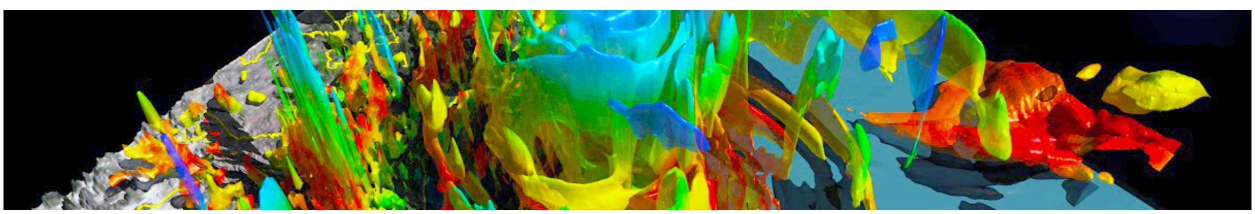
CISL Computational & Information Systems Laboratory

# Strategic Plan

## 2015–2019

## Service

## Science



## Education

NCAR

NSF

# CISL Strategic Plan, 2015–2019

*Accelerating discoveries in the Earth System sciences*

## Introduction

NCAR's Computational and Information Systems Laboratory (CISL) is a leader in supercomputing and data services necessary for the advancement of atmospheric and geospace science. CISL's mission is to remain a leader at the forefront of ensuring that research universities, NCAR, and the larger geosciences community have access to the computational resources they need for their research. CISL's portfolio also includes computational science research targeted at atmospheric sciences, in an interdisciplinary manner that incorporates computer science, mathematics, statistics, and software engineering. Further, to fulfill the need for a stronger workforce at the intersection of HPC and geoscience problems, CISL engages in education and outreach activities to inspire and attract a diverse future workforce.

CISL's 2015–2019 strategic plan addresses the computing and data challenges set forth in NCAR's grand-challenge objectives of predicting "atmospheric, chemical, and space weather hazards and their impacts on ecosystems, people, and society" and predicting and projecting "the consequences of natural and anthropogenic climate variability and change at regional and global scales," while meeting the interdisciplinary challenges of the National Science Foundation's (NSF) Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21 Vision). These science goals will require converting exploratory "proof-of-concept" simulations into routine forecasts, enhancing model resolution to study larger physical regions in more detail, successfully conducting large-scale modeling campaigns like the Climate Modeling Intercomparison Program (CMIP), and analyzing and inter-comparing massive observational and model data produced by CMIP and ever-more-capable sensors and computer systems.

The high-performance cyberinfrastructure of CISL Strategic Goal 1 directly supports NCAR Strategic Imperative 4 to develop and sustain advanced computing and data system services. CISL's Strategic Goal 2 to conduct applied computer science, mathematics, and statistics research supports NCAR Strategic Imperatives 1 to conduct innovative fundamental research to advance the atmospheric and related sciences, and 3 to develop, deliver, and support a suite of advanced community models. CISL's Strategic Goal 3 to invest in a spectrum of internship, visitor, outreach, training, and diversity programs contributes to NCAR's Strategic Imperative 6 to educate and entrain a talented and diverse group of students and early-career professionals.

The NSF's CIF21 Vision calls for pushing "beyond the current boundaries of knowledge to provide ever-deeper insights through fundamental disciplinary research by addressing increasingly complex questions, which often requires extremely sophisticated integration of theoretical, experimental, observational, and simulation and modeling results," then synthesizing "these fundamental groundbreaking efforts across multiple fields to transform scientific research into an endeavor that integrates the deep knowledge and research capabilities developed within the universities, industry, and government labs." CISL Imperatives 1.4,1.6, 2.1, and 2.2 support this vision by managing the computing, data, and collaborative challenges that must be met by CISL, NCAR, and the research community over the next five years and beyond.

CISL used many sources to help develop this strategic plan. CISL stakeholders in the atmospheric science community, respondents to a recent user survey, plus members of allocation panels, advisory panels, external reviewers, as well as the procurement Science Requirements Advisory Panel (SRAP) and the CISL Advisory Panel. Findings and Recommendations contributed significant inputs to this plan.

