Funded by the National Science Foundation, Grant #2127548

CICompass

CI CoE: CI Compass: An NSF Cyberinfrastructure (CI) Center of Excellence for Navigating the Major Facilities Data Lifecycle

Ewa Deelman
Information Sciences Institute
University of Southern California

Webinar: Introducing CI Compass 10/4/2021

USC Viterbi School of Engineering Information Sciences Institute
INDIANA UNIVERSITY
UNIVERSITY OF NOTRE DAME
THE UNIVERSITY OF UTAH
TEXAS TECH UNIVERSITY
renci

Funded by the National Science Foundation, Grant #2127548

NSF's Blueprint for a National Cyberinfrastructure Ecosystem for Science and Engineering in the 21st Century

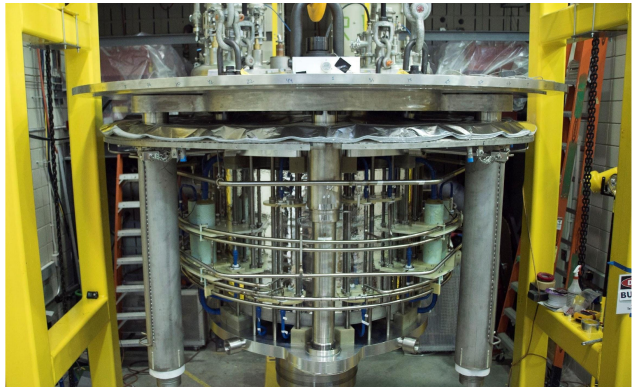
"NSF large facilities are increasingly dependent on advanced CI. Current and planned large facilities are producing increasing volumes of data and data products resulting in growing CI requirements, including computational, networking, data management, and processing requirements, and associated expertise."

https://www.nsf.gov/cise/oac/vision/blueprint-2019/ *2019*

NRAO

LIGO

ARF, RV Sikuliaq



NHMFL, MagLab

Funded by the National Science Foundation, Grant #2127548

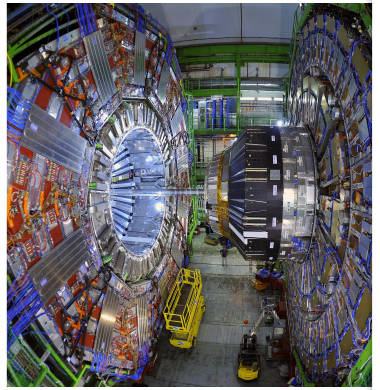
NSF Large/Major Facilities

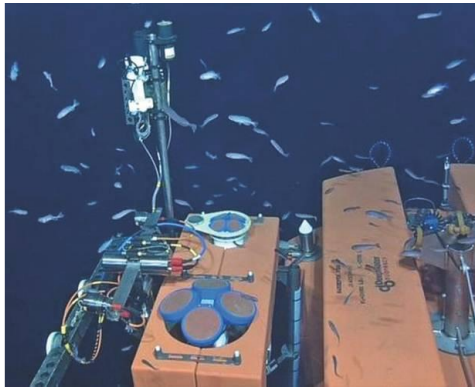
- Deliver data, modeling, computational, and physical capabilities to the broad research community and the public
- Highly diverse, complex, and heterogeneous
- Differ in types of data captured, scientific instruments used, data processing and analyses conducted, policies and methods for data sharing and use
- Rely on complex CI to the transform raw data into more interoperable and integration-ready data products that can be visualized, disseminated, and transformed into insights and knowledge.

IceCube

IRIS/SAGE

NEON


CMS


OOI

Funded by the National Science Foundation, Grant #2127548

Mission

CI Compass 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.

Funded by the National Science Foundation, Grant #2127548

2017 NSF Workshop on CI for Large Facilities

- ***Establish a center of excellence*** *(following a model similar to the NSF-funded Center for Trustworthy Scientific Cyberinfrastructure, CTSC) as a resource providing expertise in CI technologies and effective practices related to large-scale facilities as they conceptualize, start up, and operate.*
- *Foster the creation of a facilities' CI community and establish mechanisms and resources to enable* ***the community to interact, collaborate, and share****.*

