2022 NSF CYBERINFRASTRUCTURE FOR SUSTAINED SCIENTIFIC INNOVATION (CSSI) PRINCIPAL INVESTIGATOR MEETING

Towards a Sustainable Data and Software Cyberinfrastructure

July 25-26, 2022
Alexandria, Virginia

Abstract Book

First name	Alexey
Middle Initial	V
Last name	Akimov
Organization	University at Buffalo, SUNY
NSF Award Title	Elements: Libra: The Modular Software for Nonadiabatic and Quantum Dynamics
NSF Award Number	NSF-OAC-1931366
Abstract	Sustained progress in scientific endeavors in solar energy, functional, and nanoscale material sciences requires advanced methods and software components that can be used to model the complex dynamics of excited states, including charge and energy transfer. Within this project, we develop the open-source Libra software which implements a multitude of community-developed methods and computational workflows for nonadiabatic and quantum dynamics (NA/QD) calculations. We focus on trajectory surface hopping methods, decoherence schemes, state tracking algorithms, and more. We interface Libra with various electronic structure codes and excited state calculation methods, enabling it to treat bigger systems and use higher levels of theory in materials modeling. We develop a database of analytic Hamiltonians and conduct a systematic assessment of various NA/QD methods to standardize and rank the “zoo” of the presently available NA/QD methods.
Keywords	nonadiabatic dynamics, excited states, materials science, surface hopping methods, quantum dynamics
Poster session	1
Poster #	1

First name	Alexey
Middle Initial	V
Last name	Akimov
Organization	University at Buffalo, SUNY
NSF Award Title	CyberTraining: Pilot: Modeling Excited State Dynamics in Solar Energy Materials
NSF Award Number	NSF-OAC-1924256
Abstract	The design and discovery of new solar energy materials can be accelerated via computational modeling of excited state dynamics in these systems. Nonetheless, training in this area remains relatively scarce; the community is often unaware of the available cyberinfrastructure, lacks the best practice guidelines, and may experience entry barriers to employing these advanced tools. Within this pilot project, we fill the above gaps by helping young scientists proficiently use the advanced cyberinfrastructure for modeling of excited state dynamics in solar energy materials. During summer schools, we trained a diverse group of participants in a variety of advanced software, including but not limited to Newton-X, SHARC, COLUMBUS, OpenMolcas, CP2K, pyUNIxMD, as well as the Libra code developed by the PI's group. We create an extensive set of tutorials, video tutorials, and examples for Libra and related packages and distribute them widely and openly via GitHub.
Keywords	excited states, nonadiabatic dynamics, summer schools, tutorials, education
Poster session	2
Poster #	1

First name	Metin
Middle Initial	
Last name	Aktulga
Organization	Michigan State University
NSF Award Title	Collaborative Research: CDS&E: ReaxFF2: Efficient and Scalable Methods for Long-time Reactive Molecular Dynamics Simulations
NSF Award Number	1807622
Abstract	This project aims to enable long-time simulations of reactive molecular systems through efficient and scalable techniques. Long-time reactive simulations are critical for several scientific problems. However, progress on these fronts is limited because long-time simulations of large-scale systems are very difficult, if not impossible, to perform using existing methods. The Reactive Force Field (ReaxFF) method is in principle ideally suited for this purpose, but the short time steps required in current ReaxFF simulations and the computationally expensive force field formulation limit ReaxFF's temporal capabilities to narrow simulation time ranges. This project overcomes such limitations by creating ReaxFF2, which extends time scales by one to two orders of magnitude - thus making large-scale, long-time RMD simulations accessible to a wide community.
Keywords	Molecular dynamics, reactive systems, performance optimization, GPU acceleration, sparse solvers
Poster session	1
Poster #	2

First name	Daniel
Middle Initial	G
Last name	Aliaga
Organization	Purdue University
NSF Award Title	Elements: Data: U-Cube: A Cyberinfrastructure for Unified and Ubiquitous Urban Canopy Parameterization
NSF Award Number	1835739
Abstract	Urban canopy parameters (UCPs) can be used in model simulations to study the health and behavior of a city, determine the ability to sustain a growing population, and study potential impacts of extreme weather events. The project deploys innovative science-based analysis tools within an extensible, broadly-available cyberinfrastructure portal, allowing users to ingest satellite imagery and other geographic information system (GIS) data to calculate urban canopy parameters. The cyberinfrastructure would improve urban modeling and planning, particularly for extreme weather events. The tools and high-performance computing and storage resources would be usable by other researchers through a portal. The team has cultivated relationships with the research communities and stakeholders relevant to the proposed research. Through World Urban Database and Access Portal Tools (WUDAPT) project, the team is already connected to the urban planning community globally.
Keywords	urban design, urban planning, inverse modeling, procedural modeling
Poster session	1
Poster #	3

First name	Rafal
Middle Initial	A
Last name	Angryk
Organization	Georgia State University
NSF Award Title	Elements: Comprehensive Time Series DataAnalytics for the Prediction of Solar Flares and Eruptions
NSF Award Number	1931555
Abstract	Through the proposed research, we aim to achieve our interdisciplinary team's two ambitious research goals: 1) to improve scientists' understanding of solar active regions' behavior to the point that we can utilize their observations over time to predict whether and when they will flare, and 2) to enable scientists worldwide to perform comparative, reproducible, and data-driven studies on the prediction of solar explosive events through the use of publicly available big data benchmark data sets and a tested infrastructure of our own, trained machine learning models (software). The external user will be able to use our datasets and / or our software at will.
Keywords	data science, heliophysics, imbalanced data, solar flares prediction, space weather
Poster session	1
Poster #	4

First name	Ritu
Middle Initial	
Last name	Arora
Organization	UTSA
NSF Award Title	COLLABORATIVE RESEARCH: EAGER: Towards Building a CyberInfrastructure for Facilitating the Assessment, Dissemination, Discovery, & Reuse of Software and Data Products
NSF Award Number	2037661
Abstract	The main goal of this project is to develop a software infrastructure for facilitating the assessment, discovery, dissemination, and reuse of publicly accessible software and data products. As a preliminary step towards meeting this goal, this project has initiated research and development activities for prototyping: (1) iTracker: the software infrastructure for tracking the user-defined metrics of products released and deployed on different platforms & computing environments, (2) CompChecker: a license and software-stack compatibility checker for advising the users on the feasibility of integrating or interoperating with existing products, and (3) Discovery Catalog: a prototype of a catalog of NSF-funded products which can display the most recent information captured by iTracker for each product of interest and integrate CompChecker as a feature.
Keywords	interoperability, licensing, metrics, catalog, software discovery
Poster session	2
Poster #	2

First name	Raymundo
Middle Initial	
Last name	Arroyave
Organization	Texas A&M University
NSF Award Title	Elements: Software: Autonomous, Robust, and Optimal In-Silico Experimental Design Platform for Accelerating Innovations in Materials Discovery
NSF Award Number	1835690
Abstract	The objective of this project is to develop an in-silico experimental design platform to accelerate the discovery of novel materials. The platform will be built on optimal Bayesian learning and experimental design methodologies that can translate scientific principles in materials, physics, and chemistry into predictive models, in a way that takes model and data uncertainty into account. The optimal Bayesian experimental design framework will enable the collection of smart data that can help exploring the material design space efficiently, without relying on slow and costly trial-and-error and/or high-throughput screening approaches. The developed methodologies will be integrated into MSGalaxy, a modular scientific workflow management system, resulting in an accessible, reproducible, and transparent computational platform for accelerated materials discovery that allows easy and flexible customization as well as synergistic contributions from researchers across different disciplines.
Keywords	Bayesian Optimization, Materials Discovery, Materials Genome Initiative
Poster session	1
Poster #	5

First name	Berkay
Middle Initial	
Last name	Aydin
Organization	Georgia State University
NSF Award Title	Elements: Spatiotemporal Analysis of Magnetic Polarity Inversion Lines (STEAMPIL)
NSF Award Number	2104004
Abstract	This project creates an innovative and sustainable software infrastructure to detect, characterize and analyze polarity inversion lines (PILs) which are precursors to extreme space weather events. The first step toward that objective is the identification of PILs, and quantitative characterization of these multi-faceted features through image descriptors. In subsequent stages, this project analyzes the time series of these features and descriptors using advanced machine learning and data mining techniques, specifically for improving space weather forecasting capabilities. Automatically identifying and analyzing polarity inversion lines has several direct benefits: physically understanding solar magnetic shear layers and the transition from typical non-eruptive active region states to intense, eruptive ones; making contributions to forecasting of solar eruptions.
Keywords	Object detection, shape descriptors, space weather forecasting, polarity inversion lines
Poster session	3
Poster #	1

First name	Scott
Middle Initial	
Last name	Bachman
Organization	National Center for Atmospheric Research
NSF Award Title	Elements: Cyberinfrastructure for streamlining coupled, simplified climate modeling within the Community Earth System Model
NSF Award Number	2004575
Abstract	Global climate models have increased dramatically in their complexity over the last few decades, but are computationally expensive and are challenging to use for targeted experiments and investigating specific dynamical processes. Idealized simplified models of the climate system enable highly focused studies while avoiding complexity and expense. However, the climate modeling community presently lacks the software infrastructure to easily configure such models using minimal physics packages, simplified domain geometries, or more streamlined model versions, within a single framework that bridges the gap between simplified models and the fully coupled system. This project aims to fill this gap by developing a software toolchain to enable fast and seamless setup of coupled simplified models within NCAR's Community Earth System Model (CESM). This Simpler Models toolchain will serve the needs of the research and academic communities across all climate science disciplines.
Keywords	climate, CESM, toolchain, idealized, model
Poster session	1
Poster #	6

First name	Ryan
Middle Initial	S
Last name	Baker
Organization	University of Pennsylvania
NSF Award Title	Collaborative Research: Frameworks: Cyber Infrastructure for Shared Algorithmic and Experimental Research in Online Learning
NSF Award Number	DRL-1931419
Abstract	Our project supports researchers in 1) conducting large-scale field experiments in K-12 and MOOCs and 2) conducting post-hoc data mining efforts on large datasets from those sources. The project's cyberinfrastructure enables educational researchers to conduct randomized controlled experiments in authentic learning environments with fewer obstructions and costs, while establishing a robust repository of longitudinal data for privacy-protected secondary analyses on unrestricted data. Our project enables researchers to embed an online tutoring system into MOOCs, to experimentally manipulate student supports (e.g, hints, worked examples, and common wrong answer feedback) to study how to increase student learning and engagement. Our project builds upon ASSISTments, an online learning platform used by more than 300,000 K-12 students each year. Researchers are currently using the infrastructure to conduct studies in Penn MOOCs and K-12 courses and to analyze the data from those sources.
Keywords	MOOC, A/B test, data enclave, privacy protection
Poster session	4
Poster #	1

First name	Michael
Middle Initial	
Last name	Barton
Organization	Arizona State University
NSF Award Title	Frameworks: Collaborative Research: Integrative Cyberinfrastructure for Next-Generation Modeling Science
NSF Award Number	2103905
Abstract	This project is developing an Integrative Cyberinfrastructure Framework (ICF) to enable innovative next-generation modeling of human and natural systems, and build capacity in modeling science. It supports a suite of activities that integrate the human and technological components of cyberinfrastructure. 1) Software tools will be developed that augment model codebases with modern software development scaffolding to facilitate reuse, integration, and validation of model code. 2) It will provide high-throughput computing (HTC) resources for simultaneously running numerous iterations of models needed to capture stochastic variability, explore parameter space, and generate alternative scenarios; 3) New online trainings will build expertise and capacity to make effective use of the cybertools and the HTC resources; 4) The ICF will engage a global modeling science community to provide professional incentives that encourage researchers to adopt best practices and catalyze innovative science.
Keywords	modeling, social and ecological sciences, model reusability, education and training, open science
Poster session	1
Poster #	7

First name	Dino
Middle Initial	
Last name	Bektesevic
Organization	University of Washington
NSF Award Title	SI2-SSE: An Ecosystem of Reusable Image Analytics Pipelines
NSF Award Number	1739419
Abstract	Astronomy has entered an era of massive petabyte-sized data streams generated by instruments that can scan tens of thousands of square degrees of the sky across many decades. This project is to develop an open source cloud-based scalable framework for the analysis of large imaging data sets. It incorporates new or legacy image processing algorithms, support and optimize complex analysis workflows, and scale analyses to thousands of processors. This framework will be integrated with state-of-the-art image analysis algorithms developed for astronomical surveys that can be used by future telescopes and the astronomical community. Beyond astronomy, the framework will be extended to enable scientists from the physical and life sciences that make use of imaging data to focus their work on developing scientific algorithms and analyses rather than the infrastructure required to process massive data sets.
Keywords	image analytics, big data, cloud
Poster session	1
Poster #	8

First name	Mahdi
Middle Initial	
Last name	Belcaid
Organization	University of Hawaii at Manoa
NSF Award Title	SAGE3: Smart Amplified Group Environment for Harnessing the Data Revolution
NSF Award Number	2004014
Abstract	Artificial intelligence, data visualization, and dedicated cyberinfrastructure have become essential in modern research to harness the sheer volume and variety of datasets that are becoming the norm across disciplines. Nonetheless, the costs and complexity associated with implementing these tools make them only available to those with a high level of technical expertise. SAGE3 (Smart Amplified Group Environment) puts the “human in the loop” by providing scientists with an intuitive framework that brings together cyberinfrastructure, standard workflows, smart visualizations and collaboration services built using an extensible programming framework. These resources are designed to help users access, share, explore and analyze their data, come to conclusions, and make decisions with greater speed, accuracy, comprehensiveness and confidence.
Keywords	Collaboration platforms, data science environment, artificial intelligence platforms, Human-in-the-loop, decision-support systems
Poster session	1
Poster #	9

First name	Anupam
Middle Initial	
Last name	Bhar
Organization	Iowa State University
NSF Award Title	Elements: Agricultural Cyber-infrastructure support for Field and Grid Modeling, and Runtime Decision-Making
NSF Award Number	2004766
Abstract	This project develops science-driven modeling and decision-making algorithms for precision agriculture, with associated cloud-based implementations. It would increase the potential for more efficient agricultural systems through the development of more accurate and larger scale models, data collection, and ultimately a software infrastructure that farmers could use for real-time monitoring of their crops. The project is derived from scientific questions in sustainable agriculture, sensor networks, decision sciences, data science, and machine learning. The primary innovations are applications of grid-based soil parameters using incomplete data, machine learning-based time series analysis, and optimization problems to estimate the optimal mix of inputs for nutrients and irrigation. These applications are likely to have wide usage among the precis. agric. community. The CI, MyGeoHub, is a cloud-based service that users can access through web browsers and leverages an existing infra.
Keywords	Precision agriculture, MyGeoHub, Agriculture Model, Decision-Making, Optimization.
Poster session	1
Poster #	10

First name	Volker
Middle Initial	
Last name	Blum
Organization	Duke University
NSF Award Title	Collaborative Research: SI2-SSI: ELSI - Infrastructure for Scalable Electronic Structure Theory
NSF Award Number	1450280
Abstract	The overarching goal of the ELSI (ELectronic Structure Infrastructure) project is the creation of an open-source software infrastructure that integrates and helps support several leading approaches to solve or to circumvent the eigenvalue problem of electronic structure theory, particularly Kohn-Sham density-functional theory (DFT). Electronic structure theory continues to be one of the most widely important applications of scientific computing. ELSI fosters a stable, sustainable software infrastructure to solve the key bottleneck of DFT in a code-agnostic way. We review ELSI and recent developments including (1) ELPA2-GPU, a scalable, cross-node eigenvalue solver for GPU architectures, (2) a successful, robust mixed-precision approach that approximately halves the eigenvalue solver time in production DFT, and (3) very large DFT calculations without any specific sparsity approximations, carried out on Google's Tensor Processing Units (TPUs) together with a team at Google.
Keywords	Density-functional theory, electronic structure theory, computational chemistry, computational materials science, parallel matrix algebra
Poster session	1
Poster #	11

First name	Brian
Middle Initial	P
Last name	Bockelman
Organization	Morgridge Institute for Research
NSF Award Title	CSSI Elements: EWMS - Event Workflow Management Service
NSF Award Number	2103963
Abstract	Event Workflow Management System (EWMS) is a workflow management tool to allow researchers to process large numbers of independent pieces of data ("events") within the manager-worker paradigm. Events are found in many different scientific disciplines - including simulated readouts from a particle physics detector, recorded data points, or images. In a data analysis workflow, events are bunched together in files to extend the per work unit processing time to reduce the overhead on the resource scheduling mechanism. This reduces the overall throughput for a given data analysis workflow and can lead to significant resource waste. By implementing the manager-worker paradigm with existing NSF-support and commercially adopted, EWMS will allow users to reduce their time to result by processing their data on the smallest independent piece of data.
Keywords	event processing, manager-worker, High Throughput Computing, distributed computing
Poster session	3
Poster #	2

First name	James
Middle Initial	
Last name	Bordner
Organization	University of California, San Diego
NSF Award Title	Collaborative Research:Framework:Software:NSCI:Enzo for the Exascale Era (Enzo-E)
NSF Award Number	1835402
Abstract	This project is concerned with getting our adaptive mesh refinement (AMR) astrophysics and cosmology application to run efficiently on upcoming Exascale platforms. Our approach has been to develop Enzo-E, the "Extreme" scalable version of our mature ENZO structured-AMR MPI code. Two key features of Enzo-E are 1) it is built on a re-designed scalable "array-of-octree" AMR framework, Cello; and 2) it is parallelized using Charm++ rather than MPI. Charm++'s asynchronous and data-driven task-based parallel programming approach explicitly targets Exascale application development, and is well-suited for the irregular and dynamic data structures inherent in AMR applications. Weak scaling of Enzo-E is excellent, with demonstrated scaling on some of the largest NSF systems available, including Blue Waters at NCSA and Frontera at TACC. Recent accomplishments include additional physics capabilities, scalable I/O, and the first science publication citing results obtained from Enzo-E simulations.
Keywords	AMR, astrophysics, cosmology, Charm++, framework
Poster session	1
Poster #	12

First name	David
Middle Initial	
Last name	Cantu
Organization	University of Nevada, Reno
NSF Award Title	Elements: The ThYme database and identifying representative amino acid sequences that originate thioester-active enzyme families
NSF Award Number	2001385
Abstract	The thioester-active enzyme (ThYme) database includes most known enzymes that act on thioesters, and it was built and brought online to provide the scientific community with a single source for known amino acid sequences and tertiary structures of all the enzymes in the fatty acid synthesis and polyketide synthesis cycles. The goal of this project is to launch a new and updated ThYme database by identifying the current families of thioester-active enzymes, developing a new approach to identify representative sequences that maintains completeness and curtails redundancy, improving ThYme's database management scheme, and modernizing the online user interface. The new database's web interface and functionalities will be designed with input from the ThYme user community. The new database will include capabilities for seamless data search and acquisition, as well as tutorials and collaboration forums.
Keywords	Enzyme, protein, database, bioinformatics, structure
Poster session	1
Poster #	13

First name	Lei
Middle Initial	
Last name	Cao
Organization	MIT
NSF Award Title	Collaborative Research: Elements: A Self-tuning Anomaly Detection Service
NSF Award Number	2103799
Abstract	Finding anomalous behavior in data is important in many applications. A large number of anomaly detection algorithms exist, and it can be difficult to determine which algorithm is best suited to a particular domain. And once an algorithm is selected, users must tune many parameters manually to get the algorithm to perform well; this requires in-depth knowledge of the machine learning process and an understanding of the trade-offs among different algorithms to select the best performing approach. To address these difficulties, this team develops a package that can test a range of unsupervised anomaly detection techniques on a dataset, explore options to identify best-fit, and classify anomalies with higher accuracy than manual tuning. The resulting cyberinfrastructure provides tuning-free anomaly detection capabilities. It enables scientists and engineers having little experience with anomaly detection techniques to steer the anomaly detection process with domain expertise.
Keywords	self-tuning, anomaly detection, unsupervised, domain-specific, anomaly classifier
Poster session	2
Poster #	3

