Affiliation	Name	E-mail
UCAR/Unidata	Mohan Ramamurthy	mohan@ucar.edu
What percentage of the facility CI was developed in-house versus by reusing existing solutions?		
Data systems and services, software/middleware and tools; Almost all data and software from Unidata are made available freely and openly and use open source licensing, so they can be reused.		
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?		
In addition to Unidata-developed software, we also provide externally developed software to our users. Such tools are identified based on the needs of the academic users and deliberated by our governing committees.		
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?		
NetCDF is Unidata's most widely used software. The challenge is to provide support to a very large and diverse user base in almost every country in the world and all geoscience domains and sectors. The Local Data Manager and THREDDS Data Server applications also have a diverse user community in both operational and research settings. Providing support to an ever expanding community remains an ongoing challenge. Another challenge stems from the rapid growth in the volume of data, so a push approach will not not be sustainable. The increasing volume and diversity of data sources, coupled with the growing user base, also creates challenges in scaling and interoperability.		
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?		
As stated earlier, maintaining high quality of support to a growing and expanding user base in an era of shrinking or level budgets remains a challenge. There are also sociological and cultural challenges with changing technologies and adoption and use of new tools and services. Migration to cloud platforms poses challenges in developing business and cost recovery models.		
Key Risks		

The lack of NSF-funded operational cloud facilities for hosting data and delivering services remains a key gap. Also, most CI facilities are operating independently without much collaboration and partnership. In addition to sharing knowledge and expertise, a discussion on how the facilities can share other resources and infrastructure would be valuable.
What CI-related workforce development activities does your facilities engage in?
Unidata provides education and training, through workshops in Boulder and at different universities, on a regular basis to students and faculty on its products and services. In addition, Unidata hosts several interns and mentors them every summer.
What do you see as your key new CI requirements and challenges in the next 5-10 years
Exploding data volumes and scaling of CI to meet the growing needs remains a challenge. Cybersecurity is another challenging area. Entraining and retaining professionals into scientific CI areas is a challenge given that graduating students and professionals are paid much more by the IT and software industry that is thriving.
Do you have any other suggestions for the workshop?
Clearly stated goals for the workshop and more in-depth discussions on important issues (rather than many overview presentations) is likely to lead to meaningful outcomes.

Affiliation	Name	E-mail
NEON	Tom Gulbransen, Battelle	gulbransen@battelle.org
What percentage of the facility CI was developed in-house versus by reusing existing solutions?		
3 ingestion queues, 4 transformation pipelines, 2 websites. Tailored so unlikely to reuse.		
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?		
6 external host partners for community distribution and limited data product creation. AeroNet, MG-Rast, SRA, BOLD, PhenoCam, AmeriFlux		
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?		
Sensor messaging and control challenging at sites infrequently visited. Ingestion queues which can accommodate dozens of data types and sources. APIs which greatly simply powerful data access and sharing options.		
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?		
The fusion of classical IT systems development now in ntegralky relies on code written by non-IT analysts. The value of the latter was underestimated initially, and will be over-emphasized going forward during community engagement.		
Key Risks		
Sensor unreliability is a risk addressed by engineering. User diversity will create demands beyond the dev team capacity. Initial Ops period will reveal if/where/when/how cyberinfrastructure may need to automate more checks and editsbility.		
What CI-related workforce development activities does your facilities engage in?		

Lots of cyberinfrastructure recruitment and resultant learning curve climbing during construction. Scientific cosers are being herded toward conventions to promote easier interoperability and expansion through external contributions which can be evaluated.
What do you see as your key new CI requirements and challenges in the next 5-10 years
User community traceability and expansion of user's demands.
Do you have any other suggestions for the workshop?
Share registrants info.

Affiliation	Name	E-mail
Ocean Observatory Initiative (OOI)	Ivan Rodero, Rutgers University	irodero@rutgers.edu

What percentage of the facility CI was developed in-house versus by reusing existing solutions?
The infrastructure of the CI has been developed in-house following industry best practices. It includes the data lifecycle management system, and the network and system architecture distributed across two geographically distributed data centers. The customized software stack, including core data management system and user interface has been also developed. The CI architecture and best practices are available to other to reuse.
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?
The OOI CI uses a number of external services and tools, including an Apache server for raw data delivery, a THREEDS server for asynchronous data product delivery, Alfresco for document configuration management and shipboard data delivery, and a number of tools such Redmine and Confluence for documentation and configuration management, gerrit and Jenkins for continuous integration, and phpBB for forums. These tools were selected based on requirements and prioritizing open source solutions, when needed.
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?
1) On-demand data product delivery: OOI provides users with a graphical user interface (i.e., OOINet data portal) for plotting and downloading on-demand data products. The portal also provides access to live video and other data products. 2) Raw data archive: data is available for download in “raw” indicates data as they are received directly from the instrument, in instrument-specific format. 3) Machine-to-machine API: a REFTful user interface is available to access OOI CI programmatically using authentication mechanisms. We’d like to share the architecture of the enterprise-level information lifecycle management system, including networking and monitoring components which use industry best practices.

What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?
Two of the most important challenges of the OOI CI are 1) evolving requirements (e.g., data rates, services), 2) and integration of new components (e.g., new instruments). There are lessons learnt related to the implementation of industry best practices for the deployment and operation of a production-level CI.
Key Risks
One of the highest risks for the OOI CI is related to the uncertainties for keeping the funding level for operating and maintaining the core infrastructure, the software stack and fundamental services. For example, the lack of expanding the storage infrastructure in the future is a risk. A mitigation step was including expandable tape-base storage infrastructure in the information lifecycle management system.
What CI-related workforce development activities does your facilities engage in?
CI-related workforce development is at different levels. On the one hand, technical personnel are engaged with continuous training on the technologies involved in CI (e.g. Palo Alto training, Dell Compellent, Apache Cassandra, etc.). On the other hand, OOI engaged with NSF-funded CTSC for the development of a comprehensive cyber-security plan.
What do you see as your key new CI requirements and challenges in the next 5-10 years
New CI requirements/challenges in the next 5-10 are related to the expansion of the CI network with new instruments, increasing data rates and evolving data delivery mechanisms.
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
National Nanotechnology Coordinated Infrastructure (NNCI)	Azad Naeemi, Georgia Institute of Technology	azad@gatech.edu
What percentage of the facility CI was developed in-house versus by reusing existing solutions?		
Institute developed components include a self-service firewall management, and a shared access model where institute purchased equipment is provided to faculty who in return provide shared access to their purchased hardware.		
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?		
We are actively implementing the Open Science Grid, Globus, science DMZ, and perfSONAR file and networking components. In addition, we are implementing Ohio Supercomputing Center's PBS Tools, Open XDMoD from the University at Buffalo.		
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?		
1) Rapidly growing data sources. Our storage systems have grown exponentially since 2009 to 8 petabytes. 2) Utilization patterns that are many small jobs, i.e. high throughput computing (HTC) vs the few very large monolithic jobs (HPC). We aim to funnel these types of workloads to OSG, and implement hardware dedicated to running OSG computation.		
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?		
Key Risks		
Not at this time		