## CISL's Vision and Mission

*Take an already excellent supercomputing, data, and cyberinfrastructure experience and make it superb for the Earth System scientists who use our resources.*

Remain at the forefront of this vision and serve as a leader in:

- Expanding the productivity of researchers in the atmospheric, geospace, and related sciences through advanced computing and data services.
- Advancing the use of current and future computing systems by developing the computational methods and data systems needed by Earth System scientists.
- Inspiring and encouraging future generations to engage in Earth sciences by reaching out and creating excitement through discovery with hands-on computational experiences.

## CISL's Role in the Earth System Research Community

The Earth System research community depends on easy-to-use computing and data services to conduct its research. To support scientific objectives, CISL optimizes its computational services, adopts new technologies, and responds to university community needs. CISL operates a virtual organization of gateways, grids, and portals for sharing resources. CISL provides education, training, and outreach to support current and future scientists and engineers. CISL's discipline-specific mission focuses its development of tools and services for Earth System scientists to conduct their research. In all of three aspects of its mission, CISL seeks community input, collaborates with universities and other elements of NCAR, and maintains broad engagement with peer organizations operating on local, regional, national, and international scales.

## CISL Strategies

CISL's strategic goals are based on the three elements of its mission, each articulated with an overarching strategic goal that is addressed by imperatives and achieved through the actions specified. As part of NCAR, CISL supports, enhances, and extends the capabilities of the university community and the broader scientific community, both nationally and internationally. Throughout the next five years, CISL will prudently and creatively manage its budget to advance its imperatives and fulfill its mission.

### Strategic Goal 1: Advance Earth System science by expanding the productivity of researchers through high-performance computing and data services.

Computing is a foundational element of scientific research, and CISL provides a portfolio of advanced computing and data services specifically tailored for the atmospheric, geospace, and related sciences communities. At its core, CISL is a laboratory dedicated to providing advanced cyberinfrastructure to advance world-class science. CISL curates, manages, and archives a rich, growing set of digital collections to which it provides free and open access, and which attracts more than 12,000 individual researchers who annually download 1.5 petabytes of data through various pathways. Moreover, CISL creates essential, widely used software such as visualization and analysis tools; creates frameworks for implementing parallel modeling workflows; creates and operates science gateways; provides user support and training for all of these services; and integrates its resources and services with regional, national, and international cyberinfrastructure. CISL services evolve in response to changes in the underlying computational technologies and the scientific demands of the community, informed by the research and development activities performed under CISL Strategic Goal 2.

### Imperative 1.1: Acquire, deploy, and maintain cyberinfrastructure resources required by the community.

NCAR's integrated and balanced petascale computing, analysis, visualization, networking, and storage systems constitute a world-class supercomputing resource for about 2,500 researchers from institutions throughout the U.S. and abroad.

*Provide an integrated suite of production supercomputing resources*. CISL will operate and enhance its supercomputing environment at the highest levels of performance and availability to meet the data-intensive computing needs of the research community.

*Provide a test and exploration laboratory for rapidly evolving HPC technologies via the High Performance Futures Lab (HPFL)*. Through vendor partnerships, CISL has deployed cutting-edge technologies (e.g., memory/storage, processors, interconnect fabrics, and accelerators) to provide early experimental access to components in support of the large-scale NWSC-2 acquisitions. Future hardware will allow us to explore new technology impacts on application and workflows, optimize community models for better efficiency, and port models to future systems. The planned Phase II of the NWSC-2 acquisitions will require exploring emerging classes of data storage technologies. New technologies such as high-performance stacked and non-volatile memory, and hybrid solid-state and non-volatile disk technologies are rapidly transforming the storage landscape. These storage systems are poised to transform system architectures through tighter integration with computational elements. CISL will use the HPFL and vendor partnerships, such as the SGI-NCAR Joint Center of Excellence and the Intel Parallel Computing Center, to learn how to apply these technologies effectively to meet the data requirements for future systems and applications. Many-core capability is part of a planned Phase II acquisition, and HPFL efforts will evaluate competing many-core systems to guide CISL in choosing the best technologies for geoscience codes and understanding how these technologies perform in a production environment.

