation of researchers.

Observation four – Spanning all Campuses: This CI Ecosystem is becoming a fundamental requirement and capability for all campuses. CI has evolved to become part and parcel of the university structure supporting education and research for undergraduates, graduates, and faculty. Although campuses already provide, in aggregate, more CI services and capabilities than do federal facilities, this ratio will continue to grow as every student will be expected to have basic CI competence to be a productive informed citizen in the 21st Century. Since CI (integration of computer science, data science, engineering) doesn't neatly fit into the domains of a typical Vice-President for Research or a typical campus CIO or typical departmental and college leadership, the people who provide CI services to faculty and students are operating without a clear institutional home. Even among R1 universities, the CI roles vary considerably. Universities and colleges with fewer resources (including HBCUs, HSIs, and NSIs) have to pull from faculty and students and are particularly challenged to respond.

Observation five – Integrative Services: The rapidly evolving CI Ecosystem requires new approaches to offering effective services to sciences that covers all the CI components. This in turn means a persistent focus and approach to provide integrative services and support across all the CI components and emerging technologies. Since the CI Ecosystem has become more distributed (across campuses, other organizations and the commercial sector), the services need to be coordinated across multiple organizations and sectors in order to be effective. This requires interagency coordination in government, public-private partnerships, multi-university initiatives, and other collaborative arrangements in order to facilitate the increasing integration of services.

Observation six – Service and Resource Allocation: The emerging CI Ecosystem requires new approaches on how to allocate CI services, which includes computer, data, support, and other resources. Since CI services and resources are expected to continue to evolve and become more constrained (need and demand exceeding availability), methodologies and processes to manage fairness, effectiveness, and priorities of use must be developed and adopted. The allocation processes must cover all the components of the CI Ecosystem and, do so in a manner than follows the NSF mission and objectives.

Conference Findings

Constraints on Science

- The lack of data preservation, curation and sharing resources, including supporting services and capabilities are putting major science projects at risk because much of this data will not be available for re-use. This in turn incurs both high opportunity losses as well as negatively impacting future science.

- The demand for CI services is growing faster than the resources being made available for science, even with the power of Moore's law. One specific example was provided by TACC; requests for TACC resources in 2018 exceeded the amount available resources by 11:1.

- Data facilities, which include domain-specific facilities, general facilities, libraries, research center repositories, and others, all face funding constraints that fall far short of what is being promogulated for data preservation and access.

- There needs to be a stronger focus on software, from development and integration, to use and efficiency. Limits of Moore's law coupled with ever growing over-subscription of compute resources results in ever more dependence upon software capabilities and services. Efforts to look more seriously at software efficiency, performance, and quality assurance could be very productive and make more resources available to the research community.

- Overall, conference participants expressed a sense of urgency CI Ecosystem issues must be more systematically addressed to adequately meet the rapidly changing needs for CI services and to translate the rapidly evolving potential of CI to advance scientific discovery and education.

- "My science is limited by my CI Services"

Emerging Technologies

- The broader research community is not currently being ubiquitously served or supported by emerging capabilities such as AI/ML, IoT, or by the integration of CI capabilities

- There is an open debate on what role infrastructure cloud resources (especially commercial clouds) should play the overall research portfolio. Too much emphasis in current discussions is upon price and cost; issues of elasticity, or on-demand, or batch (especially since the cost curve for core hours is decreasing) and broad access to new and innovative technologies will be more productive, noting that such services will complement, not necessarily replace, existing CI services.

- Ethical issues associated with AI/ML, IoT, digital fabrication, and other advances have implications for the design, structure, and operation of CI, including standards for security, transparency, traceability, and related considerations.

Disparities and Inequities

- There is a growing gap between those who can leverage advanced CI services and those who do not.

- Small institutions find it difficult to engage with national efforts and also find it difficult to travel and participate in larger research efforts with the result that they tend to be part of the have-nots. Are there ways to incentivize national efforts to engage more directly with these communities? For example, requiring resource sharing on a more formal basis similar to what EPSCOR does for major CI resource awards?

Workforce

- NSF tends to use incremental approaches for workforce development (2-3 years) which is inadequate and to some extent, counterproductive. By the time the program gets established and underway, the funding ends. This is particularly devasting to smaller institutions or to emerging research communities. There should be a balance between impact to large numbers of users versus train the trainers type programs that more effectively scale.