What CI-related workforce development activities does your facilities engage in?
We hire undergraduate students, contribute to Linux Cluster Institute workshops and are in the process of deploying an instructional cluster.
What do you see as your key new CI requirements and challenges in the next 5-10 years
As a major technological research institution, the Georgia Institute of Technology, which includes academic units and the Georgia Tech Research Institute (GTRI), has direct experience with many of the current and emerging research challenges facing today's
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
NHERI	Tim Cockerill, University of Texas - Texas Advanced Computing Center	cockerill@tacc.utexas.edu

What percentage of the facility CI was developed in-house versus by reusing existing solutions?
Nearly all of the CI components are developed in-house by TACC and are made available as open source in github.
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?
We use the Django web framework based on our previous experiences with this and other frameworks. We also have a local implementation of the Fedora Digital Object Repository Management System for our archiving our published data.
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?
The Data Depot is our most used CI component. Our users have already uploaded more than 16TB of data in addition to the 40TB we transitioned in from the predecessor project NEES. We allow all file types and we encourage our users to upload any and all data they need to do their research - we feel that not restricting the users is key to their adoption of our CI. We worked with Mathworks to acquire a MATLAB license that enables all academic users to access MATLAB via our CI. The engineering community are heavy MATLAB users, and this has also helped with adoption. We implemented Jupyter Notebooks and are providing training on how to use them along with basic Python scripting skills. We are seeing pretty strong uptake of Jupyter. It runs pretty fast in the cloud, and users are finding it to be as capable as MATLAB.
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?

Challenge: operation of a tightly-coupled operation across hemispheres It is preliminary to speak of lessons lesson learned, as LSST is in construction. However, accurate and detailed model to effectively communicate, coordinate and maintain the ability to trace CI features to the requirements and business need. Is an area of focus which LSST feels will help meet this challenge.
Key Risks
For this project, since the CI is all at TACC, there is not much risk.
What CI-related workforce development activities does your facilities engage in?
We provide roughly monthly training webinars which are recorded and then made available persistently on YouTube. We also have summer programs for high school students - this year they built an instrumented model, experimented with that model on a shake table, and then analyzed their results using our CI.
What do you see as your key new CI requirements and challenges in the next 5-10 years
Performance is the priority, since web data transfer and remote use of interactive tools like MATLAB are slower than on a local laptop. Also expanded simulation and data analysis/visualization capabilities on the web portal so that we capture all researchers in this community.
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
LSST	Don Petraivck, NCSA - UIUC Jeff Kantor, William O'Mullane	Petravick@illinois.edu
What percentage of the facility CI was developed in-house versus by reusing existing solutions?		
R: LSST is in construction, but the following are underway, LSST has funded the development of a significant, high bandwidth network between Chile and the United States. LSST is developing QSERV, a spatially shared database which is anticipated to require 40 PB of disk provisioning, over 250 node by 2025.		
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?		
- LSST Uses HT-CONDOR for the basis of its production system. HT-Condor is a standard in thoughput computing, is used in LHC and the Dark Energy survey. HTCondor supports the various batch use cases identified in LSST. LSST has had a collaborative engagement with HTCondor for many years. LSST has used XSEDE and Blue Waters during its pre-construction phase for demonstrations of feasibility of its production system, and has used simulation data generated on the Open Science Grid. – These were the obvious choices dues to agency support and availability. LSST has built upon authentication and authorization system work that is also in use in LIGO. The reason is that the system supports a variety of authentication and authorization protocol, and interoperated with Incommon. National education and research identity federations are seen as useful source of identity information for LSST, where the class of all US and all Chilean professional astronomers have data rights. LSST's Master Information Security Plan was developed in Consultation with the CTSC. CTSC was selected due it is knowledge of contemporary security standards, as applied to NSF projects. LSST's science user interface is based on the Firefly Tool Kit developed at IPAC at Caltech. This is a commonly used advanced toolkit used within Optical Astronomy. Rucio, a component developed at CERN for the LHC is being evaluated for internal file synchronization, as is Pegasus for the production workflows. Both of these components were selected due to their use with similar use cases in other experiments.		

Jupyter is a foundational component to support internal quality assessment and to support exploitation of the data at the UN and Chilean LSST Data Access Centers. Jupyter is a well-supported method of exposing aspects of a facility in a structured way to a large group of users.

BRO is use for intrusion detection at the LSST Chilean sites, and at NCSA. BRO is selected for us utility in being an intrusion detection system where large volumes of data re transferred between sites, and sue to the body of expertise with the system at NCSA

List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?

- LSST Uses HT-CONDOR for the basis of its production system. HT-Condor is a standard in throughout computing, is used in LHC and the Dark Energy survey. HTCondor supports the various batch use cases identified in LSST. LSST has had a collaborative engagement with HTCondor for many years.

LSST has used XSEDE and Blue Waters during its pre-construction phase for demonstrations of feasibility of its production system, and has used simulation data generated on the Open Science Grid. – These were the obvious choices dues to agency support and availability.

LSST has built upon authentication and authorization system work that is also in use in LIGO. The reason is that the system supports a variety of authentication and authorization protocol, and interoperated with Incommon. National education and research identity federations are seen as useful source of identity information for LSST, where the class of all US and all Chilean professional astronomers have data rights.
LSST's Master Information Security Plan was developed in Consultation with the CTSC. CTSC was selected due it is knowledge of contemporary security standards, as applied to NSF projects.
LSST's science user interface is based on the Firefly Tool Kit developed at IPAC at Caltech. This is a commonly used advanced toolkit used within Optical Astronomy.
Rucio, a component developed at CERN for the LHC is being evaluated for internal file synchronization, as is Pegasus for the production workflows. Both of these components were selected due to their use with similar use cases in other experiments.
Jupyter is a foundational component to support internal quality assessment and to support exploitation of the data at the UN and Chilean LSST Data Access Centers. Jupyter is a well-supported method of exposing aspects of a facility in a structured way to a large group of users.
BRO is use for intrusion detection at the LSST Chilean sites, and at NCSA. BRO is selected for us utility in being an intrusion detection system where large volumes of data re transferred between sites, and sue to the body of expertise with the system at NCSA

1) Upgrading the north south network from LaSerena, Chile to NCSA in the context of a MREFC project.
2) Dealing with the evolution of processors, in particular the reduction of the amount of memory per core, and the need to increase the level of threading in LSST Codes.
3) Selecting the technologies needed to support end users in the data access center.

What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?

Challenge: operation of a tightly-coupled operation across hemispheres
It is preliminary to speak of lessons lesson learned, as LSST is in construction. However, accurate and detailed model to effectively communicate, coordinate and maintain the ability to trace CI features to the requirements and business need. Is an area of focus which LSST feels will help meet this challenge.

Key Risks

Changes in computing platforms over the remaining period of construction and operations through 2034 are a concern. LSST has data processing access and archive facilities in three continents. For each continent the pace of sustainable change will vary. For example, we expect cloud computing to lag in South America. The response to these challenges includes providing software isolation layers, for example Kubernetes, which can be deployed in locally provisioned or in commercial systems.