Funded by NSF under grant #1742969

2017 2018 2019 2020 2021

Funded by the National Science Foundation, Grant #2127548

NSF CI Center of Excellence pilot

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

Funded by NSF under grant #1842042

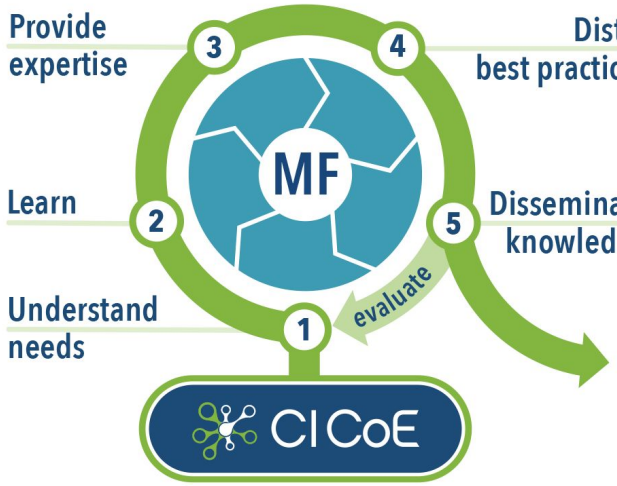
2017 2018 2019 2020 2021

Funded by the National Science Foundation, Grant #2127548

Deep Engagement

CI CoE PILOT

- Identify a topic or topics that are important and not-yet fully solved by the Major Facility (MF)
- Form working groups that includes MF and CI CoE
- 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



