CI CoE PILOT

https://tinyurl.com/mycicoe

Cyberinfrastructure Center of Excellence Pilot

Funded by the National Science Foundation
Grant #1842042

Ewa Deelman, USC (PI)

Co-PIs:
Anirban Mandal, RENCI

Jarek Nabrzyski, Notre Dame University

Valerio Pascucci and **Rob Ricci**,
University of Utah

Tom Gulbransen, NEON

Sr. Personnel

Rafael Ferreira da Silva (USC)

Steve Petruzza, University of Utah

Jane Wyngaard, UND

Charles Vardeman, UND

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

Cyberinfrastructure "consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible." [1]

[1] Craig A. Stewart, et al. 2010. "What is cyberinfrastructure?" SIGUCCS '10. ACM, New http://doi.acm.org/10.1145/1878335.1878347

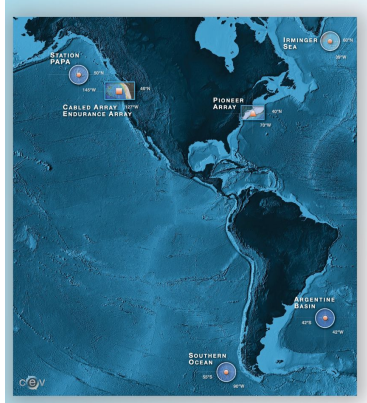
CI is a critical component of Science: Large Facilities (LFs)

Searching for gravitational waves

Understanding ocean and coastal ecosystems

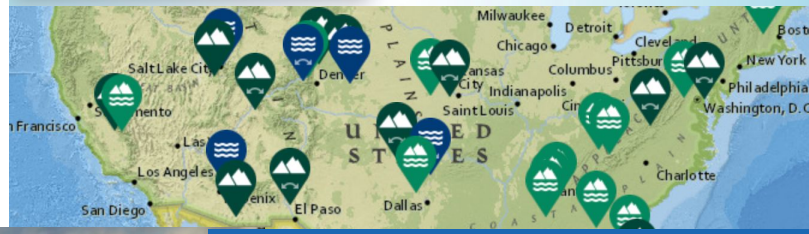
Looking for exoplanets

Studying climate



THE INFRASTRUCTURE

89 PLATFORMS
CARRYING OVER
830 INSTRUMENTS
PROVIDING OVER
100,000 DATA PRODUCTS
HAVE BEEN DESIGNED, BUILT, AND DEPLOYED.



NRAO **National Radio Astronomy Observatory**

The National Ecological Observatory Network: Open data to understand how our aquatic and terrestrial ecosystems are changing.

neon National Ecological Observatory Network

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 CoE Pilot Project Goals

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
- Key partner for the establishment and improvement of Large Facilities with advanced CI architecture designs
- Grounded in re-use of dependable CI tools and solutions
- Forum for discussions about CI sustainability and workforce development and training
- Pilot a study for a CI CoE through close engagement with NEON and further engagement with other LFs and large CI projects.

Overall Strategy

1. Recognize the expertise, experience, and mission-focus of Large Facilities
2. Engage with and learn from current LFs CI
3. Build on existing knowledge, tools, community efforts
   -Avoid duplication, seek providing added value,
4. Prototype solutions that can enhance particular LF's CI
   -Keep a separation between our efforts and the LF's CI developments
5. Build expertise, not software
6. Work with the LFs and the CI community on a blueprint for the CI CoE

Build partnerships:

- Trusted CI (identity management): share personnel
- Open Science Grid (data and workload management): share expertise
- Campus Research Computing Consortium (CaRCC): workforce development

NEON Engagement, October 2018-

- Engagement facilitated by NSF
- Engagement Goals:
  - Increase **Pilot's understanding of NEON's cyberinfrastructure** architecture and operations
  - Increase **NEON's understanding of the Pilot's goals** and expertise
  - **Select & scope mutually beneficial opportunities** to prototype or learn from CI methods
- Engagement Process
  - In-person management meeting
  - NEON shared a number of design documents
  - Team conference calls
  - Meeting with NEON
    - November 2018: Identified topics and formed working groups
    - August 2019: took stock, summarized

Working Groups and Products co-lead by the Pilot and NEON

- Data Life Cycle and Disaster Recovery
- **Data Capture -- Jane Wyngaard**
- Data Processing
- **Data Storage/Curation/Preservation -- Chuck Vanderman**
- **Data Visualization/Dissemination -- Steve Petruzza**
- Identity management
- **Workforce Development -- Rafael Ferreira da Silva**
- Engagement with Large Facilities

CI CoE PILOT

NSF

Tom Gulbransen, Battelle

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

National Ecological Observatory Network Mission

NEON provides a coordinated national system for monitoring critical ecological and environmental properties at multiple spatial and temporal scales.