*Provide a centralized high-speed data storage resource*. The Globally Accessible Data Environment (GLADE) is a centralized high-performance file system that provides workspaces common to all CISL supercomputing, data post-processing, data analysis, visualization, and HPC-based data transfer services. CISL will refine and expand the GLADE architecture to optimize users' research efficiency by aligning scientific workflows with the flow of data.

*Provide an integrated data analysis and visualization (DAV) environment*. This environment improves scientific workflows by providing state-of-the-art systems tailored for the specialized needs of data post-processing, analysis, and visualization. CISL will expand the functionality and introduce more parallelism into these scientific workflows on the DAV resources.

*Provide a capacious and reliable data archive*. The High Performance Storage System (HPSS) provides data archive services for CISL's HPC environment, as well as for the divisional servers of other NCAR laboratories. Disaster recovery for irreplaceable data (e.g. Research Data Archive (RDA) and Earth

Observing Laboratory (EOL) datasets) is achieved by maintaining HPSS infrastructure at the NCAR Wyoming Supercomputing Center (NWSC) in Cheyenne, Wyoming and at the NCAR Mesa Lab in Boulder, Colorado.

### Imperative 1.2: Provide support services to all users of CISL resources.

CISL is committed to providing robust, accessible, and innovative user services and resources to researchers in the Earth System sciences, including the broad university community, the University of Wyoming and NCAR users. CISL also offers special computing campaigns for strategic objectives and field campaigns, and a limited number of extra-large computing grants to early users of new systems.

*Provide an integrated, cohesive environment tailored to user needs*. To supplement our traditional phone and email customer-support environments, CISL is enhancing self-help resources for customers. In particular, CISL will integrate accounting, allocation requests, and usage monitoring systems with its documentation and communication tools.

*Support strategic use of CISL resources.* CISL will provide a computing ecosystem that supports the requirements of special discipline-specific campaigns including on-demand compute campaigns through customized schedulers and allocation usage.

*Optimize model performance for current and future supercomputer architectures*. CISL will maximize the efficiency of supercomputing resources by working with NCAR's community model development teams, further optimizing their applications, and finding more parallelism in applications on current and follow-on systems. CISL's Strategic Parallel and Optimization Computing initiative (SPOC) tackles this challenge by improving model performance and efficiency on many fronts that include compiler tuning, system configuration, code vectorization and parallelism, communication and I/O optimization, and statistical validation methods that accelerate the acceptance of code improvements.

### Imperative 1.3: Sustain and enhance the NCAR-Wyoming Supercomputing Center.

CISL built the NWSC to provide advanced high-performance computing (HPC), data storage, and analysis capabilities into the future. CISL regularly enhances the capability and capacity of NCAR's supercomputing center while maximizing its efficiency, sustainability, and usability.

*Efficiently and reliably operate the NWSC*. CISL will continue improving NWSC efficiency, as measured by metrics such as its power utilization efficiency (PUE). Over the next five years, CISL expects to incrementally improve the NWSC's exemplary PUE. CISL also looks for ways to improve facility reliability by increasing the redundancy in certain electrical subsystems and employing more best-practices procedures for maintenance.

### Imperative 1.4: Provide the community with Big Data services.

CISL provides data tools and services for locating, accessing, transferring, and analyzing a variety of research data collections. CISL serves data to the community through data gateways over high-speed wide-area networks and via high-speed disk and near-line tape systems.

*Provide national and international leadership in research data stewardship*. CISL's Research Data Archive (RDA) contains more than 600 curated data sets of meteorological and oceanographic observations, operational and reanalysis model outputs, and remote sensing data to support scientific studies in climate, weather, Earth System modeling, and other related sciences.

*Expand the content of and access to the RDA*. The RDA contains more than 2.1 petabytes of reference data that supports weather and climate research by more than12,000 unique users per year with ocean and atmosphere observations, analyses, reanalyses, and operational Numerical Weather Prediction (NWP) outputs. CISL will improve the RDA by enhancing and augmenting the data content and by facilitating data access with new tools and web services. The RDA is also hosted on GLADE, making reference datasets immediately available for model initialization and validation in the HPC environment. This dual-purpose usage of GLADE is the most cost-effective way to serve both HPC and web-based users.