NEON NEON and NCAR SAGE and GAGE Arecibo OOI

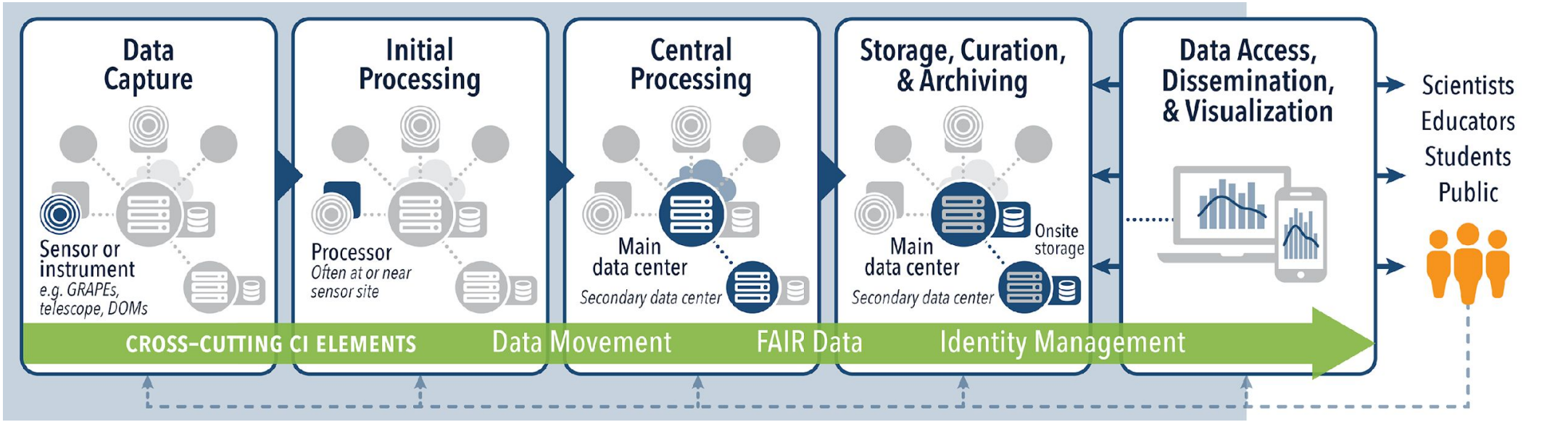
Interviews with selected MFs

2017 2018 2019 2020 2021

Funded by the National Science Foundation, Grant #2127548

MF Data Lifecycle Model development

CI CoE PILOT



2017 2018 2019 2020 2021

Funded by the National Science Foundation, Grant #2127548

Blueprint for a CI CoE that serves the NSF Major Facilities

zenodo Search Upload Communities

March 7, 2021 Report Open Access

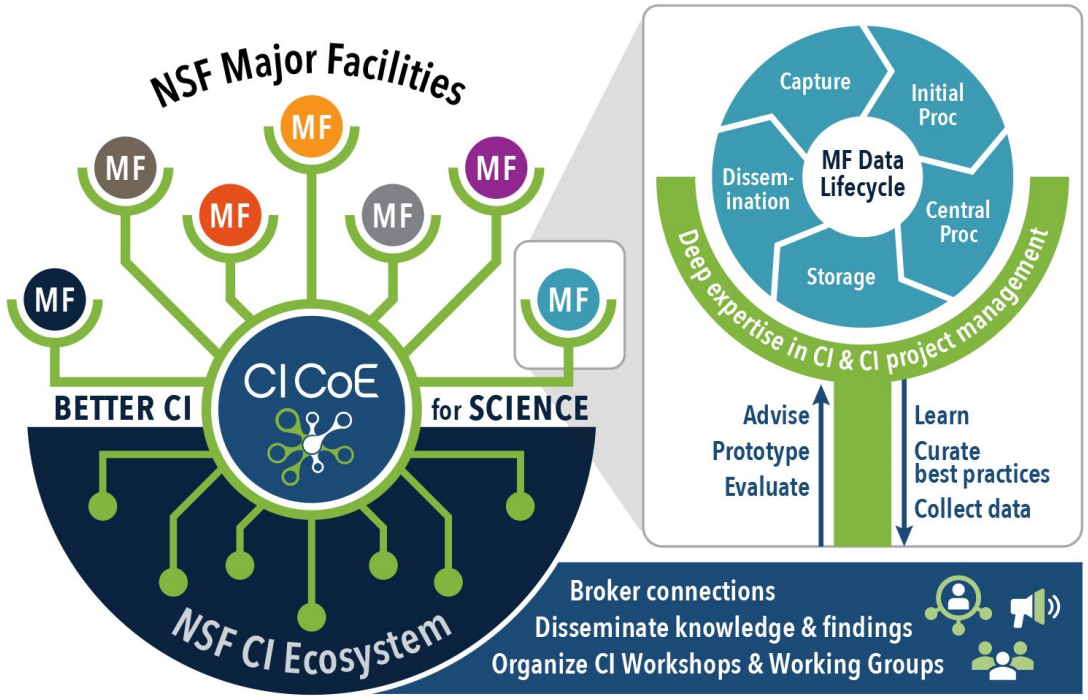
Blueprint: Cyberinfrastructure Center of Excellence

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

CI Compass Proposal



2017 2018 2019 2020 2021

Funded by the National Science Foundation, Grant #2127548

CI Compass Team

- **Deep expertise in areas critical to the MF DLC**
  - Data management, data processing, visualization,
  - identity management, systems and infrastructure
- **Experience in the management of CI projects**
  - Conceptualization, from design phase to the broad adoption
- **Highly collaborative, strong history of working together**
  - Many diverse community connections
- **Dedicated to the advancement of CI for science, engineering, and education**

USC Viterbi School of Engineering Information Sciences Institute
INDIANA UNIVERSITY
UNIVERSITY OF NOTRE DAME
THE UNIVERSITY OF UTAH
TEXAS TECH UNIVERSITY
renci

Funded by the National Science Foundation, Grant #2127548

CI Compass Team

USC:

- Ewa Deelman (PI)
- Mats Rynge
- Karan Vahi
- Loïc Pottier
- Wendy Whitcup
- Ciji Davis
- Rajiv Mayani

Automation, Resource Management, Workflows, Project Management

RENCI:

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

Resource Management, Networking, Clouds, Social Science, Evaluation

Funded by the National Science Foundation, Grant #2127548

Project Team

University of Notre Dame:

- Jarek Nabrzyski (Co-PI)
- Joanne Fahey
- Charles Vardeman
- Mary Gohsman
- Christina Clark

Workforce development, Sensors, operations, Semantic technologies, Communications and Outreach

University of Utah:

- Valerio Pascucci
- Rob Ricci
- Steve Petruzza
- Giorgio Scorzelli

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

Indiana University:

- Angela Murillo

Texas Tech University:

- Kerk Kee

Data Archiving

Communication & organization science

Funded by the National Science Foundation, Grant #2127548

Overall CI Compass Strategy

1. Recognize the expertise, experience, and mission-focus of Major Facilities
2. Contribute knowledge and expertise to the MF DLC CI and enhance the overall NSF CI ecosystem.
3. Build expertise, not software
4. Build on existing knowledge, tools, community efforts
   -Avoid duplication, seek providing added value
5. Share knowledge, lessons learned, best practices
6. Build partnerships to leverage community expertise
   - Trusted CI: cybersecurity
   - Open Science Grid/PaTh: data and workload management
   - Science Gateways Community Institute: portals
   - Engagement and Performance Operations Center (EPOC): network utilization/optimization