We currently use could services for software build and test. The EPO component of LSST has a very large cloud deployment component. Our baseline thinking allows for use of cloud services for disaster recovery, for opportunistic bulk computing, and for elastic expansion of the US Data Access centers. Our baseline may evolve as construction proceeds.

What CI-related workforce development activities does your facilities engage in?

Project staff attend workshops and conferences. At NCSA significant work in CI is performed by NCSA staff. NCSA has a program of work to develop the HPC workforce, including responding to NSF calls for proposals for training Cyber Infrastructure Professionals. Additionally, NCSA has a program of research and supporting its infrastructure, including operational security group, support for the Linux Cluster Institute (LCI), which trains Infrastructure professionals.

What do you see as your key new CI requirements and challenges in the next 5-10 years

Keeping the CI efforts in Chile and the in the US coordinated and with a like technology base. Changes in CI technologies and how CI is absorbed by the project. LSST has obligations to provide computing facilities in Chile, where for example cloud functionality is not equivalent to the functionality available in the US.
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
National Optical Astronomy Observatory (NOAO)	Sean McManus	mcmanus@noao.edu

What percentage of the facility CI was developed in-house versus by reusing existing solutions?
data reduction pipeline (DEC Community Pipeline); TADA (Telescope Automatic Data Archiver); yes these tools are mostly open-source
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?
Scientific Linux, IBM General Parallel File System, Puppet, Foreman, Libvirt, Django. The criteria used to select tools varies. For some open-source tools, there is minimal investment needed to try something, and therefore doesn't require a formal selection process. For paid software contracts, there is obviously more vetting by internal IT staff, management, and procurement. As part of normal vetting we try to look at what is working / not working for other peer organizations inside and outside of AURA.
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?
1) Mass storage: We require inexpensive storage on the multi-Petabyte scale to store astronomy data products; 2) Bandwidth: Reliable, fast bandwidth across continents is needed to move data from telescope to archive; 3) Software: The software stack must meet operational requirements but also be sustainable inside flat or shrinking budget envelope.
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?
For small departments, it is difficult to achieve a balance of experience versus motivation and familiarity with cutting edge tools. Low staff turnover can result in staff being settled on one particular technology, and lagging behind recent developments in IT. On the other hand, it's

not cost-effective to react to the latest/greatest thing that comes out every year. A balance of new versus proven tools must be made.
Key Risks
workforce reduction due to budgets, even a small one, could have significant impact.
What CI-related workforce development activities does your facilities engage in?
We budget for continuing education, but whether or not staff participate is voluntary
What do you see as your key new CI requirements and challenges in the next 5-10 years
transition from NOAO/LSST/Gemini to NCOA
Do you have any other suggestions for the workshop?
n/a

Affiliation	Name	E-mail
LIGO	Stuart Anderson, Caltech	stuart.anderson@ligo.org

What percentage of the facility CI was developed in-house versus by reusing existing solutions?

All of the following in-house CI components are available for reuse:
* LIGO Data Replicator (bulk data transfers)
* Metadata databases and tools designed for GW observations
* low-latency data distribution on large clusters
* Data Monitoring Tools
* low-latency transient event alert system
* Network Data Server
* Web and Matlab based Data Viewer tools
* GW Detector status monitoring service
* GW detection and parameter estimation pipelines
* Library of gravitational wave algorithms
* LIGO Open Science Center notebooks
* Job accounting system

What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?

* HTCondor/Pegasus/BOINC
* OSG
* Docker/Singularity/Shifter
* CVMFS/StashCache/Xrootd/GridFTP
* Shibboleth/Grouper/CILogon/Kerberos/LDAP/GSI
* Oracle HSM/ZFS/HDFS
* GitHub/GitLab/Travis/Jenkins
* JupyterHub

These tools where predominantly identified by first recognizing a need and then charging a small group to research (sometimes a self-forming group) to research what is currently available. In some cases that group takes a solution to full scale prototype (build it and they will come), and in others the alternatives are presented to a LIGO computing committee to evaluate the pros and cons first.and Matlab based Data Viewer tools*GW Detector status monitoring service*GW detection and parameter estimation pipeline*Library of gravitational wave algorithms*LIGO Open Science Center notebooks*Job accounting system

List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?
* Identity and Access Management was a challenge during the early phases of LIGO, leading to significant loss in productivity due to unnecessary barriers to efficient access to needed information and systems. Integrating Shibboleth, Grouper, InCommon, and CILogon into LIGO's CI has been a game changer. Investing in I&AM early on in a project is highly recommended. * In the early years of LIGO attempts to use OSG to run LIGO data analysis tasks failed. In the last few years this has become a major success, in part due to more mature tools for managing data intensive workflows (e.g., Pegasus, CVMFS, and containerization), and in part due to more mature gravitational wave data analysis pipelines. * LIGO initially invested in a home grown job execution environment that attempted to minimize the amount of code needed to be developed by scientists performing searches for gravitational waves.. However, that proved in practice to be insufficiently flexible and the pendulum swung over to allowing scientists to develop arbitrary a.out executables managed by HTCondor. In hind site, the optimum would have been somewhere in-between.
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?
* Integrating CI with international collaborators remains a significant challenge.. OSG has recently provided a major breakthrough for providing a uniform interface to plan and execute LIGO workflows on international computing resources. However, international federated I&AM remains a significant challenge for LIGO. * Finding the right set of CI to support both tightly controlled production data analysis and allowing creative new ideas be developed is a challenge.
Key Risks
* Funding for CI experts that support scientific personnel to use existing CI * Sustainability of CI and being able to effectively identify new CI that will be available in the long-term before investing limited internal resources.
What CI-related workforce development activities does your facilities engage in?
* Sending students to summer schools and similar training opportunities. * Sending professional staff to conferences and workshops. * Inviting external experts to provide training at internal scientific meetings.

What do you see as your key new CI requirements and challenges in the next 5-10 years
* Inter-federation agreements that comply with international privacy laws while still releasing enough information to be useful for international scientific collaborations. * Training the teachers. As most of the workforce comes from academic research groups how do we train academic faculty to be able to train their new students to use modern CI. * long-term stability of software packaging and distribution that will allow reproducibility of scientific results on an interesting time scale.
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
LIGO	Albert Lazzarini, Caltech	lazz@ligo.caltech.edu
What percentage of the facility CI was developed in-house versus by reusing existing solutions?		
Please see white paper submitted by Stuart Anderson for all attendees from LIGO		
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?		
Please see white paper submitted by Stuart Anderson for all attendees from LIGO		
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?		
Please see white paper submitted by Stuart Anderson for all attendees from LIGO		
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?		
Please see white paper submitted by Stuart Anderson for all attendees from LIGO		
Key Risks		
Please see white paper submitted by Stuart Anderson for all attendees from LIGO		
What CI-related workforce development activities does your facilities engage in?		
Please see white paper submitted by Stuart Anderson for all attendees from LIGO		
What do you see as your key new CI requirements and challenges in the next 5-10 years		
Please see white paper submitted by Stuart Anderson for all attendees from LIGO		
Do you have any other suggestions for the workshop?		