First name	Jonathan
Middle Initial	E
Last name	Carifio
Organization	Center for Astrophysics | Harvard & Smithsonian
NSF Award Title	CDS&E: AAG: "glupyter": Enabling multi-dimensional linked data visualization with glue in the browser
NSF Award Number	1908419
Abstract	The open-source glue package allows exploring relationships within and across related datasets, making it easy to create multi-dimensional linked visualizations of datasets, select subsets in up to 3 dimensions, and see selections propagate live across visualizations. A unique feature of glue is that datasets from multiple sources can be linked with user-defined relationships between data components, allowing selection across datasets. The primary way to use glue is its desktop application, but a new proof-of-concept interface in the Jupyter ecosystem, called "glupyter," is under development. Jupyter notebooks are a popular data analysis tool, as they allow storing workflows using narrative text, code, and visualizations. The recently-released JupyterLab environment is ideal for developing a fully-featured glupyter application, allowing users to construct visualization dashboards. This grant will allow for the transformation of glupyter into a fully featured, science-grade application.
Keywords	data,visualization,multi-dimensional,python,jupyter
Poster session	1
Poster #	14

First name	Kyle
Middle Initial	
Last name	Chard
Organization	University of Chicago
NSF Award Title	Globus Automate: A Distributed Research Automation Platform
NSF Award Number	1835890
Abstract	Automation is increasingly important within scientific research in order to enhance reliability, productivity, and reproducibility. To address the specialized needs of research resources and activities, we have developed a new set of automation services within the Globus research data management platform for executing what we call flows: sequences of actions triggered by external events. These Globus automation services feature: 1) cloud hosting for reliable execution of even long-lived flows despite episodic failures; 2) an extensible asynchronous action provider API for invoking arbitrary actions; and 3) integration with the Globus Auth identity and access management service for secure invocation of actions on remote resources. Together, these methods permit researchers to outsource and automate the management of a broad range of research tasks to a reliable, scalable, and secure cloud platform.
Keywords	Automation, Globus, FaaS, Transfer
Poster session	3
Poster #	41

First name	Kyle
Middle Initial	
Last name	Chard
Organization	University of Chicago
NSF Award Title	funcX: A Function Execution Service for Portability and Performance
NSF Award Number	2004894
Abstract	funcX is a federated Function as a Service (FaaS) platform that enables flexible, scalable, and high performance remote function execution. Unlike centralized FaaS platforms, funcX allows users to execute functions on heterogeneous remote computers, from laptops to campus clusters, clouds, and supercomputers.
Keywords	FaaS, federated computing, elastic scalability, remote execution
Poster session	4
Poster #	2

First name	Guoning
Middle Initial	
Last name	Chen
Organization	University of Houston
NSF Award Title	CDS&E: Multi-scale Coherent Structure Extraction and Tracking For Modern CFD Data Analysis
NSF Award Number	2102761
Abstract	In turbulent flows, coherent structures are closely related to a diverse range of physical phenomena, and understanding their behavior is crucial for characterizing, predicting and controlling these flows. This project brings together experts from both the data visualization and fluid mechanics communities to investigate novel solutions to multi-scale coherent structure extraction, separation, tracking, and visualization. It aims at significantly advancing the ability to analyze large datasets of turbulent flows stemming from computational fluid dynamic (CFD) simulations in a wide range of engineering and scientific applications.
Keywords	data visualization, turbulence flows, CFD simulations, coherent structure, multi-scale
Poster session	online only

First name	Yong
Middle Initial	
Last name	Chen
Organization	Texas Tech University
NSF Award Title	Elements:Software:NSCI: Empowering Data-driven Discovery with a Provenance Collection, Management, and Analysis Software Infrastructure
NSF Award Number	1835892
Abstract	Scientific breakthroughs are increasingly powered by advanced computing and data analysis capabilities delivered by high performance computing (HPC) systems. In the meantime, many scientific problems have moved to a level of complexity that the ability of understanding the results, auditing how a result is generated, and reproducing the important experiments or simulation results, is critical to scientists. Enabling such a capability in HPC systems requires an infrastructure for provenance data. In this project, we focus on provenance of jobs, users, nodes/hosts, files and the relationships among them. We developed to collect jobs from schedulers including Slurm and UGE (a component has been integrated into Dell HPC Omnia repository). We leverage jobstats and changelogs to gather file system provenance. We collected data from a 467-node cluster and conducted visual analytics. This project expects to improve the productivity of science in complex HPC simulation and analysis cycles.
Keywords	Provenance, high performance computing, data collection, data analysis, visual analytics
Poster session	1
Poster #	15

First name	Lydia
Middle Initial	
Last name	Chilton
Organization	Columbia University
NSF Award Title	Elements: Decision Engine for Socioeconomic Disaster Risk (DESDR) - Data Collection, Fusion, and Analysis to Protect Vulnerable Populations From Extreme Weather
NSF Award Number	2103794
Abstract	Natural disasters can have long-lasting financial consequences for vulnerable populations - farmers can be driven into debt by a single drought season, and hurricanes and floods can cause long-lasting damage to social and physical infrastructure. A major challenge is that these natural disasters have disparate impacts on populations even in similar regions. However, socioeconomic risk modelers lack the software tools to address these complex problems in a holistic way – most disaster risk models currently rely exclusively on satellite data or satellite-based crop models, which fail to capture the actual hazard to affected populations due to the lack of, and difficulty to acquire, such data. We are building an extensible and open source infrastructure to address this class of problems. This includes: 1) tools to fill data voids by crowdsourcing data collection from farmers over SMS 2) Tools for local official to clean and manage data and 3) Tools to co-design risk financing.
Keywords	agriculture, finance, software, co-design, open-source
Poster session	2
Poster #	5

First name	In Ho
Middle Initial	
Last name	Cho
Organization	Iowa State University
NSF Award Title	Elements: Development of Assumption-Free Parallel Data Curing Service for Robust Machine Learning and Statistical Predictions
NSF Award Number	1931380
Abstract	Large, incomplete datasets create major challenges for statistical prediction in research. This project will develop a data curing service that is able to manage large, incomplete, and diverse datasets, and would provide uncertainty measures for the cured data. The project identifies and collaborates with several communities where this data service is central to scientific research, including civil engineering, building science, urban energy, and social science.The effort creates a parallel data curing service, provides uncertainty measures for the cured data, and develops supplementary imputing algorithms. The team develops a data curing platform with imputation for incomplete, heterogeneous data; robust machine learning (ML) and statistical predictions would be established by developing an easy-to-use, general-purpose, large data-friendly imputation program.
Keywords	Missing data curing, big incomplete data, hybrid data, imputation, fractional hot deck imputation
Poster session	1
Poster #	16

First name	Eunseo
Middle Initial	
Last name	Choi
Organization	The University of Memphis
NSF Award Title	Elements: Developing an integrated modeling platform for tectonics, earthquake cycles and surface processes
NSF Award Number	2104002
Abstract	This project aims to develop a user-friendly and sustainable code that can simulate lithospheric deformations coupled with landscape evolution and earthquake cycles. One of the challenges facing the Earth Sciences community is how to link geological time scales of thousands to millions of years to human time scales. Natural hazards such as earthquakes, flooding, landslides, volcanoes sit right at the intersections of geological and human time scales. Capability to simulate the interactions among diverse geological processes provides a baseline to study how global changes are perturbing these processes on human time scales. The goal of the project is aligned with the on-going research trend in Earth Sciences to view earthquake cycles and landscape evolutions as a system coupled with long-term lithospheric deformation. Multiple research communities including geomorphology, structural geology and earthquake physics will benefit from the new modeling capability this project enables.
Keywords	tectonics, earthquake cycle, modernization, parallelization, user experience
Poster session	3
Poster #	3

First name	Sean
Middle Initial	B
Last name	Cleveland
Organization	University of Hawaii - System
NSF Award Title	Frameworks: Project Tapis: Next Generation Software for Distributed Research
NSF Award Number	1931575
Abstract	Tapis, is a new platform for distributed computational experiments that 1) provides production-grade support for sensors and streaming data, 2) maximizes application portability, allowing flexible scheduling of computational workloads across geographically distributed providers, and 3) provides science-as-a-service HTTP-based RESTful APIs to enable multi-facility, decentralized deployments that are both secure and scalable. Working alongside a diverse set of domain researchers to drive real-world use cases. Tapis aims to be the underlying cyberinfrastructure for computational workflows and science gateways.
Keywords	Science Gateways, Compute, Containers, Microservices, Distributed workflows
Poster session	2
Poster #	6

First name	Brianna
Middle Initial	D
Last name	Corsa
Organization	University of Colorado Boulder
NSF Award Title	GeoSCIFramework Project
NSF Award Number	1835791
Abstract	Differential Interferometric Synthetic Aperture Radar (DInSAR) and Global Navigation Satellite System (GNSS) document comprehensive ground motions or ruptures at or near the Earth's surface. These datasets may be independently applied to detect and analyze natural hazard phenomena, such as ground motion leading to major volcanic eruptions. Additionally, when combined into a single dataset, assimilated DInSAR + GNSS deformation results have improved accuracy. Here we attempt to generate a synthetic dataset that best matches our integrated DInSAR + GNSS time series results over Hawaii from November 2015 through January 2022 by comparing multiple volcanic deformation source models, including a Mogi, diking and distributed source models. We then demonstrate how the synthetic- and real- time series will be streamed through machine learning algorithms to help identify precursor motion leading to major eruptions, contributing towards improved early warning systems.
Keywords	DInSAR, GNSS, time series, modeling, volcanic deformation
Poster session	1
Poster #	17

First name	Aaron
Middle Initial	
Last name	Costin
Organization	University of Florida
NSF Award Title	Elements: Cyberinfrastructure Service for IoT-Based Construction Research and Applications
NSF Award Number	2004544
Abstract	This project develops a robust cyberinfrastructure (CI) system and service for construction research and applications to address the current challenges faced in the construction industry. The outcomes and services that this proposal aims to provide are 1) a distributed SDN-managed and AI-assisted IoT-based system that can be adapted and extended based on needs of the research and application; 2) identification of the data and data security requirements needed to address the challenges in the construction industry and potential technologies that can provide those data; 3) evaluation of reliable real-time multi-sensor fusion techniques for ruggedness, usability, and limitations of IoT-based components deployed in the dynamic construction environments; 4) robust prototype system for real-time safety monitoring based on the IoT system framework; and 5) recommendations of potential configurations of the system with the appropriate technology and sensors to meet the needs of the application.
Keywords	Internet of Things, Construction, Wireless Sensor Network, Safety Monitoring
Poster session	1
Poster #	18

First name	Daniel
Middle Initial	
Last name	Crawford
Organization	Molecular Sciences Software Institute/Virginia Tech
NSF Award Title	S2I2: Impl: The Molecular Sciences Software Institute
NSF Award Number	2136142
Abstract	Since 2016, the Molecular Sciences Software Institute has served as a nexus for the broad computational molecular sciences community by providing software expertise, community leadership, and education. Through an array of software infrastructure projects, teaching workshops, and community outreach, the MolSSI catalyzes the scientific advances needed to solve emerging scientific computing Grand Challenges. The MolSSI's Software Scientists – drawn from the molecular sciences, computer science, and applied mathematics – promote improved interoperability of community codes, easier deployment on heterogenous computing architectures, and greater scalability of theoretical models. The MolSSI's Education Initiative annually reaches thousands of students worldwide and its Software Fellowship program, which has already benefitted more than 100 graduate students and postdocs across the U.S.
Keywords	molecular sciences, best practices, education, interoperability, high-performance computing
Poster session	4
Poster #	3

First name	Peter
Middle Initial	
Last name	Crozier
Organization	Arizona State University
NSF Award Title	Elements: Collaborative Research: Community-driven Environment of AI-powered Noise Reduction Services for Materials Discovery from Electron Microscopy Data
NSF Award Number	2104105
Abstract	Developing a fundamental understanding of atomic level structure and dynamics is critical for transformative advances in materials science. Aberration-corrected transmission electron microscopy is a primary tool to accomplish this goal. Unfortunately, the information content of microscopy data may be severely limited by poor signal-to-noise ratios. This is particularly true for radiation sensitive materials and experiments where high time resolution is required to investigate dynamic kinetic processes. AI methodology can exploit prior information about material structure by training deep neural nets with extensive simulations. These approaches may significantly outperform existing state-of-the-art methods, especially for non-periodic structures, including defects, interfaces, and surfaces. The proposed cyberinfrastructure (CI) will provide AI noise reduction services which will yield immediate advances and impacts for many areas of materials.
Keywords	atomic level structure and dynamics, denoising, catalysis, machine learning
Poster session	3
Poster #	4

First name	Sajal
Middle Initial	K
Last name	Das
Organization	Missouri University of Science and Technology
NSF Award Title	CSSI:CANDY: Cyberinfrastructure for Accelerating Innovation in Network Dynamics
NSF Award Number	2104078
Abstract	Analysis of dynamic networks, represented as graphs, has applications in diverse disciplines, such as data mining (big data), social networks, bioinformatics, epidemiology, cyber-physical infrastructures, transportation networks, and cybersecurity. However, in contrast to the numerous software available for analyzing static networks, there exists no comprehensive cyberinfrastructure to support complex challenges of analyzing network dynamics. This project aims to address this gap by proposing to develop a novel framework, Cyberinfrastructure for Accelerating Innovation in Network Dynamics (CANDY), for facilitating research on efficient processing of large-scale, dynamically changing networks. The CANDY project will deliver the first parallel, scalable, extendable, and user-friendly software platform for updating important properties of dynamic networks, as well as provide requisite functionalities and tools to modify existing algorithms or create new ones.
Keywords	Parallel dynamic network analysis, open source software framework, complex networks
Poster session	3
Poster #	5

First name	Ewa
Middle Initial	
Last name	Deelman
Organization	University of Southern California
NSF Award Title	SI2-SSI: Pegasus: Automating Compute and Data Intensive Science
NSF Award Number	1664162
Abstract	The project sustains and enhances the Pegasus Workflow Management System, which enables scientist to orchestrate and run data- and compute-intensive computations on diverse distributed computational resources. Enhancements focus on the automation capabilities provided by Pegasus to support workflows handling large data sets, as well as usability of Pegasus that lowers the barrier of its adoption. This effort expands the reach of the advanced capabilities provided by Pegasus to researchers from a broader spectrum of disciplines that range from gravitational-wave physics to bioinformatics, and from earth science to material science.
Keywords	workflow management, distributed computing, resource management
Poster session	1
Poster #	19

First name	Robert
Middle Initial	L
Last name	DeLeon
Organization	SUNY at Buffalo
NSF Award Title	Elements: Development and Dissemination of a Slurm Simulator
NSF Award Number	2004954
Abstract	We are developing a slurm simulator that can be used to properly configure a slurm instance without disturbing the production installation. We will provide this utility to the community with full documentation and instructions to make it a useful tool.
Keywords	slurm, simulator, scheduler,throughput,work-load
Poster session	4
Poster #	4

First name	Leszek
Middle Initial	
Last name	Demkowicz
Organization	The University of Texas at Austin
NSF Award Title	Elements:Software A Scalable Open-Source hp-Adaptive FE Software for Complex Multiphysics Applications
NSF Award Number	2103524
Abstract	Computer models can be used to augment, inform, and even replace expensive experimental measurements in science and engineering. However, complex models of engineering applications can quickly exceed computational capability, driving the need for advanced simulation tools. Applications in high-frequency wave simulation--such as submarine sonar (acoustics), fiber optics (electromagnetics), and structural analysis (elastodynamics)--pose a significant challenge for large-scale simulation. This project advances computational modeling capabilities through the development, documentation, and dissemination of a leading-edge simulation software. The effort builds on decades-long research and code development by the investigators and their project collaborators. Distributed as open-source, the software is accessible to the broader scientific community, thereby contributing to fundamental research and education for computer modeling in science and engineering.
Keywords	Finite Elements, Open Source, DPG Method, Multigrid, Adaptivity
Poster session	2
Poster #	7

First name	Brian
Middle Initial	
Last name	Demsky
Organization	UC Irvine
NSF Award Title	SI2-SSE: C11Tester: Scaling Testing of C/C++11 Atomics to Real-World Systems
NSF Award Number	1740210
Abstract	We have long relied on increased raw computing power to drive technological progress. However, processors are reaching their limits in terms of raw computing power, and continuing progress will require increased productivity in developing parallel software. Fully leveraging multi-core processors in many cases require developers to make use of low-level "atomic" operations such as those provided by the C11 and C++11 languages to make fine-grained optimizations to their code, and take advantage of the computing power these processors offer. Unfortunately, using C/C++ atomics is difficult to do correctly and it is easy to introduce bugs in their use. Testing for concurrency bugs in code using C/C++11 atomics can be difficult as a bug can depend on the schedule, the state of the processor's memory subsystem, the specific processor, and the compiler. The C11Tester project will develop tools for testing concurrent code that makes use of C/C++11 atomics and make these tools available.
Keywords	Fuzzing, concurrency, atomics, memory models
Poster session	online only

First name	Peter
Middle Initial	
Last name	Diener
Organization	Louisiana State University
NSF Award Title	Collaborative Research: Frameworks: The Einstein Toolkit ecosystem: Enabling fundamental research in the era of multi-messenger astrophysics
NSF Award Number	2004157
Abstract	The Einstein Toolkit is a community driven framework for numerical astrophysics simulations. It features state of the art codes to simulate compact objects of astrophysical interest such as black holes, and neutron stars in general relativity and theories of gravity beyond general relativity. This CSSI project focuses on developing a new mesh refinement driver CarpetX for the Einstein Toolkit based on the AMReX library, improving the code generation framework NRPy+ in the toolkit, and expanding the set of gravity theories supported by the Canuda framework.
Keywords	Numerical Relativity, Compact Objects, Einstein Toolkit, Adaptive Mesh Refinement, High Performance Computing,
Poster session	1
Poster #	21

First name	Peter
Middle Initial	
Last name	Elmer
Organization	Princeton University
NSF Award Title	S2I2: Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP)
NSF Award Number	OAC-1836650
Abstract	IRIS-HEP is a software institute funded by the National Science Foundation. It aims to develop the state-of-the-art software cyberinfrastructure required for the challenges of data intensive scientific research at the High Luminosity Large Hadron Collider (HL-LHC) at CERN, and other planned HEP experiments of the 2020's.
Keywords	Physics, Particle Physics, Software
Poster session	1
Poster #	23

First name	Keivan
Middle Initial	
Last name	Esfarjani
Organization	University of Virginia
NSF Award Title	ELEMENTS: Anharmonic formalism and codes to calculate thermal transport and phase change from first-principles calculations
NSF Award Number	2103989
Abstract	In this project, we are developing a new software capability to accurately compute thermodynamic and thermal properties of anharmonic crystals at finite temperatures from first-principles. We will implement the powerful self-consistent phonon approximation which explicitly takes into account anharmonicity to compute the effect of thermal fluctuations on lattice structural and dynamical properties. We will extend this method to incorporate structural phase change and coupling of atomic vibrations to other degrees of freedom of electronic origin. The output of these codes will consist of equilibrium crystal structure, crystal phonon dispersion, lifetimes, thermal conductivity, dielectric susceptibility, thermal expansion coefficients, free energy, all as a function of temperature. These timely tools will enable materials scientists to predict or understand thermophysical properties of anharmonic, complex and multifunctional materials at arbitrary temperatures.
Keywords	self-consistent phonons, thermal conductivity, force constants, phase transitions, equilibrium phase
Poster session	3
Poster #	6