Funded by the National Science Foundation, Grant #2127548

Advisory Committee

Stuart Anderson
Caltech

Adam Bolton
Aura/NoirLab

Bonnie Hurwitz
University of Arizona

Miron Livny
University of Wisconsin-Madison

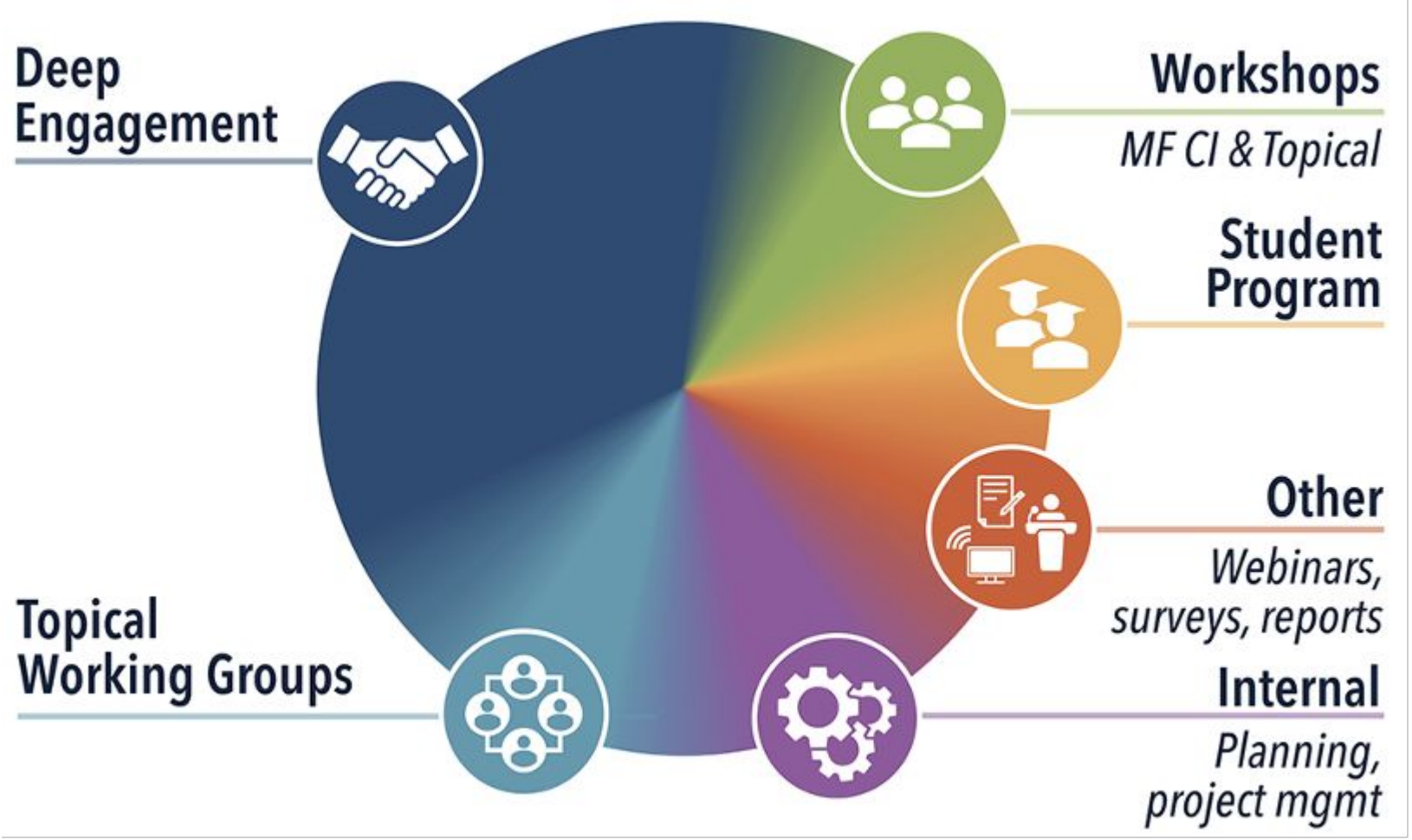
Ellen Rathje
University of Texas at Austin

Von Welch
Indiana University

Michael Zentner
San Diego Supercomputer Center

Funded by the National Science Foundation, Grant #2127548

CI Compass activities



Funded by the National Science Foundation, Grant #2127548

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, work together on particular challenges
- Work products: documents/papers, proofs of concept, 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

