Funded by the U.S. National Science Foundation, Grant #2127548

CI Compass: Navigating the U.S. National Science Foundation's Major and Midscale Facilities Data Lifecycle

Ewa Deelman
University of Southern California
CI Compass, PI
deelman@isi.edu

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 U.S. National Science Foundation, Grant #2127548

Mission

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

ci-compass.org

Funded by the U.S. National Science Foundation, Grant #2127548

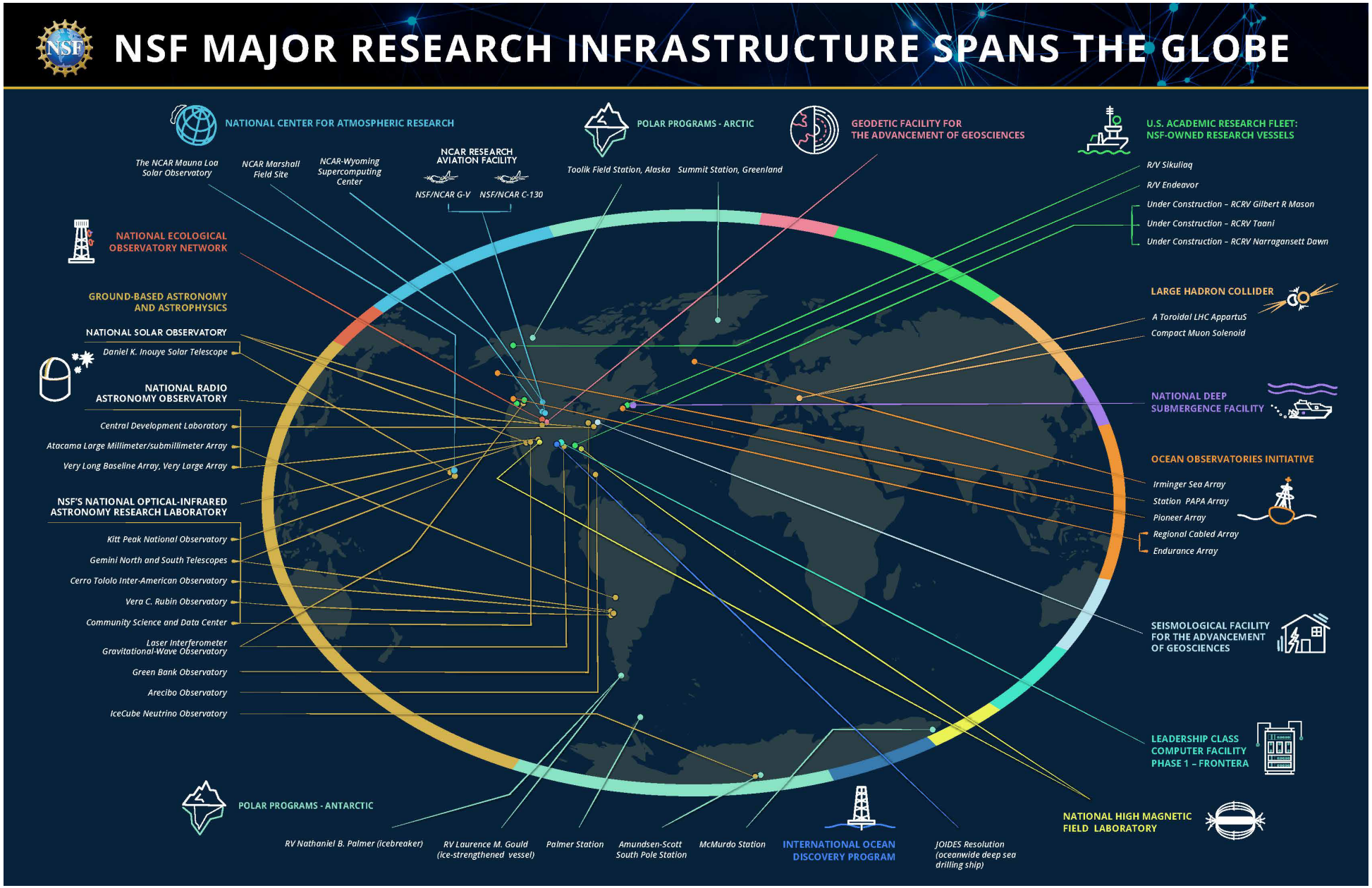
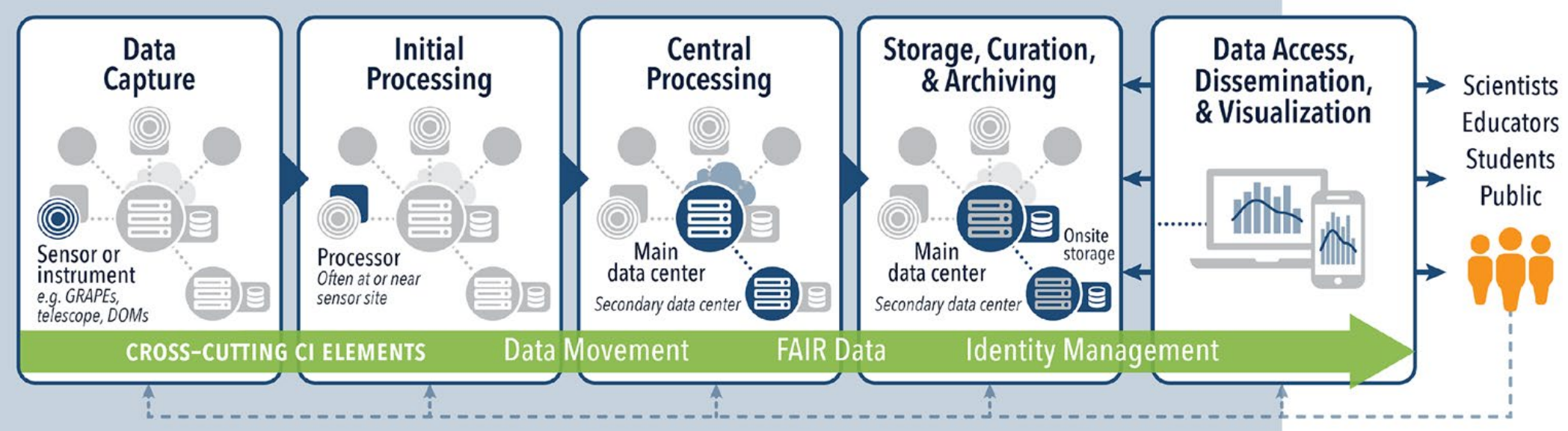