…transformative science …workforce development








20 ecoclimatic domains distinct landforms, vegetation, climate, and ecosystem dynamics.

Terrestrial sites: terrestrial plants, animals, soil, and the atmosphere,

Aquatic sites: aquatic organisms, sediment and water chemistry, morphology, and hydrology.

Data collection over 30 years

27 Relocatable terrestrial sites
13 Relocatable aquatic sites

NEON Experience

Impact: converted willingness to change into design & strategy commitments

- Sensor data acquisition re-architecture è messaging serialization
- Data Portal identity mgt è Auth0 + CI Logon
- High volume geospatial data use constraints è ViSUS interactivity
- Data discovery metadata FAIRness è new Schema.org subcommittee
-

What enabled conversion from pain to action? First trust, then…

- Expanded awareness of leading-edge methods and community practices
- Affirmations which strengthened confidence of four DevOps teams
- Parallel prototyping in academic settings
- Expanded community participation, i.e. talks, committees, papers

NEON Experience

How are the CI CoE interactions valued?

- Data acquisition labor will become more efficient for ~20 O&M staff, plus 4 developers
- Processing pipeline scientific QA/QC smarter, faster, placed with optimal staff
- Many users initial UX improved
- Unimagined interactive capability to publish remote sensing results, avoided ~3 FTE devs

NEON Experience

Did CI CoE alter our trajectory?

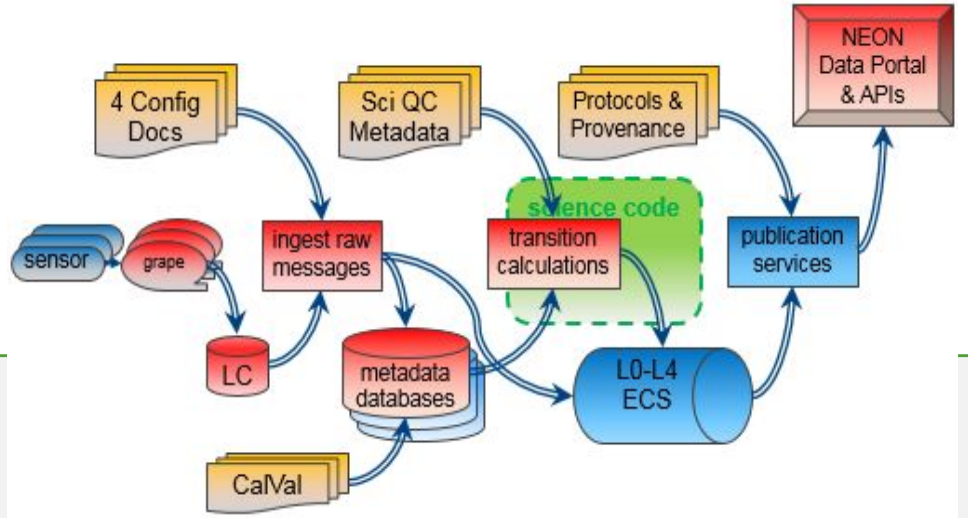
- More aggressive involvement in semantics and ontology coalitions
- Sooner full alignment of schema throughout lifecycle from sensor capture to curation
- Redeployment of developers to priorities other than high-volume geospatial publication
- Increased ability for researchers to lead development of CI components
- Earlier consensus on expectations for disaster recovery preparedness

CI CoE Pilot Benefits to NEON Thus Far

- Short ramp-up due to receptivity/readiness to change
- Broadened network of expert CI colleagues
- Major upgrade to Data Portal's remote sensing visualization
- Accelerated Data Portal completion plan
- Affirmed strategies for workflow, messaging, & DR
- Raised critical mass of attention on semantics & schema.org
- Excited software developers
- Escalated accountability of CI
- More coming