*Provide institutional leadership through the NCAR-wide Data Stewardship and Engineering Team (DSET).* CISL will play a key role in enhancing user access and discovery of NCAR's inventory of digital assets. In collaboration with NCAR Laboratory partners, CISL will pioneer the identification, prototyping, and deployment technologies needed to support discovery and access across the asset base.

*Advance the application of ensemble Data Assimilation (DA).* CISL's Data Assimilation Research Testbed (DART) software is increasingly important in geoscience initiatives that combine large sets of observations across ensembles of models. DART is a community facility for researchers at universities, government laboratories, and within NCAR's DA initiative for managing ensemble-based data assimilation. DART is also a significant research and development activity in CISL (see Imperative 2.1).

*Build and deploy science gateway and data sharing and analysis services*. CISL builds and operates science gateways (e.g., the Earth System Grid) that provide diverse scientific communities with access to data-sharing infrastructure. CISL gateways span climate science (e.g., Coupled Model Intercomparison Project (CMIP) Analysis Platform), regional climate change, arctic science, solar science, digital preservation, and international efforts to develop metadata and knowledge infrastructure. The CMIP Analysis Platform provides a one-stop shop for analysis and visualization of CMIP data on CISL's DAV resources. CISL is exploring opportunities to enhance this resource and other large data sets with web-based capabilities (e.g., server-side analysis, microservices, specialized queries) to enhance the overall usability and impact of this platform and other data resources across NCAR.

*Provide advanced visualization services to the scientific community.* CISL provides individual visualization services to help scientists better understand and communicate scientific findings to their peers, stakeholders, and the general public. These workflows must be efficient, able to handle large data sets, and use state-of-the-art visualization techniques. CISL will investigate and deploy appropriate collaborative visualization technologies in its Visualization Laboratory to enable distributed scientific teams to work together more effectively.

*Enhance the functionality and reach of data analysis and visualization software*. The core of CISL's visualization strategy further broadens the interdisciplinary and big-data impact of its widely used tools designed for the atmospheric and related sciences. These include the NCAR Command Language (NCL) data analysis tool with two-dimensional visualization capabilities, and the Visualization and Analysis Platform for Ocean, Atmosphere, and Solar Researchers (VAPOR) tool that provides interactive three-dimensional visualization and analysis capabilities. CISL's visualization strategy for these tools is to: 1) integrate NCL and VAPOR more closely to provide an intelligent storage format for scientific data; 2) integrate NCL with the massive scientific Python ecosystem, making NCL software accessible to a much wider user base while continuing to improve its performance and parallelism capabilities for big data applications; and 3) more tightly integrate the distribution, support, and training functions of the two tools.

## Imperative 1.5: Advance enterprise IT at NCAR and UCAR.

CISL's expertise in designing and provisioning cyberinfrastructure is essential to effectively managing and operating enterprise IT, supercomputing, cybersecurity, and communication services. CISL's robust networking technology is the literal backbone of IT resources and services that support UCAR's scientific mission.

*Provide networking services*. Networking infrastructure is vital to NCAR's HPC efforts and UCAR's ability to support scientific inquiry. CISL plans, engineers, installs, operates, maintains, develops strategy, and performs research for fast, reliable, and flexible networks that support data services at the campus, metropolitan, regional, national, and international levels.

*Formalize and enhance UCAR's cybersecurity capabilities*. UCAR's IT environment is a large and diverse set of compute, data, web, and network servers. These systems comprise vital scientific research platforms as well as business application systems. Over the next five years CISL will introduce new security practices aligned with community best practices, such as those defined by the Federal Information Security Management Act (FISMA).

*Deliver modern IT infrastructure, services, and platforms.* CISL supports the hardware and software infrastructure for many of UCAR's IT services (e.g., domain name service, web content management system, software revision control systems, and systems administration support). CISL will evolve these services to enable platform-as-a-service models; enhance and enable storage, data protection, and archive services; and provide and promote virtualized computing services. CISL will also identify, promote, and implement common software development tools (e.g., GIT, Wiki, and automation tools).