First name	John
Middle Initial	A
Last name	Evans
Organization	University of Colorado Boulder
NSF Award Title	Collaborative Research: Elements: EXHUME: Extraction for High-Order Unfitted Finite Element Methods
NSF Award Number	2104106
Abstract	The overarching objective of this project is to construct a novel software library, EXHUME (EXtraction for High-order Unfitted finite element MEthods), to enable the use of classical finite element codes for unfitted finite element analysis. Unfitted finite element methods allow for the simulation of physical systems that are difficult if not impossible to simulate using classical finite element methods, and they also streamline the construction of design optimization technologies that optimize the geometry and material layout of an engineered system based on prescribed performance metrics. However, the computer implementation of an unfitted finite element method remains a challenging and time-consuming task even for domain experts. EXHUME overcomes this barrier by generating data structures that can be leveraged to transform classical finite element codes relying on element formation and assembly into unfitted finite element codes with little implementation effort.
Keywords	Unfitted Finite Element Analysis, Immersogeometric Analysis, Extraction, Finite Element Software
Poster session	4
Poster #	5

First name	Marc
Middle Initial	
Last name	Fehling
Organization	Colorado State University
NSF Award Title	Collaborative Research: Frameworks: Software: Future Proofing the Finite Element Library Deal.II -- Development and Community Building
NSF Award Number	OAC-1835673
Abstract	deal.II is a widely used, open source library providing finite element and linear algebra support for individual research applications. This project aims at broadening support for parallel computations and at building a large user community.
Keywords	finite elements, linear algebra, software
Poster session	1
Poster #	24

First name	Renato
Middle Initial	
Last name	Figueiredo
Organization	University of Florida
NSF Award Title	Collaborative Research: Elements: EdgeVPN: Seamless Secure Virtual Networking for Edge and Fog Computing
NSF Award Number	2004441
Abstract	Edge computing encompasses a variety of technologies that are poised to enable new applications across the Internet that support data capture, storage, processing and communication near the edge of the Internet. Edge computing environments pose new challenges, as devices are heterogeneous, widely distributed geographically, and physically closer to end users, such as mobile and Internet-of-Things (IoT) devices. This project develops EdgeVPN, a software element that addresses a fundamental challenge of networking for edge computing applications: establishing Virtual Private Networks (VPNs) to logically interconnect edge devices, while preserving privacy and integrity of data as it flows through Internet links. This project addresses technical challenges in creating virtual networks that self-organize into scalable, resilient systems that can significantly lower the barrier to entry to deploying a private communication fabric in support of existing and future edge applications.
Keywords	edge computing, fog computing, virtual networks, software-defined networks
Poster session	1
Poster #	25

First name	Feliciano
Middle Initial	
Last name	Giustino
Organization	University of Texas, Austin
NSF Award Title	Frameworks: An Interoperable Software Ecosystem for Many-Body Electronic Structure Calculations
NSF Award Number	2103991
Abstract	The overarching aim of this project is to create an interoperable software ecosystem to model and design materials at the atomic scale using many-body field-theoretic approaches beyond ground-state density functional theory (DFT). To this aim, we will expand and combine the complementary strengths of three electronic structure software packages developed in our groups, namely EPW, BerkeleyGW, and SternheimerGW, into a user-centric, containerized simulation laboratory with shared data formats and builtin compatibility layers for major DFT codes. The proposed cyber-infrastructure (CI) will enable predictive, systematic, and reproducible ab initio calculations of excited-state phenomena including finite temperature effects and phonon-assisted quantum processes, and will accelerate CI development by distributing curated, reusable, and interoperable opensource software, modules, and libraries.
Keywords	Computational materials science, electronic structure calculations, many-body methods, excited-state phenomena
Poster session	3
Poster #	7

First name	Andreas
Middle Initial	W
Last name	Goetz
Organization	University of California, San Diego
NSF Award Title	Elements: Software: NSCI: Efficient GPU Enabled QM/MM Calculations: AMBER Coupled with QUICK
NSF Award Number	1835144
Abstract	The goal of our project is to improve our software cyberinfrastructure for solving important molecular-level problems in catalysis, drug design, energy conversion. Combined quantum mechanical/molecular mechanical (QM/MM) models have enabled significant advances in our understanding of chemical reactivity. The shortcoming of QM/MM models when using ab initio or density functional theory (DFT) methods is the computational expense, which limits QM/MM modeling. The performance of QM methods has been greatly improved over the years through algorithmic and hardware improvements. We describe the enhancements and performance of our open-source multi-GPU enabled Quantum Interaction Computational Kernel (QUICK) QM program, which we have coupled to the AMBER molecular dynamics package to enable high-performance QM/MM simulations in condensed phase under periodic boundary conditions.
Keywords	QM/MM, DFT, GPU, open-source software, QUICK, AMBER
Poster session	1
Poster #	26

First name	Boyce
Middle Initial	
Last name	Griffith
Organization	The University of North Carolina at Chapel Hill
NSF Award Title	Multiphase Fluid-Structure Interaction Software Infrastructure to Enable Applications in Medicine, Biology, and Engineering
NSF Award Number	OAC 1931516
Abstract	IBAMR is software for simulating fluid dynamics, solid mechanics, and fluid-structure interaction (FSI). It offers leading community implementations of the immersed boundary (IB) method and its extensions, and it achieves high performance through its support for block-structured adaptive mesh refinement (AMR). This project is advancing the biological and biomedical applications that have traditionally been the focus of IBAMR's development while also enabling new engineering applications that benefit from the same classes of numerical methods. It is extending IBAMR to solve much larger problems, and is improving the accuracy, efficiency, and fidelity of IBAMR-based models through novel fluid solver and FSI coupling algorithm technologies. All project developments are motivated by and anchored in specific applications, including cardiovascular, esophageal, and pulmonary medicine and biology; wave energy converters; and additive manufacturing.
Keywords	fluid-structure interaction, adaptive mesh refinement, high performance computing
Poster session	1
Poster #	27

First name	Michael
Middle Initial	
Last name	Groeber
Organization	Ohio State University
NSF Award Title	Elements: Data Science Methods for Real-Time Resource Allocation During Characterization of Dynamic Systems
NSF Award Number	2005012
Abstract	The project will create an openly available control package to enable dynamic experiments informed by real-time modeling.The developments will make efficient use of scarce beam-time at national synchrotron user facilities, enabling higher throughput for in-situ experiments probing the response of materials under load. The effort evaluates the hypothesis that rare material events can be predicted from a small number of features describing evolving local material states using ML solutions. The project focuses on dynamic resource allocation for hyper-efficient data collection at the CHESS, a national user facility. The project integrates these toolsets to detect precursor signatures through real-time processing of data and suggests resource allocations during study of early stages of stochastic events in dynamic materials systems. The goal is to develop ML techniques to inform the allocation of limited detector resources.
Keywords	machine-learning, experiments, materials, modeling, rare events
Poster session	online only

First name	Thomas
Middle Initial	
Last name	Hacker
Organization	Purdue University
NSF Award Title	Elements: Data: Integrating Human and Machine for Post-Disaster Visual Data Analytics: A Modern Media-Oriented Approach
NSF Award Number	1835473
Abstract	Structures can be impacted by events (e.g. earthquakes) over a short time, or over a longer time (e.g. aging infrastructure such as bridges). Civil Engineers collect visual data that contains valuable information about the structure and use these images to help assess the state of structures. Distilling insights from visual data is complicated by the large number and variety of images collected. To help address this challenge, we have developed web-based applications that focus on enhancing the experience of analyzing damage from images collected during the inspection of infrastructure with different levels of damage likely caused by natural hazards. We have built a platform approach (VISER) to host two applications: the Automated Reconnaissance Image Organizer (continued work from a prior NSF grant) and the Automated Bridge Image Reporting Tool; and are investigating computing infrastructure approaches that seek to facilitate access to the capabilities provided by these applications.
Keywords	Infrastructure assessment, Structural damage, Earthquake reconnaissance data, Bridge inspection, Image classification
Poster session	1
Poster #	28

First name	Thomas
Middle Initial	
Last name	Haine
Organization	Johns Hopkins University
NSF Award Title	Collaborative Research: Framework: Data: Toward Exascale Community Ocean Circulation Modeling
NSF Award Number	1835640
Abstract	The goals of the project are to deliver: 1. A benchmark, accurate global ocean circulation solution at O(1) km horizontal resolution. 2. Open-source software tools enabling efficient storage, indexing, and analysis of petabyte-scale ocean/atmosphere/climate datasets. 3. A Data Access Portal that deploys these tools, together with custom-built high-performance storage and computing resources, to provide scalable interactive analyses and visualizations of the benchmark solution to the climate and computer science communities. 4. Explorations of machine learning frameworks for automated identification of important events and data compression, with working prototypes for use in coupled climate models. 5. A foundation and a path to migrate computational oceanography to exascale. 6. A fully-functioning instance of the sort of cyberinfrastructure that will increasingly be needed by next generation simulation software in geosciences and beyond.
Keywords	ocean circulation, ocean models, community data analysis, computational oceanography
Poster session	2
Poster #	9

First name	Kathryn
Middle Initial	R
Last name	Hamilton
Organization	Drake University
NSF Award Title	Elements: NSCI-Software -- A General and Effective B-Spline R-Matrix Package for Charged-Particle and Photon Collisions with Atoms, Ions, and Molecules
NSF Award Number	1834740
Abstract	This project concerns the development and subsequent distribution of a suite of computer codes that can accurately describe the interaction of charged particles (mostly electrons) and light (mostly lasers and synchrotrons) with atoms and ions. The results are of importance for the understanding of fundamental collision dynamics, and they also fulfil the urgent practical need for accurate atomic data to model the physics of stars, plasmas, lasers, and planetary atmospheres. With the rapid advances currently seen in computational resources, such studies can now be conducted for realistic systems. In particular, it has become possible to describe very complex targets, such as transition metals and other open-shell systems. The source code resulting from this project will be made publicly available. A website devoted to user-developer interaction will be developed and maintained together with the necessary code documentation and training materials.
Keywords	R-matrix, B-spline, electron-atom collisions, open source, parallel computing
Poster session	2
Poster #	10

First name	Chad
Middle Initial	R
Last name	Hanna
Organization	Penn State
NSF Award Title	An A+ Framework for Multimessenger Astrophysics Discoveries through Real-Time Gravitational Wave Detection
NSF Award Number	2103662
Abstract	The first direct detection of ripples in space, known as gravitational waves in 2015 opened a new window on the universe and provided an unprecedented ability to study distant astronomical phenomena that could otherwise not be seen with conventional telescopes. The subsequent 2017 detection of merging neutron stars, through gravitational waves with LIGO combined with the light detected by conventional telescopes, opened a new era whereby scientists hope to routinely study the universe using information analogous to both sight and sound. This project will directly enable future detections of gravitational waves through the development of robust signal processing software and an ecosystem of cyberinfrastructure services designed to analyze LIGO data in real time. This work contributes to the national cyberinfrastructure as a core data-producing component for astronomy and will be relied upon by thousands of scientists globally.
Keywords	LIGO, gravitational waves, black holes, neutron stars, real-time, multi-messenger, signal processing, stream-processing
Poster session	2
Poster #	11

First name	Robert
Middle Initial	J
Last name	Harrison
Organization	Stony Brook University
NSF Award Title	Production quality Ecosystem for Programming and Executing eXtreme-scale Applications (EPEXA)
NSF Award Number	1931387
Abstract	EPEXA is creating a production-quality, general-purpose, community-supported, open-source software ecosystem that attacks the twin challenges of programmer productivity and portable performance for advanced scientific applications on modern high-performance computers. Of special interest are irregular and sparse applications that are poorly served by current programming and execution models, and making effective use of heterogeneous architectures with dynamically changing performance. EPEXA builds upon prior work performed by the TESSE (Task-based Environment for Scientific Simulation at Extreme Scale) project, and in particular is completing the design and provide robust implementation of TESSE's main C++ API, the Template Task Graph (TTG) that provides a powerful data-flow programming model.
Keywords	Parallel programming, data flow, exascale
Poster session	2
Poster #	12

First name	Pedram
Middle Initial	
Last name	Hassanzadeh
Organization	Rice University
NSF Award Title	Collaborative Research: Framework: Improving the understanding and representation of atmospheric gravity waves using high-resolution observations and machine learning
NSF Award Number	2005123
Abstract	The objective of this project is to leverage unprecedented observations from Loon high altitude balloons and use specialized high resolution computer simulations and machine learning techniques to develop accurate, data-informed representation of atmospheric gravity waves.
Keywords	climate modeling, machine learning, parameterization, climate change, atmospheric dynamics
Poster session	2
Poster #	13

First name	Bryna
Middle Initial	J
Last name	Hazelton
Organization	University of Washington
NSF Award Title	Collaborative Research: Elements: Software: Accelerating Discovery of the First Stars through a Robust Software Testing Infrastructure
NSF Award Number	1835421
Abstract	The birth of the first stars and galaxies 13 billions years ago is one of the last unobserved periods in the history of the Universe. Scientists are working to observe the 21cm radio light emitted by the primeval neutral hydrogen fog as the first stars reionized the universe. One of the biggest challenges is the presence of bright astrophysical foregrounds that obscure the 21cm signal, requiring extraordinarily precise modeling and calibration of radio telescopes performing these observations. The 21cm cosmology community is rapidly developing new techniques for instrument calibration, foreground removal and analysis, but thorough testing and integration into existing data analyses has been slow. This project provides a software infrastructure that enables rigorous testing of novel algorithms within a unified framework, significantly improving the reliability and reproducibility of these analyses and accelerating the integration of new developments into production level code.
Keywords	radio astronomy, interferometry, cosmology, testing, open source software
Poster session	2
Poster #	14

First name	Hendrik
Middle Initial	
Last name	Heinz
Organization	University of Colorado Boulder
NSF Award Title	Collaborative Research: Frameworks: Cyberloop for Accelerated Bionanomaterials Design
NSF Award Number	1938517
Abstract	The project supports the understanding of the dynamical evolution of biological and materials systems from the atomic scale to the microscale by means of developing new cyberinfrastructure for simulations. Hereto, Cyberloop integrates three existing successful platforms for soft matter and solid state simulations (IFF, OpenKIM, and CHARMM-GUI) into a single unified framework. These l systems, including a new Nanomaterial Modeler module in CHARMM-GUI, work together to enable users to set up complex bionanomaterial configurations, select reliable validated force fields, generate input scripts for popular simulation platforms, and assess the uncertainty in the results. The tools support groundbreaking advances in the health sciences, materials sciences, energy conversion, sustainability, and improving overall quality of life.
Keywords	Atomistic simulation, databases, force fields, web-based interface, biomolecular simulation, materials simulation, molecular dynamics
Poster session	2
Poster #	15

First name	Chris
Middle Initial	
Last name	Hill
Organization	MIT/DJ4Earth
NSF Award Title	Collaborative Research: Frameworks: Convergence of Bayesian inverse methods and scientific machine learning in Earth system models through universal differentiable programming
NSF Award Number	2103942, 2103791, 2104009, 2104068, 2103804, 2147601
Abstract	Our project, DJ4Earth (https://dj4earth.github.io) focusses on differentiable Julia programming for Earth science. The projects connects computational/computer science expertise with practicing domain science researchers interested in oceanic and cryosphere modeling. The project spans Julia programming language developers to open-source simulation model architects. The broad aim of the project is to advance Earth system modeling capabilities to autmotically synthesize models and data through the broad family of Bayesian thinking inspired tools. The Julia programming language offers some potentially novel, language-native, opportunities around dynamic program introspection, meta-programming techniques and high-level abstraction. These are being applied to create new open-source tools for ice-sheet, sea-ice and ocean science work that can learn from rich new data streams in ways that have been too technologically complex to fully realize their full potential in the past.
Keywords	open-source, earth-science, machine-learning, uncertainty-qualification, learning-models-from-data
Poster session	4
Poster #	7

First name	Jeff
Middle Initial	
Last name	Horsburgh
Organization	Utah State University
NSF Award Title	Collaborative Research: Elements: Advancing Data Science and Analytics for Water (DSAW)
NSF Award Number	1931297
Abstract	Scientific challenges in the water domain are multi-disciplinary, requiring synthesis of data from multiple domains. Many data analysis tasks performed by water scientists are difficult because datasets are large and complex; standard formats for common data types are not always agreed upon nor mapped to an efficient structure for analysis; and scientists sometimes lack training in methods needed to efficiently tackle large and complex datasets. This project is advancing Data Science and Analytics for Water (DSAW) by developing: (1) an advanced object data model that maps common water-related data types to high performance Python data structures based on standard file, data, and content types established by the CUAHSI HydroShare system; and (2) new Python packages that enable scientists to automate retrieval of water data, loading it into high performance memory objects, and performing reproducible analyses that can be shared, collaborated around, and formally published for reuse.
Keywords	hydrology, data, Python, data science, software
Poster session	2
Poster #	16

First name	Nancy
Middle Initial	M
Last name	Ide
Organization	Brandeis University
NSF Award Title	Towards a Robust Cyberinfrastructure for NLP-based Search and Recoverability over Scientific Literature
NSF Award Number	2104025
Abstract	Keeping up with the ever-expanding flow of data and publications in the sciences is untenable and poses a fundamental bottleneck to scientific progress. Current search technologies do not extract and organize the information content of these documents or suggest new scientific hypotheses based on this content. Natural Language Processing (NLP) strategies are a recognized means to approach this problem, but most scientists do not have the expertise and/or time required to take advantage of them. There is currently no open, easy-to-use platform for accessing and mining information from scientific texts that provides access to a wide array of software, computing resources, and publication data in one place. This project marries two major NSF-funded projects that have developed complementary capabilities: the Language Applications (LAPPS) Grid, a platform for development of NLP applications and University of Wisconsin's xDD, a digital library of over 15 million scientific publications.
Keywords	Information retrieval, text mining, scientific literature, natural language processing
Poster session	4
Poster #	8

First name	Philip
Middle Initial	
Last name	Ilten
Organization	University of Cincinnati
NSF Award Title	Elements: Machine Learning Quark Hadronization
NSF Award Number	2103889
Abstract	Particle physics explores the fundamental building blocks of nature and their interactions. Experimentally, particle beams are collided, producing many particles that are measured by detectors. Event generators simulate these collisions, and are critical for particle physics. Event generators simulate collisions in three steps: a high energy collision, evolution of the event to low energies, and hadronization into bound particles. Hadronization is particularly challenging, since it cannot be directly calculated from first principles. This project applies machine learning to hadronization to understand the underlying description of data by current event generators and provide a fast and accurate simulation of hadronization. Hadronization models are improved by this project, fulfilling a critical step for understanding the scientific impact of upcoming large-scale particle, neutrino, and nuclear physics projects.
Keywords	Monte Carlo, machine learning, particle physics, simulation
Poster session	2
Poster #	17