Tom Gulbransen

NEON



Slide courtesy of Tom Gulbransen, NEON

USC Viterbi School of Engineering Information Sciences Institute

UNIVERSITY OF NOTRE DAME

THE UNIVERSITY OF UTAH

CoE Pilot Working Groups

CI CoE PILOT

NSF

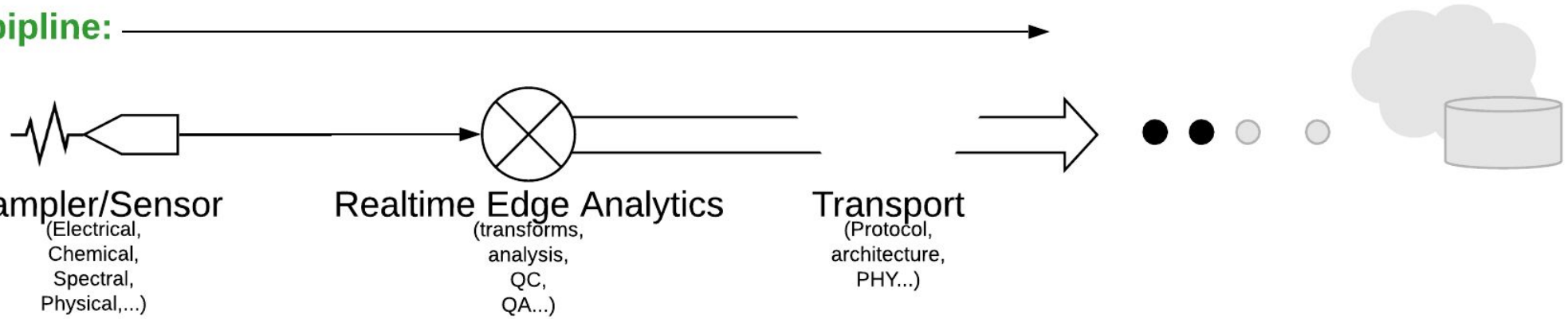
Data Capture

Jane Wyngaard, University of Notre Dame

Data Capture

Understanding and supporting LF data capture challenges at the edge.

Data Capture pipline:



Edge Data Operational Pipeline: Design, Development, Deployment, Monitoring, Upgrading, Error Handling, Retiring

Data Capture

Lessons from challenges faced by NEON

- Operationally managing old and new generation deployments
- Deployments outliving device manufacturer support
- Remote monitoring and upgrade in power/network scare environments
- Harsh environmental conditions
- Corrupt storage devices
- Code porting and maintenance
- High data rates
- Synchronising data streams with unreliable networks
- Barrage of ‘shiny’ new tech (IoT et al)

CI CoE PILOT

NSF

Data Storage/Semantic Annotation

Chuck Vanderman, University of Notre Dame

Data Storage

- Schema.org dataset markup for search discovery
- Evaluation of ESIP science-on-schema.org recommendations relative to NEON data products (Earthcube P418/P419 projects)
- Development of competency questions relative to what domain experts would use as search criteria for dataset discovery
- Development of "Research Site" as a subtype of schema:Place
- Formation of ESIP schemas cluster and geoschemas.org for geo extensions to schema.org
- Connection of schema.org to other ontologies (ENVO, SWEET, SOSA)
- Research Data Alliance schema effort across earth science repositories