*Host all NCAR and UCAR enterprise computers in the Mesa Lab Computing Facility (MLCF)*. To achieve energy and other efficiencies, CISL is consolidating computer rooms across the organization. CISL will continue to refurbish and operate the MLCF as a centralized co-location facility for the growing numbers of smaller-scale systems deployed across NCAR, UCAR, and member universities. CISL expects this effort to require about three years, and it will begin in early FY2017.

## Imperative 1.6: Lead and participate in the broader cyberinfrastructure community.

CISL's strategy is to engage the community in numerous activities to advance high performance computing (HPC), data storage, and networking technologies, and to improve regional, national, and international cyberinfrastructure. These activities are aligned with NCAR's leadership role as a Federally Funded Research and Development Center (FFRDC).

*Lead regional CI engagements*. CISL will actively initiate and lead networking partnerships and collaborations and engage in regional HPC activities. For example, the Wyoming-NCAR Alliance, the Rocky Mountain Advanced Computing Consortium (RMACC), and the Pacific Research Platform are regional HPC activities. CISL will provide the technical leadership for operating several regional networking projects that are tightly integrated with national networks. Examples include the Front Range GigaPop (FRGP) and the Bi-State Optical Network (BiSON).

*Lead national CI engagements*. CISL contributes strategic direction for nationwide networks because active participation helps ensure access to the high-speed wide-area network connectivity and services required to carry out its mission. CISL's involvement with networking consortia includes Internet2 (I2), the National Oceanic and Atmospheric Administration (NOAA) science network NWAVE, and the

Department of Energy (DOE) Energy Science Network (ESnet). CISL will continue participating as a Level 2 service provider and collaborator in NSF's Extreme Science and Engineering Discovery Environment (XSEDE), and will also help develop national-scale data CI through programs such as NSF's EarthCube and the Department of Energy's (DOE) Earth System Grid Federation (ESGF).

*Lead international CI activities*. CISL shares its strong technical competencies in supporting international developments, advisory functions, best practices, capacity building, and research projects. CISL provides technical leadership to the WMO's Global Information System (WIS), the system designed for real-time global weather data exchanges, as well as formal data exchanges and data provisioning agreements with peer centers in supercomputing, climate, and weather, and with international projects such as ESGF.

*Engage in partnerships with vendors in HPC, networking, security, and related industries to provide the best technologies for atmospheric and related scientific demands*. CISL informs vendors about the computational requirements of Earth System sciences and tracks technology trends that will inform future supercomputing at NCAR. CISL's HPCFL is one-way CISL keeps abreast of technological developments and innovations. In addition, CISL staff members serve on vendor advisory boards and work with vendors through joint centers of excellence and other forms of persistent collaboration.

## Strategic Goal 2: Enhance the effective use of current and future computational systems by improving mathematical and computational methods for Earth System models and related observations.

The goal of CISL's research activities is to sustain progress in the Earth System sciences by combining powerful supercomputing resources with the latest computational science research in algorithms, mathematical techniques, and statistical methods. CISL helps produce significant and transformative impacts on geoscience by aligning its computational resources with the research objectives of NCAR's other laboratories and the requirements of data-centric science.

Guided by the NCAR strategic plan, CISL research improves predictions of weather and climate and estimations of their impacts. CISL tackles the challenges of developing new data-centric approaches, combining numerical models with observations, interpreting heterogeneous data, and quantifying the uncertainty in predictions in ways that are useful for decision making and policy. CISL also adapts scientific computing in innovative ways like accelerating computation through new algorithms and exploiting new technologies such as coprocessors. This basic computational and data science research supports NCAR's first strategic imperative, and it also benefits the community model development central to NCAR's Imperative 3. Finally, CISL research on the use of new architectures and technologies plays a central role in NCAR's Strategic Imperative 4, developing new computational resources.

