USC Viterbi
School of Engineering
Information Sciences Institute

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

Connecting CI, Connecting People

Award #1842042

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

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

Project Team

University of Notre Dame

Jarek Nabrzyski (Co-PI)
Jane Wyngaard
Charles Vardeman
Mary Gohsman

Workforce development, Sensors, operations, Semantic technologies

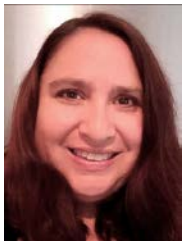
University of Utah

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

Data management, visualization, clouds, large-scale CI deployment

Indiana University

Angela Murillo



Data Archiving

Trusted CI

Josh Drake

Cybersecurity

Texas Tech U

Kerk Kee

Communication & organization science

USC Viterbi School of Engineering Information Sciences Institute
University of Notre Dame
The University of Utah
renci
Trusted CI The NSF Cybersecurity Center of Excellence

How we work

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.

NEON Engagement

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

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

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

COMPLETE

2018/2019

Engagement with NEON and NCAR

Goals:

2020

- 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

COMPLETE

- Engagement in system design

IN PROGRESS

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): 2020-
  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

IN PROGRESS

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

COMPLETE

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.

COMPLETE

2020

Arecibo Observatory

Goal: help with data dissemination and cloud data processing considerations:

- Metadata analysis
- Data organization
- Repository technologies
- Supporting data

IN PROGRESS

Identity Management WG
(in Collaboration with Trusted CI)

2020 Cyberinfrastructure/Cybersecurity Workshop

August 18 and August 20, 2020
https://cics-workshop.org

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:

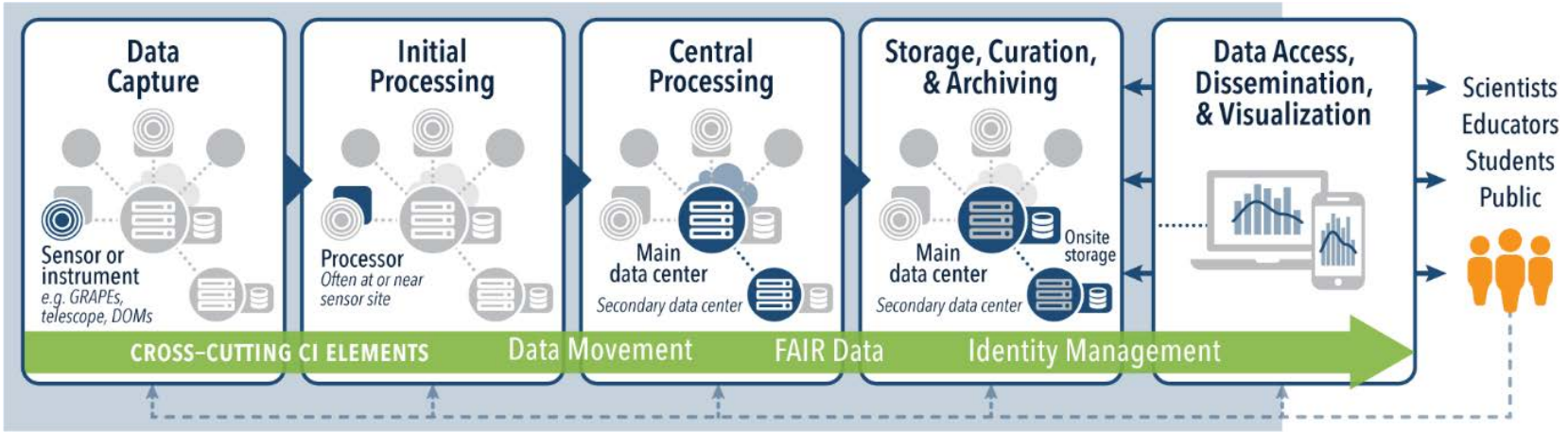
- Big data visualization
- Cloud migration
- ConOps
- Fair data
- MF data lifecycle
- Science workflows
- Workforce development

Community Interviews and Data Lifecycle Model

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.



CI CoE PILOT 2018-2021

zenodo Search Upload Communities

March 7, 2021 Report Open Access

Blueprint: Cyberinfrastructure Center of Excellence

Deelman, Ewa; Mandal, Anirban; Murillo, Angela P.; Nabrzyski, Jarek; Pascucci, Valerio; Ricci, Robert; Baldin, Ilya; Sons, Susan; Christopherson, Laura; Vardeman, Charles; Ferreira da Silva, Rafael; Wyngaard, Jane; Petruzza, Steve; Rynge, Mats; Vahi, Karan; Whitcup, Wendy R.; Drake, Josh; Scott, Erik

In 2018, NSF funded an effort to pilot a Cyberinfrastructure Center of Excellence (CI CoE or Center) that would serve the cyberinfrastructure (CI) needs of the NSF Major Facilities (MFs) and large projects with advanced CI architectures. The goal of the CI CoE Pilot project (Pilot) effort was to develop a model and a blueprint for such a CoE by engaging with the MFs, understanding their CI needs, understanding the contributions the MFs are making to the CI community, and exploring opportunities for building a broader CI community. This document summarizes the results of community engagements conducted during the first two years of the project and describes the identified CI needs of the MFs. To better understand MFs' CI, the Pilot has developed and validated a model of the MF data lifecycle that follows the data generation and management within a facility and gained an understanding of how this model captures the fundamental stages that the facilities' data passes through from the scientific instruments to the principal investigators and their teams, to the broader collaborations and the public. The Pilot also aimed to understand what CI workforce development challenges the MFs face while designing, constructing, and operating their CI and what solutions they are exploring and adopting within their projects. Based on the needs of the MFs in the data lifecycle and workforce development areas, this document outlines a blueprint for a CI CoE that will learn about and share the CI solutions designed, developed, and/or adopted by the MFs, provide expertise to the largest NSF projects with advanced and complex CI architectures, and foster a community of CI practitioners and researchers.