Image courtesy of NSF

Funded by the U.S. National Science Foundation, Grant #2127548

CI Compass Services focus on Major Facilities' Data Lifecycle



Evaluate CI plans, Help architect new solutions, Develop proofs of concept,
Assess applicability/performance of existing solutions, Help leverage existing technologies

CI Compass Team: Who we are

Deep expertise in several CI areas critical to the MFs

- Data management, data processing, visualization, archiving, semantic technologies
- Automation, resource management, workflows, sensors
- Networking, clouds, systems and infrastructure
- Large-scale CI deployment and operations, IdM
- Social science, understand the organization structures and culture of MFs

Experience in the management of CI projects

- Conceptualization, design phase, broad adoption
- Project Management and Evaluation
- Organizational science
- Communications & Outreach

Highly collaborative, strong history of working together and with the CS and CI Communities

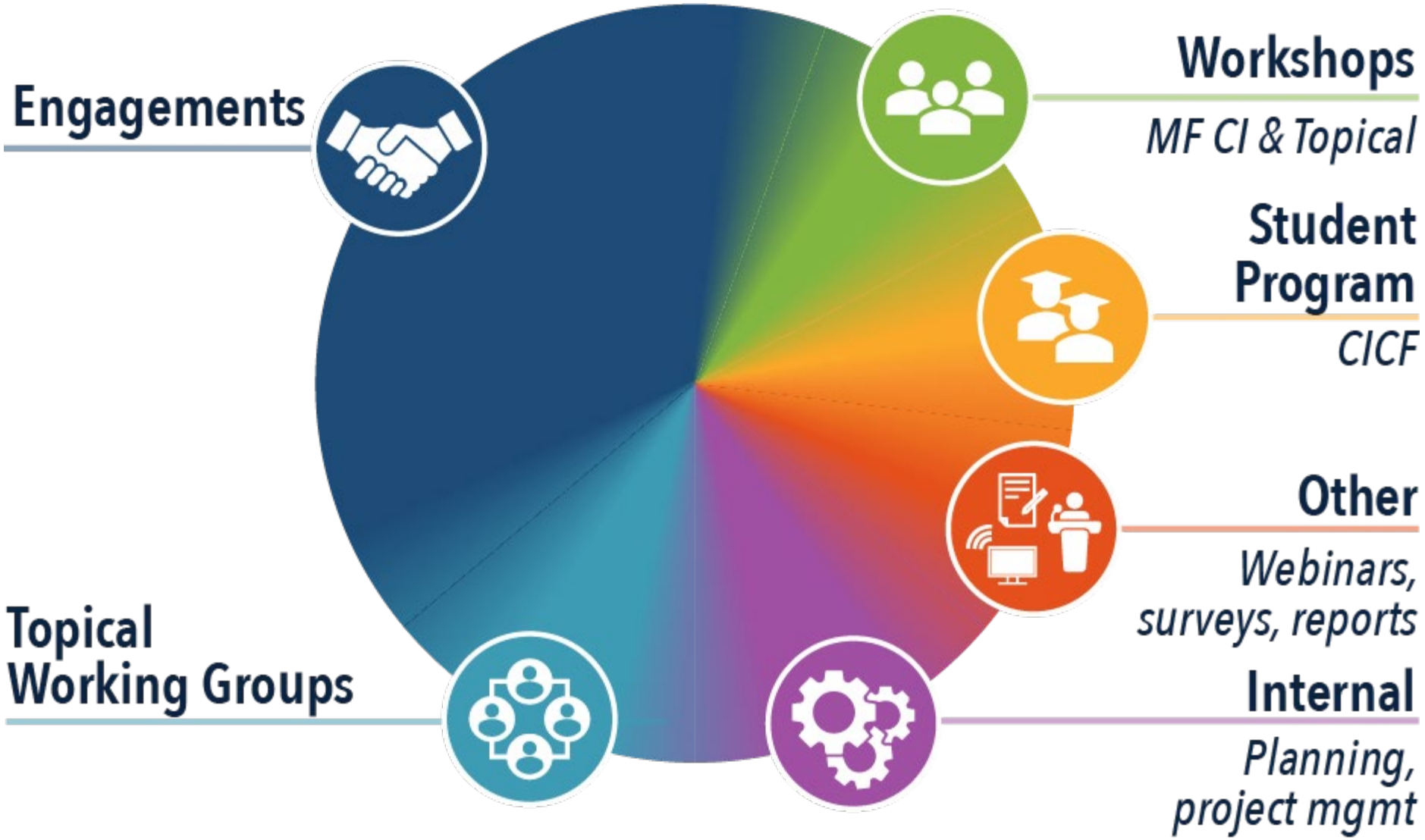
- Many diverse community connections in astronomy, earth science, physics

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 U.S. National Science Foundation, Grant #2127548

CI Compass activities



Funded by the U.S. National Science Foundation, Grant #2127548

Regional Class Research Vessel (RCRV) Engagement:

Shipboard CI/network plan review

Planned RCRV vessels

January - March 2022

CICompass

experienced on the vessel. Systems of this sort range from high-resolution video conferencing setups to full-scale virtual reality. As the degree of immersiveness increases, so does the demand on the CI capacity. This portion of the CI is still under design.

Review of Network Architecture

The basic architecture of the on-board network is a switched hub-and-spoke model. The central hub

"One of the primary concerns identified by the review was that the planned 1GbE switch ports in the ship's computer lab should be supplemented with 10 GbE and higher to support deployment of visiting equipment with high-speed network interfaces. "

