NSF CI Center of Excellence Pilot

Developing a model and a plan for a Cyberinfrastructure Center of Excellence

Ewa Deelman
Information Sciences Institute
University of Southern California

Presentation to the Council of Data Facilities
5/14/2021

Award #1842042

10/2018–9/2021
Develop a model and a plan for a Cyberinfrastructure Center of Excellence

- Dedicated to the enhancement of CI for science
- Platform for knowledge sharing and community building
- Forum for discussions about CI sustainability and workforce development and training
- Key partner for the establishment and improvement of large-scale projects with advanced CI architecture designs
- Partnering with other community efforts (TrustedCI, ResearchSOC, SGCI, OSG,..) to support science

[Link to the website: http://cicoe-pilot.org/]
Project Team

**USC**
Ewa Deelman (PI)
Mats Rynge
Karan Vahi
Loïc Pottier
Rafael Ferreira da Silva
Wendy Whitcup

Automation, Resource Management, Workflows, Project Management

**RENCI**
Anirban Mandal (co-PI)
Ilya Baldin
Laura Christopherson
Erik Scott

Resource Management, Networking, Clouds, Social Science
University of Notre Dame
Jarek Nabrzyski (Co-PI)
Jane Wyngaard
Charles Vardeman
Mary Gohsman

University of Utah
Valerio Pascucci, Rob Ricci (Co-PIs)
Steve Petruzza

Indiana University
Angela Murillo

Texas Tech U
Kerk Kee

Trusted CI
Josh Drake

Data Archiving
Cybersecurity
Communication & organization science

Workforce development, Sensors, operations, Semantic technologies
Data management, visualization, clouds, large-scale CI deployment
Deep engagement:
- Identify a topic or topics that are important and not-yet fully solved by the Major Facility (MF)
- Form working groups
- Conduct focused discussions, mix of virtual and in-person presence, hands-on work
- Work products: documents/papers, prototypes, schema implementations, demos
- Document and evaluate the collaboration and outcomes

- **Topical discussions:** Identify a topic that is important to a number of MFs
  - Facilitate discussions, sessions at conferences, collect and share experiences, distill best practices

- **Community building:** Identify related efforts
  - Collect information and disseminate information about the broad community activities
  - Host community activities

*Each engagement has a working group with a leader and a set of work products.*
Goals

• Learn about MFs (operations, enhancements)
• Understand how a CI CoE can help with planned CI enhancements
• Inform the model for a CI CoE
• Address issues related to sensor data collection and processing
• Help with NEON data annotation and discovery
• Visualize AOP data and enhance access
• Design an IDM solution

Products: software prototypes, documents, schema designs, presentations, videos, publications

The Ci CoE had four types of profound influence on NEON developers. First, as we transitioned from construction to operations, our developers benefited from greater awareness of the wider NSF CI community practices. Second, deep engagement with CI CoE experts produced three major technologies insertions into NEON CI, remarkably within 6 months. Third, open dialog and prototyping with CI CoE experts affirmed our workflow-based sensor message handling strategy and built our confidence to invest in this novel method. NEON’s CI and Data Sciences team mission includes advancing methods and ecological science; interaction with CI CoE nudged our efforts ahead significantly through community workgroup involvement, presentations and publications.

– Tom Gulbransen, NEON

2018/2019
Goals:

• Combine NEON ecosystem data with NCAR atmospheric and land modeling capabilities
• Inspire new discoveries with integrated data from NEON and NCAR modeling
• Use cloud technologies to enable data modeling and wide community access
• Consulted on cloud technologies, including data, containers, etc.
• Helped inform NEON/NCAR’s proposal to NSF
• Engagement in system design

Overall, the CI CoE was very proactive in enabling and advancing constructive conversations during the development of the NCAR-NEON cyberinfrastructure technical plan and in providing feedback that improved the quality of the plan. The CI CoE facilitated weekly telecons among NCAR, NEON, and the CI CoE to discuss the proposed project and to provide guidance. Discussions during the telecom were mostly high level, helping to identify needs and priorities for the cyberinfrastructure collaborations, but also identifying gaps in the technical plan as well as providing overviews of various alternative implementation strategies. The discussions were always constructive, collaborative, and respectful of all participants. The CI CoE also undertook written feedback on the technical plan at several stages of development. Again, the comments were constructive and improved the quality of the final document. The CI CoE’s engagement was critical to preparing the final technical plan, in part because of the CI CoE’s familiarity with NEON’s data cyberinfrastructure, but also as external computer science experts familiar with the computational needs of the modeling (NCAR) and the data (NEON) and serving to bridge the two different areas of expertise.

– Gordon Bonan, National Center for Atmospheric Research (NCAR)
SAGE & GAGE Engagement

- Initial whole team presentations to understand the context and goals
- **Embedded** Pilot members embedded in CCP working groups (initial list): Data flows and use cases, Conops: High-level requirements: Platform design:
  - Close collaboration, even virtually the teams have come together quickly
- CI CoE Pilot
  - gained knowledge from a large-scale, extremely complex, real-time system
  - gained understanding of important design considerations for cloud migration
  - contributed to CCP documents with platform design schematics, use-case suggestions, reviews and document organization

**Products**: internal and public documents, video presentations

The engagements are helping to further inform the building of the planned CI CoE
Other Deep Engagements

**OOI**

**Goal:** help with the new datacenter provider selection

Participated in:
- RFP review
- Evaluation criteria definition
- Evaluation of proposals
- Made recommendation to the OOI PMO

**GAGE (in collaboration with Trusted CI)**

**Goal:** Recommending a federated IdM platform for GAGE to identify and track usage of data available through their research portal.

**Arecibo Observatory**

**Goal:** help with data dissemination and cloud data processing considerations:
- Metadata analysis
- Data organization
- Repository technologies
- Supporting data

2020

COMPLETE

IN PROGRESS
Identity Management WG
(in Collaboration with Trusted CI)

**Goal:** Disseminate IDM information
- Monthly meetings with speakers and discussions on topics relevant to MFs: e.g. CILogon
- Engagements, primarily focusing on federated identity management
- Issues of identifying data usage and enabling reporting

**Engaging with the community to discuss issues of:**
- Big data visualization
- Cloud migration
- ConOps
- Fair data
- MF data lifecycle
- Science workflows
- Workforce development
Data Lifecycle

Goal: Understand the MF DLC to design CI CoE

Interviews with 22 representatives from 9 MFs participated in the interviews: CHESS, GAGE, IceCube, LHC-CMS, NHERI/Designsafe, NHERI/RAPID, NOIR, OOI, and SAGE.
Blueprint: Cyberinfrastructure Center of Excellence

Proposal:
An NSF Cyberinfrastructure Center of Excellence for Navigating the Major Facilities Data Lifecycle

2018-2021
2021-2026 (proposed)

https://tinyurl.com/cicoe-blueprint

COMPLETE
CI Compass Vision:
Be the leader in supporting and enhancing the national CI ecosystem that includes people, practical knowledge, and processes to facilitate knowledge sharing and discovery across NSF Major Facilities.

CI Compass Mission Statement:
CI CoE provides expertise and active support to cyberinfrastructure practitioners at NSF Major Facilities in order to accelerate the data lifecycle and ensure the integrity and effectiveness of the cyberinfrastructure upon which research and discovery depend.
Data Lifecycle Services:

- Evaluate CI Data lifecycle Plans
- Help Design New Solutions
- Develop Proofs of Concept
- Assess Applicability/Performance of Existing Solutions/Help Leverage CI Solutions

We look forward to building broader community connections!

Ewa Deelman: deelman@isi.edu

http://cicoe-pilot.org/