CI CoE PILOT

NSF

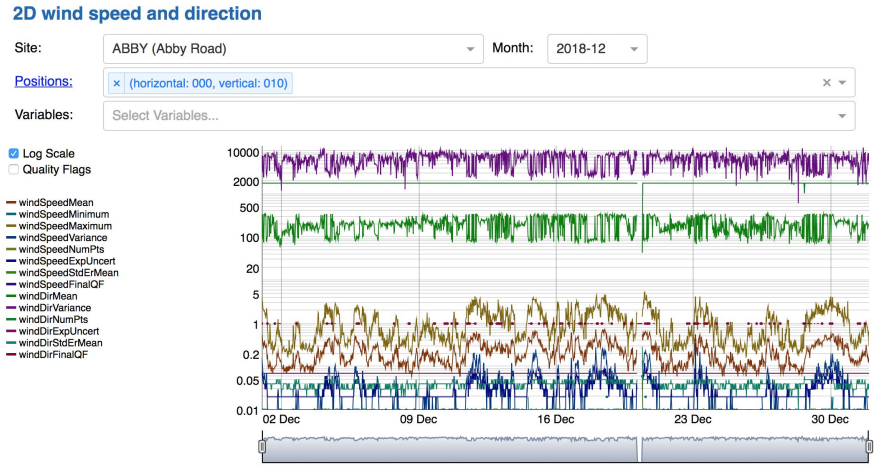
Data Visualization/Presentation

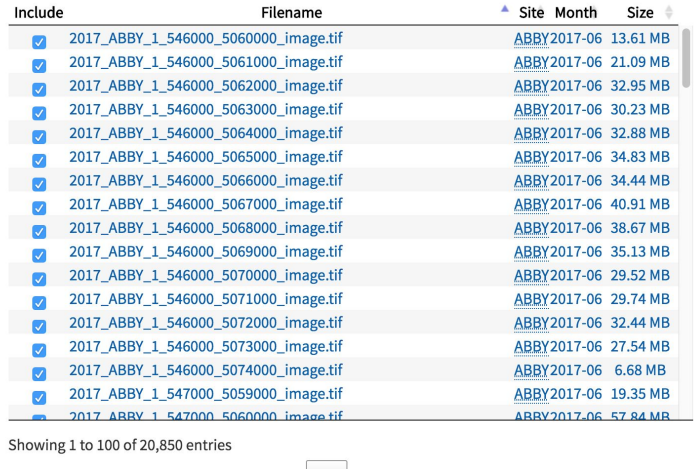
Steve Petruzza, University of Utah

NEON data collections and AOP data

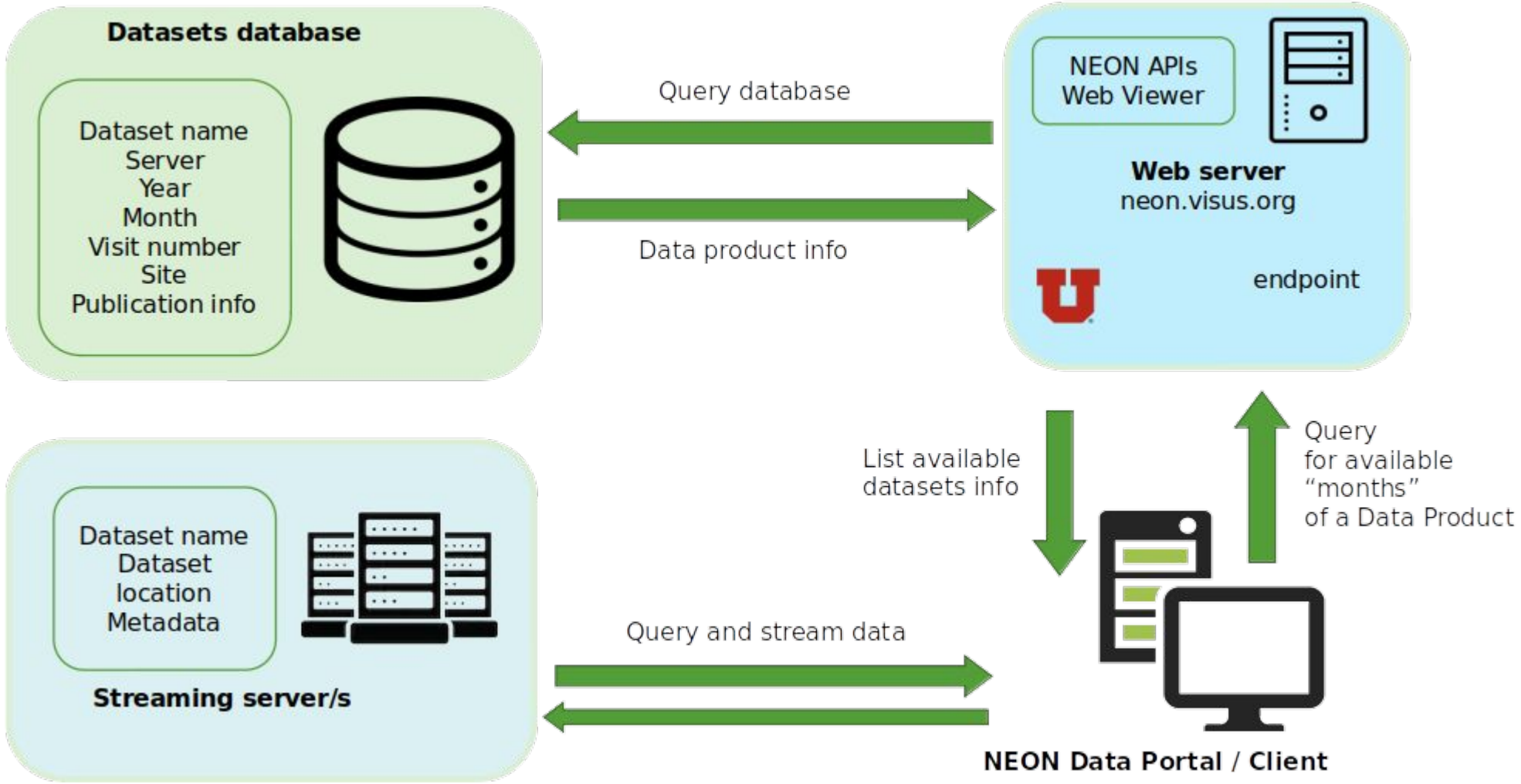
- NEON has a large amount of data that is shared with the community through their **data portal**
- There exist **APIs** to download those data in bulk (per site, per year, per data product, now also by area)
- For some data, such as sensor measurements, the portal provides an **interactive** navigation system
- For others, like **Airborne Observation Platforms** data, there is a long list of image files…