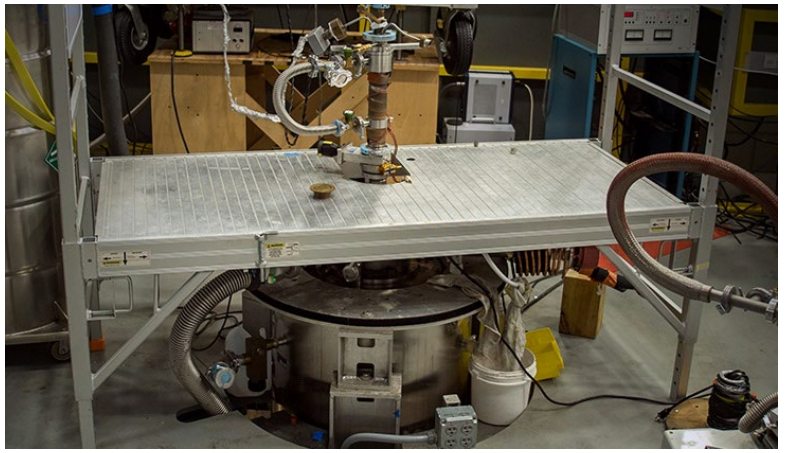
From Chris Romsos, RCRV, OSU : "Thank you for identifying this as something to address now before delivery of the vessel. We planned for future upgrades like this and have sufficient fiber between the network core and the computer lab to support the upgrade…
Sikuliaq recently upgraded their edge switching throughout the vessel… . A nice piece of corroborating evidence there with Sikuliaq! "

Field report from ARF, RV Sikuliaq

Funded by the U.S. National Science Foundation, Grant #2127548

Engagement: MagLab

- **Goal**: Provide MagLab assistance in implementing FAIR data practices for the facility and community of users.
- **Focus**: FAIR instrument descriptions leveraging Research Data Alliance (RDA) Persistent Instrument Identifiers (PIDINST) working group recommendations. Implementation will allow scientists to specify resolvable instrument identifiers in publications.
- **Community**: Engaging Earth Science Information Partners (ESIP) science on schema.org cluster to extended Dataset and Repository with Instrument recommendation. Connecting with ORCID and NSF FAIR Instrument RCN.
- **Deliverable**: Example FAIR instrument documents in JSON-LD along with documentation.



17 Tesla Superconducting Magnet for Optics
Credit: National MagLab

Princeton Instruments
SpectraPro HRS-750

Funded by the U.S. National Science Foundation, Grant #2127548

FAIR Topical Working Group

Interest group for people interested in FAIR and data management at Major Facilities.

- Facilitates conversations on FAIR between MFs
- Discussion on best practices and emerging issues
- Open membership with representation from MFs, and Mid-scale facilities, and others
- Engagement with NSF FAIROS RCN groups
- Building collaborations and doing outreach where it makes sense

Other working groups:

- Identity Management: IdM Cookbook (with Trusted CI)
- Cloud Infrastructure: Report about adopting clouds for MFs
- DLC Guidebook (planned)