- The growing network of individuals who serve as facilitators for research computing lack clear career paths within research institutions, which results in high turnover and incomplete opportunities for professional development beyond the entry level content.

- The hiring of research computing facilitators by the private sector is, on the one hand, highly disruptive to the research computing enterprise and, on the other hand, an important service to industry.

- The development of a strong coordinated CI community is under appreciated; it is fundamental requirement for the CI Ecosystem.

Research Data

- Large research instruments (telescopes, colliders, ships, etc.) are increasingly seen as data sets with attached instruments – requiring increased staffing, storages, access, and communities of practice centered on the data.

- Despite the language of open access, not all data can or should be made open. The movement toward increased openness does, however, pose challenges and opportunities for researchers, publishers, professional societies, and others – with broad implications for the research CI.

CI Administration

- More attention must be paid to effective use of CI resources

- Discoverability of available CI services and resources needs to be improved

- Multi-agency coordination of a national portfolio of CI services and resources is needed, but there is no compelling vision or strategy to make it happen. Despite this, there needs to be multi-agency coordination with input from the broader research community to address the critical needs of the growing and diversifying CI that exists today and is continually evolving.

- The allocation process needs to evolve beyond the current HPC-and-node-hour-centric approach to include data, software, network, and people as well as eliminate the double jeopardy issue that results from giving an award for a research effort without making sure the project has the resources to be successful. That is, it would be desirable for all CI services to be reviewed and awarded with the original proposal.

- There was considerable discussion about doing a "decadal" type study of the CI resources required by the research community and identify the gaps in the CI Ecosystem that need to be addressed. This approach has well served the Astronomy community for a number of years, but it isn't clear how the broad research community can accomplish this. Moreover, the continuing rapid change of the CI Ecosystem suggests that a multi-year long review would not be effective. Yet, efforts are needed to seed these efforts.

- Overall, advances in technology have required complementary advances involving agile teams and agile organizations. The challenge now facing the NSF and other agencies is the challenge of agile institutions (still providing the needed stability and continuity, while being responsive to emerging threats and opportunities).[1]

[1] "For three decades we have become used to ever-growing, larger and larger machines, but that doesn't seem to be the winning approach for creating effective science in the post-exascale and post-Moore era." John Shalf, Berkeley Labs: HPC Wire, June 27, 2019 -- https://www.hpcwire.com/2019/06/27/berkeley-labs-john-shalf-ponders-the-future-of-hpc-architectures/

Key Questions regarding National Scale CI Coordination Services

1. **What are the essential elements of, and relationships between services? What are the key application drivers (and usage modes)?**

   While the focus of NSF CI services traditionally has been on access and use of large compute resources, the growth and changing landscape of CI requires a much broader focus on access and use of the entire CI Ecosystem. The growth, evolution, and expansion of compute and data resources is creating a broad and robust distributed CI Ecosystem which will be serving an equally broad and growing science and research community. Basic services, such as authorization and authentication, are essential to support users who will be accessing a wide range of resources owned and operated by others (including commercial entities). For small campuses, authentication type services may have to be initially supported, but should be migrated to services such as InCommon that are already available in the marketplace.

   Other basic services required are those that enable users to access and use the broad and evolving CI Ecosystem. A focus on portability and inter-operability (such as use of containers), common approaches and integration of tools that can be used on a wide range of distributed resources and data becomes essential. Support of services that provide best practices, easy-to-use tools, software packages (including science-type applications), libraries, and identification of canonical workflows is critical to leverage the new CI Ecosystem capabilities. This becomes especially important for the next generation of researchers and scientists who expect to be able to use a range of applications to do their science without having to become experts in how to modify the software in order to use CI resources. The services also need to be robust and widely available allowing users to adopt or adapt templates and a wide range of 'applications" to conduct their research.

2. **What are the appropriate models for architecting, implementing, managing and delivering services, individually and collectively? What are the relative priorities and scales of these services?**

   The appropriate models for architecting, implementing, managing, and delivering services need to reflect the new CI Ecosystem which is highly distributed. The services have to help reduce obstacles and increase access to the wide array of resources for individual researchers as well as support dynamic teams (in many cases, multidisciplinary teams) that coalesce to meet specific or emerging research problems. The models for services need to provide support much closer to end users and the campuses as that is where the bulk of science will be conducted. This in turn implies that the service model should be distributed rather than centralized, with a priority on "train-the-trainer" models to provide scale and develop a vibrant workforce.