What is the appropriate scale and relationship among large NSF computing facilities, computing facilities that are part of e.g., physics large facilities and MRI resources provided to individual collaboration institutions? Does NSF have a policy on these?

Affiliation	Name	E-mail
ARF	Jon C. Meyer, UC San Diego	jmeyer@ucsd.edU
What percentage of the facility CI was developed in-house versus by reusing existing solutions?		
we are in the process of developing data delivery via modern message queue and welcome the opportunity to collaborate and have others reuse.		
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?		
Some vendors' tools are used due the demand for certain types of data to be regularly produced during a seagoing mission		
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?		
Uninterrupted Internet connectivity. Research vessels at sea need consistent, reliable communication paths to be able to produce scientifically interesting data in near to real time.		
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?		
Key Risks		
What CI-related workforce development activities does your facilities engage in?		
Some specialized and general computing-related training.		
What do you see as your key new CI requirements and challenges in the next 5-10 years		

High-speed, realtime delivery of data from the ocean. Ability to interact with field researchers seamlessly from
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
Gemini	Chris Morrison, Gemini Observatory	cmorrison@gemini.edu

What percentage of the facility CI was developed in-house versus by reusing existing solutions?
none (note that we do not include software in our definition of CI)
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?
Google apps for business; Amazon web services; zoom conferencing services. Identified in all cases by industry surveys & best practices; selection via requirements analysis, in some cases usability analyses, and value for money.
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?
Challenges: 1. Netapp storage. Large impact if this redundant system fails. 2. Backup storage infrastructure. Expensive, complex and requires significant expertise. 3. Remote access connectivity. Brings user management and security concerns. Best practices: 1. Gemini infrastructure has significant redundancy, as a result of lessons learned in previous failures. 2. Use of cloud service (AWS) for large-scale data archiving and access. 3. CI replacement policy on equipment at end of warranty.
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?
Challenges & gaps: see above. Lessons to share: Redundancy (storage, networking, VM clusters, connectivity). Lessons to learn in the meeting: offsite storage methods & data retention.

Key Risks
Dependencies: Access to Google (for business applications); AWS (for archive storage) - low likelihood, high impact risks. Mitigation: Redundant network links in Hawaii and Chile. Backup plan for an extended outage of AWS would be to bring the archive in house temporarily until service restored.
What CI-related workforce development activities does your facilities engage in?
Enterprise specialist training courses and certifications.
What do you see as your key new CI requirements and challenges in the next 5-10 years
Challenge: Integration of Gemini CI into a larger Center, and aligning services with other Programs in that Center. We do not see significant changes in the technical challenge for Gemini CI, as the telescopes will not fundamentally change the way they operate at night.
Do you have any other suggestions for the workshop?
1. Future role of NSF in coordinating or providing CI through grant funding. 2. Large-scale science data storage and access via cloud services - best practices.

Affiliation	Name	E-mail
DKIST, NSO	Steve Berukoff and Eric Cross, NSO	sberukoff@nso.edu ecross@nso.edu

What percentage of the facility CI was developed in-house versus by reusing existing solutions?
For the DKIST telescope Built In-House • Instrument Control Systems • Facility Control Systems • Telescope • Enclosure • Environmental • Adaptive Optics, Wavefront Control • Coude • Safety Systems • Are these useful to other CI organizations? Unclear if they would be useful elsewhere.
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?
• Open Source software; given budgetary constraints DKIST CI is leveraging Open Source where applicable. The deployment of Open Source is centered within the Infrastructure layers. • Globus GridFTP will be ustilized to move data from the telescope on Maui to the Boulder Data Center. • CEPH object storage for long-term data storage
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?
• Complexity of DKIST Instruments has driven a flexible but customizable approach to instrument controls. • Data network management has provided a challenge to DKIST. We have network Interconnects between the DKIST Facility on Maui, the University of Hawaii, the University of Colorado, and also leveraging Internet2. • Complexity of DKIST Instruments has driven a flexible but customizable approach to

instrument controls. • Data network management has provided a challenge to DKIST. We have network Interconnects between the DKIST Facility on Maui, the University of Hawaii, the University of Colorado, and also leveraging Internet2. • The combination of Petascale data volume under a very constrained budget challenges the ability of the CI to support its community. Best Practices • Because of the distributed nature of the program with multiple product owners following Systems Engineering practices for developing effective requirements and interface controls.
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?
• Ensuring the end to end CI design from Facility Control, Data Acquisition and end-user distribution is built-in to the overall design and budget.
Key Risks
• Operational funding levels should allow appropriate maintenance to be completed with appropriate personnel. • Long-Term operational lifetimes mandate avoidance of monolithic architectures. Mitigation • Ability to build infrastructure building blocks by developing a roadmap for DIBBS awards.
What CI-related workforce development activities does your facilities engage in?
• Professional development conferences
What do you see as your key new CI requirements and challenges in the next 5-10 years
• Ensure we can deliver the scope that we need to support our community.
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
ARF	Suzanne Carbotte, Columbia University	carbotte@ldeo.columbia.edu

What percentage of the facility CI was developed in-house versus by reusing existing solutions?
R2R has developed a network file system for storage of data and documents; a relational database for storage of associated metadata; a Web portal for search, browse, and download; scripted tools for data cataloging, archiving, processing, and assessment; and a suite of Web services for interoperability. Most are built on existing open-source software such as PostgreSQL, Apache HTTP/Tomcat, MapServer, etc. Selected tools for data processing have been released in the public domain via GitHub.
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?
R2R uses commercial provisioning in selected cases for Web service hosting (Linode.com), domain services (Site5.com), and deep storage (Amazon Glacier).
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?
1. R2R's network file system is the heart of its daily operation, used for both internal processing workflows and serving content to the Web. The file system is built on a suite of FibreChannel storage arrays, switches, and Linux servers. 2. R2R's "NavManager" software package is used routinely to create a suite of quality-controlled shiptrack navigation products, which are reused by downstream QA processes and Web services. 3. R2R's "Linked Data" server disseminates the Cruise Catalog in a standards-compliant format, which is harvested by other geoscience data repositories as well as by global search indexes such as Google. What aspects about the facility CI and its operation would you like to share as best practices? It is not uncommon to revisit old(er) data packages, in order to extract additional information and/or refine quality assessment. Maintaining data packages on spinning disk for a 5 or more-year sliding window has proven advantageous, and can be sustained using (less expensive) HDDs rather than SSDs. Every digital resource published online (vessel, cruise, dataset, document, sample, person,

award, etc) should have a globally unique persistent identifier. This enables interoperability with other repositories, reliable citation, and linking to the scientific literature.
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?
The volume of environmental sensor data being produced by modern research vessels, is increasing faster than the disk storage capacity that can be deployed with affordable enterprise-grade local equipment. Commercial provisioning provides an affordable solution for deep storage, but not for local data processing or egress. Academic provisioning via systems like XSEDE is difficult because the resources are disjointed and constantly evolving, and carry the risk of abrupt termination when the grant period ends. Data transfer is also hampered by local campus network bandwidth. While progress has been made toward standardization, the US. academic fleet still produces data in a very heterogeneous manner. Each cruise is unique. Significant manpower is still required to stay abreast of changing directory structures and file formats, and to recover from operator errors.
Key Risks
Maintaining local server, storage, and network infrastructure remains an ongoing challenge, especially with the increased need to provide monitoring, metrics, and network security. Commercial provisioning shifts resources from a local to a remote location, but does not eliminate the need for a system administrator and does not reduce costs.
What CI-related workforce development activities does your facilities engage in?
R2R staff attend annual community meetings such as ESIP, RDA, and RVTEC, to stay abreast of emerging technologies. Junior staff work in tandem with senior staff, receiving on-the-job training.
What do you see as your key new CI requirements and challenges in the next 5-10 years
The ability to store and move large volumes of data as environmental sensors continue to evolve faster than storage/network resources; the lack of "smart" self-documenting sensors; and the lack of designated long-term archives for some data types remain significant challenges.
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
National Center for Atmospheric Research (NCAR)	Aaron Andersen, UCAR	aaron@ucar.edu