There is a need to present all AOP data interactively, where the users can preview, navigate, and select/access/download the data they need

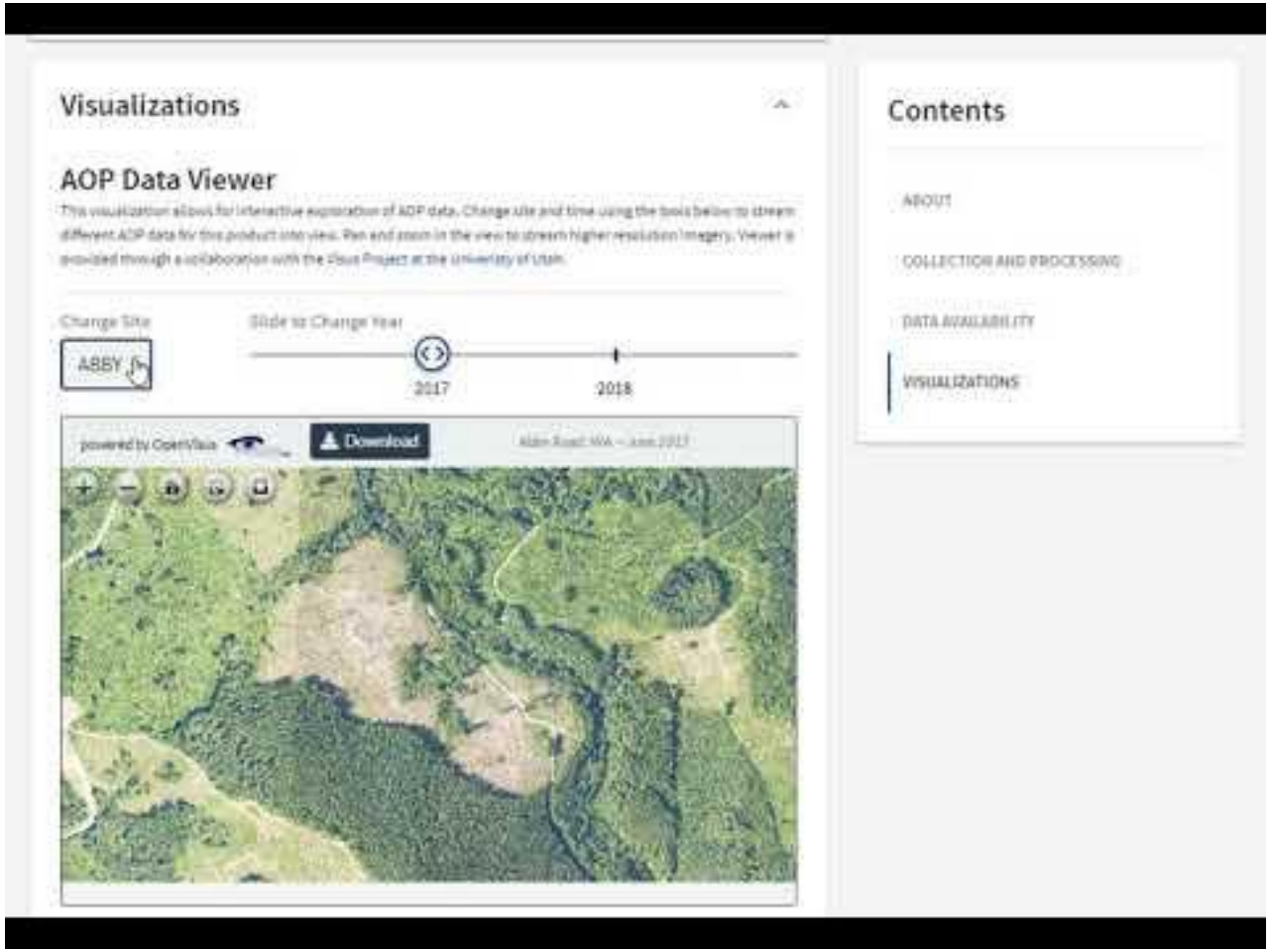




Interactive exploration of NEON AOP data: system Architecture



Integration with NEON Data Portal



https://data.neonscience.org/data-products/DP3.30010.001

References

References

- Live integration:
  https://data.neonscience.org/data-products/DP3.30010.001
- Utah endpoint:
  https://neon.visus.org/neonapi/products/{productCode}
- Endpoint and web viewer source code repository:
  https://github.com/sci-visus/neon-visus
- Documentation and other use cases of the visualization framework:
  www.visus.org

CI CoE PILOT

NSF

Workforce Working Group

Rafael Ferreira da Silva, University of Southern California

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

Workforce Working Group

Challenges

- 2019 NSF Workshop on Connecting Large Facilities (LFs) and CI
  - "LFs face an ever-growing challenge in workforce training, development, and retention"
  - "Competition with industry"
  - "There are no curricula or certifications in the area of CI"
  - "Funding uncertainty, differences between research and industry skillsets, differences between research and industry work environments"

Workforce Working Group

Some of the initiatives concerned with career paths

ACI-REF Virtual Residency	Campus Research Computing Consortium (CaRCC)	Campus Champions	Research Software Alliance (ReSA)
UK Research Software Engineer Association	US Research Software Engineer Association (US-RSE)	SIGHPC Education Chapter	Coalition for Academic Scientific Computation
CyberAmbassadors	Software & Data Carpentry	Science Gateways Community Institute (SGCI)	US Research Software Sustainability Institute Conceptualization (URSSI)

Workforce Working Group