   A focus on supporting and leveraging small consortiums or geographic regions of universities to provide the services would put the services much closer to the end user and help build local capabilities and capacities. Supporting new "gateway" projects that focus on providing access and use of common infrastructure approaches (including cloud usage), rather than providing access and use of major compute resources, will build models that can help scale usage. The new "gateway" projects should focus on tools and libraries that can be used on a diverse set of resources; as such, they are likely also to involve formal and informal relationships with resource providers and the science communities developing new approaches for integration, software packages and data exchange.

Care and feeding of the distributed science community is necessary to address issues associated with supporting common approaches so users in one region/consortium can use resources and capabilities in other regions and the nation. Support of periodic workshops and a national conference would help facilitate the development of a vibrant community and encourage sharing and staff development opportunities. Lessons learned from efforts such as ACI REF will be useful as the effort needs to both provide leadership and direction as well as building direct end-user support. This National Conference should function more like a clearinghouse and convener for the community rather than providing technical leadership.

3. **What are the opportunities, mechanisms, and use cases for these services and elements to operate in an integrated manner? What are effective mechanisms for ensuring that these services evolve and that innovations and new services are introduced?**

While supporting multiple distributed service entities may sound counterproductive to a goal of operating in an integrated manner, if the program is structured to encourage and require adoption of common approaches from the outset, the opportunity to provide integrated services becomes quite feasible. The set of distributed service entities, along with a national entity that would focus on evolving services and innovation, should be charged to develop a rolling road map of services to be deployed and supported. Providing a common set of integrated services across the nation would help with scaling efforts; having a national group providing coordination and leadership in integration will help the services evolve and support the deployment of new tools and approaches. Lastly, having geographically distributed service entities closely working with researchers across campuses helps ensure that the services are meeting the existing and emerging needs of the science community.

4. **What are the opportunities and use cases for integrating NSF-funded services with complementary investments by other agencies in the US and internationally? What is the role of emerging relevant service offering from industry?**

The potential to integrate NSF-funded services with complementary investments by other agencies is very high in cloud services and capabilities including support of open and secure data sharing approaches. Although other Federal agencies (especially the "mission" agencies) have some degree of mandated efforts, the use of cloud service offerings for on-demand and elasticity requirements is a common need across all the agencies. Low barrier-to-entry and use of resources, services and data exchange are also very important. Finding common approaches for security and privacy are also areas that will resonant across multiple agencies. Lastly, a consistent focus on developing and supporting common vehicles and platforms to support science and research will be of benefit to everyone.

Recommendations

Near term recommendations

1. Define a focused effort to pilot and test the deployment of a national connected network of repositories for data that is highly reused to help address the existing and growing problems central to data and what has been termed the data tsunami. *(The pilot should be enabled to support an allocation process for persistent (and likely distributed) research-related data resources on the 100 PB to Exascale size with a ~5-year time scale)*

2. Invest in developing facilitation capabilities with a focus on integrated CI and persistence
    a. Incentivize the development of new/ongoing efforts that bring together CI professionals to learn from one another and that generate community efforts to identify and improve leading practices.
    b. Support regional collaborations with under-resourced institutions to achieve shared support and training of researchers and facilitators.
    c. Support the development of a strong coordination CI community.

3. Requests for building Research CI resources must always address sustainable plans for support and enablement for researchers and users.

4. Define a focused effort to explore approaches that will increase the quality of existing software and increase computational efficiency to help free up CI resources for other science projects and users. *(TACC reported an 11 to 1 oversubscription for resource requests; several compute centers also report a number of inefficient software packages that use more resources than necessary)*

5. Support the development of new approaches to use commercial cloud resources to address such issues as elasticity and on-demand to support research that doesn't fall into traditional categories. This should also include broad access to new/innovative technologies and providing resources for the huge range of science computations for emerging communities and users.

6. Further develop the role of stimulating investment and sharing of CI services across campuses nationwide via the CC* program to build a comprehensive national CI Ecosystem.

Medium-Term recommendations

1. Evolve the allocations process to include the use of all CI services (not just HPC) required to do the proposed research. This includes computational, software, data, network, and the people resources to conduct the research. (*Linking CI services required to the proposed research up front will also eliminate the double jeopardy issues research proposals face when the proposed research is approved separately from the compute resources required.)*

2. Create a production scale portfolio of the related infrastructure persistent programs that supports the innovation in the NSF ecosystem and helps reduce the growing gap between those that have access to the CI Services and those that do not.