First name	Zachary
Middle Initial	G
Last name	Ives
Organization	University of Pennsylvania
NSF Award Title	mProv: Provence-Based Data Analytics Cyberinfrastructure for High-frequency Mobile Sensor Data
NSF Award Number	1640813
Abstract	Mobile sensors continuously capture data, and have the potential to address problems in a range of scientific and engineering domains. This effort addresses assembling and analyzing health data collected through mobile sensors and apps. Improvements to the usefulness of extremely noisy, distributed data can serve many communities, and the components are extensible outside the human health domain. The project addresses several challenges: Variability is addressed by providing detailed annotation with provenance and context for reasoning about the metadata. The system captures provenance metadata alongside streaming data, and propagates this information from one stage to the next. This creates cyberinfrastructure that 'replays' mobile device data with different configurations, to comparatively benchmark algorithms or diagnose erroneous output. This project develops open-source provenance cyberinfrastructure, facilitating the sharing of the mobile sensor data itself.
Keywords	provenance, big data, sensors
Poster session	2
Poster #	18

First name	Eric
Middle Initial	
Last name	Jankowski
Organization	Boise State University
NSF Award Title	Collaborative Research: NSCI Framework: Software for Building a Community-Based Molecular Modeling Capability Around the Molecular Simulation Design Framework (MoSDeF)
NSF Award Number	1835593
Abstract	In this project, nine research groups from eight universities are combining their expertise to create a software environment, called the Molecular Simulation Design Framework (MoSDeF) that will enable the automation of molecular-based computer simulations of soft materials (such as fluids, polymers, and biological systems) and will enable Materials Genome Initiative-style screening of such systems. MoSDeF is open source and the use of MoSDeF will enable reproducibility in molecular-based computer simulations, because all simulation steps, all input data, and all codes used will be publicly accessible to anyone to reproduce a published simulation. MoSDeF will contribute to reproducibility through standardization and maintaining the provenance of forcefields, one of the most common sources of irreproducibility in molecular-based simulations.
Keywords	molecular simulation, reproducibility, high performance computing, thermodynamics
Poster session	2
Poster #	19

First name	Kenneth
Middle Initial	E
Last name	Jansen
Organization	University of Colorado Boulder
NSF Award Title	NSCI SI2-S2I2 Conceptualization of CFDSI: Model, Data, and Analysis Integration for End-to-End Support of Fluid Dynamics Discovery and Innovation
NSF Award Number	1743178
Abstract	The conceptualized institute will make a wide variety of powerful simulation, data, and analysis resources available to the fluid dynamics research community by lowering or eliminating barriers associated with the adoption and use of these resources. The software infrastructure will have a number of positive impacts on the fluid dynamics research community. To do so, CFDSI will connect the best research in fluid dynamics to the best research in data science/analytics within a highly sustainable software development environment.
Keywords	Fluid Dynamics, Data Science, CFD, experiments
Poster session	3
Poster #	23

First name	Shantenu
Middle Initial	
Last name	Jha
Organization	Rutgers University
NSF Award Title	Elements: RADICAL-Cybertools: Middleware Building Blocks for NSF's Cyberinfrastructure
NSF Award Number	1931512
Abstract	Sophisticated and scalable workflows have become essential for advances in computational science. In spite of the many successes of workflow systems, there is an absence of a reasoning framework for end-users to determine which systems to use, when and why. Workflows are increasingly a manifestation of the algorithmic and methodological advances; workflow users and workflow system developers are often the same. Workflow systems must be easily extensible so as to support diverse functionality and the proverbial “last mile customization”. We advance the science of workflows and prevent workflow system “vendor lock-in” by formulating a building blocks approach to middleware for workflow systems grounded on four design principles of self-sufficiency, interoperability, composability, and extensibility. We will pursue developments, driven by the need to scale the number of software components, user, and supported platforms; and improve performance, engineering processes, and sustainability.
Keywords	Middleware, software and performance engineering, workflows
Poster session	2
Poster #	20

First name	Shantenu
Middle Initial	
Last name	Jha
Organization	Rutgers University
NSF Award Title	RHAPSODY: Runtime for Heterogeneous Applications, Service Orchestration and Dynamism
NSF Award Number	2103986
Abstract	The end of traditional scaling paradigms coupled with innovations in machine learning are driving unprecedented changes in the formulation of scientific applications as well as the nature of high-performance computing (HPC) software and application ecosystems. As a consequence, scientific applications increasingly depend on heterogeneous components with diverse computational characteristics and performance challenges, coordinated in agile and innovative ways. The Runtime for Heterogeneous APplications, Service Orchestration and DYnamism (RHAPSODY) addresses challenges arising from heterogeneity and need for performance. It will enable the effective and efficient execution of scientific applications at unprecedented scale and on a variety of current and upcoming HPC platforms.
Keywords	runtime system, high performance computing, cloud computing, workflows, heterogeneity
Poster session	2
Poster #	20

First name	Chaitanya
Middle Initial	
Last name	Joshi
Organization	Tufts University
NSF Award Title	Elements: Morpho-Cyberinfrastructure for scientists and engineers studying shape change
NSF Award Number	2003820
Abstract	We present Morpho, an open-source programmable environment for finite element modeling of shape, including shape optimization and shape shifting. We showcase example applications for a number of diverse systems like liquid crystal tactoids, swelling hydrogels and wetting on curved surfaces. We highlight our advances in the optimization algorithms under the hood as well as in the incorporation of GPU acceleration. Lastly, we summarize our code-base improvements, outreach efforts including the first Morpho workshop, and the adoption of Morpho by independent scientific researchers. *This material is based upon work supported by the National Science Foundation under grant OAC-2003820
Keywords	shape optimization, shapeshifting, soft matter, biophysics, open-source
Poster session	1
Poster #	29

First name	Ian
Middle Initial	
Last name	Joughin
Organization	University of Washington
NSF Award Title	Elements: Software. icepack: an open-source glacier flow modeling library in Python
NSF Award Number	1835321
Abstract	This project aims to support the development and maintenance of icepack, a new software package for simulating the flow of ice sheets and glaciers. Nearly all glaciologists, from specialists in remote sensing to ice core geochemistry, need to use simulations of ice flow. Using first-principles physics-based models, however, has largely been the province of experts in numerical simulation for much of the past 40 years. The goal of icepack is to lower the barrier to entry to using sophisticated physics models for all glaciologists, even those who are not experts in numerical simulation as such. This goal has become attainable in the past few years due to recent developments that combine symbolic and numerical computing approaches to discretizing and solver partial differential equations. In this presentation, we will describe some successes in broadening the swathe of the glaciological community that can productively use simulation tools.
Keywords	glaciology, simulation, numerics, data assimilation, climate change
Poster session	2
Poster #	21

First name	Mahmut
Middle Initial	T
Last name	Kandemir
Organization	Penn State
NSF Award Title	Frameworks: Re-Engineering Galaxy for Performance, Scalability and Energy Efficiency
NSF Award Number	1931531
Abstract	This project consists of four complementary tasks, which follow a logistic progression as follows: Task-I focuses on redesigning existing Galaxy tools with GPU/FPGA support and integrate them to Galaxy tool-chains; Task-II provides containerization support for the tools and accelerator-aware orchestration for running Galaxy on cloud platforms; Task-III implements specific policy driven scheduling schemes for Task-I and Task-II; and finally, Task-IV redesigns Galaxy storage to speed up execution and reduce bottlenecks related to data transfer. On the broader impact and outreach/educational front, this project impacts the performance and energy efficiency of Galaxy tools and applications and improves the productivity of a typical Galaxy user.
Keywords	Galaxy, GPU, biomedical, cloud, storage
Poster session	2
Poster #	22

First name	Peter
Middle Initial	
Last name	Kasson
Organization	University of Virginia
NSF Award Title	SCALE-MS - Scalable Adaptive Large Ensembles of Molecular Simulations
NSF Award Number	1835780
Abstract	The SCALE-MS project seeks to enable high-level algorithms for molecular simulation and simplify their execution on high-performance cyberinfrastructure. Increasingly, molecular simulation problems are best formulated not as a single simulation or even a static pipeline but as ensembles of simulations where the logic may adapt to intermediate results. The SCALE-MS project has designed high-level APIs for specifying this logic and links API execution to a capable runtime framework leveraging the RADICAL stack that runs on most NSF high-performance computing clusters. Our sets of problems and target codes include both biomolecular simulations with Gromacs and materials simulations with LAMMPS, but the API framework and software is intended to be general in nature.
Keywords	ensemble simulations, molecular dynamics simulations, adaptive workflows
Poster session	2
Poster #	23

First name	Latifur
Middle Initial	R
Last name	Khan
Organization	University of Texas at Dallas
NSF Award Title	Elements: Data: Sustaining Modern Infrastructure For Political And Social Event Data
NSF Award Number	1931541
Abstract	In this work, we propose deep-learning based frameworks to obtain state-of-the-art results for extracting structured events from natural language text in political and social sciences domains. We do so by exploring three main directions: (i) automatically extending the external dictionaries and knowledge bases utilized in the current event coders through knowledge extraction techniques; (ii) developing an innovative deep neural network design by combining state-of-the-art language representation models with multi-task learning technique to efficiently extract events in a structured format from multilingual corpus, and (iii) introducing ConfliBERT, a domain-specific pre-trained language model to monitor conflict and violence on a global scale.
Keywords	Deep Learning, Event Coding, BERT, Conflict, Violence
Poster session	2
Poster #	24

First name	Hyesoon
Middle Initial	
Last name	Kim
Organization	Georgia Tech
NSF Award Title	OAC CORE: ENABLING RAPID, TARGETED OPTIMIZATION OF PDE SOLVERS VIA HARDWARE
NSF Award Number	2103951
Abstract	The proposed framework offers end-to-end simulation capability, including both models based and learning based algorithms, as well as verification before hardware prototyping on different UAV (drone) platforms. Each of the components: flight scenarios, flight controllers, accelerators can be substituted with a simulation module that allows software/hardware in a loop (CPS simulation). The framework will allow the following research areas: (1) evaluation of new UAV flight control algorithms such as reinforcement algorithms and efficient way of implementing them (2) Evaluation of the requirement of computing power requirement for new algorithms (3) End-to-end performance and flight time/response time evaluations with proposed new accelerators.
Keywords	Multi-drone simulation, computing evaluation, FPGA, CPS simulation
Poster session	4
Poster #	42

First name	Andreas
Middle Initial	
Last name	Kloeckner
Organization	University of Illinois
NSF Award Title	SHF: Small: Collaborative Research: Transform-to-perform: languages, algorithms, and solvers for nonlocal operators
NSF Award Number	SHF-1911019
Abstract	Non-local operators such as layer potentials have long played an important role in mathematical modeling across many disciplines. Robust software systems for these currently lag far behind. This project will extend UFL, a well-known language for finite element discretizations of PDE, to include a robust range of non-local operators. As a result, application scientists will be able to concisely describe combined local/non-local models and obtain high-performance implementations.
Keywords	Non-local operator, Convolution, Potential, FEM/BEM coupling, Fast Algorithm, Domain-Specific Language
Poster session	2
Poster #	25

First name	Andreas
Middle Initial	
Last name	Kloeckner
Organization	University of Illinois
NSF Award Title	Elements: Transformation-Based High-Performance Computing in Dynamic Languages
NSF Award Number	OAC-1931577
Abstract	Multidimensional arrays (sometimes called 'tensors') are a foundational data structure for much of scientific computing. Through a polyhedrally-based program transformation tool, we will provide separation between mathematical intent and program optimization. In the proposed project, we will develop means for more efficient on-chip communication, code generation for prefix sums, reuse and abstraction in program transformation, and for expressing array computations in user programs.
Keywords	Array, GPU, Program transformation, Polyhedral, Separation of Concerns
Poster session	3
Poster #	9

First name	Fanwei
Middle Initial	
Last name	Kong
Organization	University of California, Berkeley
NSF Award Title	SI2-SSI: Collaborative Research: The SimCardio open source multi-physics cardiac modeling package
NSF Award Number	1663671
Abstract	Cardiovascular simulations have become a crucial component of fundamental research in surgical planning, device design diagnosis, and disease mechanisms. The project team has previously developed SimVascular, which is currently the only open source software package providing a complete pipeline from medical image data segmentation to patient specific blood flow simulation and analysis in arteries and veins. The SimCardio open source project will extend and enhance the functionality of SimVascular to the realm of heart modeling, providing the first fully integrated computer model of cardiac physiology and function to study numerous diseases affecting heart function. This computer modeling software will enable researchers to build models of the heart and vascular anatomy from medical imaging data, which can be used for personalized treatment planning and medical device design, ultimately leading to new treatments for patients with cardiovascular disease.
Keywords	Cardiovascular, Finite Element, SimVascular, Fluid Structure Interaction
Poster session	3
Poster #	10

First name	Alice
Middle Initial	E
Last name	Koniges
Organization	University of Hawai'i
NSF Award Title	Elements: ALE-AMR Framework and the PISALE Codebase
NSF Award Number	2005259
Abstract	This project creates software for modeling Partial Differential Equations (PDEs) on High Performance Computing Platforms and applies it to simulate complex groundwater flow processes in Hawaiian islands characterized by highly heterogeneous volcanic rocks and dynamic interaction between freshwater and seawater. Hawaii's water is pumped from volcanic aquifer systems since the islands are surrounded by ocean. Freshwater accumulates on top of saltwater, making it susceptible to anthropogenic activities and intrusion induced by sea and volcanic events. Groundwater flow modeling plays an important role in predicting sustainable yields for aquifer systems and planning groundwater resources for sea rise. Roughly half the population in the US lives near coastal areas where groundwater supplies much of the water. The software uses advanced mathematical techniques to dynamically adapt the grids and special Lagrangian-flow methods that can reproduce the sharp freshwater-seawater interfaces.
Keywords	Flow in Porous Media, Adaptive Mesh Refinement, Arbitrary Lagrangian Eulerian, PISALE
Poster session	2
Poster #	26

First name	Anthony
Middle Initial	
Last name	Kougkas
Organization	Illinois Institute of Technology
NSF Award Title	ChronoLog: A High-Performance Storage Infrastructure for Activity and Log Workloads
NSF Award Number	CSSI 2104013
Abstract	Modern applications generate massive amounts of data at unprecedented rates. Beyond simply storing data, one increasingly common requirement is to store activity data, also known as log data, which describe things that happen rather than things that are. The fast growing of activity data stresses current data management systems beyond their capability. This project develops ChronoLog, a novel system for organizing and storing activity data effectively and efficiently. ChronoLog uses physical time to provide a synchronization-free data distribution and total data ordering. It leverages multiple storage tiers, such as storage-class memories and new flash storage, to transparently auto-tier the log. ChronoLog will enable better fraud detection in financial transactions, faster and more accurate weather simulations, reduced time-to-insight for medical and bioengineering data, autonomous computing, and more secure web and mobile services.
Keywords	distributed log, shared log, tiered storage, activity data, high-performance storage
Poster session	4
Poster #	9

First name	Boris
Middle Initial	
Last name	Kozinsky
Organization	Harvard University
NSF Award Title	Elements: FLARE infrastructure for reproducible active learning of Bayesian force fields for ex-machina exascale molecular dynamics
NSF Award Number	2003725
Abstract	Molecular dynamics (MD) simulations are an indispensable tool for materials research for resolving microscopic details of atomic motion and predict thermodynamics, reaction kinetics and ionic diffusivities. Accurate interatomic forces at each MD step can be obtained using density functional theory (DFT), but this approach is computationally costly and applicable only when the simulation size is small and time is short. Acceleration can be achieved by replacing expensive DFT force calculations by a surrogate machine learning (ML) model. The central challenge is to be able to reliably and quickly train accurate ML models and quantify their uncertainty. Principled uncertainty quantification is especially critical for prediction of non-equilibrium dynamics, where rare important events, such as the breaking of bonds or atomic migration, determine the material's performance but involve atomic configurations that are unlikely to be in the unbiased training set.
Keywords	molecular dynamics, materials simulation, machine learning
Poster session	1
Poster #	30

First name	JoAnn
Middle Initial	C
Last name	Kuchera-Morin
Organization	University of California, Santa Barbara
NSF Award Title	Elements: Cyber-infrastructure for Interactive Computation and Display of Materials Datasets
NSF Award Number	2004693
Abstract	We are developing TINC, a new toolkit for interactive computation and visualization, applying it to materials simulations approaches, interfacing this toolkit with the software package CASM, designed to enable first-principles statistical mechanics studies of complex crystalline materials through a custom application built on top of TINC. TINC is being developed for a fundamental study of ion transport and intercalation processes in the electrode materials of sodium (Na) ion batteries. TINC ties computation and display to data through user interaction, a general-purpose platform useful in different domains. Researchers performed rapid iterations of Grand Canonical and Kinetic Monte Carlo simulations generated on the HPC cluster, facilitating verification of simulations in a near-to-real-time environment, a faster path to insight and validation. TINC can analyze data generated with Bayesian learning tools such as STAN. Bayesian regression is used in all disciplines of engineering.
Keywords	interactive computation, computational materials science, interactive visualization, complex crystalline materials research, automated model fitting
Poster session	1
Poster #	31

First name	Chris
Middle Initial	
Last name	Kuhlman
Organization	University of Virginia
NSF Award Title	Collaborative Research: Framework: Software: CINES: A Scalable Cyberinfrastructure for Sustained Innovation in Network Engineering and Science
NSF Award Number	1916805
Abstract	Networks are ubiquitous and are a part of our common vocabulary. Network science and engineering that emerged as a formal field over the last 20 years has seen explosive growth. Ideas from network science played a central role in the formation of companies such as Akamai, Twitter, Google, Facebook, and LinkedIn. The concepts have also been used to address fundamental problems in diverse fields (e.g., epidemiology and marketing), and are now part of most university curricula. Network science is multi-disciplinary, yet resources for doing network science are largely dispersed and standalone, of small scale, home-grown for personal use, and/or do not cover the broad range of operations that need to be performed on networks, much less compose these operations.
Keywords	Network Science, cyberinfrastructure, engineering
Poster session	4
Poster #	10

First name	David
Middle Initial	
Last name	Lange
Organization	Princeton University
NSF Award Title	Elements: C++ as a service - rapid software development and dynamic interoperability with Python and beyond
NSF Award Number	1931408
Abstract	A key enabler of innovation and discovery for many scientific researchers is the ability to explore data and express ideas quickly as software prototypes. Tools and techniques that reduce the "time to insight" are essential to the productivity of researchers. Conversely, performance-focused languages, such as C++, are a critical infrastructure component for many scientific fields that have either large computing challenges or the need for low latency for results. The productivity of data scientists is increased by an easy to use dynamic programming and development environment, together with a fully featured interoperability layer. The CaaS project provides a dynamic C++ execution environment and enables runtime language interoperability between C++ and other languages, such as Python, through a native-like, dynamic environment.
Keywords	llvm, interoperability, interactivity
Poster session	3
Poster #	11

First name	Christopher
Middle Initial	
Last name	League
Organization	Long Island University
NSF Award Title	Bifrost: A CPU/GPU Pipeline Framework for High Throughput Data Acquisition and Analysis
NSF Award Number	2103771
Abstract	Bifrost is a Python/C++/CUDA pipeline framework for developing high-throughput data acquisition and analysis systems. This project aims to improve the usability of Bifrost, its performance, and its applicability to other disciplines. We intend to increase the data rates that Bifrost is capable of handling, and to support more sophisticated visualization, verification, and debugging of pipelines. We are also working with collaborators to incorporate Bifrost in telescopes and other instruments and disciplines.These improvements will increase the scientific return of current and future radio telescopes. The power and flexibility of Bifrost has allowed us to develop a broadband imaging correlator with which to better study meteor radio afterglows, which are providing new insights into ionospheric physics. We seek to engage the Remote Sensing, Geophysics, and other scientific communities in the use of Bifrost for data capture and analysis.
Keywords	GPU, Python, radioastronomy, data acquisition, big data analysis
Poster session	2
Poster #	27