- Goals of the working group
  - **Development, training, retention, career paths, and diversity, as well as the overall career paths for CI-related personnel**
  - Identify and build partnership with current initiatives
  - Keep an up-to-date calendar of activities related to workforce development

Engagement with Large Facilities Future Plans

1. **Developing a blueprint for the CI CoE:**
    a. Community needs
    b. Defining areas of focus
2. Reaching out to other large facilities and the community
3. Discovering opportunities for CI sharing
4. Defining new working groups and discussion topics
    a. Collecting Architectures & Established Best Practices
    b. Collecting Pain Points
    c. Investigating Emerging technologies
    d. Demonstrators?
    e. Training?

Potential CI CoE Functions

- Provide CI discovery and sharing of existing solutions, services, training resources, best practices
- Evaluate new technologies and provide training
- Maintain expertise in specialized areas (e.g., Internet of Things, workflows, data modeling, data archiving)
- Provide assistance in science-driven blueprinting
- Foster communication, collaboration, and community across LFs and CI projects
- Assist facilities in overcoming workforce challenges

https://tinyurl.com/mycicoe

Discussion

- Other Menti

1. What do you think the role of CI CoE should be?
2. What other communities should we know about?
3. What are your pain points?
4. Do you use primarily off the shelf or custom inhouse created solutions?
5. What is the role of commercial entities in your facility's work?

CI CoE PILOT

To find out more

NSF

CI CoE Pilot: http://cicoe-pilot.org

ci-coe-pilot@isi.edu

NEON: https://www.neonscience.org/

2019 NSF Connecting LF CI workshop: http://facilitiesci.org

Share your thoughts: https://tinyurl.com/mycicoe