3. Advance the science of science institutions (a complement to the National Academies' initiatives on the science of science teams), with agile CI institutional arrangements as a leading application.

Longer-Term Recommendations

1. NSF should develop a NSF-scale plan for CI Ecosystem including determining the needs of research to help guide and prioritize broadly defined CI resource provisioning and services. As part of this, efforts should aim to collaborate with other agencies to develop a full national plan that spans from campuses to the nation.

2. Develop an oversight effort across federal agencies to help coordinate CI services and resources to support open science*. (NSF has an ACCI; what or how can other agencies participate and engage to support the science community?)*

National Cyberinfrastructure Coordination Service Conference Agenda

Day 1: Thursday, June 27

7:00	Breakfast	
8:15	Welcome, introductions, goals, vision, process	Alan Blatecky, Program Committee Bob Chadduck
9:00	NSF Science and Advanced CI	Fleming Crim
9:30	CISE	Jim Kurose
9:45	OAC Vision - Transforming Science through Advanced Infrastructure	Manish Parashar
10:15	Break	
10:30	8 lightning talks (7 minute presentation; 5 minute discussion) --draft	
	ALMA, VLA and ngVLA	David Halstead
	Future of Bio-compute	Alex Feltus
	The future ACI in the Polar Sciences	Paul Morin
	XSEDE	John Towns
	OSG	Miron Livny
	XDMOD	Tom Furlani
	Facilitators/Facilitation	Lauren Michael
	ACI-REF - CaRCC / BioSim	Thomas Cheatham
12:15	Lunch	
1:00	Technologies Trajectories	Sam Adams
1:30	Other agencies (10 min including Q&A)	
	NOAA	Frank Indiviglio
	NASA	Tsengdar Lee
	NSA	Olga Ratismor
	DOE/Fermilab	Liz Sexton-Kennedy
	Compute Canada	Greg Newby

2:30	Discussion of conference questions and the charge for all three Teams • Providing coordinated CI services • Developing the next generation of researchers and technical workforce • Integrating emerging CI capabilities and supporting a balanced portfolio of advanced CI capabilities	
3:15	Break	
3:30	Breakout Team Sessions; open discussion, brainstorming, etc	
5:00	Break	
6:00	Catered Dinner - informal discussion	

Day 2: Friday, June 28

7:00	Breakfast	
8:15	Agenda review	Alan Blatecky
8:45	Second Breakout Team session (identify key findings and requirements, define problem space, chunk into categories to report out to full conference)	
10:15	Break	
10:30	3 Team reports on areas of focus, categories, issues to be discussed	
11:15	Third breakout session (group develops a set of findings and recommendations	
12:15	Working Lunch (groups continue to meet)	
1:15	Team summary reports Providing coordinated CI services	Angel Hedberg
1:30	Developing the next generation of researchers and technical workforce	Dana Brunson
1:45	Integrating emerging CI capabilities and supporting a balanced portfolio of advanced CI capabilities	Thom Cheatham

2:00 – wrap up, items and areas that should be further explored
2:30 – Adjourn

Conference Questions for each Team

Meta-Issues for each team

- How can science leverage the rapidly changing compute and data landscapes and investments being made by industry and government?
- How can public and private technology investments be integrated, leveraged, or coordinated into a coherent CI ecosystem that is more valuable to advance science and engineering research than the sum of the individual resources or capabilities?

Providing coordinated CI services: Team 1

Facilitators: Angel Hedberg, Joel Gershenfeld (first day only)
Team 1 members: Dan Fay (Microsoft), Tom Furlani (Rochester), Frank Indiviglio (NOAA), Klara Jelinkova (Rice), David Lassner (Hawaii), Tsengdar Lee (NASA), Matt Mayernick (NCAR), Michael Miller (Johns Hopkins), Howard Pfeffer (Internet2), Elizabeth Sexton-Kennedy (Fermi), John Towns (NCSA), Frank Wuerthwein (UCSD)

- What types of coordinated CI services are needed and what should NSF support?
- Who needs the services; who would benefit?
    - Students, researchers and scientists
    - Universities, research institutions, major research projects
    - Commercial sector
- How should these services differ from, relate and/or co-exist with existing CI services being provided by campuses, regional organizations, commercial cloud providers or other agencies?
    - What is the rationale for these services? Can some of these services be handed off (such as the case of NSF support of networking services in the 90s which are now being provided by institutions)?
    - How does the rapid growth and capabilities associated with cloud computing change the research and service landscape?
    - What types of CI services should not be provided at the national level? What types and level of services should campuses and research organizations be expected to support?
    - How should the services be migrated to provide leading-edge services?
    - What should the scope services entail (level of support, length of time)?
- What role should NSF play in providing coordinated and integrated CI services?
    - How should these services fit with or be coordinated with other government agencies such as DOE, NASA, NIST, and DOD?
- What are some opportunities to provide useful and sustained efforts that are not being done currently?
- What would a national services support structure should look like?