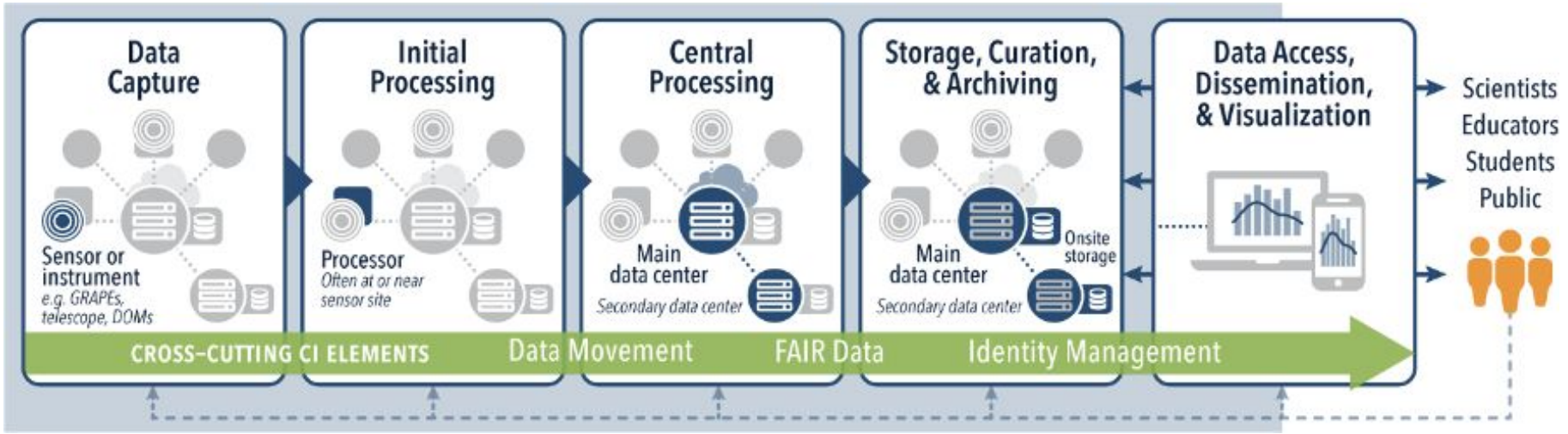
Community Building

- Identify related efforts
- Collect information and disseminate information about the broad community activities
- Host community activities: workshops, training

Deep Engagements

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

Funded by the National Science Foundation, Grant #2127548



Evaluate CI DLC Plans, Help Design New Solutions, Develop Proofs of Concept, Assess Applicability/Performance of Existing Solutions, Help Leverage CI Solutions

Examples of services

Data Capture: Discuss sensor data annotations: apply community sensor data models to the MFs data, explore messaging systems, support interoperability across MFs, and connect to industry standards and to similar communities of practice within scientific domains.

Central Processing: Help leverage existing testbeds to evaluate new software stacks and configurations during CI design/enhancements.

Data access, dissemination & visualization: Assist in designing data visualization solutions (tools, formats, proofs of concept, etc.) for enhancing data discoverability and accessibility.

Funded by the National Science Foundation, Grant #2127548

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

2018/2019

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

CICompass

Funded by the National Science Foundation, Grant #2127548

Data Presentation – New interactive data access

Steve Petruzza, Utah

Before

NEON SCIENCE | DATA PORTAL | BIOREPOSITORY

neon

› Data Collection › Airborne Remote Sensing › AOP Data to Hard Drive Request

AOP Data to Hard Drive Request

There are several ways, users can access airborne data:

- Download the data from the NEON data portal (recommended for amounts of data)
- Programmatically access the data with the NEON Data API or using the NEON Utilities GitHub repo (>1 GB downloads)
- Mail in a hard drive to receive your data

Please fill out the form below if you are interested in receiving a hard of AOP data, and we will respond with a recommended hard drive siz well as mailing instructions.

After

Change Site: LIRO

Slide to Change Year: 2017

powered by OpenVisus | Download | Little Rock Lake, WI -- September 2017

Select a resolution to download

Download | Close

Photos - DP3.30010.001-D05-2017_LIRO_1-L3-Camera-Mosaic-V01_2170_1153_1_uint8[3]_t0_lvl31.png