What percentage of the facility CI was developed in-house versus by reusing existing solutions?

A number of components of the CI were developed in house. A few concrete examples include:
- Research Data Archive services - public interface can be found at: https://rda.ucar.edu/
- Parallel Python tools for post production of NetCDF files and specifically climate data: https://www2.cisl.ucar.edu/tdd/asap/parallel-python-tools-post-processing-climate-data
- System Accounting Manager (SAM) on HPC systems https://www2.cisl.ucar.edu/user-support/systems-accounting-manager (currently NCAR specific)
- VAPOR is the Visualization and Analysis Platform for Ocean, Atmosphere, and Solar Researchers. VAPOR provides an interactive 3D visualization environment that can also produce animations and still frame images https.://www.vapor.ucar.edu/
- NCAR Command Language - NCL is an interpreted language designed specifically for scientific data analysis and visualization.

All tools were primarily developed with the needs of the Atmospheric science community in mind. All components are available for reuse except for SAM. SAM could be customized and utilized by others but would require some generalization or site specific customization.

What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?

A good number of external CI capabilities and/or externally developed tools are in use at NCAR within the Computing and Information Systems Lab (CISL).. Highlights include:
- NCAR Data Sharing Service - Globus Toolkit - https://www.globus.org/
- NCAR also utilizes XDMoD as part of the suite of tools used to manage the HPC resources - http://open.xdmod.org/

Within the NCAR Wyoming supercomputing center two commercial packages are in use to control, manage and monitor the facility.
- The core of the facility utilizes Building Automation, hardware, software and sensors from Johnson Controls Inc. based on the Metasys Building Automation System http://www.johnsoncontrols.com/buildings/building-management/building-automation-systems-bas
- More recently NCAR has deployed an advanced system to allow higher fidelity sampling of

the electrical infrastructure. Those components were provided by Schneider Electric Software LLC. under their Wonderware brand.
These two commercial packages were purchased utilizing a formal RFP process and were evaluated by a technical team, business team and pricing team. Technical requirements were developed in partnership with external engineering firms.

List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?

The three most used CI components are the High Performance Computing systems, High Performance Disk Storage (GLADE) and the tape archive HPSS. The HPC systems are regularly see greater than 90% system utilization. GLADE similarly has been exceptionally popular providing common shared space across HPC, data analysis and visualization platforms. Finally the HPSS based archive system is still the cornerstone of data archival at NCAR and in some respects is too popular:
- HPC systems utilize test and development hardware that is much smaller scale but provides capabilities to not impact production work while upgrading, patching or adding new tools to the user environment. Once changes to the test environments are stable the teams can then upgrade or change the large HPC environments. Here complexity and scale provide significant challenges.
- The GLADE environment is technically challenging providing a very large (50PB) high performance InfiniBand storage environment. However the technical challenges are only one component of the environment, user retention policies and management of quotas are equally as challenging.
- HPSS presents a more financial challenge. Historical archival storage policies were predicated on computing being expensive but storage being cheap. Currently those economic assumptions are no longer valid and CISL has embarked on modifications to storage policies. That effort is too new but may become a best practice.

What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?

We see human capital as possibly one of our most challenging areas currently. Expertise in HPC, large data storage and IT environments are in high demand. We often find recruiting staff a challenge especially where some areas like data analytics and data science are in significant demand in the commercial as well as research sectors. Keeping pace with salaries in a challenging federal environment is proving difficult.
Closer to the facility operation level we are seeing highly dynamic HPC energy consumption based on computing workloads. All HPC vendors are actively pursuing power saving capabilities all the way down to the chip level, turning down clocks or components on demand. Overall this is a good thing as computing systems of the past were notoriously

wasteful. However, computing components that turn up and down on computing timescales (sub seconds) may not be a match for traditional building automation systems or more broadly utility providers. Large changes in electrical demand influence mechanical cooling systems as well as the capacity of the utility. The NWSC has a highly energy efficient design that adapts to the demands of the CI housed in the facility.
Key Risks
Workforce development, recruiting and retention are a significant risk.
What CI-related workforce development activities does your facilities engage in?
NCAR has a number of efforts underway as we see workforce development as critical. The NWSC has been utilized as a teaching laboratory with 7 summer interns over the last 5 years working within the facility. Within that timeframe, 3 women and 2 minority students have been through three-month intensive summer internships. All but two of those students have remained in fields engaged with large CI. CISl also manages the Summer Internships in Parallel Computational Science (SIParCS). The goal of the SIParCS program is to make a long-term, positive impact on the quality and diversity of the workforce needed to use and operate 21st century supercomputers. Graduate students and undergraduate students (who have completed their sophomore year by summer 2017) gain significant hands-on experience in high-performance computing and related fields that use HPC for scientific discovery and modeling. More recently the Operations Manager at the NWSC has been engaged as part of the state of Wyoming Workforce Development Council. Wyoming in particular is looking to develop greater inroads specific to large computing facilities with more traditional trades, community colleges and non-traditional students.
What do you see as your key new CI requirements and challenges in the next 5-10 years
Specific to modeling and simulation we see a highly disruptive CI environment with significant computing architecture diversity on the horizon and new clear winners. Heterogeneous computing architectures are now commonplace but the complexity and scale remain challenging. There is also an explosion of data and data resources that has long been promised but we are starting to see with greater clarity. New methods such as machine learning offer some promise but there are many paths and options. NCAR certainly doesn't have the capability to explore all possible paths and will need to partner across many disciplines to find answers.
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
Incorporated Research Institutions for Seismology (IRIS)	Tim Ahern, University of Washington	tim@iris.washington.edu

What percentage of the facility CI was developed in-house versus by reusing existing solutions?
Most components have been developed in house over the 30 years life of the DMC. Of course commercial and open source software systems are used when appropriate such as DBMS software. Much of our infrastructure is somewhat domain specific such as reception of real time data and tools that work with domain specific data.
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?
We use commercial software for virtualization (VmWare), PostgreSql for DBMS software, commercial geolocation software. All external tools were acquired using IRIS purchasing guidelines, multiple bids etc.
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?
1) Web services, methods to abstract time series and metadata access both internally and externally 2) storage RAID indexing scheme to improve access to commodity RAID 3) Synchronization of data versions across multiple storage systems (1 primary and 1 secondary at each of the DMC and the ADC)
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?
Scalability. Access to seismological data can be episodic especially after earthquakes. Also certain preprocessing services can exceed our internal capabilities. The promise of cloud resources has potential but not yet realized.
Key Risks

Loss of key personnel and their knowledge. NSF budgets are making facilities like our more and more vulnerable.
What CI-related workforce development activities does your facilities engage in?
Both NSF and commercially sponsored training courses. We participate as time and financial resources allow
What do you see as your key new CI requirements and challenges in the next 5-10 years
Reducing the cost to maintain our infrastructure and finding external resources perhaps cloud, that can meet our demands and fit our way of doing business not theirs.
Do you have any other suggestions for the workshop?
Nothing at this time, not able to spend much time on this.....