- Should it be centralized, be regionally based, or be widely distributed?
- How should universities, colleges, and other institutions of higher education be effectively involved since many scientists, educators, students, as well as campus cyberinfrastructure investments, are located in these institutions?
- How may private and public sector entities be appropriately involved?

Developing the next generation of researchers and technical workforce: Team 2

Facilitators: Dana Brunson
Members: Alex Feltus (Clemson), Lauren Michael (UW-Madison), Olga Ratismor (NSA), David Halstead (NRAO), Wende Huntoon (Kinber), Gwen Jacobs (Hawaii), Doug Jennewein (USD), Susan Winter (U Maryland), Michael Zenter (Purdue)

Focus on CI Professionals:

- How best to optimize the development of the CI workforce in each of the domains from which it comes (undergraduates discovering the field, community colleges, domain scientists reinventing themselves, librarians and curators, professional development initiatives by consortia, and others)?
- How can people from underserved institutions and communities become part of the new CI technical workforce?
- What funding models are best matched to workforce development investments -- so they are both innovative and sustainable? What can we learn from funding models in other nations and from past initiatives that have been more and less successful?
- What deeply embedded assumptions guide the NSF in its investments in new research that are either helpful or problematic when it comes to workforce development?
- How best to build CI expertise that is always at the forefront of advances in technology (in comparison to the need for stable systems in IT)?
- What systems architecture (modular, interoperable, extensible, scalable) might enable the many diverse professional development offerings to operate effectively in the broader ecosystem? A non-exclusive list includes: Carpentries, Linux Cluster Institute, ACI-REF Virtual Residency, XSEDE, OSG, CaRCC, and Campus Champions.
    - Internships (national labs, companies)Nano-degrees
    - Certified boot camps,
    - Land Grant Cooperative Extension centers to provide education/training
    - MOOCs
    - Apprentice-type programs at computing centers (including NSF/DOE/NASA -funded projects)
    - ACI-REF initiatives
    - Campus Champions,
    - CaRCC initiatives
    - Regional approaches
    - Development of local expertise, campus expertise, and training for researchers (outside of standard curriculum/for-credit courses).
    - Events; Tiger Teams, community events, regional events, national events (SCXX)

Focus on Researchers Utilizing CI:

- What does the next generation researcher need to know about CI?
  - What constitutes a working knowledge and appropriate foundation of CI?
  - How can emerging CI capabilities and technologies be incorporated into community college and university courses (which are too often not at the state of the art))?
  - How best to couple CI-awareness and knowledge with Domain-centric expertise?
  - How to develop practicums or "labs" to provide experience in how to use CI and develop meaningful teaming and partnering? (Rapid evolving CI tools and approaches requires a mindset that on-going collaborations will be norm for science).
  - How to expand CI internship opportunities within major science projects?
- What needs to be done to establish and support a viable pipeline of next generation researchers with a focus on continuity and sustainability?

Additional Meta Issue: The need for research cyberinfrastructure on the campuses is exploding and expanding into new domains where expertise in the use and applicability of CI is lacking. This emphasizes the need and demand for technical experts or facilitators who can bridge between the researchers and the technology to train and support in the effective use of cyberinfrastructure resources.

Integrating emerging CI capabilities and supporting a balanced portfolio of advanced CI capabilities: Team 3

Facilitators: Tom Cheatham
Members: Sam Adams (RTI), Laura Bevin (DOE), Kate Keahy (Argonne), Dave Lifka (Cornell), Miron Livny (UW-Madison), Paul Morin (U Minnesota, Polar Program), Greg Newby (Compute Canada), Sanjay Padhi (Amazon), Phil Papadapolous (UC-Irvine), Dan Stanzione (TACC)

- How can the nature and implementation of advanced cyberinfrastructure services strategically evolve, be supported, and be sustained to enable science in the 21st Century?
- How can emerging technologies be introduced, enabled, and supported for science and engineering research, for example emerging edge technologies and capabilities such as AI and IoT?
- How can services, capabilities or technologies being developed or deployed in one context be leveraged, re-packaged, evaluated and assessed, and/or "transitioned to practice"?
- How can these capabilities and technologies be seamlessly inserted into the larger CI ecosystem, especially from developments emerging from NSF supported CSSI, CC* awardees and/or other efforts?