See all photos | Add to | Edit & Create | Share

Funded by the National Science Foundation, Grant #2127548

NEON/NCAR Engagement

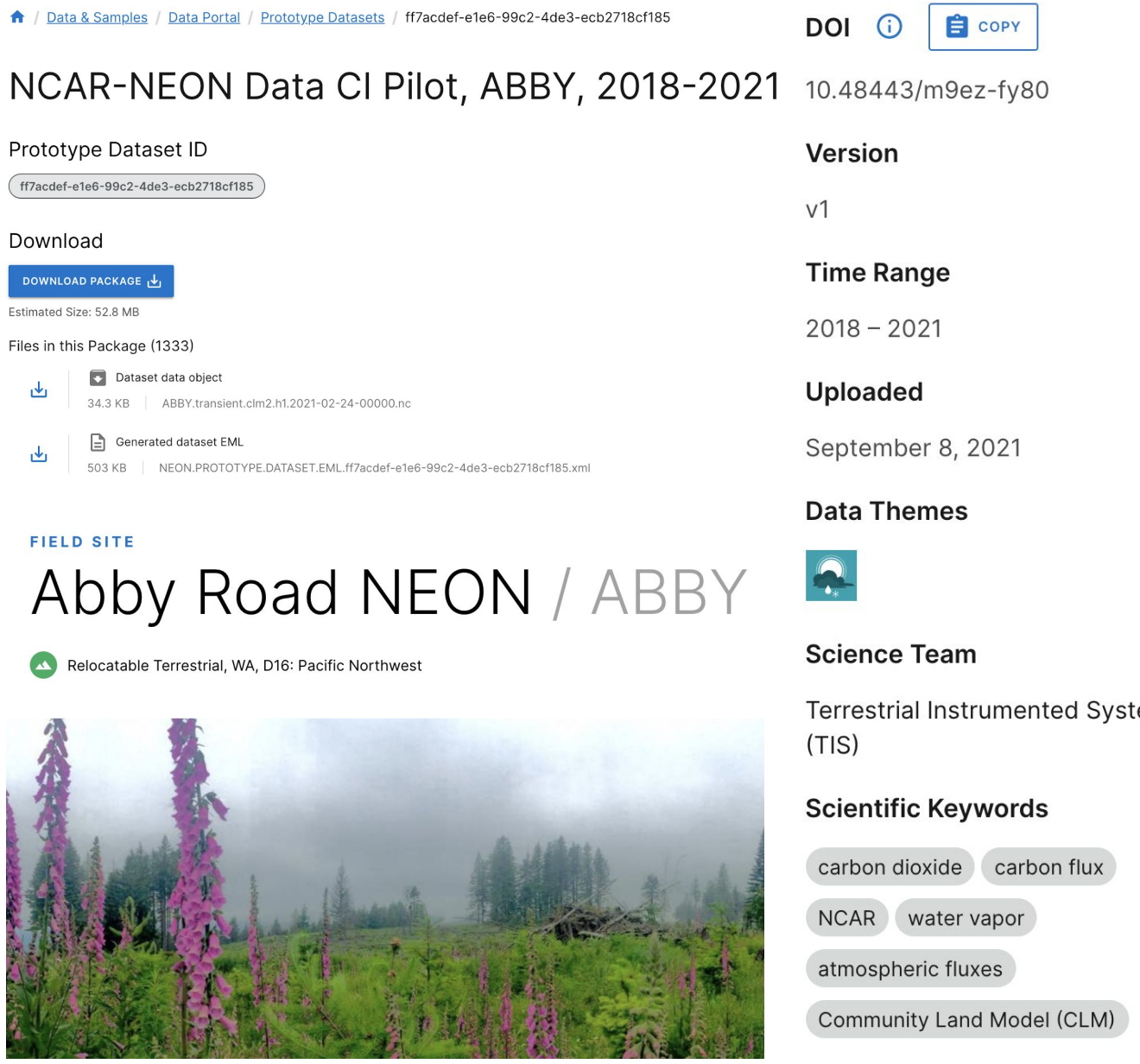
Goals

- 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 containers
- Helped with container testing
- Consulted on FAIR data and visualization

Engagement in system design



Funded by the National Science Foundation, Grant #2127548

SAGE (Seismological Facilities for the Advancement of Geoscience)/
GAGE (Geodetic Facility for the Advancement of Geoscience)

Common Cloud Platform Engagement

- Initial whole team presentations to understand the context and goals
- Embedded team members in CCP working groups (example groups)
  - **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

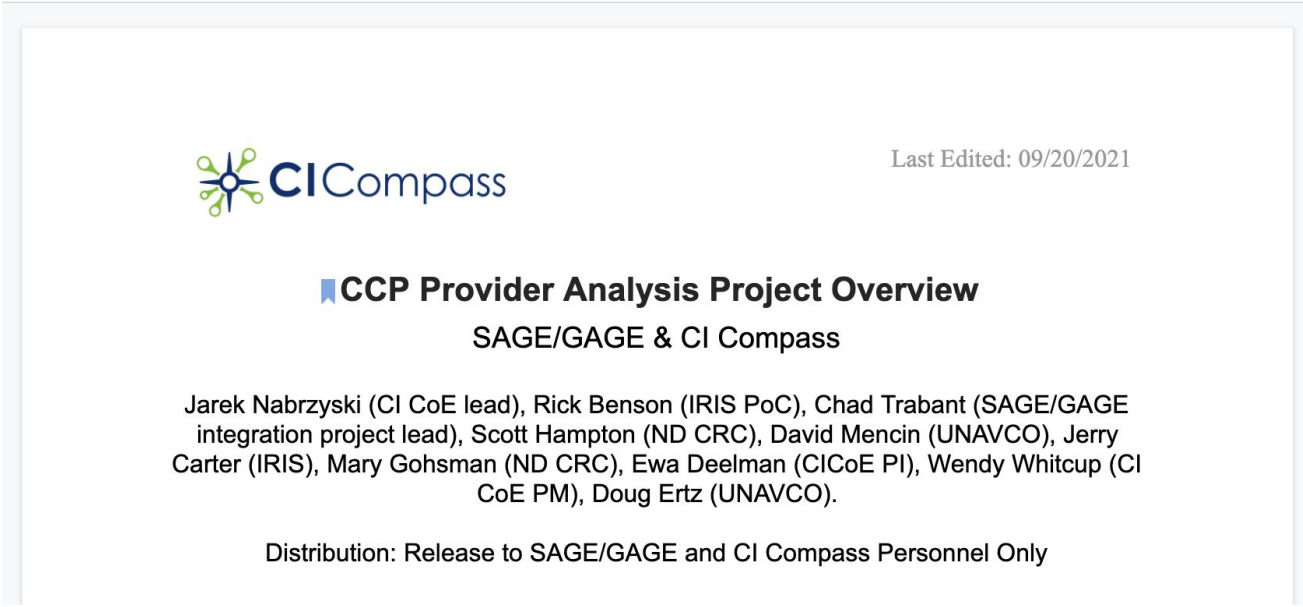
Products: internal and public documents, video presentations