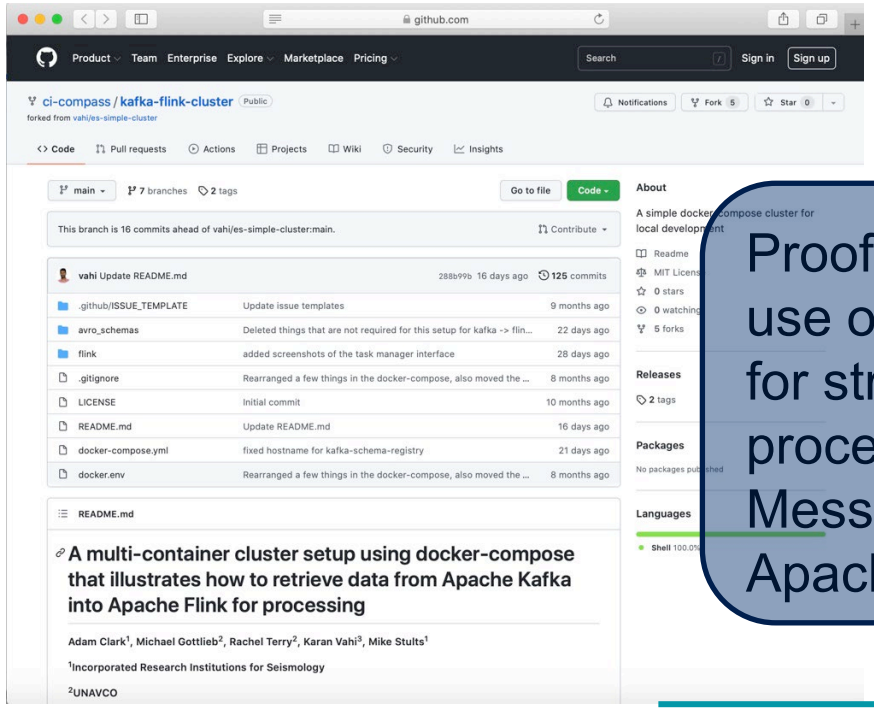
doi:10.5281/zenodo.10481410

CICompass

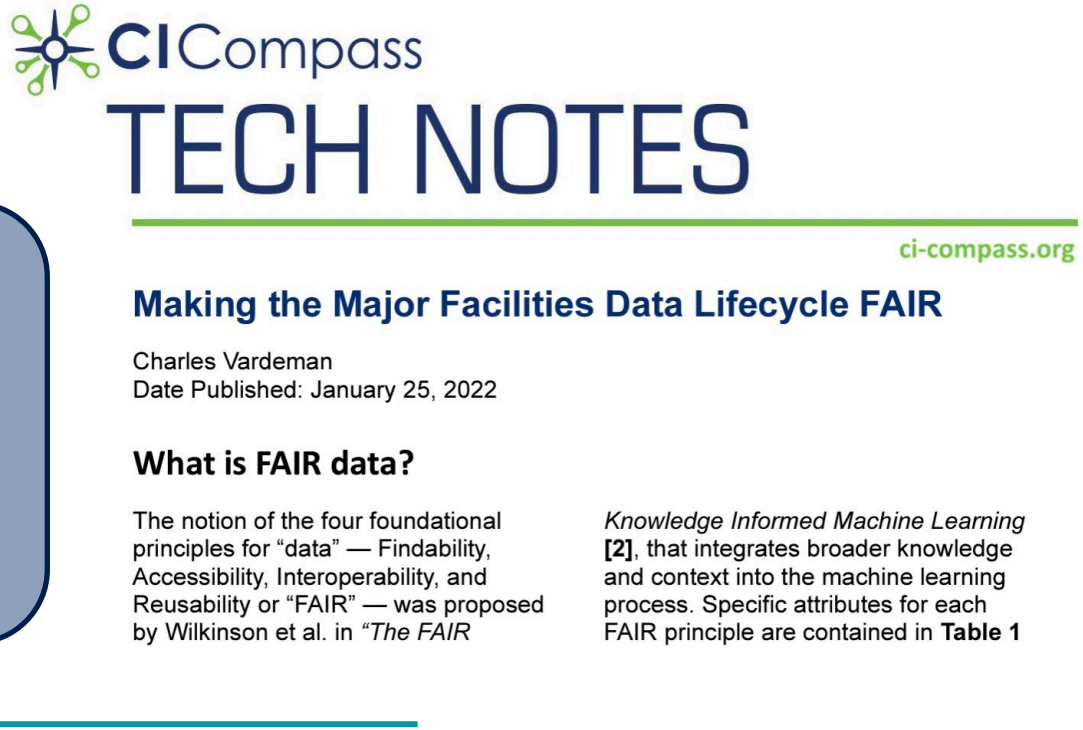
NSF Major Facilities Cloud Use Cases and Considerations

Public Products

Funded by the U.S. National Science Foundation, Grant #2127548



Proof of concept on use of Apache Flink for stream-based processing of Messages out of Apache Kafka



Webinars

CICompass

Upcoming Webinar

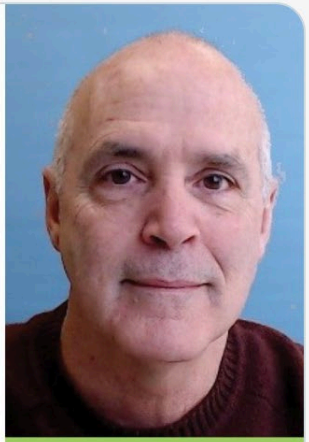
Flexible Data Engineering Pipelines ft. Chris Bontempi