Affiliation	Name	E-mail
UNAVCO	Fran Boler, UNAVCO	fboler@unavco.org
What percentage of the facility CI was developed in-house versus by reusing existing solutions?		
Essentially all components of UNAVCO's CI have been developed in house. This includes data handling for data arriving at UNAVCO from multiple varieties field instrumentation and from a variety of providers, archiving, and distribution functions. Most of the CI that aids in data handling is not available for reuse since it is highly customized. An exception is the GNSS preprocessing software tool called "teqc", which is widely shared with the community. Selected CI components have been developed in partnership with other institutions and are shared with them including SAR web services developed via the NASA SSARA project is shared with the Alaska Satellite Facility; and the Geodesy Seamless Archive Centers open source software was developed with NASA ACCESS support by UNAVCO with UCSD and NASA's Crustal Dynamics Data Information Systems. GSAC is widely shared.		
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?		
Certain proprietary software provided by sensor manufacturers for handling raw data are part of UNAVCO's CI. These are prescribed when a manufacturer is selected as a sensor provider. Much of UNAVCO's SAR data handling infrastructure is currently being migrated to the XSEDE cloud. Commercial cloud storage is employed as one of our backup strategies.		
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?		
The data systems that we operate (software and hardware) that receive, handle and deliver GNSS data to our external customer base have the largest user base and are used 24/7. We have been "saved" many times over by having failover systems at the ready for the inevitable hiccups in systems.		
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?		
A gap is lack of adequate resources to keep software and to a lesser extent hardware up to date. Functionality is regularly added through time as new component software systems, and this functionality is developed with technologies reflecting the era during which it was		

developed, with some attempt to see into the future; these components tend to remain part of operational infrastructure (we call them legacy components, but they are still key to accomplishing our tasks). All along the way technical debt is incurred, and of course technology moves ahead. This is a further challenge to moving capabilities to the cloud. We are trying to slowly and on a trial basis move components to the cloud. Legacy components are a further risk as it becomes increasingly difficult to find programmers with appropriate skillsets to maintain them. The priority is almost never to rebuild these older systems as long as they continue to operate. Another challenge is the wide variety of technologies in use in the Earth Sciences to meet CI needs of various domains. Trying to cover all bases is nearly impossible; trying to identify which technologies will emerge as most useful is a challenge for all. The EarthCube initiative is clearly exposing/highlighting this.
Key Risks
Key risks are related to the technical debt described in a previous section. Another key risk is looming retirement of staff members with decades of domain knowledge and in-depth knowledge of our CI components. Further, there is strong competition in our geographic area for skilled CI workers.
What CI-related workforce development activities does your facilities engage in?
We send staff members to training. We engage interns.
What do you see as your key new CI requirements and challenges in the next 5-10 years
Making use of the cloud (with appropriate return on investment). Continuing to track and identify trends in technologies and being able to respond nimbly. Managing functionality demands under resource constraints.
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
IceCube	Gonzalo Merino, University of Wisconsin Madison	gonzalo.merino@icecube.wisc.edu

What percentage of the facility CI was developed in-house versus by reusing existing solutions?
1) Data management software, handling data archive, transfer from the south pole and replication to long term archives. 2) Software framework to manage distributed workloads. Used to manage and bookkeep all the IceCube simulation production. In both cases, others could use, but this does not happen yet.
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?
1) South Pole broadband satellites SPTR, DSCS and Skynet. Provided by NASA, through USAP. This is the only available service for daily bulk data transfer from the South Pole. ~100Gbytes/day. 2) Tape storage for long term data archive. Provided by collaborating institutions NERSC and DESY-Zeuthen. These institutions already operate large scale automated tape facilities for several experiments. The service is offered as in-kind contribution to the Collaboration. 3) Open Science Grid. Providing access to millions of CPU hours in opportunistic resources. Also, operating core Grid services that provide us access to IceCube collaborating sites in Europe and Canada. We have been participating in OSG for several years. Distributed computing, and in particular opportunistic computing, represents a big advantage in our field where a lot of the data processing and analysis is pleasantly parallel. 4) XSEDE. Part of the IceCube simulation chain relies on GPUs. We started requesting allocations in GPU-capable XSEDE resources in 2016 to enlarge the computing capacity available for IceCube and increase the analysis potential. 5) Globus data transfer service (globus.org). Convenient data transfer service used to schedule/steer data transfers from UW-Madison to archive locations: NERSC and DESY-Zeuthen. Selected because it provided the needed functionality (integrity, retries, etc) currently at no cost. Also, interested in ongoing developments to interface more efficiently the HPSS tape system at NERSC with Globus (file integrity, performance).
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?

1) Main data processing cluster at UW-Madison. Large CPU and GPU cluster coupled to a multi-petabyte filesystem (Lustre) used by ~300 researchers to analyze the IceCube data. The most challenging part to operate is the storage, including monitoring, accounting, etc. However, operating our own Lustre cluster seems to still be the most cost effective solution for our size (~6 Petabytes of disk). 2) User-friendly scalable/elastic computing infrastructure: OSG and HTCondor have provided great capabilities so far in this front. However, we still see a lot of room for improvement in the user experience: higher efficiency, ease of use, interface to cloud resources, etc.
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?
Every time we have been able to leverage existing 3rd party services to build our infrastructure around them, we have seen benefits in doing that. From large archive storage facilities, to data transfer services, to workload management services, our lesson learnt is that it seems worth for us to invest on having a solid interface with existing services rather than trying to replicate them, or reinvent the wheel.
Key Risks
With the use of external services, there comes dependencies and risk. Mitigation strategies are therefore an important topic. In our case, several of these external services are coming from the academic ecosystem, so some coordination inside or between agencies could address part of the risk. Part of it would be ensuring that those common services that many researchers depend on, are sustainable.
What CI-related workforce development activities does your facilities engage in?
Assisting to various workshops and conferences in the field: NSF cyberinfrastructure, Open Science Grid, National Data Service ...
What do you see as your key new CI requirements and challenges in the next 5-10 years
Understanding how to best adapt IceCube analysis code to new emerging computing architectures and software frameworks such as manycore, GPU, FPGA, machine learning and data analytics frameworks, etc and engage the workforce with the required skills that we need to make this happen. Hiring and retaining this personnel is getting increasingly difficult as we compete head-on with the IT private industry.
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
NSCL	Andreas Stolz, Michigan State University	stolz@nscl.msu.edu