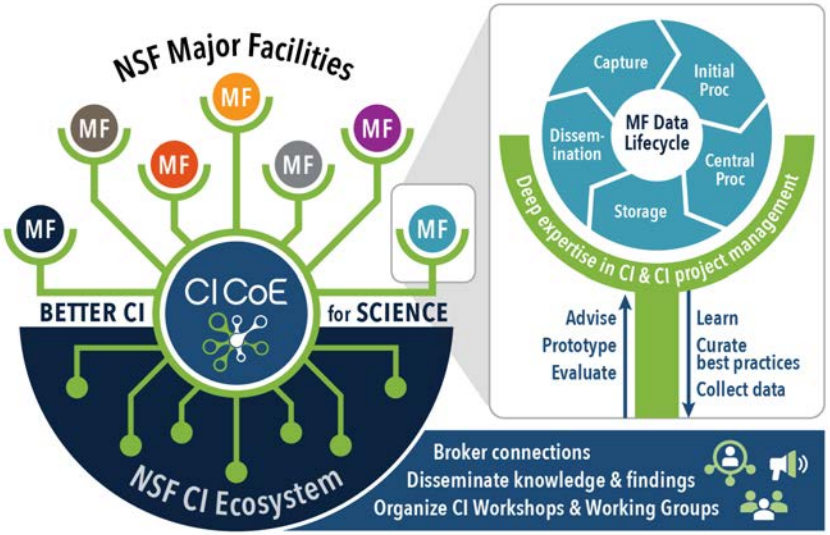
https://tinyurl.com/cicoe-blueprint

COMPLETE

CI Compass 2021-2026 (proposed)

REVIEW

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



USC Viterbi School of Engineering Information Sciences Institute
UNIVERSITY OF NOTRE DAME
THE UNIVERSITY OF UTAH
renci
TRUSTED CI THE NSF CYBERSECURITY CENTER OF EXCELLENCE

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:

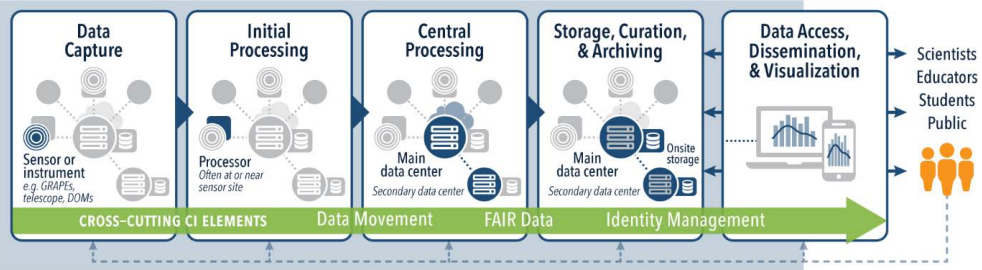
CI CoE provides expertise and active support to cyberinfrastructure practitioners at NSF Major Facilities in order to accelerate the data lifecycle and ensure the integrity and effectiveness of the cyberinfrastructure upon which research and discovery depend.

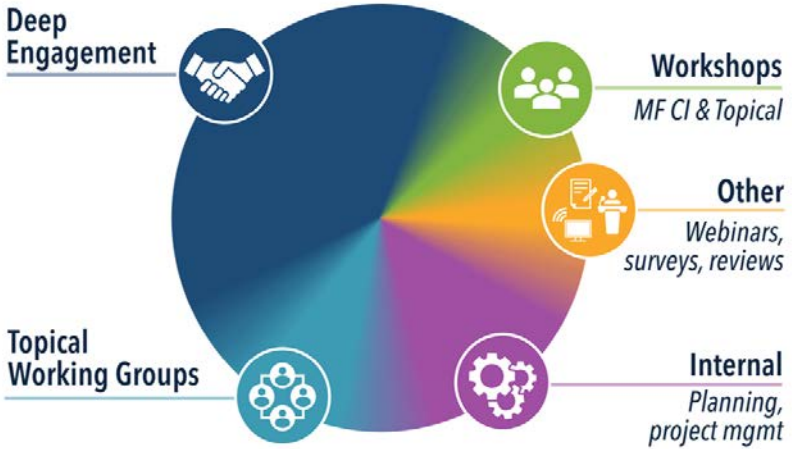
CICompass *An NSF CI CoE for Navigating the Major Facilities Data Lifecycle*

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

REVIEW





Activities

We look forward to building broader community connections!

Ewa Deelman: deelman@isi.edu

http://cicoe-pilot.org/

USC Viterbi School of Engineering Information Sciences Institute

UNIVERSITY OF NOTRE DAME

THE UNIVERSITY OF UTAH

renci

TRUSTED CI THE NSF CYBERSECURITY CENTER OF EXCELLENCE