Director of IT, Network for Advanced NMR

Wednesday, May 17
9 am PST / Noon EST, via Zoom

To learn more and register, please visit:
ci-compass.org/news-and-events

ALT



NAN

CICompass

Webinar Series

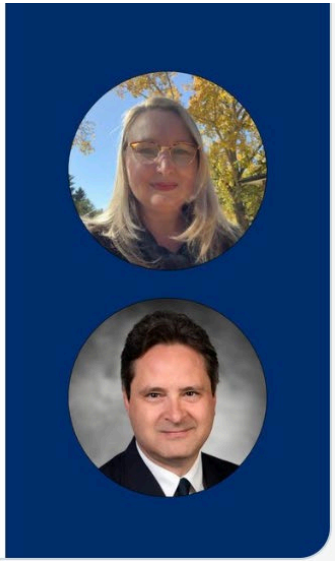
NSF, the Nelson Memo, and Persistent Identifiers

*ft. **Martin Halbert**, NSF Advisor for Public Access, and **Shawna Sadler**, Head of Outreach and Partnership at ORCID*

January 31, 2023
2 p.m. EST / 11 a.m. PST, via Zoom

To learn more and register, please visit:
ci-compass.org/news-and-events

ALT



Webinar Series

CICompass

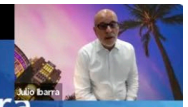
DesignSafe: Data Challenges for National Scale Cyberinfrastructure

Monday, December 9, 2024
2 - 3 p.m. EST / 11 a.m. - Noon PST

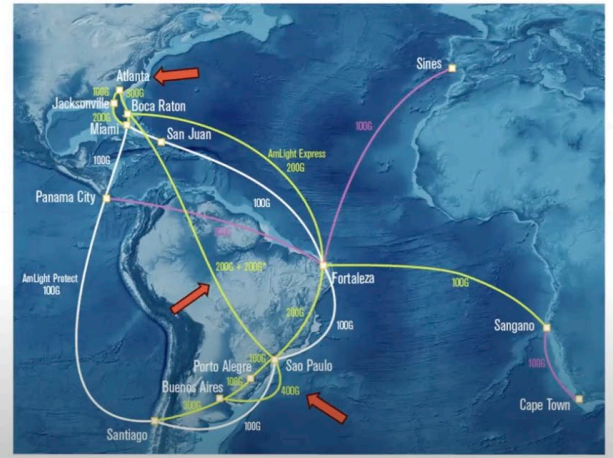
Featuring Chris Jordan, lead of the Data Management and Collections group at the Texas Advanced Computing Center (TACC)

Funded by the U.S. National Science Foundation, Grant #2127548

AmLight International R&E Network: FIU, PI: Julio Ibarra



- 600G of upstream capacity between the U.S., Latin America, Caribbean and 100G to Africa
  - Increasing to 1.1 Tbps between Florida and Brazil
- OXPs: Florida(3), Atlanta, Brazil(2), Chile, Argentina, Puerto Rico, Pana th Africa
- Production SDN Infrastructure since 2014
- Deeply programmable across the network stack
  - Programmable P4 Data Plane
  - Open Source SDN Controller
  - Fine-grained telemetry
  - Run-time network verification
  - Closed-Loop Orchestration
- Highly instrumented
  - PerfSonar, sFlow, Juniper Telemetry Interface (JTI), In-band Network Telemetry (INT)