The engagements helped to further inform the building of CI Compass

GAGE Operated by UNAVCO
SAGE Operated by IRIS

Funded by the National Science Foundation, Grant #2127548

Jarek Nabrzyski,
Notre Dame,
CI Compass


CICompass

Last Edited: 09/20/2021

CCP Provider Analysis Project Overview

SAGE/GAGE & CI Compass

Jarek Nabrzyski (CI CoE lead), Rick Benson (IRIS PoC), Chad Trabant (SAGE/GAGE integration project lead), Scott Hampton (ND CRC), David Mencin (UNAVCO), Jerry Carter (IRIS), Mary Gohsman (ND CRC), Ewa Deelman (CICoE PI), Wendy Whitcup (CI CoE PM), Doug Ertz (UNAVCO).

Distribution: Release to SAGE/GAGE and CI Compass Personnel Only

Collaboration with INTERNET2

The document is internal but a version with public findings will be released.

CICompass

Arecibo Observatory

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

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

Community effort:

"The Arecibo Observatory, UCF, TACC, the University of Puerto Rico, EPOC, Globus, CICoE Pilot partner to move telescope data to Ranch system"

Data from the collapsed Arecibo radio telescope are being safely moved and preserved to the Texas Advanced Computing Center of UT Austin. A multi-institutional partnership formed to safely transfer over three petabytes of data from 50 years of Arecibo sky scans. The data have helped scientists make Nobel-prize-winning discoveries about the heavens. This data move will ensure future generations of astronomers an opportunity to make new discoveries from Arecibo's data. (Credit: Arecibo Observatory).

https://www.tacc.utexas.edu/-/continuing-arecibo-s-legacy

IVOA

Connected with the International Virtual Observatory to help develop solutions based on community standards

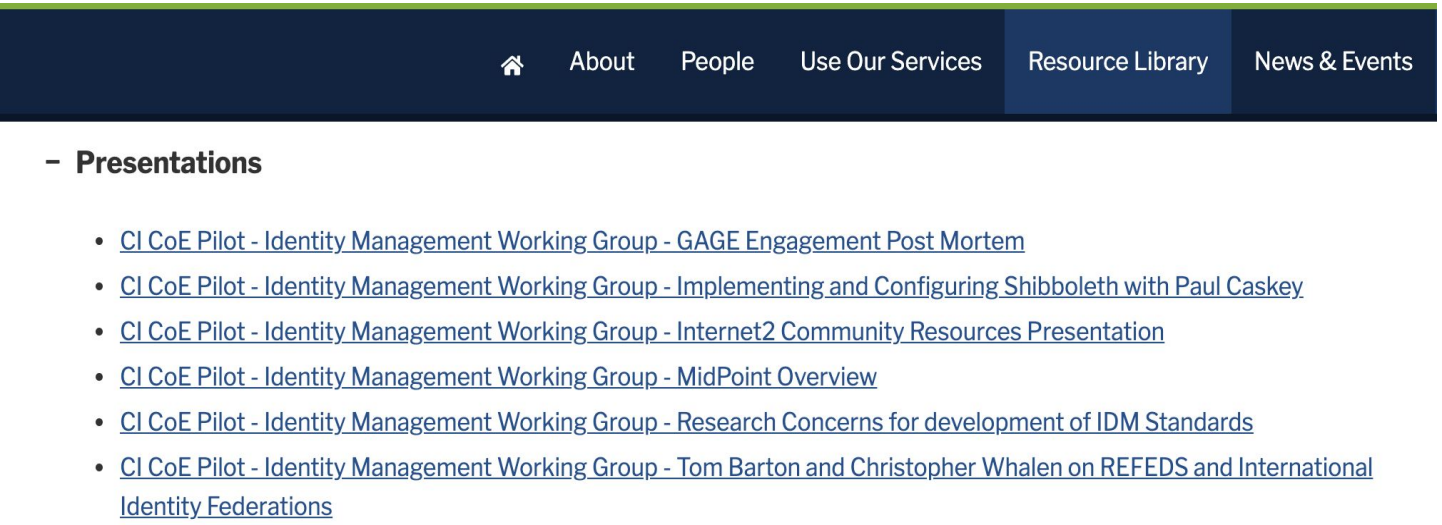
Funded by the National Science Foundation, Grant #2127548