First name	Sanjiva
Middle Initial	K
Last name	Lele
Organization	Stanford University
NSF Award Title	Elements: AMR-H: Adaptive multi-resolution high-order solver for multiphase compressible flows on heterogenous platforms
NSF Award Number	2103509
Abstract	This project develops a general computational framework combining high-order, high accuracy, solution-adaptive discretizations of partial differential equations (with emphasis on flows of non-ideal fluids) tailored to the physics they represent. The discretization is optimized for high resolving efficiency, and high utilization of the computer resources due to its high arithmetic intensity and data locality. Adaptive mesh refinement in combination with high-order multi-resolution compact scheme allows for easy pre-processing and meshing for complex-geometry problems. Co-designing the numerical framework with new developments in the Legion framework would allow for automated, optimized runtime scheduling of tasks involving computational kernels and data movement across memory hierarchies. Efficient leveraging of Kokkos, would free the computational scientist/engineer from hardware specific programming models and allow exascale computations on heterogeneous computers.
Keywords	high-order discretization, compressible flow simulations, turbulence, AMR, Exascale computing
Poster session	2
Poster #	29

First name	Zhen
Middle Initial	
Last name	Li
Organization	Clemson University
NSF Award Title	Collaborative Research: Elements: SciMem: Enabling High Performance Multi-Scale Simulation on Big Memory Platforms
NSF Award Number	2103967
Abstract	The project will enable high performance multiscale simulations on big memory platforms through more efficient utilization of large and heterogeneous memory machines. Specifically, it will replace computations with pre-computed and stored in memory data on a heterogeneous computing systems. The developed tool named SciMem will be integrated and tested with the popular parallel molecular dynamics (MD) simulator, LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator). The developed improvements in the use of computational resources will allow more accurate models of complex physical phenomena to be carried out on the emerging hardware systems. SciMem aims to bring a 10x performance improvement for certain larger-scale multi-scale simulations widely applied in the fields of computational chemistry and material science, e.g., quantum mechanical/molecular mechanical-based MD simulation of catalysis.
Keywords	multiscale simulation, molecular dynamics, big memory, heterogeneous computing
Poster session	3
Poster #	12

First name	Xu
Middle Initial	
Last name	Liang
Organization	University of Pittsburgh
NSF Award Title	CyberWater—An Open and Sustainable Framework for Diverse Data and Model Integration with Provenance and Access to HPC
NSF Award Number	1835785
Abstract	This project develops CyberWater, a community-driven open and sustainable modeling software framework that integrates a wide range of models and datasets across disparate temporal and spatial scales, to address a high priority need for water research communities: interoperability among a wide variety of data sources and models, and integration of different computational models into water research communities. The CyberWater framework allows scientists to discover and access heterogeneous data online for diverse models, to assess and evaluate model interactions and outcomes, and to use these coupled models to test comprehensive hypotheses and alternate processes. Moreover, CyberWater enables not only easy and incremental integration of diverse data sources and models but also seamless and on-demand access to various HPC (high performance computing) resources. Its core components include the Meta-Scientific-Modeling (MSM) core, Data Agents, Model Agents, and generic model agent toolkit.
Keywords	Open data and modeling, graphical workflows, model integration, HPC access on-demand, reproducibility
Poster session	2
Poster #	30

First name	Guoyu
Middle Initial	
Last name	Lu
Organization	Rochester Institute of Technology
NSF Award Title	Elements: A Deep Neural Network-based Drone (UAS) Sensing System for 3D Crop Structure Assessment
NSF Award Number	2104032
Abstract	Unmanned aerial systems (UAS) provide a holistic view of crop growth, thereby facilitating the ability of agriculture professionals to accurately and rapidly identify growth risks and optimize management inputs. Existing UAS solutions mainly involve expensive sensors, required to assess crop physiology and structure. The overarching objective of this project is to develop a cost-effective and accurate/precise 3D reconstruction sensing system that can be deployed on UAS to be used by agricultural researchers, growers, and service providers to rapidly assess crop growth. Our approach relies on a specialized AI system for large area, fine-scale agriculture 3D reconstruction applications. The UAS sensing system will be able to estimate crop structure for a large (field-level and beyond) coverage area at a much lower cost (in terms of both time and monetary perspectives) than current standards that rely on active LiDAR or even structure-from-motion approaches.
Keywords	3D Sensing, Crop Assessmen, LiDAR-camera fusion, Low cost UAS system, Precision Agriculture
Poster session	3
Poster #	13

First name	Bertram
Middle Initial	
Last name	Ludaescher
Organization	University Illinois, Urbana-Champaign
NSF Award Title	CC*DNI DIBBS: Merging Science and Cyberinfrastructure Pathways: The Whole Tale
NSF Award Number	1541450
Abstract	Whole Tale is an open-source software platform for computational transparency and reproducibility. Its goal is to simplify the process of creating, publishing, and verifying the transparency and reproducibility of computational research artifacts, which we call tales. Tales are research objects that capture the code, data, workflow, documentation, results and information about the computational environment typically associated with a published manuscript. They are intended to be executable (i.e., a workflow can be executed to obtain consistent results) as well as publishable and transparent. We operate a production instance of the platform on the NSF Jetstream Cloud infrastructure, but it is also locally installable for interested communities. Since 2018, over 1300 users across various scientific domains have created over 2300 tales. Whole Tale is used primarily for research and teaching and is currently being evaluated for use in journal reproducibility audit workflows.
Keywords	reproducibility, transparency, research objects, verification workflows
Poster session	2
Poster #	31

First name	Xiaogang (Marshall)
Middle Initial	
Last name	Ma
Organization	University of Idaho
NSF Award Title	Elements: Software: HDR: A knowledge base of deep time to facilitate automated workflows in studying the co-evolution of the geosphere and biosphere
NSF Award Number	1835717
Abstract	This project will result in the creation of a software that will support research in the Earth's deep time history. The co-evolution of the geosphere and biosphere is one of the fundamental questions for the 21st century Earth science. However, the shortage of efficient methods for accessing and synthesizing multi-source datasets hamper the data-intensive co-evolution research. Geologic time is an essential topic in the co-evolving geosphere and biosphere, and can be used as a common reference to connect various parameters among the data silos. This project will improve the machine readability and alignment of various global, local and regional geologic time standards and build a knowledge base of deep time and its service on the Web.
Keywords	Geologic time, Knowledge graph, Workflow systems, Co-evolution, Open science
Poster session	3
Poster #	14

First name	B.S.
Middle Initial	
Last name	Manjunath
Organization	UC Santa Barbara
NSF Award Title	SI2-SSI: LIMPID: Large-Scale IMage Processing Infrastructure Development
NSF Award Number	1664172
Abstract	Scientific imaging is ubiquitous: From materials science, biology, marine science to remote sensing and medicine, much of the big data science is image centric. Currently, interpretation of images is usually performed within isolated research groups either manually or as workflows over narrowly defined conditions with specific datasets. The project will create an image processing marketplace for use by a diverse community of researchers, enabling them to discover, test, verify and refine image analysis methods within a shared infrastructure. The potential impacts of the project are significant: from wide dissemination of novel processing methods, to development of automatic methods that can leverage data and human feedback from large datasets for software training and validation. For the broader scientific community, this immediately provides a resource for joint data and methods publication, with provenance control and security.
Keywords	Computer vision, machine learning, image processing, software infrastructure
Poster session	2
Poster #	32

First name	T. Andrew
Middle Initial	
Last name	Manning
Organization	National Center for Supercomputing Applications
NSF Award Title	Frameworks: MUSES, Modular Unified Solver of the Equation of State
NSF Award Number	2103680
Abstract	MUSES is a large collaboration project that is developing cyberinfrastructure and novel tools for the scientific community to answer critical interdisciplinary questions in nuclear physics, gravitational wave astrophysics, and heavy-ion physics. Initial work is primarily converting, rewriting, and upgrading existing code, mostly as high-performance C++ libraries, to create the core modules of the Calculation Engine. We are building a cloud-native deployment system based on the latest container technologies, leveraging the power of Kubernetes to create a reproducible environment compatible with a wide range of hosting platforms. Individuals and organizations can deploy their own MUSES instance on-premises, utilizing their own backend computing solutions and maintaining full control over their data.
Keywords	high performance computing, equation of state, nuclear physics, gravitational wave astrophysics, heavy-ion physics
Poster session	2
Poster #	33

First name	Reed
Middle Initial	
Last name	Maxwell
Organization	Princeton University
NSF Award Title	Collaborative Research: Framework: Software: NSCI : Computational and data innovation implementing a national community hydrologic modeling framework for scientific discovery
NSF Award Number	2054506 / 1835903
Abstract	HydroFrame is a platform to facilitate easy interaction with large computationally intensive hydrologic models and massive simulated outputs. Our tools enable users to subset model inputs and outputs for any watershed in the US, run their own simulations, and visualize and analyze existing model outputs or newly generated results. We also develop free educational tools and lesson plans to teach students of all ages about groundwater and the hydrologic cycle. https://hydroframe.org
Keywords	hydrologic modeling, community platforms, data and more sharing, model performance
Poster session	4
Poster #	11

First name	Kenton
Middle Initial	
Last name	McHenry
Organization	University of Illinois at Urbana-Champaign
NSF Award Title	Collaborative Research: CSSI: Framework: Data: Clowder Open Source Customizable Research Data Management, Plus-Plus
NSF Award Number	1835834
Abstract	Sharing and reusing diverse collections of data is essential to scientific discoveries in many domains, in particular where data and tools must span multiple domains. To support these needs effectively new methods are required that simplify the amount of effort needed by researchers to find and utilize data, support community accepted data practices, and bring together the breadth of standards, tools, and resources utilized by a community. Clowder addresses these challenges by distributing the data curation overhead throughout the lifecycle of the data, interacting with and utilizing a variety of community tools, while also supporting different data governance and ownership requirements. The effort will transition and expand the grassroots Clowder user community into a distributed and organized open source community, governed by the stakeholders, providing a long term sustainable software resource supporting convergent research data needs.
Keywords	data management, active curation, analytics, convergent research
Poster session	3
Poster #	16

First name	Charles
Middle Initial	
Last name	Meneveau
Organization	Johns Hopkins University
NSF Award Title	Frameworks: Advanced Cyberinfrastructure for Sustainable Community Usage of Big Data from Numerical Fluid Dynamics Simulations
NSF Award Number	2103874
Abstract	This project will develop and implement an advanced cyberinfrastructure for turbulence databases. The system will enable ground-breaking research on fluid turbulence in various engineering, atmospheric and ocean flows. In most high-fidelity computer simulation-based research, the prevailing data service approaches have not been able to democratize access to the vast amounts of data generated even through substantial computational efforts are being invested. An existing cyberinfrastructure that serves as a predecessor has seen growing usage but has focused on engineering turbulent flows and its services have programming limitations and are difficult to scale up. The system will dramatically expand services as well as user communities in geophysics and in engineering fluid dynamics. Broader impacts arise from the community, educational use and outreach efforts enabled by a cybersystem that facilitates access to massive numerical turbulence datasets of great interest.
Keywords	Big data, turbulence, computational fluid dynamics, geophysics, numerical laboratories
Poster session	2
Poster #	35

First name	Tim
Middle Initial	
Last name	Menzies
Organization	NC State
NSF Award Title	Elements: Can Empirical SE be Adapted to Computational Science?
NSF Award Number	1931425
Abstract	This project will apply data miners, hyperparameter optimizers and active learning to project data from the computational science community to reduce the associated cost (time, money, etc.) required to handle many of the large and more tedious aspects of software development. This will free up more time of the computational scientists, and allow them to focus on core scientific issues.
Keywords	Software engineering, empirical, computational science
Poster session	3
Poster #	17

First name	Jelena
Middle Initial	
Last name	Mirkovic
Organization	USC/ISI
NSF Award Title	Elements: Software: Distributed Workflows for Cyberexperimentation
NSF Award Number	1835608
Abstract	This project will enable rigorous and repeatable experimentation on testbeds by enabling experiments to be defined as executable workflows, that can then be repeatably run, as well as managed. This project will develop and deploy a software framework called Elie. Elie consists of an experiment representation capability, along with several supporting services that make experimentation workflows robust, fault tolerant, easily sharable and reusable. Elie's technologies provide testbed support throughout an experiment lifecycle. All of Elie's technologies are extensible, and open-sourced, to allow the research community to contribute to their development and customize them to their needs. Outcomes of this project will significantly improve ease of testbed experimentation, by offloading tedious, repetitive and detail-sensitive tasks to testbeds. This will shorten experiment duration and improve quality and reliability of results.
Keywords	testbeds, experiment lifecycle, repeatability, reproducibility, experiment workflow
Poster session	3
Poster #	18

First name	Dmitry
Middle Initial	Y
Last name	Mishin
Organization	University of California San Diego
NSF Award Title	Pacific Research Platform
NSF Award Number	ACI-1541349
Abstract	The Pacific Research Platform (PRP) is a multi-institutional extensible deployment that establishes a science-driven high-capacity data-centric 'freeway system.' The PRP spans all 10 campuses of the University of California, as well as the major California private research universities, four supercomputer centers, and several universities outside California. Fifteen multi-campus data-intensive application teams act as drivers of the PRP.
Keywords	kubernetes, network, monitoring, machine learning, measurements
Poster session	3
Poster #	19

First name	Dane
Middle Initial	
Last name	Morgan
Organization	University of Wisconsin – Madison
NSF Award Title	Machine Learning Materials Innovation Infrastructure
NSF Award Number	1931298
Abstract	The goal of this work is to support the development of use of machine learning models in materials science and engineering through a new ecosystem centered on accessible formatted data, containerized models, and robust quality assurance. We will do this through (i) a cloud service of formatted machine-learning ready databases and containerized callable machine learning models, (ii) accessible methods of error and domain assessment for models to support robust use of models in complex workflows, (iii) engaging the relevant communities to grow this ecosystem. We have made significant progress on a cloud-based data and model service and practical methods for error and domain guidance. Further work is needed to make the cloud services robust enough for broad community submission, deliver the error and domain tools in a simple open source package, and engage the relevant communities.
Keywords	Machine learning, cloud resource, materials data, error bars, model domain
Poster session	3
Poster #	20

First name	David
Middle Initial	
Last name	Morse
Organization	Univ. of Minnesota
NSF Award Title	Elements: Open-source tools for block polymer phase behavior
NSF Award Number	2103627
Abstract	Block polymers are polymers containing two or more covalently connected blocks containing distinct monomer repeat units. Thermodynamic incompatibility between different monomers drives micro-phase separation into a wide variety of spatially periodic equilibrium structures. This project aims to provide a versatile package of tools for accurate field-theoretic simulation of such systems, including tools for both self-consistent field theory (SCFT), the current work-horse method, and more accurate stochastic field theoretic simulation (FTS) methods. The package is written in C++ using CUDA for GPU acceleration.
Keywords	Polymer, Copolymer, Field Theoretic Simulation
Poster session	2
Poster #	36

First name	Nick
Middle Initial	
Last name	Murphy
Organization	Center for Astrophysics | Harvard & Smithsonian
NSF Award Title	Collaborative Research: Frameworks: An open source software ecosystem for plasma physics
NSF Award Number	1931388
Abstract	The mission of the PlasmaPy project is to foster the creation of a fully open source software ecosystem for plasma research and education. The PlasmaPy package is being developed to include the common core functionality needed by plasma physicists across disciplines. The functionality includes object-oriented representations of particles, a subpackage with commonly needed plasma formulae, wave dispersion relationship solvers, and tools to analyze laboratory plasma experiments and simulations. PlasmaPy prioritizes code readability, consistency, and maintainability while using best practices for scientific computing such as open development, version control, continuous integration testing, and code review. PlasmaPy is being developed to be compatible with Astropy as well as packages from the Python in Heliophysics Community.
Keywords	plasma, fusion energy, Python
Poster session	3
Poster #	21

First name	Dan
Middle Initial	
Last name	Negrut
Organization	University of Wisconsin-Madison
NSF Award Title	Elements:Software:NSCI: Chrono - An Open-Source Simulation Platform for Computational Dynamics Problems
NSF Award Number	CISE1835674
Abstract	This project seeks to augment modeling and solution methods employed by Chrono, an open-source computer simulation platform for multi-body dynamics and fluid-solid interaction problems. Chrono is used and further developed by other users and has an active forum with close to 500 registered users. This project enhanced the richness of Chrono's modeling features and ability to leverage emerging hardware architectures. Chrono has been used in extraterrestrial applications, machine learning, image processing, computer vision, mechanical watch design, architectural studies, autonomous vehicles, wind turbine dynamics, next generation space suit design, oil extraction and accident mitigation, hardware-in-the-loop simulation, etc. This project engaged high-school students from under-represented groups in a six-day residential camp run and will train a group of undergraduate students from California State University at University of Wisconsin-Madison through a new two-week residential program.
Keywords	rover simulation, simulation in robotics, fluid-solid interaction, open source
Poster session	1
Poster #	42

First name	Ge
Middle Initial	
Last name	Ou
Organization	University of Florida
NSF Award Title	Elements: Open Access Data Generation Engine for Bulk Power System under Extreme Windstorms
NSF Award Number	2004658
Abstract	This project addresses that gap in knowledge in dealing with high wind disaster events for preventative and restorative resiliency of electric power networks. By integrating multiple data sources into a robust simulation tool, researchers can design new methods to mitigate the impact of hurricanes and other extreme wind events on power system operations. The outcomes of the research would positively impact reliability, resiliency, and delivery of electric power to US population centers.The framework is unique and will enable inter-disciplinary research between atmospheric sciences, civil engineering, and electric power engineering. In addition to three deliverable test cases (Texas, New York, and Florida), the project will design automated tools to create new test cases in the future for other researchers. The power system digital surrogate, correlation of datasets, and overall data generation engine will enhance the national cyberinfrastructure ecosystem.
Keywords	Power System, extreme weather; resilience; hurricanes; civil infrastructure
Poster session	1
Poster #	32

First name	Shrideep
Middle Initial	
Last name	Pallickara
Organization	Colorado State University
NSF Award Title	Frameworks: Collaborative Proposal: Software Infrastructure for Transformative Urban Sustainability Research.
NSF Award Number	OAC-1931363
Abstract	The United States is highly urbanized with more than 80% of the population residing in cities. Cities draw from and impact natural resources and ecosystems while utilizing vast, expensive infrastructures to meet economic, social, and environmental needs. The NSF has invested in several strategic research efforts in urban sustainability that generate, collect, and manage large volumes of spatiotemporal data. Voluminous datasets exist in domains such as climate, ecology, health, and census. These data can spur exploration of new hypotheses, particularly across traditionally disparate disciplines, and offer opportunities for discovery and innovation. However, the data are encoded in diverse formats and managed using multiple data management frameworks -- that inhibit discovery. A scientist must reconcile not only the encoding and storage frameworks, but also negotiate authorizations. This project, SUSTAIN, facilitates and accelerates discovery by alleviating data-induced inefficiencies.
Keywords	big data, spatial data, data mining, analytics, modeling
Poster session	3
Poster #	24