- When it comes to CI services, what are some opportunities and use cases for complementary investments that advance effective national scale science and interoperability (including CI resources, and support for researchers, CI professionals, and large instruments and projects)?
- What are the usage modes and use cases that would motivate and support an integrated future CI services support structure in which work flows can move seamlessly from campuses to regions to the larger ecosystem?
- What types of services are appropriate to be provided to the research and education community, including as consequence of NSF efforts?
    - Services that help integrate the various elements of CI into a coherent CI ecosystem of compute, data, and advanced networking services
    - Services that provide the "glue" that enable users the ability to use the ecosystem without inventing it anew for each research effort
- How to ensure continued training and user-support for established researchers, while at the same time providing support for the next-generation of researchers and CI professionals?
- How best to support and develop campus expertise and capabilities while at the same time supporting a national coordinated program?
- How can NSF best partner with the established leaders in the community on a competitive basis and at the same time develop an investment strategy that will 1) lift up resource constrained institutions who are less able to compete initially and 2) sustain long-term resources that will be undercut by continued competition?
- How best to balance investments that build infrastructure along with investments to advance discovery?

Meta Issues: Need to bring out contrast between 2-3 year programs vs. long term efforts; new and novel temporary activities vs. sustained (inherently less novel or innovative) activities; and types of services, training, and support models; balance between "production" vs experimental/leading edge CI.

Conference Participants

Adams	Sam	sadams@rti.org	RTI
Biven	Laura	laura.biven@science.doe.gov	DOE
Blatecky**	Alan	arblatecky@gmail.com	RTI International
Brunson*	Dana	dbrunson@internet2.edu	Internet2
Cheatham*	Thomas	tec3@utah.edu	Utah
Cutcher-Gershenfeld*	Joel	joelcg@brandeis.edu	Brandeis
Fay	Dan	Dan.Fay@microsoft.com	Microsoft
Feltus	Alex	ffeltus@clemson.edu	Clemson
Furlani	Tom	furlani@ccr.buffalo.edu	Rochester
Halstead	David	dhalstea@nrao.edu	NRAO
Hedberg*	Angelique	ahedberg@rti.org	RTI International
Huntoon	Wendy	huntoon@kinber.org	Kinber
Indiviglio	Frank	frank.indiviglio@noaa.gov	NOAA-HPC
Jacobs	Gwen	gwenj@hawaii.edu	Hawaii
Jelinkova	Klara	klaraj@rice.edu	Rice
Jennewein	Doug	Doug.Jennewein@usd.edu	USD
Keahey	Kate	keahey@anl.gov	ANL
Lassner	David	david@hawaii.edu	Hawaii
Lee	Tsengdar	tsengdar.j.lee@nasa.gov	NASA-HPC
Lifka	David	lifka@cornell.edu	Cornell
Livny	Miron	miron@cs.wisc.edu	OSG
Mayernik	Matt	mayernik@ucar.edu	NCAR
Michael	Lauren	lmichael@cs.wisc.edu	Wisconsin
Miller	Michael	mim@cis.jhu.edu; BME-Director@jhu.edu	Johns Hopkins
Morin	Paul	lpaul@umn.edu	Polar
Newby	Greg	gbnewby@computecanada.ca	Canada
Padhi	Sanjay	sanpadhi@amazon.com	Amazon
Papadopoulos	Hawaii	ppapadop@uci.edu	UCI
Pfeffer	Howard	hpfeffer@internet2.edu; tgiannettino@internet2.edu	Internet2

Ratismor	Olga	oratsimor@ltsnet.net	NSA
Sexton-Kennedy	Elizabeth	sexton@fnal.gov	FNC
Stanzione	Dan	dan@tacc.utexas.edu	TACC
Towns	John	jtowns@illinois.edu	NCSA
Winter	Susan	sjwinter@umd.edu	U of Maryland
Wuerthwein	Frank	fkw888@gmail.com	UCSD
Zentner	Michael	mzentner@purdue.edu	Purdue

Program Committee Chair **
Program Committee member*