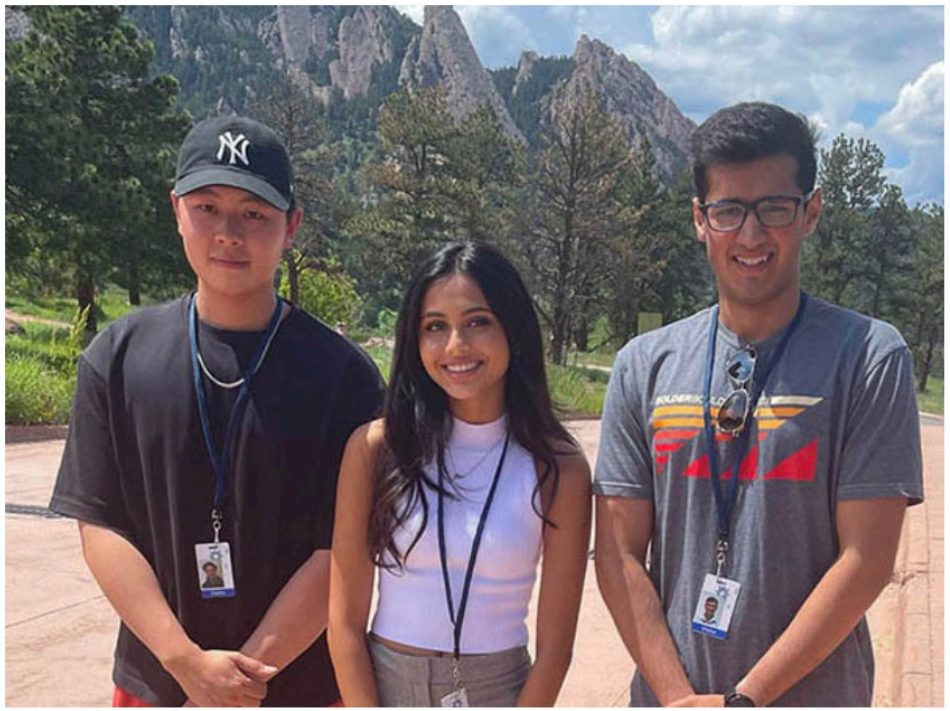


AmLight ExP
Americas Lightpaths Express & Protect

If you would like to give a seminar or suggest a topic related to DLC, please contact us:
contact@ci-compass.org

CICompass
Funded by the U.S. National Science Foundation, Grant #2127548

CI Compass Fellowship Program (CICF)



CI Compass Student Fellows Edward Lin, Mahee Shah, and Raja Allmdar Tariq Ali stand together in Boulder, Colorado, where they spent Summer 2023 working with the National Center for Atmospheric Research (NCAR).

Goal: Broaden student participation in CI research, development, deployment, and operations

Virtual Spring Program

- Free to undergraduate students. Possibility of course credit.
- *Technical Skills Component*: Students are taught technical skills relevant to CI.
- *Data Lifecycle Component*: Students research MFs and the data lifecycle to understand the importance and context of MFs, and the related data and CI. They present their results at the end of the Spring Program.

(Optional/Invited) Summer Program

- We collaborate with MFs to provide CI-related summer projects for some of our student fellows.
- In-person or virtual, depending on the MF/project.
- Students are paid for their participation.

All Students receive a scholarship to attend a professional conference of their choice

20 undergraduate students selected per year (~130 applications)

Funded by the U.S. National Science Foundation, Grant #2127548

CICF curriculum

- We meet with the students for an hour on Tuesdays and Thursdays, with office hours immediately following class time.
- Students also watch a ~1hr pre-recorded Technical Skills lecture before the following Technical Skills lab (i.e. flipped classroom)
- Students do a group project researching the CI and data lifecycle of one MF and present their work at the end of Spring.
- Below is the Syllabus for the 2024 Spring Program. The Technical Skills Component is highlighted in blue, and the Data Life Cycle Component is highlighted in green

2024 SPRING PROGRAM

Week	Tuesday	Thursday
Week 1	Major Facilities, Cyberinfrastructure, and the Data Life Cycle	Command Line
Week 2	Overview of scientific computing	Guest Speaker from TACC
Week 3	Scientific computing, Python, Jupyter, Python Data Analysis Packages	FAIR Data
Week 4	Best Practices in Software Development, Part 1	Guest Speaker from MagLab
Week 5	Best Practices in Software Development, Part 2	MF/DLC Research Project Group Work
Week 6	Cloud Computing, Part 1	Guest Speaker from ORCID
Week 7	Cloud Computing, Part 2	MF/DLC Research Project Group Work
Week 8	Spring Break	Spring Break
Week 9	Software Architecture and Systems; Digital Archives	Guest Speakers from OOI and NEON
Week 10	Data Workflows	Professional Skills, Networking
Week 11	Machine Learning (Classical)	Neural Networks
Week 12	Research Project Group Presentation Day 1	Research Project Group Presentations Day 2