First name	Dhabaleswar
Middle Initial	K
Last name	Panda
Organization	The Ohio State University
NSF Award Title	Collaborative research: Frameworks: Designing next-generation MPI libraries for emerging dense-GPU Systems
NSF Award Number	1931537
Abstract	Modern HPC platforms are using multiple CPU, GPUs and high-performance interconnects per node. Unfortunately, state-of-the-art production quality implementations of the popular Message Passing Interface (MPI) programming model do not have the appropriate support to deliver the best performance and scalability for applications (HPC and DL) on such dense GPU systems. The project involves a synergistic and comprehensive research plan, involving computer scientists from OSU and OSC and computational scientists from TACC, SDSC and UCSD. The proposed innovations include: 1) Designing high-performance and scalable communication operations that fully utilize multiple network adapters and advanced in-network computing features for GPU and CPU; 2) Designing novel datatype processing and unified memory management; 3) Designing CUDA-aware I/O; 4) Designing support for containerized environments; and 5) Carrying out integrated evaluation with a set of driving applications.
Keywords	HPC, Deep Learning, GPU, MPI, CUDA-Aware MPI
Poster session	3
Poster #	25

First name	Daniele
Middle Initial	
Last name	Panozzo
Organization	New York University
NSF Award Title	Elements:Software:Open-Source Robust Geometry Toolkit for Black-Box Finite Element Analysis
NSF Award Number	1835712
Abstract	The numerical solution of partial differential equations (PDEs) is ubiquitous in science and engineering applications. Ideally, a PDE solver should be a ``black box'': the user provides as input the domain boundary, boundary conditions, and the governing equations, and the code computes the value of the solution at a set of user-specified points of the input domain. This is surprisingly far from being the case for all existing open-source or commercial software, despite the research efforts in this direction and the large academic and industrial interest. This project introduces an integrated pipeline, considering meshing and element design as a single challenge, and developing a software platform to enable black box analysis on complex geometric models represented as point clouds, triangle meshes, or CAD (Computer Aided Design) models, opening the door to new shape design technique to a wide range of new applications in sciences and engineering.
Keywords	Simulation, Finite Element Method, Shape Design, Partial Differential Equation, Robust Computing
Poster session	3
Poster #	26

First name	Anant
Middle Initial	K
Last name	Paravastu
Organization	Georgia Institute of Technology
NSF Award Title	Element: Computational Toolkit to Discover Peptides that Self-assemble into User-selected Structures
NSF Award Number	OAC-1931430
Abstract	The goal of this project is to develop an open software toolkit that enables the identification of peptide sequences that are capable of assembling into user-selected β-sheet-based structures. The user will be able to screen potentially thousands of peptide sequences that assemble spontaneously into the structure of their choosing and rank order them according to their stability. Discontinuous molecular dynamics (DMD) simulation software will also be provided to enable analysis of the kinetics of the assembly of the designed structures. To establish efficacy and a basis for future improvement of computational tools, selected designs will be validated experimentally using biophysical characterization techniques and solid-state nuclear magnetic resonance spectroscopy. The software toolkit, which will be the first of its kind for β-sheet assemblies, will be a “plugin” available on the NSF-sponsored Molecular Simulation and Design Framework (MoSDeF).
Keywords	peptide assembly, peptide design, beta-sheet, discontinuous molecular dynamics, solid-state NMR
Poster session	3
Poster #	27

First name	Jignesh
Middle Initial	
Last name	Patel
Organization	University of Wisconsin
NSF Award Title	Elements: Software: Towards Efficient Embedded Data Processing
NSF Award Number	1835446
Abstract	Embedded databases are ubiquitous; for example, every smart phone today has many embedded databases, which implies that billions of people worldwide carry dozens of databases in their phones/pockets every day. Many of these databases are powered by data processing technology that has not kept up with the pace with which the underling hardware in phones have evolved. As a result, data processing is slow, and consumes more energy than needed. The focus of this proposal is on developing new data processing technology for mobile devices that targets a 10X efficiency and performance improvements. The aims of the project go beyond more efficient data processing on phones to also include more efficient processing in embedded environments, which also includes databases running on laptops. Thus, the project aims for a broad impact on database across a spectrum of mobile devices.
Keywords	databases, mobile, embedded, sensors, SQLite3
Poster session	3
Poster #	28

First name	Abani
Middle Initial	K
Last name	Patra
Organization	Tufts University
NSF Award Title	Ghub: A new community-driven data-model resource for ice-sheet scientists
NSF Award Number	2004302
Abstract	Predicting future ice sheet change requires a tremendous effort across a range of disciplines in ice sheet science, including expertise in observational data, paleoglaciology, numerical ice sheet modeling, and widespread use of emerging methodologies for learning from the data. However, significant knowledge and disciplinary barriers make collaboration between data and model groups the exception rather than the norm. We are developing a community-building scientific and educational cyberinfrastructure framework that includes models and data processing tools, online simulation and collaboration support ghub.org to enable coordination and synergistic exchange between ice sheet scientific communities. These tools target a wide range of audience, ranging from ice sheet modeling community efforts such as the Ice Sheet Model Intercomparison Project for CMIP6 (ISMIP6) to more specialized process orientated investigations.
Keywords	Ice sheet models, Ice sheet observation data, glaciology and cryosciences
Poster session	1
Poster #	33

First name	Michele
Middle Initial	
Last name	Pavanello
Organization	Rutgers University - Newark
NSF Award Title	Collaborative Research: Elements: Flexible & Open-Source Models for Materials and Devices
NSF Award Number	1931473
Abstract	The aim of this collaborative research project is to implement, disseminate and apply a flexible and modular implementation of multiscale approaches for first-principles condensed-matter simulations. Our focus is along the following three directions: the development of a core density-functional-theory engine in python for condensed phase systems; the development of python-based embedding libraries for continuum and density embedding models; the combination of these two embedding schemes into a single flexible approach. Current progress along these three directions will be presented.
Keywords	Density Functional Theory, Density Embedding, Multiscale Models, Continuum Models, Electronic structure
Poster session	3
Poster #	29

First name	Daniel
Middle Initial	R
Last name	Phillips
Organization	Ohio University
NSF Award Title	Frameworks: Bayesian Analysis of Nuclear Dynamics
NSF Award Number	2004601
Abstract	Nuclear physicists seek a quantitative description of strongly-interacting matter. Sophisticated models of how neutrons and protons interact in the nucleus, extreme environments, and collisions between nuclei have been key to progress made towards this goal. These models typically describe extant data well, but often yield divergent predictions for future experiments. The Bayesian Analysis of Nuclear Dynamics (BAND) framework will make broadly available a set of computational tools built through collaboration between statisticians, computer scientists and nuclear physicists. It will combine the results of several models, incorporating prior knowledge and data for each, to produce a full assessment of the uncertainty in nuclear-physics predictions. This will enable quantitative evaluation of the impact of future experiments. It will also help quantify uncertainties for terrestrially inaccessible environments, e.g., the core of neutron stars or the first microsecond after the Big Bang.
Keywords	Bayesian statistics; uncertainty quantification; nuclear physics; experimental design; computationally expensive models.
Poster session	1
Poster #	34

First name	Carlo
Middle Initial	
Last name	Piermarocchi
Organization	Michigan State University
NSF Award Title	Elements: Software: NSCI: A Quantum Electromagnetics Simulation Toolbox (QuEST) for Active Heterogeneous Media by Design
NSF Award Number	1835267
Abstract	Designing novel optical materials with enhanced properties would impact many areas of science and technology, leading to new lasers, better components for photonics, and to a deeper understanding of how light interacts with matter. This project develops software that simulates how light would propagate in yet to be made complex optical materials. The final product will be a software toolbox that computes the dynamics of each individual light emitter in the materials rather than calculating an average macroscopic field. This toolbox permits the engineering and optimization of optical properties combining heterogeneous components at the nanoscale. The software will be disseminated widely to enable scientists worldwide to conduct research on this area and will provide a blueprint for broader applications to magnetic materials and ultrasound acoustics.
Keywords	Nonlinear optics, computational electromagnetics, light-matter interaction
Poster session	3
Poster #	31

First name	Jim
Middle Initial	
Last name	Pivarski
Organization	Princeton University
NSF Award Title	Awkward Arrays - Accelerating scientific data analysis on irregularly shaped data
NSF Award Number	2103945
Abstract	Awkward Array is a Python library for manipulating arrays of JSON-like data with NumPy-like idioms. Like JSON, the data may contain variable-length lists and strings, nested records, missing values, and even mixed data types. Unlike JSON, the data are encoded in an efficient "columnar" format that data analysts and software developers can restructure and perform computations on using an interface that generalizes Python's popular NumPy library. Awkward Array was developed for particle physics, but applies to any problem domain with large datasets of complex data types. Our project is to identify these domains, help experts apply Awkward Array to their projects, and learn from their feedback about how it can be improved for their use-cases. In particular, we will be expanding Awkward Array's interoperability with the rest of the scientific Python ecosystem, starting with a new Dask collection type for Awkward Arrays, and a CuPy-based backend to enable computations on GPUs.
Keywords	Python, NumPy, JSON-like data, data analysis, ragged array, jagged array, Apache Arrow
Poster session	3
Poster #	32

First name	Serban
Middle Initial	
Last name	Porumbescu
Organization	University of California, Davis
NSF Award Title	SI2-SSE: Gunrock: High-Performance GPU Graph Analytics.
NSF Award Number	1740333
Abstract	We present a plan to develop the Gunrock programmable, high-performance graph analytics library for programmable graphics processors (GPUs) from a working prototype to a robust, sustainable, open-source component of the GPU computing ecosystem. We believe Gunrock is the most fully-featured and highest-performing programmable graph library today for single-GPU and single-node, multiple-GPU graph analytics. However, we see numerous exciting opportunities for a coherent program of research and development to improve Gunrock and make it more valuable for the computing community, including supporting greater scalability, expanding the Gunrock stack with both higher-level APIs and additional Gunrock back ends, adding more core operations and primitives to Gunrock, and exploring graph data structure challenges. In addition, as the grant is coming to an end we hope to present some lessons learned.
Keywords	graph, analytics, gpu, sparse, parallel
Poster session	4
Poster #	12

First name	Loic
Middle Initial	
Last name	Pottier
Organization	University of Southern California
NSF Award Title	Simulation-driven Evaluation of Cyberinfrastructure Systems
NSF Award Number	2103508
Abstract	Most scientific breakthroughs and discoveries are now preconditioned on performing complex processing of vast amounts of data as conveniently, reliably, and efficiently as possible. This requires software systems to automate the processing on these resources, an enormous amount of effort has been invested in producing such "cyberinfrastructure" software systems. This project transforms the way in which these systems are evaluated, so that their capabilities can be developed and evolved judiciously. This project uses simulation to transform the way in which cyberinfrastructure systems are evaluated as part of their long-term development cycles. This is achieved via software elements for enhancing production cyberinfrastructure systems with simulation capabilities so as to enable quantitative evaluation of these systems for arbitrary execution scenarios.
Keywords	simulation, cyberinfrastructure, scientific workflows, distributed computing, wrench
Poster session	2
Poster #	37

First name	Saptarshi
Middle Initial	
Last name	Purkayastha
Organization	Indiana University
NSF Award Title	CICI: SSC: Development of a Secure and Privacy-Preserving Workflow Architecture for Dynamic Data Sharing in Scientific Infrastructures
NSF Award Number	1839746
Abstract	The project aims to provide a secure architecture on JetStream, the first public cloud funded by the NSF and used by many researchers. The secured infrastructure provides comprehensive multi-level protection for the Protected Health Information (PHI) and its workflows through user authentication, fine-tuned data access control, confidentiality, integrity, and traceability. The project implements role-wise passwordless authentication and authorization, cryptography-based hierarchical access control, dual-level key management, and secure digital provenance integrity protection. By employing these, JetStream VMs can guarantee the security, privacy, and integrity of scientific workflows and associated data, thus protecting data and computing resources from internal and external attacks. When applied to healthcare and life-science cyberinfrastructures, it enables sensitive health data to be shared securely, which is an essential requirement for accelerating life science research.
Keywords	healthcare workflows, role-based authentication, hierarchical access control, JetStream, OpenMRS
Poster session	4
Poster #	13

First name	Rajiv
Middle Initial	
Last name	Ramnath
Organization	The Ohio State University, Computer Science and Engineering
NSF Award Title	EAGER: Bridging the Last Mile; Towards an Assistive Cyberinfrastructure for Accelerating Computationally Driven Science
NSF Award Number	1945347
Abstract	This exploratory project investigates whether effective, intelligent guidance can be extracted from past experience on specific applications using artificial intelligence techniques and then provided in a tailored manner to the individual researcher. This will be done within the software tools being used by the researcher, and provided by the underlying cyberinfrastructure itself, so that anyone who uses these tools can benefit. This project will explore the feasibility of AI-based approaches for providing such assistance, and prototype, pilot and validate a range of capabilities, and widely disseminate results.
Keywords	Cyberinfrastructure, HPC, AI4CI, Machine-Learning
Poster session	4
Poster #	14

First name	Dave
Middle Initial	
Last name	Randall
Organization	Colorado State University
NSF Award Title	Collaborative Research: Frameworks: Community-Based Weather and Climate Simulation With a Global Storm-Resolving Model
NSF Award Number	2005137
Abstract	The open-source Community Earth System Model (CESM) is both developed and applied to scientific problems by a large community of researchers. It is is managed by the National Center for Atmospheric Research. The CESM includes sub-models of the atmosphere, ocean, land surface, and sea ice. EarthWorks is a five-year project to develop a global coupled model of the atmosphere, ocean, and land surface, based on the CESM. The model uses a ~4 km global grid for all three components. The atmosphere and ocean sub-models use closely related dynamical cores for the atmosphere and ocean that have been developed at NCAR and at the Los Alamos National Laboratory, respectively, and are well-suited for applications that require very high spatial resolution. All components of our model will use the same very high-resolution grid. The use of kilometer-scale resolution makes it possible to eliminate the particularly troublesome parameterization of deep cumulus convection (i.e., thunderstorms).
Keywords	Atmosphere, Ocean, Weather, Climate, Fluids
Poster session	2
Poster #	38

First name	Chris
Middle Initial	
Last name	Rapier
Organization	Pittsburgh Supercomputing Center
NSF Award Title	Elements: HPN-SSH
NSF Award Number	2004012
Abstract	HPN-SSH is a series of modifications to OpenSSH, the predominant implementation of the ssh protocol. It was originally developed to address performance issues when using ssh on high speed long distance networks. By taking advantage of automatically optimized receive buffers HPN-SSH could improve performance dramatically on these paths. Later advances include; disabling encryption after authentication to transport non-sensitive bulk data, parallelization of the AES-CTR cipher, more detailed connection logging. Current work has led to the development of the automatic resumption of failed transfers, the use of on-die AES instructions to accelerate the performance of the parallelized AES-CTR cipher, in line network telemetry based on the TCP_INFO struct, improvements to message authentication performance, performance improvements related to buffer sizing, research into a parallelized ChaCha20 cipher, and the development of binary packages for Debian and Fedora.
Keywords	ssh, networking, security, data transfer, performance
Poster session	1
Poster #	35

First name	Dmitrij
Middle Initial	
Last name	Rappoport
Organization	University of California Irvine
NSF Award Title	libkrylov, a Modular Open-Source Software Library for Extremely Large Eigenvalue and Linear Problems
NSF Award Number	OAC-1835909
Abstract	Strongly coupled linear equation systems or eigenvalue problems with extremely large numbers of unknowns are a critical bottleneck for computational solutions to many grand challenges in science and engineering. This project aims to develop, implement, test, and deploy libkrylov, a robust, efficient, and general open-source library of "on-the-fly" Krylov space methods suitable for solving such extremely large, dense problems. libkrylov will deliver the latest innovations in Krylov-space methods to the scientific and engineering communities by providing a uniform, reproducible, and user-friendly software standard. This project will promote computational literacy through student training and workforce education at University of California, Irvine and San Diego State University, and enhance national software infrastructure through collaboration with the NSF-funded Molecular Sciences Software Institute (MolSSI) in Blacksburg, VA.
Keywords	Numerical Linear Algebra, Iterative Methods, High-Performance Computing, Electronic Structure
Poster session	4
Poster #	15

First name	Amy
Middle Initial	
Last name	Roberts
Organization	CU Denver
NSF Award Title	Collaborative Research: Elements: Shared Data-Delivery Infrastructure to Enable Discovery with Next Generation Dark Matter and Computational Astrophysics Experiments
NSF Award Number	2104003
Abstract	Direct-detection dark matter experiments and astrophysical simulations offer an incredible wealth of information about our world. This project works to make that information accessible to everyone. Much of this data is inaccessible: scientists within an experiment or community access data with custom-built programs that are often difficult to use and maintain. Switching to a standard format is usually not an option, especially for existing data. The Personal Data-Delivery infrastructure (PONDD) makes data accessible both within and across collaborations. Existing efforts to improve data access makes this project possible: yt is software that provides uniform access to simulation data; Kaitai is a data-description language that enables easy access to any data format; Rucio and other tools provide a standard interface that allows data downloads; and ServiceX can identify, subset and process data with little effort from the end user.
Keywords	Data description, data format, data accessibility, cross-experiment data analysis, dark matter, simulation comparison
Poster session	3
Poster #	33

First name	Wissam
Middle Initial	A
Last name	Saidi
Organization	University of Pittsburgh
NSF Award Title	Elements: DeepPDB: An open-source automated framework to enable high-fidelity atomistic simulations in unexplored material space
NSF Award Number	2003808
Abstract	The primary goal of the proposed Deep Potential DataBase (DeepPDB) is to offer an open-source toolkit for deep neural network potentials (DNP), which will allow for the simulations of materials with the accuracy of first-principles methods but with a fraction of their cost. DeepPDB will: (1) develop automated workflows that given a target composition, will run the necessary first-princples calculations, train DNPs, validate against metrics imposed by the training data, identify the input-data space with the largest uncertainty and iterate until an optimal DNP is trained; (2) openly disseminate the training data along with the pre-trained DNPs; (3) develop transparent automated validation that encompasses both traditional DNP based methods as well as fully integrated tests that include target metrics. The resulting open-source toolkit will serve as a foundational tool to investigate the properties of hitherto-unseen materials at length- and time-scales previously not possible.
Keywords	Materails Modeling, first-principles, machine learning, atomistic potentials, large data
Poster session	1
Poster #	36

First name	David
Middle Initial	T
Last name	Sandwell
Organization	UCSD
NSF Award Title	Elements: Software - Harnessing the InSAR Data Revolution: GMTSAR
NSF Award Number	OAC-1834807
Abstract	GMTSAR is an open source InSAR processing system for generating wide-area mapping of the deformation of the surface of the Earth using repeated synthetic aperture radar (SAR) images collected by spacecraft (https://topex.ucsd.edu/gmtsar/). The major deformation signals of interest are associated with earthquakes, volcanoes, glacier flow, and subsidence due to withdrawal of crustal fluids (e.g., water and hydrocarbons).
Keywords	Earth surface deformation, radar interferometry, processing software
Poster session	online only

First name	Linda
Middle Initial	
Last name	Schadler
Organization	University of Vermont
NSF Award Title	Collaborative Research: Framework: Data: HDR: Nanocomposites to Metamaterials: A Knowledge Graph Framework
NSF Award Number	OAC-1835677
Abstract	This project led by 5 academic institutions is developing an open-source, extensible, semantic cyberinfrastructure (CI) following FAIR principles for Materials research (MaterialsMine), with specific application to polymer nanocomposites (NanoMine) and metamaterials (MetaMine). It is comprised of a database to store semi-automated curated data with associated validation protocols, a Knowledge Graph (KG) for ontology-enabled exploration and discovery, design tools and custom user dashboards. Recent enhancements in CI include a modularized backend for increased speed and performance, augmented KG flexibility pairing the semantic context of metadata (SPARQL query) with the visual context (Vega-Lite specification), and a new search and visualization interface for users with varying levels of expertise. Concurrent development in image analysis, NLP-enabled curation, and implementation of design tools for discovery of processing-structure-property relationships are also highlighted.
Keywords	SPARQL, Vega-Lite, knowledge graph, materials design, polymer nanocomposites, metamaterials
Poster session	2
Poster #	39