What percentage of the facility CI was developed in-house versus by reusing existing solutions?
Data acquisition and analysis software framework (NSCLDAQ/SpecTcl/DDAS), available to others. Controls software (EPICS) development, available to others. Business process software; custom and customized applications.
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?
Data acquisition (DAQ) and experimental data analysis on Linux based infrastructure. Commodity PCs/Servers. Storage using commodity hardware and ZFS/Linux. This is widely used, freely available software and low cost. DAQ is developed in-house. Analysis applications are typical freely available physics applications (GEANT, ROOT, etc.) Business process: ERP (IFS software), Sharepoint workflows and document management. Engineering software? Solidworks etc. Networking/Internet – external access provided by MSU
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?
Infrastructure – virtualization: Normal for enterprise infrastructure, but does require expertise for support. Sharepoint: Used for business processes, collaboration etc. Again requiring developer and administrator expertise. Security: Network and systems security including technical controls themselves and the workload around maintaining and documenting same. Adopting configuration management tools and testing deployment processes. System configuration – maintaining stable operations along with ongoing software changes and security updates.

What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?
Security is ongoing challenge.
Key Risks
Main risks are similar to any enterprise: security and disaster recovery.
What CI-related workforce development activities does your facilities engage in?
Participating in relevant workshops. CI Security training for all users.
What do you see as your key new CI requirements and challenges in the next 5-10 years
Providing increased data access to outside visitors and experimenters in face of increasing dataset sizes and security restrictions. Future DAQ systems for FRIB experiments.
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
International Ocean Discovery Program (IODP)	Jim Rosser, Texas A&M University	jrosser@tamu.edu

What percentage of the facility CI was developed in-house versus by reusing existing solutions?
Several CI components are developed and maintained in-house: instrument host data uploaders, web services, web science applications, databases, business applications (procurement, inventory, crew tracking). Yes, these are available to others for reuse, but, in most cases, would require extensive effort.
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?
Our approach is to focus on JRSO core competencies and leverage commodity services from other organizations when possible. For example, Texas A&M University provides many shared services that we use to support JRSO operations, including email; directory services; storage services; web conferencing; video streaming; software training; cloud storage; financial, travel and HR management systems; cybersecurity assessment tools; software procurement; project management assistance, etc.
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?
1. WAN (including VSAT) operations and support. Sustaining highly available WAN services is quite challenging when the research vessel (JR) operates globally. 2. Oracle ODAs. Oracle ODAs significantly increased JRSO database engine performance. However, there has been a steep learning curve for configuring and maintaining this capability. 3. Cybersecurity. Minimizing security risk while supporting international customers who bring many different personal devices onboard the JR and expect assured access to the ship's portfolio of science lab services (e.g., LAN, server storage, application and database services).
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?

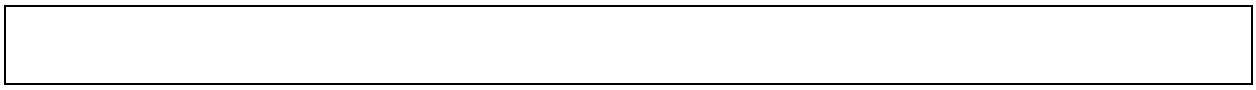
Minimizing security risk while supporting international customers who bring many different personal devices onboard the JR and expect assured access to the ship's portfolio of science lab services (e.g., LAN, server storage, application and database services).
Key Risks
Commercially available tools are increasingly cloud-based (e.g., Adobe Creative Suite, macOS apps, etc.). Our meager communication bandwidth supporting the JR rules those out. Yet, many software publishers provide no alternative. This issue is probably unique to facilities operating in low bandwidth, high latency environments, and probably also applies to organizations, such as DoD, that operate isolated networks (SIPRNet, JWICS, etc). This is a growing problem that continues to challenge us.
What CI-related workforce development activities does your facilities engage in?
Technology specific training for all aspects of infrastructure, software development and data management.
What do you see as your key new CI requirements and challenges in the next 5-10 years
Better WAN link for the JR. Adoption of automation/configuration management tools, such as Chef, Ansible, Salt, etc. Making data more discoverable.
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
CHESS	Werner Sun, Cornell University	wms8@cornell.edu

What percentage of the facility CI was developed in-house versus by reusing existing solutions?
Our high-availability clusters and Compute Farm were developed using commodity hardware and open-source software, assembled and configured in-house to meet the requirements of our facility. These configurations could be shared with other facilities.
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?
We provide CHESS users with remote data download capabilities using Globus. We selected this tool for its excellent performance and because of its widespread adoption in the NSF Large Facility community.
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?
High-availability Linux server clusters form the backbone of our CI. We use them for our central file systems, core infrastructure services, web and database servers, and hardware control systems. In commissioning these clusters, we gained experience with selecting free and open-source software and commodity hardware solutions without sacrificing reliability and performance. The CHESS data acquisition system is a central repository that receives raw data from multiple input streams and provides access for offline analysis and processing. We developed backup, archive, and rotation procedures to ensure disk access to two run-cycles' worth of data and tape retrieval for all previous data.
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?

We would be interested in learning about methods for provisioning temporary accounts and implementing fine-grained authorization for CHESS users.
Key Risks
We face an increasingly challenging cybersecurity threat landscape. We are always seeking ways to balance securing our facility control systems while maintaining usability, access, and productivity.
What CI-related workforce development activities does your facilities engage in?
Online tutorials, managerial and technical trainings.
What do you see as your key new CI requirements and challenges in the next 5-10 years
Upgrades to the scientific capabilities of the CHESS facility will result in increased data throughput and volumes, which will eventually exhaust a single system's ability to both serve as the data store and the access point. We may need multiple ingress and separate analysis systems.
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
PSC/CMU	James A. Marsteller	jam@psc.edu
What percentage of the facility CI was developed in-house versus by reusing existing solutions?		
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?		
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?		
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?		
Key Risks		
What CI-related workforce development activities does your facilities engage in?		
What do you see as your key new CI requirements and challenges in the next 5-10 years		
Do you have any other suggestions for the workshop?		