### Imperative 2.1: Advance data-centric research.

CISL takes a broad view of data: observations from instruments, outputs from models, and derived "data products" from analyses. In the past, HPC in the geosciences has focused on modeling. CISL research also focuses on how data transformation and data analysis also benefit from HPC resources. CISL research includes applying data assimilation and spatio-temporal statistical methods for combining models and data, creating gridded climate data products with companion measures of uncertainty, and developing statistical and visualization approaches that can handle the very large volumes of data now common in the geosciences.

*Advance data assimilation science*. CISL will lead NCAR's development of cross-component data assimilation (DA) for Earth System models, such as the Community Earth Systems Model (CESM). This research will require novel implementations of ensemble assimilation algorithms with the goal of running very large DA problems on NCAR's future supercomputers in an efficient manner. New DA algorithms and development will be studied within the Data Assimilation Research Testbed (DART) because it can accommodate a wide variety of observations and geophysical models and leverage state-of-the-art DA methods. Research using DART also provides a fast way to disseminate new results to the data assimilation community because this test bed has a base of active users.

*Develop statistical methods to interpret geophysical data and improve model experiments*. CISL researchers will investigate ways to adapt and approximate Bayesian hierarchical statistical models for the practical analysis of large data sets. They will create scalable methods for blending data products and heterogeneous observations, design numerical experiments to tune and develop geophysical models, and retrieve physical fields and processes from indirect measurements.

*Develop and apply novel data science techniques for regional climate change studies using high-resolution research data sets.* A grand challenge for Earth System science is to translate the influence of global processes that affect our climate into specific regional and local impacts. The research in CISL combines knowledge of Earth System models, downscaling methods, scientific workflows for large data sets, and the needs and constraints of local stakeholders. This effort therefore integrates CISL expertise in data science and impact assessment with the goal of transferring climate science into useful products for decision making in adaptation research and risk analysis.

*Exploit many-core processors for data analysis to allow interactive exploration of large data sets*. CISL will leverage the high performance of many-core processors to accelerate the analysis of spatial and spatiotemporal observations and model output wherever possible. One focus will be integrating emerging statistical and machine learning tools with significant data archives such as the CMIP analysis platform.

*Research optimal ways to visualize large scientific data sets.* In response to the large volumes of data generated as model output, CISL will build on The Visualization and Analysis Platform for Ocean, Atmosphere, and Solar Researchers (VAPOR) progressive-access data model, explore ways to represent more generalized data types, and incorporate these efficiency techniques into other tools and services. This will reduce the cost of I/O and storage.

### Imperative 2.2: Develop efficient and scalable numerical methods for Earth System modeling through collaborations with CGD, MMM, and HAO.

Numerical models are the basis for assessing, understanding, and forecasting the complex interactions among geophysical processes and human activities. CISL focuses on the numerical algorithms and computational science that will accelerate the simulation rate of Earth System models by crafting scalable numerical algorithms that can take advantage of large numbers of processors and coprocessors.

*Extend the High Order Method Modeling Environment (HOMME) to include conservative schemes at non-hydrostatic scales.* Next-generation atmospheric climate models will depend on numerical methods that scale to large numbers of processors, and discontinuous Galerkin methods provide one route to meet the need for high-resolution dynamical cores. Numerical development within HOMME is strategic because it is not only a robust numerical test bed, but it is also a framework to transfer numerical methods to the CESM.

*Explore the benefits of meshless numerical methods for geophysical modeling*. Meshless methods such as Radial Basis Functions (RBFs) offer a highly accurate yet computationally cost-effective way to model geophysical processes, particularly for problems that require geometric flexibility and local refinement for small features. RBF methods are also simple to code and so may be particularly useful to harness the power of many-core processors.

### Imperative 2.3: Advance applied computational science research.

Meeting the grand challenges in simulating the Earth System requires much more than migrating old algorithms to new computing platforms. Novel, more efficient or scalable algorithms, new parallel computing paradigms, and new features of hardware architectures must all be mastered to reach the model resolution and complexity levels necessary to support future scientific breakthroughs. Thus, it is imperative that these new technologies each be evaluated as they emerge, in order to determines their fit for geoscience applications.