First name	Matthew
Middle Initial	
Last name	Scotch
Organization	Arizona State University
NSF Award Title	Collaborative: Elements: Cyberinfrastructure for Pedestrian Dynamics-Based Analysis of Infection Propagation Through Air Travel
NSF Award Number	1931483
Abstract	Despite commercial airlines mandating masks, there have been multiple documented events of COVID-19 superspreading on flights. Conventional models typically do not consider the movement of passengers during the flight, boarding, or deplaning. Understanding the risks for each of these aspects could provide insight into effective mitigation measures. We modeled infection risk from seating and fine-grained movement patterns – boarding, deplaning, and inflight movement. We validated the model and the impact of interventions using available data from three flights. Our results show that the inclusion of passenger movement better explains the infection spread patterns than conventional models do. We also found that FFP2/N95 mask usage would have reduced infection by 95-100%, while cloth masks would have reduced it by only 40-80%. Leaving middle seat vacant is effective in reducing infection, and the effectiveness increases when combined with good quality masks.
Keywords	COVID-19, air travel, Infection modeling, pedestrian dynamics
Poster session	3
Poster #	22

First name	Robert
Middle Initial	E
Last name	Settlage
Organization	Virginia Tech
NSF Award Title	Frameworks: Software NSCI-Open OnDemand 2.0: Advancing Accessibility and Scalability for Computational Science through Leveraged Software Cyberinfrastructure
NSF Award Number	1835725
Abstract	Reducing barriers that limit the adoption of high performance computing (HPC) addresses an important problem that broadly affects the science, engineering, and humanities communities. This effort builds on existing capabilities with large and varied user communities, and on national scale cyberinfrastructure and high-performance computing resources. The project develops Open OnDemand 2.0, an open-source software that enables access to high-performance computing, cloud, and remote computing resources via the web, and lower the barriers to access HPC systems. The project combines two widely used HPC resources Open OnDemand and XDMoD. The integrated platform will enhance resource utilization visibility, extend to more resource types and institutions, and support a smooth and easy utilization of HPC resources with intuitive web interfaces.
Keywords	HPC, Open OnDemand, XDMoD, High-performance computing
Poster session	4
Poster #	17

First name	Krister
Middle Initial	
Last name	Shalm
Organization	University of Colorado at Boulder
NSF Award Title	RAISE TAQS: Randomness expansion using a loophole-free Bell test
NSF Award Number	1839223
Abstract	Our team is building the a publicly accessible randomness service that can provide both certifiable and auditable randomness. Our random number generation is based on using measurements made on entangled particles under a set of strict conditions, and the security is guaranteed by the fact that it is impossible to send information faster than the speed of light. Current experimental setups to generate this randomness are large, spanning hundreds of meters. Our team has made progress shrinking those distances to 5 meters with innovative new quantum measurement techniques.
Keywords	Randomness generation, entanglement, quantum, auditable, certifiable
Poster session	3
Poster #	8

First name	Naveen
Middle Initial	
Last name	Sharma
Organization	Rochester Institute of Technology
NSF Award Title	Citizenly: Empowering Communities by Democratizing Data Science
NSF Award Number	1943002
Abstract	This project seeks to research and prototype Citizenly, a cyber-infrastructure, to achieve the visions for democratizing data science and community engagements in civic progress. Citizenly provides knowledge and tools to foster citizen engagement through participatory data collection, analysis and actions thus enabling “smart communities” to be built together with empowered (“smart”) citizens.
Keywords	urban, community engagement, data, learning, smart-city
Poster session	4
Poster #	19

First name	Robert
Middle Initial	S
Last name	Sinkovits
Organization	University of California, San Diego
NSF Award Title	Elements: Spatial Ecology Gateway
NSF Award Number	2104104
Abstract	Spatial ecology is the study of how landscape characteristics influence the distribution and movement of organisms within their environment. The Spatial Ecology Gateway (SEG) enables researchers, students and wildlife managers to upload biotelemetry data, typically GPS readings, and construct home ranges that allow them to interpret animal space use. Applications of the SEG can include classroom projects, basic research into problems in wildlife ecology, environmental impact studies and mitigation of adverse outcomes such as habitat fragmentation or increased human-wildlife interaction resulting from new development. The SEG insulates users from the underlying computational details so that they can focus on their science rather than mastering the technology. Users of the SEG have the option to generate two-dimensional (2D) or, where applicable, three-dimensional (3D) home ranges. Users are also provided with less computationally intensive tools to perform exploratory analyses.
Keywords	Spatial ecology, science gateway, home range, HubZero, Brownian Bridge Movement Model
Poster session	3
Poster #	34

First name	Michael
Middle Initial	D
Last name	Sokoloff
Organization	University of Cincinnati
NSF Award Title	Collaborative Research : Elements : Extending the physics reach of LHCb by developing and deploying algorithms for a fully GPU-based first trigger stage
NSF Award Number	2004364
Abstract	The development of the Standard Model of particle physics is a major intellectual achievement. However, the Standard Model leaves open many questions, The LHCb experiment at CERN has been designed to explore Beyond the Standard Model physics. The LHCb sensor arrays produce about 100 TB/s and close to a zettabyte per year. Even after drastic data-reduction performed by custom-built read-out electronics, the data volume is still about 25 exabytes per year. To reduce the persisted data set to manageable size, we use a two-stage, real-time software "trigger" to select a small subset of the data to write to disk and tape. The first stage, called Allen, will process 25 exabytes each year using high-level computing algorithms executing in GPUs. This software will make it possible to explore potential explanations for dark matter, and the matter/anti-matter asymmetry of our universe using data that would be otherwise inaccessible due to trigger-system limitations.
Keywords	GPUs, real-time, particle physics, machine learning
Poster session	1
Poster #	37

First name	Edgar
Middle Initial	
Last name	Solomonik
Organization	University of Illinois at Urbana-Champaign
NSF Award Title	Collaborative Research: Frameworks: Scalable Modular Software and Methods for High-Accuracy Materials and Condensed Phase Chemistry Simulation
NSF Award Number	1931258
Abstract	This project aims to build fast and accurate computer programs to simulate new materials as part of models developed for the Materials Genome Initiative. This research includes novel parallel implementations of low-scaling, high-accuracy methods for electronic structure. In particular, new techniques for mean-field calculations have been developed, which act as groundwork for periodic coupled-cluster and quantum Monte Carlo. State-of-the-art techniques in sparsity and tensor decomposition have been developed to achieve good system-size scaling while retaining accuracy within each of these numerical schemes. The methods use efficient high-level software abstractions, implemented as Python-level modules within PySCF that leverage the Cyclops library for massively-parallel execution. New library software infrastructure includes automatic differentiation for tensor optimization, as well as to enable execution of sparse kernels on emerging GPU-based supercomputing architectures.
Keywords	quantum chemistry, high-accuracy simulation, low-scaling methods, tensors, scalable software systems
Poster session	4
Poster #	20

First name	Carol
Middle Initial	X
Last name	Song
Organization	Purdue University
NSF Award Title	Framework: Data: HDR: Extensible Geospatial Data Framework towards FAIR (Findable, Accessible, Interoperable, Reusable) Science
NSF Award Number	1835822
Abstract	Scientists in geospatial data-driven fields often spend significant efforts in “wrangling data”, i.e., accessing and processing data to make them usable in modeling and analysis tools. This project has created GeoEDF, an extensible geospatial data framework, to reduce this barrier by creating seamless connections among platforms, data and tools, making large distributed geospatial datasets directly usable in models and tools. Through an extensible set of community contributed, modular and reusable data connectors and processors, GeoEDF abstracts away the complexity of acquiring and utilizing remote datasets. Researchers can string them together into a workflow for execution in various environments including a well-established science gateway MyGeoHub, JupyterHub-based deployments, and as a Docker container on laptops. By bringing data to the science, GeoEDF helps accelerate data-driven discovery and improve FAIR science practices.
Keywords	geospatial data, data framework, data connector, data processor, workflow
Poster session	4
Poster #	21

First name	Raymond
Middle Initial	
Last name	Speth
Organization	Massachusetts Institute of Technology
NSF Award Title	CSSI Framework: Extensible and Community-Driven Thermodynamics,Transport, and Chemical Kinetics Modeling with Cantera: Expanding to Diverse Scientific Domains
NSF Award Number	1931391
Abstract	Cantera is an open-source software package for problems involving chemical kinetics, thermodynamics, and species transport, with broad applications in chemistry, materials science, and engineering. Written in C++, Cantera provides a modern, object-oriented, extensible programming tool for multi-species and/or multi-phase phenomena. It provides a library of generalizable classes and modules that automate calculations to support flexible modeling of complex thermo-kinetic and mass transport phenomena. This project aims to extend the utility of Cantera beyond the combustion community, where it is widely used, to fields such as electrochemistry and catalysis, where its ability to efficiently handle complex chemistry can accelerate scientific discovery and technology development.
Keywords	chemical kinetics, combustion, electrochemistry, catalysis, mass transport
Poster session	4
Poster #	22

First name	Dan
Middle Initial	C
Last name	Stanzione
Organization	University of Texas at Austin
NSF Award Title	Elements: Cognitasium - Enabling Data-Driven Discoveries in Natural Hazards Engineering
NSF Award Number	2103937
Abstract	Numerical modeling plays a critical role in assessing and mitigating risks posed by natural hazards. Accurately predicting these hazards requires modeling the multi-scale nature of these problems, covering a range of physical scales from microscopic to kilometer-scale interactions. With the advent of community data repositories such as the DesignSafe CyberInfrastructure, there is an as-yet untapped opportunity to effectively use these large datasets to develop new data-driven models to solve multi-scale problems. Cognitasium, an Artificial Intelligence (AI)-powered cyberinfrastructure, addresses these challenges by automatically extracting the hazard analysis workflows, augmenting large community datasets with relevant information for analysis, and enabling AI models to discover new theories from massive datasets. Cognitasium is being developed as an open-source framework and can be easily adapted to a variety of communities.
Keywords	Natural Hazards, AI, Open Source
Poster session	3
Poster #	35

First name	Petr
Middle Initial	
Last name	Sulc
Organization	Arizona State University
NSF Award Title	Elements: Models and tools for on-line design and simulations for DNA and RNA nanotechnology
NSF Award Number	1931487
Abstract	We develop infrastructure for online design, simulation and analysis of DNA and RNA nanostructures, as well as repositories for sharing designs between researchers. The goal of the developed set of tools and webservers is to greatly simplify in-silico design of nanostructures and make it accessible to experimentalists as well as researchers from other fields.
Keywords	Coarse-grained modeling, DNA nanotechnology, RNA nanotechnology, molecular dynamics, online database
Poster session	4
Poster #	23

First name	Xian-He
Middle Initial	
Last name	Sun
Organization	Illinois Institute of Technology
NSF Award Title	Framework: Software: NSCI: Collaborative Research: Hermes: Extending the HDF Library to Support Intelligent I/O Buffering for Deep Memory and Storage Hierarchy Systems
NSF Award Number	OCI-1835764
Abstract	As we move towards the exascale era, I/O bottleneck is a must to solve performance bottleneck facing the HPC community. Deep Memory-Storage Hierarchies (DMSHs) with multiple levels of memory/storage layers offer a feasible solution but are very complex to use effectively. There is a need to enhance and extend current software systems to support data access and movement transparently and effectively under DMSHs. This project proposes the development of Hermes, a heterogeneous aware, multi-tiered, dynamic, and distributed I/O buffering system that will significantly accelerate I/O performance. The deliveries of this project include an enhanced HDF5 library, a set of extended HDF technologies, and a group of general I/O buffering and memory system optimization mechanisms and methods, in addition to Hermes. The Beta version of Hermes has been released under the Hierarchical Data Format (HDF) technologies. It is readily available and is an effective tool to improve HPC I/O performance.
Keywords	HPC I/O system, memory-storage hierarchy, data buffering system, HDF, data management
Poster session	4
Poster #	24

First name	Dingwen
Middle Initial	
Last name	Tao
Organization	Washington State University
NSF Award Title	CDS&E: Collaborative Research: HyLoC: Objective-driven Adaptive Hybrid Lossy Compression Framework for Extreme-Scale Scientific Applications
NSF Award Number	2003624
Abstract	Today's extreme-scale scientific simulations and instruments are producing huge amounts of data that cannot be transmitted or stored effectively. Lossy compression, a data compression approach leading to certain data distortion, has been considered as a promising solution, because it can significantly reduce the data size while maintaining high data fidelity. However, the existing lossy compression methods may not always work effectively on all datasets used in specific applications because of their distinct and diverse characteristics. Moreover, the user objectives in compression quality and performance may vary with applications, datasets or circumstances. This project aims to develop a hybrid lossy compression framework to automatically construct the best-fit compression for diverse user objectives in data-intensive scientific research.
Keywords	HPC, I/O and communication, data movement, lossy compression, scientific data
Poster session	1
Poster #	38

First name	Dingwen
Middle Initial	
Last name	Tao
Organization	Washington State University
NSF Award Title	Collaborative Research: Elements: ROCCI: Integrated Cyberinfrastructure for In Situ Lossy Compression Optimization Based on Post Hoc Analysis Requirements
NSF Award Number	2104024
Abstract	Today's simulations and advanced instruments are producing vast volumes of data, presenting a major storage and I/O burden for scientists. Error-bounded lossy compressors, which can significantly reduce the data volume while controlling data distortion with a constant error bound, have been developed for years. However, a significant gap still remains in practice. On the one hand, the impact of the compression errors on scientific research is not well understood, so how to set an appropriate error bound for lossy compression is very challenging. On the other hand, how to select the best fit compression technology and run it automatically in scientific application codes is non-trivial because of strengths and weaknesses of different compression techniques and diverse application and dataset characteristics. This project aims to develop a cyberinfrastructure for data-intensive domains, which can select and run the best fit lossy compressor automatically at runtime.
Keywords	HPC, storage and I/O, compression, scientific data, post hoc analysis
Poster session	2
Poster #	40

First name	Carl
Middle Initial	
Last name	Tape
Organization	University of Alaska Fairbanks
NSF Award Title	Collaborative Research: Frameworks: Seismic COmputational Platform for Empowering Discovery (SCOPED)
NSF Award Number	2104052
Abstract	The SCOPED project establishes cyberinfrastructure that provides fast access to large seismic archives from a suite of containerized open-source computational tools for big data analysis, machine learning, and high-performance simulations. The project offers compute performance for both model- and data-driven seismological problems. Hundreds of terabytes of waveform data are directly accessible both to modelers—for data assimilation problems—and to data scientists for processing, analysis, and exploration. It seeks to establish a gateway with ready-to-run (or adapt) container images and data as a service for the seismological community. It develops computational tools that promote the democratization of HPC/Cloud with cutting-edge data processing and modeling software through their scalability from laptops to HPC or Cloud systems and through their portability with containerization.
Keywords	seismology, software containers, cloud computing, data assimilation, high-performance computing
Poster session	3
Poster #	36

First name	David
Middle Initial	G
Last name	Tarboton
Organization	Utah State University
NSF Award Title	CSSI Framework: HydroShare: Cyberinfrastructure for Advancing Hydrologic Knowledge through Collaborative Integration of Data Science, Modeling and Analysis
NSF Award Number	1664061
Abstract	HydroShare is a data and model repository operated by the Consortium of Universities for the Advancement of Hydrologic Science Inc. (CUAHSI) to advance hydrologic science by enabling individual researchers to more easily and freely share data and models from their research. HydroShare supports Findable, Accessible, Interoperable and Reusable (FAIR) principles. It is comprised of two sets of functionalities: (1) a repository for users to share and publish data and models in a variety of formats, and (2) tools (web apps) that can act on content in HydroShare and support web-based access to compute capability. Together these move us towards a platform for collaboration and computation that integrates data storage, organization, discovery, and analysis through web applications (web apps) and that allows researchers to employ services beyond the desktop to make data storage and manipulation more reliable and scalable, while improving their ability to collaborate and reproduce results.
Keywords	Water data management, CUAHSI, Collaborative Environment
Poster session	4
Poster #	25

First name	Michela
Middle Initial	
Last name	Taufer
Organization	The University of Tennessee
NSF Award Title	Collaborative Research: Elements: SENSORY: Software Ecosystem for kNowledge diScOveRY - a data-driven framework for soil moisture applications
NSF Award Number	2103845
Abstract	To trust findings in computational science, scientists need workflows that trace the data provenance and support results explainability. In this talk, we address the scientists' need by proposing a computational environment that automatically creates a workflow execution's record trail and invisibly attaches it to the workflow's output, enabling data traceability and results explainability. We demonstrate the capabilities of our environment with the study soil moisture predictions from 27 km resolution satellite data down to fine-grain 10 m resolution necessary for practical use in policymaking and precision agriculture. We identify the causes of different accuracy for predicted soil moisture values in different resolutions of the input data, and link different results to different machine learning methods used during the soil moisture downscaling, all without requiring the scientist to know aspects of workflow design and implementation.
Keywords	scientific workflows, provenance, scientific computing, traceability, explainability, soil moisture predictions
Poster session	2
Poster #	41

First name	George
Middle Initial	K
Last name	Thiruvathukal
Organization	Loyola University Chicago
NSF Award Title	CDSE: Collaborative: Cyber Infrastructure to Enable Computer Vision Applications at the Edge Using Automated Contextual Analysis
NSF Award Number	2104319
Abstract	Digital cameras are deployed as network edge devices, gathering visual data for such tasks as autonomous driving, traffic analysis, and wildlife observation. Analyzing the vast amount of visual data is a challenge. Existing computer vision methods require fast computers that are beyond the computational capabilities of many edge devices. This project aims to improve the efficiency of computer vision methods so that they can run on battery-powered edge devices.
Keywords	computer vision, edge devices, low-power, digital cameras
Poster session	4
Poster #	26

First name	Yuan
Middle Initial	
Last name	Tian
Organization	University of California, Los Angeles
NSF Award Title	CICI: RDP: Enforcing Security and Privacy Policies to Protect Research Data
NSF Award Number	1920462
Abstract	Advances in computer systems over the past decade have laid a solid foundation for data collection at a staggering scale. Data generated from end-user devices has tremendous value to the research community. While ground-breaking discovered may occur, malicious attacks or unintentional data leaks threaten the research data. Such a threat is hard to predict and difficult to recover from once it happens. Preventative and defensive measures should be taken where data is generated in order to protect private, valuable data from the attackers. In this project, the investigators propose to interpret the documented policies and enforce them in research projects, in order to protect the privacy of research data. This work can significantly reduce researchers' overhead in implementing policy-compliant code and reduce the complexity of protecting research datasets.
Keywords	Data privacy, policy enforcement, program analysis, IRB
Poster session	4
Poster #	27

First name	Frank
Middle Initial	
Last name	Timmes
Organization	Arizona State University
NSF Award Title	Collaborative Research: SI2-SSI: Modules for Experiments in Stellar Astrophysics
NSF Award Number	ACI-1663684
Abstract	The overall goals of this project is to support the Modules for Experiments in Stellar Astrophysics (MESA) software project, the evolving and flexible needs of its expanding user community, and engagement with the wider stellar astrophysics software ecosystem.
Keywords	Astronomy, Astrophysics, MESA Software Instrument
Poster session	4
Poster #	28