CICF impact

Testimonials from Student Fellows

- *"I didn't know anything about scientific computing, cyberinfrastructure, or Major Facilities before the program. The fellowship introduced me to these and made me curious about how to potentially have a career in scientific computing or cyberinfrastructure and I am more open to working at a Major Facility."*
- *"I was not at all interested by undergraduate research or a graduate degree before starting. I am now researching potential research opportunities in my field and am now interested in learning more about what earning a graduate degree would entail and how it could benefit my future and ability to make an impact."*
- *Before I was very nervous and overwhelmed because it didn't really feel like I was meant to pursue research or further schooling. Now, I feel confident in my ability to become a scientist/academic and have a detailed plan on how I can take next steps towards that goal. Without this program, I would likely still be in a nervous/overwhelmed/paralyzed state and would never independently have learned as much as I did this Spring.*

Institution (17 schools from 11 states)	State
California Polytechnic State University, San Luis Obispo	CA
Merced College	CA
UC Riverside	CA
University of California Merced	CA
University of California, Berkeley	CA
Ball State University	IN
University of Notre Dame	IN
Carroll College	MT
University of North Carolina at Chapel Hill (3 students)	NC
New Jersey Institute of Technology	NJ
Eastern New Mexico University	NM
Brooklyn College	NY
Rochester Institute of Technology (2 students)	NY
The Ohio State University	OH
University of Oklahoma Norman	OK
Villanova University	PA
Texas Tech University	TX

Funded by the U.S. National Science Foundation, Grant #2127548

Virtual Workshop: Data Management: From Instrument to First Storage

January 22-24, 2025

http://ci-compass.org

Presentation about Oceans Network Canada by Ben Biffard

Wednesday (01/22)	
9:00 AM	Workshop introduction, Ewa Deelman
9:15 AM	Talk 3: Ted Habermann, Metadata Game Changers
9:45 AM	Talk 2: Franko Bayer, ZEUS
10:15 AM	Talk 1: Chris Bontempi, NAN
10:45 AM	Talk 4: David Butcher, MagLab
11:15 AM	Day 1 wrap up

Thursday (01/23)	
9:00 AM	Workshop introduction
9:15 AM	Talk 1: Dan Cox, NHERI
9:45 AM	Talk 2: Benedikt Riedel, IceCube
10:15 AM	Talk 3: Jameson Rollins, LIGO
10:45 AM	Talk 4: William O'Mullane, LSST
11:15 AM	Talk 5: Jeff Glatstein, OOI
11:45 AM	Day 2 wrap up

Friday (01/24)	
9:00 AM	Workshop introduction
9:15 AM	Talk 1: Christine Laney, NEON
9:45 AM	Talk 2: Bob Tawa, NSO
10:15 AM	Talk 3: Ben Biffard, ONC
10:45 AM	Talk 4: Rob Casey, EarthScope
11:15 AM	Day 3 wrap up

CICompass — Funded by the U.S. National Science Foundation, Grant #2127548

https://researchinfrastructureoutreach.com/

Research Infrastructure OUTREACH

HOME | WORKSHOP | KNOWLEDGE GATEWAY | COMMUNITIES OF INTEREST | RIG UPDATE | RESEARCH INFRASTRUCTURE OFFICE | CONTACT

Cyberinfrastructure and Cybersecurity sessions

How Are Major Facilities Preparing for AI? How to Prepare Data for AI Use?
How Do You Track Data Usage at Your Major Facility ?
How Much User Experience Plays in the Delivery of Data?
How Do You Vet and Integrate New Technologies or Data Processing Methods into Your CI Environment?

Funded by the U.S. National Science Foundation, Grant #2127548

Potential Collaborations

- Discussions around the data lifecycle and CI planning
  - FAIR data, instrument identifiers, data archival, cloud adoption
  - Handling large datasets and real time requirements
  - Preparing Research Infrastructures for AI
  - Developing CI Plans
- Best practices and what not to do
- 1-1 and group discussions, participation in topical working groups, webinars, community events
- Collaboration on student CI curriculum, potentially on student internships
- Others?

Funded by the U.S. National Science Foundation, Grant #2127548

Connect with us!

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

Contact the CI Compass Team with questions or requests by emailing

contact@ci-compass.org

Ewa Deelman
deelman@isi.edu

Social media

Twitter
Follow @CICompass

YouTube
Subscribe to our channel
CI Compass

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