Topical Working Groups

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



Cloud Infrastructure WG

- Understand the current practices for clouds used by MFs
- Research alternative solutions and keep up to date with emerging cloud technologies
- Develop a general set of best practices that can inform the MFs

Funded by the National Science Foundation, Grant #2127548

Community Workshops

2020 Cyberinfrastructure/Cybersecurity Workshop

August 18 and August 20, 2020

https://cics-workshop.org

2020

Engaging with the community to discuss issues of:

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

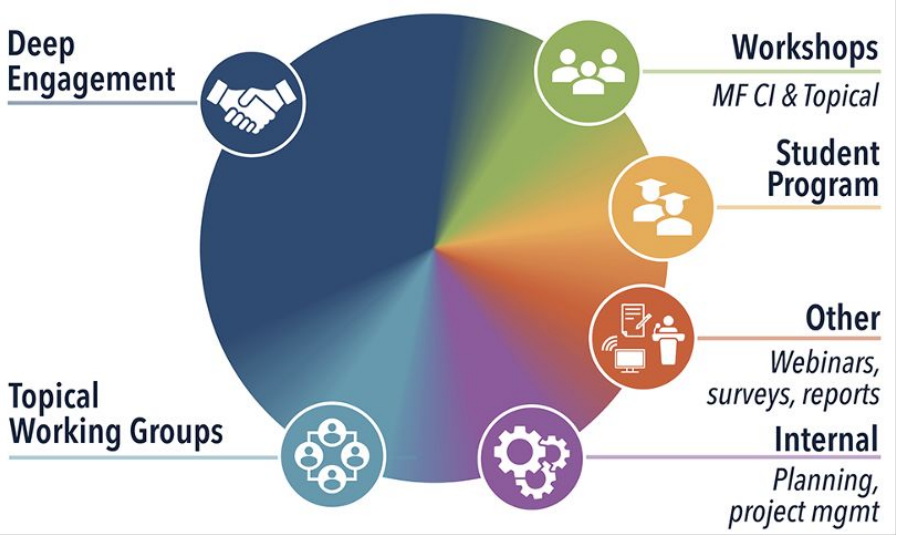
https://doi.org/10.5281/zenodo.5523235

Funded by the National Science Foundation, Grant #2127548

New CI Compass Activities

- Capitalizing on current engagements and developing lessons learned and other community products
  - Technical reports: conducting cloud risk/cost/benefit analysis
  - Topical webinars: FAIR data, cloud migration considerations
- Soliciting new deep engagements, call in the Spring 2022
- Workshop on CI for Major Facilities, February 2022, LA?
  - Soliciting topics and volunteers: https://tinyurl.com/ci-compass-mfci-2022

https://ci-compass.org/



Funded by the National Science Foundation, Grant #2127548

Student Internship Program

The program's goal is to broaden student participation in CI research, development, deployment, and operations.

CI Compass Student Interns will:

- learn about cyberinfrastructure and Major Facilities (MFs),
- develop skill sets important to the work of MFs, and
- engage directly with MFs through on-site and/or virtual internship programs.

Soliciting project ideas: https://tinyurl.com/ci-compass-students

Year 1: Create a Pilot Framework

Creating program protocols, procedures, and guidelines including:

- Student recruitment, application, and selection,
- Student onboarding and training,
- MF call for participation and project selection,
- Student intern activities with CI Compass, MFs, and other student interns, and
- Pilot program with four student interns and selected MFs

Funded by the National Science Foundation, Grant #2127548

Join the Conversation!

Contact the CI Compass Team with questions or requests by emailing

contact@ci-compass.org

To learn more about CI Compass services, leadership, news, upcoming events and our resource library,
please visit ci-compass.org

Connect on social media

Twitter
Follow @CICompass

LinkedIn
Connect with us
linkedin.com/company/ci-compass

YouTube
Subscribe to our channel
CI Compass