First name	Juan
Middle Initial	P
Last name	Tolento
Organization	UC Irvine
NSF Award Title	Advanced Lossless and Lossy Compression Algorithms for netCDF Datasets in Earth and Engineering Sciences
NSF Award Number	OAC-2004993
Abstract	For decades, the dominant geoscience and engineering array-oriented data storage format, netCDF/HDF, supported only one open source lossless compression algorithm, DEFLATE. We introduce the Community Codec Repository (CCR), a shared library framework that enables netCDF/HDF datasets and applications to transparently employ modern open source compression methods. These methods accelerate compression/decompression and improve compression ratios, which thereby reduce storage requirements, power consumption, and greenhouse emissions. Four CCR codecs have been integrated directly into the netCDF library for widespread use. These include two project-developed lossy codecs that quantize data 20-40% more efficiently than before. Geoscience and engineering datasets (and archives like CMIP7), can now store data in half the amount of space required with DEFLATE while still preserving all scientifically meaningful information.
Keywords	Compression Algorithms, netCDF, CCR
Poster session	4
Poster #	29

First name	Bill
Middle Initial	
Last name	Tolone
Organization	University of North Carolina at Charlotte
NSF Award Title	Virtual Information-Fabric Infrastructure (VIFI) forData-Driven Decisions from Distributed Data
NSF Award Number	1640818
Abstract	VIFI presents a transformative infrastructure that empowers data users to discover, analyze, transform, and evaluate distributed, fragmented data without direct access to or movement of large amounts of data, enabling analyses that are otherwise impossible, infeasible, or impractical. VIFI supports the full lifecycle of inquiry, analysis, and data-driven discovery. VIFI enables data owners to share data indirectly and facilitates automated methods for data-driven insights to broader communities. The VIFI architecture enables novel support through distributed orchestration that migrates analyses to the site of the data rather than migrating data to a site for analyses. VIFI removes barriers to data through a generalized, discipline-agnostic approach. VIFI demonstrates its capabilities within multiple science and engineering domains, including Earth Science, Astronomy, and Sustainable Human-Building Ecosystems.
Keywords	Virtual information fabrics, distributed orchestration, data-driven discoveries, big data infrastructures
Poster session	4
Poster #	30

First name	Stanimire
Middle Initial	
Last name	Tomov
Organization	University of Tennessee
NSF Award Title	SI2:SSE: MAtrix, TEnsor, and Deep-Learning Optimized Routines (MATEDOR)
NSF Award Number	1740250
Abstract	The MAtrix, TEnsor, and Deep-learning Optimized Routines (MATEDOR) project defines a standard interface for batched operations and provides a performance-portable software library (MAGMA) that demonstrates batching routines for a significant number of kernels. This is critical, given that the performance opportunities inherent in solving many small batched matrices often yield more than a 10x speedup over the current classical approaches. Working closely with affected application communities, MATEDOR developed a Batched BLAS and LAPACK standard with modular, optimizable, and language-agnostic interfaces. This modularity provides application, compiler, and runtime system developers with the option to use a single call to a routine from the new batch operation standard and allows the entire linear algebra community to collectively attack a wide range of small matrix or tensor problems in deep learning, data mining, astrophysics, image and signal processing, hydrodynamics, and more.
Keywords	Batched BLAS, LAPACK, MAGMA, tensor computations, deep learning
Poster session	4
Poster #	31

First name	Greg
Middle Initial	
Last name	Tucker
Organization	University of Colorado Boulder
NSF Award Title	Collaborative Research: Frameworks: OpenEarthscape—Transformative Cyberinfrastructure for Modeling and Simulation in the Earth-Surface Science Communities
NSF Award Number	2104102
Abstract	As the sciences that probe Earth's changing surface become more quantitative and prediction-oriented, they increasingly rely on computational modeling and model-data integration. This project develops OpenEarthscape: an integrated suite of community-developed cyber resources for simulation and model-data integration, focusing on nine high-priority geoscience frontiers. Products and activities include a JupyterHub server providing easy access to models, tools, and libraries; new capacity for creating and sharing reproducible analyses; and major enhancements to current programming libraries for model construction and coupling. OpenEarthscape catalyzes efficiency with improved performance and an extended version of the Basic Model Interface API standard to address parallel architecture and coupling. OpenEarthscape fosters research productivity with improved library capabilities for data I/O and visualization.
Keywords	geoscience, modeling, Python, earth, water
Poster session	3
Poster #	37

First name	Robert
Middle Initial	A
Last name	van de Geijn
Organization	The University of Texas at Austin
NSF Award Title	Beyond the BLAS: A framework for accelerating computational and data science
NSF Award Number	CSSI-2003921
Abstract	The BLAS-like Library Instantiation Software (BLIS) is a framework for the rapid instantiation of basic linear algebra functionality targeting modern CPUs. The current project refactors parts of the framework so as to support new precisions, the mixing of precisions and/or domains, and the synthesis of new BLAS-like operations by users including the tensor contractions of importance to (for example) computational chemistry and BLAS-like operations encountered in machine learning. This open source software has been embraced by major hardware vendors and various Linux distributions. It is supported by an active developer community and industry. This project and its predecessors incorporate massive outreach through Massive Open Online Courses (MOOCs) with ~250,000 registrations to date.
Keywords	BLIS, BLAS-like functionality, linear algebra, tensor contraction, mixed data types
Poster session	4
Poster #	32

First name	Dave
Middle Initial	
Last name	Vieglais
Organization	University of Kansas
NSF Award Title	Collaborative Research: Frameworks: Internet of Samples: Toward an Interdisciplinary Cyberinfrastructure for Material Samples
NSF Award Number	2004815
Abstract	The Internet of Samples (iSamples) project aims to integrate large, diverse, cross-discipline sample data repositories and improve the find-ability, accessibility, interoperability and reuse of material samples and sample data. iSamples is achieving this through development and deployment of light-weight infrastructure components that can be deployed close to collections and as aggregators across collections.
Keywords	physical samples, collection, identifier, distributed, linked data
Poster session	2
Poster #	28

First name	Natalia
Middle Initial	
Last name	Villanueva Rosales
Organization	The University of Texas at El Paso
NSF Award Title	ELEMENTS: DATA: HDR: SWIM to a Sustainable Water Future
NSF Award Number	1835897
Abstract	The goal of the SWIM project is to advance water sustainability research capabilities through the integration, execution and interpretation of water models and participatory reasoning processes. Using the meaning of data and models, SWIM currently facilitates interpretations of water availability in the in the Middle Rio Grande valley of US Southwest and North of Mexico for stakeholder-driven analysis from the socio-environmental perspective.
Keywords	cyberinfrastructure, water sustainability, model-to-model integration, participatory reasoning, stakeholder-driven analysis
Poster session	online only

First name	Jorge
Middle Initial	
Last name	Vinals
Organization	University of Minnesota
NSF Award Title	RAISE: A Materials Science Gateway for X-ray Imaging and Modeling of Microstructures
NSF Award Number	OAC 2037773
Abstract	This program develops an open cyberinfrastructure in the form of a public Science Gateway to serve the Cornell High Energy Synchrotron Source. The gateway is based on the Galaxy framework. Raw data is accumulated locally at the beam lines, and ingested into a Galaxy instance using standard web application programming interfaces. Data is typed, metadata is attached, and is made available through a shared data library. The gateway then provides the infrastructure for user defined transformations, including data reduction, image reconstruction, and feature analysis, while retaining metadata and provenance information. Workflows including visualization and modeling are also supported. The gateway is extensible by users of the facility and the broader community as new analysis and modeling tools are developed.
Keywords	Galaxy, High Energy Diffraction Microscopy, Grain Growth, Microstructure
Poster session	4
Poster #	34

First name	Guang
Middle Initial	
Last name	Wang
Organization	Florida State University
NSF Award Title	Private Data Analytics Synthesis, and Sharing for Large-Scale Multi-Modal Smart City Mobility Research
NSF Award Number	2003874
Abstract	Given the trend towards urbanization, understanding real-time human mobility in urban areas has become increasingly important for many research areas from Mobile Networking, to Urban Planning, Emergency Response to recent Pandemic Mitigation. Many analytical models have been proposed to understand human mobility based on mobility data. However, most of these data are proprietary and cannot be accessed by the research community at large. Fortunately, based on the latest expansion of urban infrastructures, such mobility data has been collected by city government agencies and some companies that are willing to share the data for social good. However, a key challenge is the privacy concern and security issues. To address this issue, the project aims to generate realistic yet synthetic mobility data through machine learning based on the real mobility data analytics and then share these realistic synthetic data with the research community.
Keywords	Data integration, synthetic data, privacy, data sharing, smart city
Poster session	1
Poster #	39

First name	Yinzhi
Middle Initial	
Last name	Wang
Organization	Texas Advanced Computing Center
NSF Award Title	Elements: PASSPP: Provenance-Aware Scalable Seismic Data Processing with Portability
NSF Award Number	1931352
Abstract	Massive Parallel Analysis System for Seismologists (MsPASS)is a framework for seismic data processing. The framework was designed to enable new scientific frontiers in seismology by providing a means to more effectively utilize supercomputers or cloud resources to handle the increasingly large data volume available today. MsPASS leverages several existing technologies: (1) scalable parallel processing frameworks, (2) a NoSQL database management system, and (3) containers. The system leans heavily on the widely used ObsPy toolkit. It automates many database operations and provides a mechanism to automatically save the processing history for reproducibility. The synthesis of these components can provide flexibility to adapt to a wide range of data processing workflows. We demonstrate the system with a basic data processing workflow applied to USArray data. Through extensive documentation and examples, we aim to make this system a sustainable, open-source framework for the community.
Keywords	parallel computing, database, seismology, container
Poster session	4
Poster #	35

First name	Yang
Middle Initial	
Last name	Wang
Organization	Carnegie Mellon University
NSF Award Title	Collaborative Research: Element: Development of MuST, A Multiple Scattering Theory based Computational Software for First Principles Approach to Disordered Materials
NSF Award Number	1931525
Abstract	The effect of disorder in materials is of great fundamental and technological interest. Disorder disrupts the periodic arrangement of atoms in perfect materials.Being able to understand, control, and predict the disorder effects in real physical systems is essential for the development of new structural and functional materials for future technological applications. This project involves building computer software, called MuST, that is aimed to enable the study of disorder effects using the principles of quantum mechanics and to accelerate the discovery of materials essential for industry and information technology applications. MuST will create a scalable approach for first principles studies of quantum materials that efficiently utilizes petascale and future high-performance computing resources. It will expand the user community by enabling researchers within academia and industry to perform calculations that are presently out of reach for most users.
Keywords	Disorder, density functional theory, multiple scattering theory, first principles, alloys
Poster session	4
Poster #	36

First name	Michael
Middle Initial	
Last name	Widom
Organization	Carnegie Mellon University
NSF Award Title	Elements: Cyberinfrastructure for spin and charge transport calculation of partially disordered alloys
NSF Award Number	2103958
Abstract	Commonly used metals are usually alloys - mixtures of multiple chemical elements - with the specific choice of composition optimized for their intended use. The precise arrangement of different species cannot be precisely controlled. Rather, the crystal lattice positions are randomly occupied by statistical distributions of atomic species. The randomness is tempered by preferential chemical interactions that create correlations among the occupation of nearby sites. While the overall composition greatly impacts alloy properties, the disorder has additional impact, either beneficial or detrimental, and the degree of order or disorder should be known and controlled in optimally designed new materials. This proposal develops the cyberinfrastructure needed to predict the degree of order present in alloys, and to calculate its impact on electrical and spin conductivity.
Keywords	Density functional theory, KKR-CPA, conductivity
Poster session	3
Poster #	39

First name	Peter
Middle Initial	K G
Last name	Williams
Organization	American Astronomical Society / Center for Astrophysics
NSF Award Title	Elements: FastTract: Web-Based Exploratory Visualization of Gigapixel Astronomical Images
NSF Award Number	2004840
Abstract	Like scientists in virtually every field, astronomers are struggling under a "data deluge". While the ever-larger images obtained by the newest observatories lead to cutting-edge science, they also overwhelm traditional tools designed to work with files hundreds or thousands of times smaller than the state-of-the-art. In particular, astronomers are rapidly losing the ability to simply look at the images of the sky that are coming out of their telescopes. The FastTract Project will solve this problem by marrying existing Web-based visualization technologies in the AAS WorldWide Telescope software system with new features and tools needed to efficiently work with the large astronomical images of the 2020's, enabling US astronomers to fully exploit their world-class facilities, in particular the ones that open new Windows on the Universe.
Keywords	visualization, astronomy, web applications, big data
Poster session	4
Poster #	37

First name	Yinghui
Middle Initial	
Last name	Wu
Organization	Case Western Reserve University
NSF Award Title	Elements: Crowdsourced Materials Data Engine for Unpublished XRD Results
NSF Award Number	2104007
Abstract	This project CRUX, a crowdsourced data infrastructure, to curate, discover, share, and recommend unpublished XRD data and analytical results. CRUX promotes underutilized high-quality material science data by allowing the sharing and exploration of unpublished datasets with state-of-the-art knowledge harvesting, and machine learning (ML) techniques. CRUX enables a materials knowledge graph (KG) model, automatic data integration, and an exploratory query engine that support "Why" and "What-if" analysis for XRD analysis. CRUX will enable an open, collaborative, and sustainable platform that can facilitate exchanging of unpublished XRD data and unlock new research problems (e.g., prediction of materials compositions with multi-phase data), and inspire the novel design of ML pipelines for data-driven materials science. CRUX will make materials data resources available for a broad community including materials scientists, data analysts, developers, and the general public.
Keywords	material data science, XRD, knowledge graph, machine learning, data provenance
Poster session	3
Poster #	40

First name	Yuanzhe
Middle Initial	
Last name	Xi
Organization	Emory University
NSF Award Title	CDS&E: Collaborative Research: Hierarchical Kernel Matrices for Scientific and Data Applications
NSF Award Number	2003720
Abstract	Kernel matrices in machine learning and scientific computing describe the relationships between collections of points which may represent various types of information. The increasing size of data sets in various disciplines and the increasing computational capability of computer hardware make it essential that our algorithms and software for kernel matrices are scalable, and that the time it takes for their execution grows linearly or close to linearly, with the problem size. Otherwise, such large-scale data problems may not be tractable. This project addresses the scaling bottlenecks associated with handling the kernel matrix by exploiting a hierarchical structure that is often found in these matrices. By accelerating computations with kernel matrices, this research enables large-scale data analysis and scientific simulation in diverse areas such as uncertainty quantification, integral equation problems, particle simulations, and geostatistics.
Keywords	Hierarchical matrix, kernel methods, Gaussian process, low rank compression, Brownian dynamics simulations
Poster session	1
Poster #	40

First name	Shouhuai
Middle Initial	
Last name	Xu
Organization	University of Colorado Colorado Springs
NSF Award Title	CICI: UCSS: ACSP4HR: Assuring Cyber Security and Privacy for Human Resilience Research: Requirements, Framework, Architecture, Mechanisms and Prototype
NSF Award Number	2115134
Abstract	This project investigates a technical solution to modernizing and accelerating the scientific discovery process of Human Resilience research, by tackling a range of technical challenges including: (i) specifying the requirements of a competent cyberinfrastructure; (ii) defining a solution framework to adequately address these requirements; (iii) designing a comprehensive system architecture to fulfill the framework; (iv) investigating novel mechanisms and supporting techniques; and (v) developing a prototype system and demonstrating its usefulness. The project bolsters the scientific collaboration in the Human Resilience research community by encouraging the adoption of security and privacy into its unique scientific workflows and by pioneering a holistic security cyberinfrastructure environment spanning the entire Human Resilience research data-sharing ecosystem.
Keywords	Human resilience research (HRR), cybersecurity, privacy-preservation, ransomware-resistance, automatic policy-compliance, cyberinfrastructure
Poster session	4
Poster #	38

First name	Chaowei
Middle Initial	
Last name	Yang
Organization	NSF Spatiotemporal Innovation Center/George Mason
NSF Award Title	Collaborative Research: Elements: Data: HDR: Developing On-Demand Service Module for Mining Geophysical Properties of Sea Ice from High Spatial Resolution Imagery
NSF Award Number	1835507
Abstract	Sea ice acts as both an indicator and an amplifier of climate change. At present, there are multiple sources of sea ice observations which are obtained from a variety of networks of sensors (in situ, airborne, and space-borne). By developing a smart cyberinfrastructure element for the analysis of high spatial resolution (HSR) remote sensing images over sea ice, the science community is better able to extract important geophysical parameters for climate modeling. The project contributes new domain knowledge to the sea ice community. This is accomplished by integrating HSR images that are spatiotemporally discrete to produce a more rapid and reliable identification of ice types, and by a standardized image processing that allows creating compatible sea ice products. The cyberinfrastructure module is a value-added on-demand web service that can be naturally integrated with existing infrastructure.
Keywords	arctic, climate change, sea ice, classification, AI/ML
Poster session	4
Poster #	39

First name	Yanfang (Fanny)
Middle Initial	
Last name	Ye
Organization	University of Notre Dame
NSF Award Title	CICI: SSC: SciTrust: Enhancing Security for Modern Software Programming Cyberinfrastructure
NSF Award Number	OAC-2218762
Abstract	Nowadays, software has played a vital role supporting scientific communities. Unlike conventional approaches, modern software programming cyberinfrastructure (CI), consisting of online discussion platforms (e.g., Stack Overflow) and social coding repositories (e.g., Github), has offered an open-source and collaborative environment for distributed scientific and engineering communities to expedite the process of software development. Despite the apparent benefits, its potential security-related risks have been largely overlooked - insecure or malicious codes could be easily embedded (e.g., through forking or committing) and distributed (e.g., through copy-and- paste), which could severely damage the scientific credibility of CI. To address this imminent issue, this CICI project seeks to explore innovative links between Artificial Intelligence (AI) and cybersecurity to enhance the security of modern software programming CI.
Keywords	Modern Software Programming Cyberinfrastructure, Security, AI, Scientific Credibility
Poster session	4
Poster #	40

First name	Michael
Middle Initial	
Last name	Zentner
Organization	SDSC
NSF Award Title	S2I2: Impl: The Science Gateways Community Institute (SGCI) for the Democratization and Acceleration of Science
NSF Award Number	1547611
Abstract	The Science Gateways Community Institute provides a variety of services to assist in the development, operation, and sustainability of science gateways.
Keywords	science gateways, high performance computing, sustainability, SGCI
Poster session	4
Poster #	41

First name	Liqing
Middle Initial	
Last name	Zhang
Organization	Virginia Tech
NSF Award Title	Frameworks: Developing CyberInfrastructure for Waterborne Antibiotic Resistance Risk Surveillance (CI4-WARS)
NSF Award Number	2004751
Abstract	CI4-WARS is a cyberinfrastructure (CI) enabling an integrated network of wastewater treatment plants (WWTPs) incorporating CI for antibiotic resistance (AR) surveillance. AR, the ability of bacteria to survive antibiotic treatment, poses a grand challenge that calls for an interdisciplinary approach to combat its spread. Combining next generation sequencing with CI for monitoring AR in WWTPs is a promising way to achieve this. WWTPs aggregate antibiotics excreted by all people in a community. Identifying AR anomalies in WWTPs could help identify potential outbreaks before they occur. We aim to (1) Generate sequencing data for a local WWTP; (2) Develop computational tools for identifying AR indicators and abnormal patterns; and (3) Integrate the tools into CI4-WARS, establishing it as a one-stop service for evaluating and reporting AR risk indicators. Broader impact activities include student and professional training, workshops, tutorials, and making CI4-WARS freely available online.
Keywords	Antibiotic resistance, wastewater treatment plants, surveillance, anomaly detection, metagenomic sequencing
Poster session	1
Poster #	41