Affiliation	Name	E-mail
National Radio Astronomy Observatory (NRAO)	Brian Glendenning, NRAO	bglenden@nrao.edu
What percentage of the facility CI was developed in-house versus by reusing existing solutions?		
100% (based on open source software), yes		
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?		
Amazon AWS (modest), NSF XSEDE (experimental); Convenience/capability (AWS), cost (XSEDE)		
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?		
1. The CASA data reduction package is a large (2M SLOC) package both used for internal operations use and downloaded by facility users (2k downloads per year). 2. Our "pipelines" embed expert knowledge in a python scripting framework for automated science production. 3. Our computing infrastructure has multiple "archive" storage clusters, with attached Lustre and computational clusters for data processing. We have to take the long view - we have usable data from 40 years ago, our software packages live for decades.		
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?		
Keeping software packages reasonably high-performance over decades is an issue for us.		
Key Risks		

Durable agreements with HPC facilities, IaaS research clouds, International compatibility with user authentication mechanisms etc.
What CI-related workforce development activities does your facilities engage in?
Ph.D. student / Post-doc engagement with writing research codes. Summer / co-op students.
What do you see as your key new CI requirements and challenges in the next 5-10 years
See final bullet points in white paper.
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
Ocean Networks Canada	Benoit Pirenne	bpirenne@oceannetworks.ca

What percentage of the facility CI was developed in-house versus by reusing existing solutions?
The Oceans 2.0 was entirely developed in house, starting in 2005. The code is not in the public domain owing to the decision made by ONC to pursue commercial applications of the system.
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?
External tools include standard tools such as OS (Linux), Java, Javascript and attendant libraries; Oracle as an RDMS, Cassandra for non-relational data... ERDDAP was integrated to provide standard access to specific data types. Jira for supporting all aspect of the development, including time sheets and billing on a per project basis Confluence for internal and external documentation
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?
Until recently, the challenging elements included: - Cassandra: performance issues with the tool and the complexity of the fine-tuning required , Java memory allocation issues, difficulty with profiling complex code to understand where memory and time are actually spent, despite having an advanced test environment
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?
Continuously evolving the technology and the services available and getting the continued funding for the required manpower. Providing easy to use data discovery interfaces that will be addressing user needs in the face of growing instrumentation, observing locations and expanding time
Key Risks

Risks include-maintaining the level of funding to enable continuous improvements to the facility: a CI is never over! Mitigation requires making management and funding agencies understand that.
What CI-related workforce development activities does your facilities engage in?
We have had large fractions of the team of 20+ software engineers attend classes in: - the Agile Scrum methodology - usability - Kaisen
What do you see as your key new CI requirements and challenges in the next 5-10 years
- As the facility continues to grow, a continuous emphasis on verification of our scalability, and possible adaptation will be necessary. - The support of multiple clients, re-organizing into a multi-project based entity - Need to support critical customers (e..g, Public Safety) with defined SLAs
Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
Oregon State University, College of Earth, Ocean, and Atmospheric Sciences, Regional Class Research Vessel Program	Christopher Romsos	cromsos@coas.oregonstate.edu

What percentage of the facility CI was developed in-house versus by reusing existing solutions?
The most significant CI component built in-house is our "datapresence" system. In a nutshell, the datapresence system captures and archives data from resident (or visiting) sensors, replicates the information shoreside, and presents the information to both the shipboard and shoreside science parties for use/consumption. The datapresence system includes functionality for data quality assessment, flagging, alert and user notification. Other CI components developed in-house include several databases for project management including a risk-register database application. Yes, these components are available for others to use.

What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?
There is a high likelihood that the most if not all RCRVs shall be provisioned with satellite service through HiSeasNet at UCSD (https://hiseasnet.ucsd.edu/), though some UNOLS ships are experimenting with going out and negotiating their own contracts for satellite service opting (out of the HighSeasNet program in areas where better deals can be struck such as the Gulf of Mexico). We, the RCRV datapresence developers, are currently formalizing an MOU with Leidos Antarctic Support contractors to share components of our acquisition and visualization code. Part of this process includes choosing an open source license under which to distribute software. Lastly, we've incorporated data and map services (hosted locally aboard the ship) from the Marine Geoscience Datasystem at Lamont-Doherty Earth Observatory (LDEO) into our real-time displays for scientific situational awareness. Specifically, the Global Multi-Resolution Topography Data Synthesis provides our base layer for the map interface http://www.marine-geo.org/portals/gmrt/ Other sources of thematic background information for this interface are provided by NOAA Fisheries, Office of Coast Survey, USGS, and various academic sources.

List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?

1) Ship to shore (and back) data replication over high latency, low bandwitdh satellite networks. This problem, akin to the Long Fat Network problem of high bandwidth-delay product, is the most challenging issue that we are working on. We've had good success in increasing our throughput by optimizing the TCP window and buffer sizes and are now looking at managed WAN optimizatoin solutions to provide this service.
2) Cybersecurity is another challenge for the project. The RCRVs shall be equipped with integrated monitoring control systems to cover everything from bridge to engine room systems. Securing these online systems is a priority and a challenge.

What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?

At this project phase (construction) we don't yet have lessons learned to share.

Key Risks

Key risks include security and expertise. As indicated the RCRVs shall present a significant CI advancement from current. To mitigate each of these risks we have an operations plan that includes support and oversight (budget and personnel) from a Class Management Office. However, the level of expertise for the technical support personnel (Marine Technicians) that sail with the ships will have to rise. Evidence to support this expertise risk can be gleaned from organizations that have recently taken operations responsibility for new research vessels.

What CI-related workforce development activities does your facilities engage in?

Ah, a perfect follow-up question. A key component of our operations plan during transition to operations and post-delivery under Class Management will be technology transfer and training for new operators. We expect much of this initial ' workforce development' to take the form of hands on work during transition but additional training will be made possible through the Class Management Office during operations. In addition to periodic training we have staff that shall travel to each vessel on a rotating schedule (multiple visits per year) to inspect sensor systems, perform calibrations and maintenance, as well as conduct specific training while on a site visit.

What do you see as your key new CI requirements and challenges in the next 5-10 years

BYOD IoT sensors - We must keep abreast of security and integration issues these devices present.
On-Prem IaaS and PaaS - These industry trends or options are attractive but difficult to implement under the current model of support and operations (see expertise risk above).
Cybersecurity - Particularly as it applies to on-board integrated monitoring and control systems.

Do you have any other suggestions for the workshop?
Not at this time

Affiliation	Name	E-mail
Florida International University	Julio Ibarra	julio@fiu.edu
What percentage of the facility CI was developed in-house versus by reusing existing solutions?		
N/A		
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?		
N/A		
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?		
N/A		
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?		
N/A		
Key Risks		
N/A		
What CI-related workforce development activities does your facilities engage in?		
NA/		
What do you see as your key new CI requirements and challenges in the next 5-10 years		
N/A		

Do you have any other suggestions for the workshop?
N/A

Affiliation	Name	E-mail
2-Dimensional Crystal Consortium, Pennsylvania State University	Yuanxi Wang	yow5110@psu.edu
What percentage of the facility CI was developed in-house versus by reusing existing solutions?		
N/A		
What external CI capabilities and services and/or externally developed tools (if any) does the facility use and who provides them? How were these tools identified and what criteria was used to select the tools?		
N/A		
List up to 3 of your most and least favorite CI components with a 1 sentence explanation for each. What aspects about the facility CI and its operation would you like to share as best practices?		
N/A		
What aspects of the facility CI and its operation do you see as challenges/gaps? Are there any pitfalls/mistakes you would like to share? What aspects would you be interested in outsourcing?		
N/A		
Key Risks		
N/A		
What CI-related workforce development activities does your facilities engage in?		
N/A		
What do you see as your key new CI requirements and challenges in the next 5-10 years		
N/A		
Do you have any other suggestions for the workshop?		

N/A