*Explore many-core and accelerator-based architectures*. CISL will collaborate with NCAR's science laboratories to provide new tools for exploiting many-core architectures such as general-purpose graphics processing units. This allows us to increase model performance on advanced many-core architectures, such as those in NCAR's next supercomputer. CISL also plans to give our users access to advanced systems by acquiring a many-core cluster. Using codes developed under this initiative, our scientific users will gain experience with production systems composed of these emerging technologies.

*Develop more efficient approaches for data processing and compression.* Scientific workflows need to better exploit parallel techniques and tools such as the Message Passing Interface (MPI) for Python to analyze large data sets. Another research focus to pursue is novel, aggressive data compression techniques that have the potential to substantially reduce the storage and bandwidth needed for numerical experiments.

### Imperative 2.4: Foster research collaborations.

The biggest research impact of CISL's in-laboratory research program can be realized through the involvement of the external research and technical communities. Accordingly, regular workshops, short courses, and an active visitor program are important ways to draw mathematical and computer science community to NCAR, engage them in research problems, and augment CISL research.

*Pursue a vigorous visitor program.* The CISL Visitor Program (CVP) brings university faculty, researchers, and students to NCAR to foster collaborations with CISL staff by providing travel and living expenses for stays of up to three months. These extended visits establish strong relationships that can lead to long-term collaborations. CISL also supports diversity by providing travel support for students from underrepresented groups, typically up to four per event, to attend CISL's conferences and training workshops.

## Strategic Goal 3: Reach out and inspire new generations of computational geoscientists through creative hands-on educational opportunities.

CISL integrates research and education to teach the skills that students and faculty need to use advanced cyberinfrastructure effectively. These programs also promote diversity, enhance CISL's culture of teaching and mentorship, and stimulate collaborations with the university community, all in support of NCAR's Strategic Imperative 6. CISL accomplishes these objectives through internships and externships, as well as visitor, workshop, and training programs that supplement education efforts throughout UCAR and at other institutions. A collaboration with NCAR's Director of Diversity, Education, and Outreach will follow up with the Summer Internships in Parallel Computational Science (SIParCS) interns and produce a formal assessment of their research as one evaluation of CISL's education activities.

### Imperative 3.1: Integrate research and education.

CISL serves community needs by integrating research and education in the computational Earth System sciences. An interdisciplinary focus enables CISL to complement and supplement related programs within NCAR, UCAR, and at universities and other centers.

*Provide internships and externships that support CISL research*. The Summer Internships in Parallel Computational Science (SIParCS) program provides hands-on research experiences to students with backgrounds in computational science, applied mathematics, computer science, or the computational geosciences. CISL's goal is to support 12 or more internships each summer to create a critical of mass of intern peers and to increase the program's visibility outside NCAR. SIParCS externships provide research opportunities to students who cannot spend their entire summer at NCAR. The creation of SIParCS assessment efforts will help CISL determine the program's impact on participants.

*Engage the mathematical and computer sciences communities through workshops and schools.* CISL's Institute for Mathematics Applied to Geosciences (IMAGe) Theme of the Year (TOY) is a yearly series of research and education events that enriches both applied mathematics and the geosciences through a common scientific topic. TOY will use a new topic each year to build interdisciplinary communities. CISL's Technology Development Division (TDD) will promote the exchange of technical knowledge about many-core processors at peer centers through an annual Many-core Workshop that is designed to spark new collaborations.

### Imperative 3.2: Train the scientific computing community.

CISL provides training opportunities for researchers to improve their understanding and usage of NCAR's petascale resources. CISL delivers content synchronously through conferences, workshops, tutorials, webinars, and training classes, and asynchronously through web-based content such as online documentation and recorded lectures. CISL also leverages the training and student professional development opportunities provided by regional and national HPC consortia such as the Rocky Mountain

Advanced Computing Consortia (RMACC) and the eXtreme Science and Engineering Discovery Environment (XSEDE).

*Train users and interns in computing at NCAR*. Through its face-to-face and online training efforts, CISL helps prepare students, postdocs, early-career scientists, and other users to investigate questions in the Earth System sciences using its high-performance systems.

*Provide training in geoscientific data analysis and visualization tools*. Effective analysis and visualization of the vast data produced by CISL resources requires thoughtful outreach, engagement, and training activities. CISL offers numerous multi-day training workshops each year for the NCAR Command Language (NCL), a free, interpreted language designed for 2D geoscientific data analysis and visualization. The VAPOR team offers tutorials and webinars on interactive 3D visualization and analysis with CISL's VAPOR package.

*Support community workshops, tutorials, and summer schools*. CISL builds community awareness of and catalyzes progress on a variety of current technical problems through tutorials, workshops, and summer schools. Many of these events are designed to help early-career scientists and students share ideas about emerging trends in computational science.

### Imperative 3.3: Perform community outreach.

CISL's outreach and informal educational activities efforts must be broad-based to engage students, the general public, stakeholders, and diverse underrepresented groups. Therefore, CISL's outreach efforts take several forms to reach different audiences.

*Communicate NCAR science using visualizations.* CISL's Visualization Laboratory (VisLab) serves a vital outreach function for hundreds of student, government, and scientific visitors each year. The Vislab communicates NCAR's scientific achievements – and CISL's role in supporting them – through the use of visualizations. The VisLab's state-of-the-art presentations and visualizations serve hundreds of staff and visitors each year, and its virtual meeting capabilities support advanced collaboration environments and geographically distributed research.

*Maximize the impact of the NWSC facility as a teaching and public outreach tool.* To achieve these objectives, the NWSC Visitor Center will combine current exhibits and innovative activities with engaged staff and excellent docents to provide informative visitor experiences.

*Perform outreach at regional, national, and international levels*. CISL exhibits at scientific, technical, and diversity-oriented conferences (e.g., the yearly Supercomputing conference sponsored by the Institute of Electrical and Electronics Engineers (IEEE)), where CISL staff demonstrates supercomputing capabilities, scientific visualization, and NCAR research in science, computational science, computing technology, and applied mathematics and statistics for the geosciences.

*Communicate via other media.* CISL will strategically use social media and print and television to communicate the technical and scientific accomplishments of CISL staff, the users of CISL resources, and the visitors and interns who engage in collaborative and professional development activities with our laboratory.

### Imperative 3.4: Broaden the diversity of education, outreach, and training activities.

CISL supports a spectrum of local, regional, national, and international activities designed to increase diversity and inclusion.

*Expand diversity-tailored education and training opportunities*. CISL lowers barriers for students and staff from underrepresented groups and underfunded states through participation at diversity-oriented conferences, staffing a full-time Diversity Specialist, providing student travel support to attend training courses, workshops and conferences, and offering diversity-focused internships and externships.

## Summary

This strategic plan lays out a vision for tackling the computing and data challenges set forth by NCAR's grand-challenge objectives. The pace of change is increasing, and this is disruptive to traditional ways of producing science. The way that simulation was conducted ten years ago and the need to extract meaning from increasing amounts of data is changing profoundly along with underlying technologies. Given this changing environment, our challenge is to create and adapt ways that ensure we are keeping current and providing continuing opportunities to the research community. Over the next five years CISL is excited to lead the challenges that these changes present on behalf of NCAR and its partners.

CISL must meet two significant challenges to achieve these goals:

- Disruptive new technologies such as many-core processors, stacked and non-volatile memory, and hybrid disks will present opportunities for advancement and challenges for adaptation. In response, CISL must develop and implement new systems, algorithms, and software to enable enhanced assimilation, simulation, and analysis capabilities.
- CISL must attract, develop, and retain a diverse technical staff with the skills necessary to meet the first challenge. This goal requires broad outreach and engagement with students in STEM fields, and extensive educational and research interactions with early-career scientists and engineers.

At its core, our mission is to ensure that the broad geosciences community has the computational and data infrastructure needed to meet their research challenges. Organized by its three pillars of service, science, and education, CISL's strategic objectives form a practical and inter-supporting roadmap for supporting the geosciences community in achieving their research goals. CISL's bottom line is to take an already excellent supercomputing experience and make it superb for the researchers who use our resources.