CI CoE PILOT

Introducing the Cyberinfrastructure Center of Excellence (CI CoE) Pilot

Ewa Deelman (PI) – University of Southern California
Anirban Mandal (co-PI) – RENCI/UNC Chapel Hill
Jaroslaw Nabrzyski (co-PI) – Notre Dame University
Valerio Pascucci (co-PI) – Utah University
Rob Ricci (co-PI) – Utah University

Develop a model and a plan for a Cyberinfrastructure Center of Excellence (CI CoE)
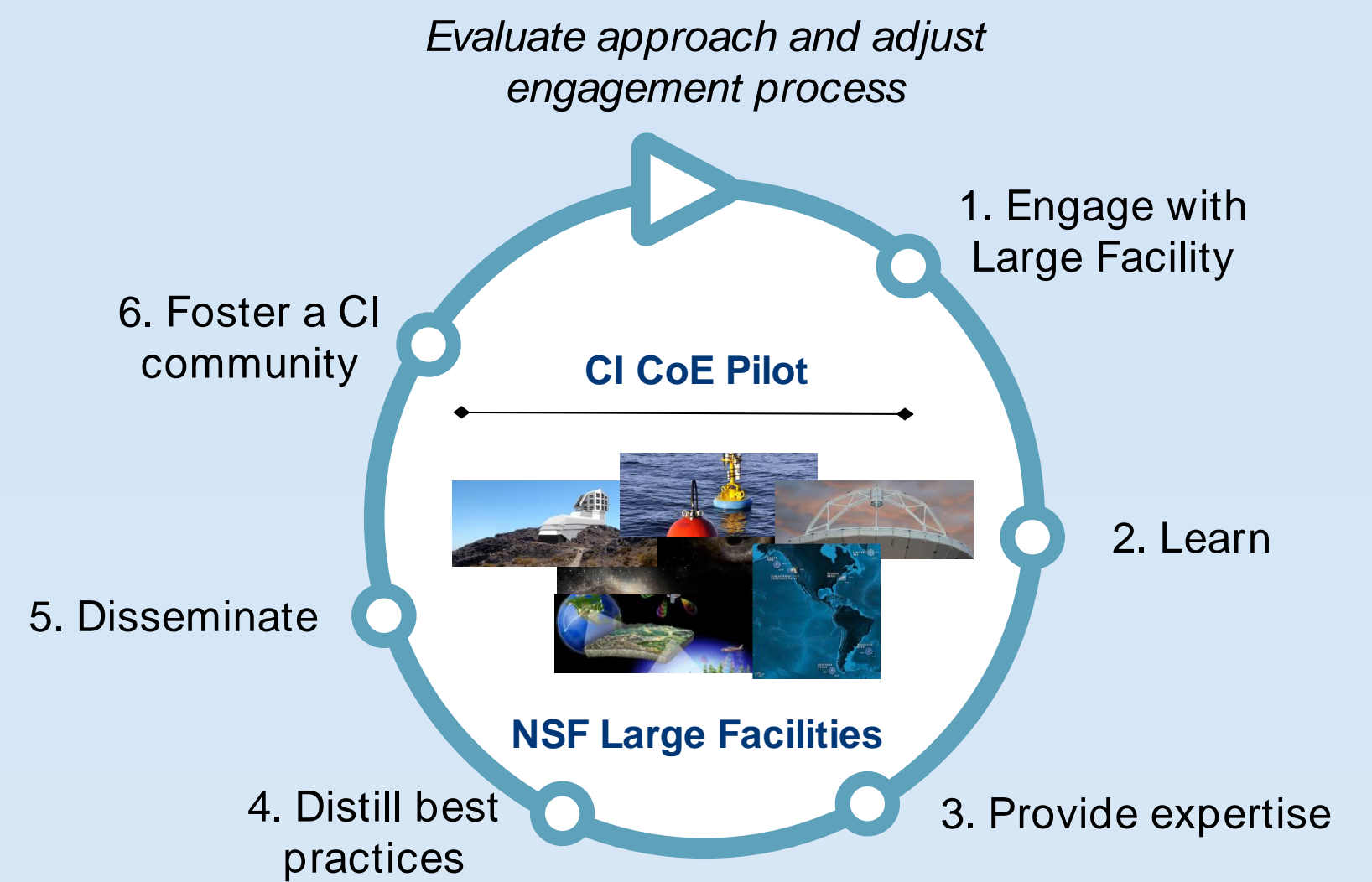
This poster introduces a new NSF Center of Excellence pilot project, with the goal of developing a model for a Cyberinfrastructure Center of Excellence (CI CoE) that facilitates community building and knowledge sharing and applies the understanding of effective practices and innovative solutions for large facility CI. The pilot project explores how such a center would facilitate CI improvements for existing facilities and for the design of new facilities by employing advanced CI architecture designs and leveraging and establishing tools and solutions. The pilot project will also catalyze a key function of an eventual CI CoE – to provide a forum for exchange of experience and knowledge among CI experts. Another area of focus is to gather best practices for large facilities, with the aim of enhancing individual facility CI efforts in the broader CI context. The discussion forum and planning effort for a future CI CoE will also address training and workforce development by expanding the pool of skilled facility CI experts and forging career paths for CI professionals. Initial engagements with NEON, IceCube, and LSST have been focusing on exploring and comparing approaches for identity management, disaster recovery, and data lifecycle management.

Intellectual Merit
- **Amplify NSF investments** in CI in LFs and large CI projects
- **Foster interactions** for technical experience sharing
- **Collect and disseminate** information about modern, robust CI software, services, and platforms to enhance national CI
- **Provide a community forum** for the discussions of social and management issues related to project management and workforce development
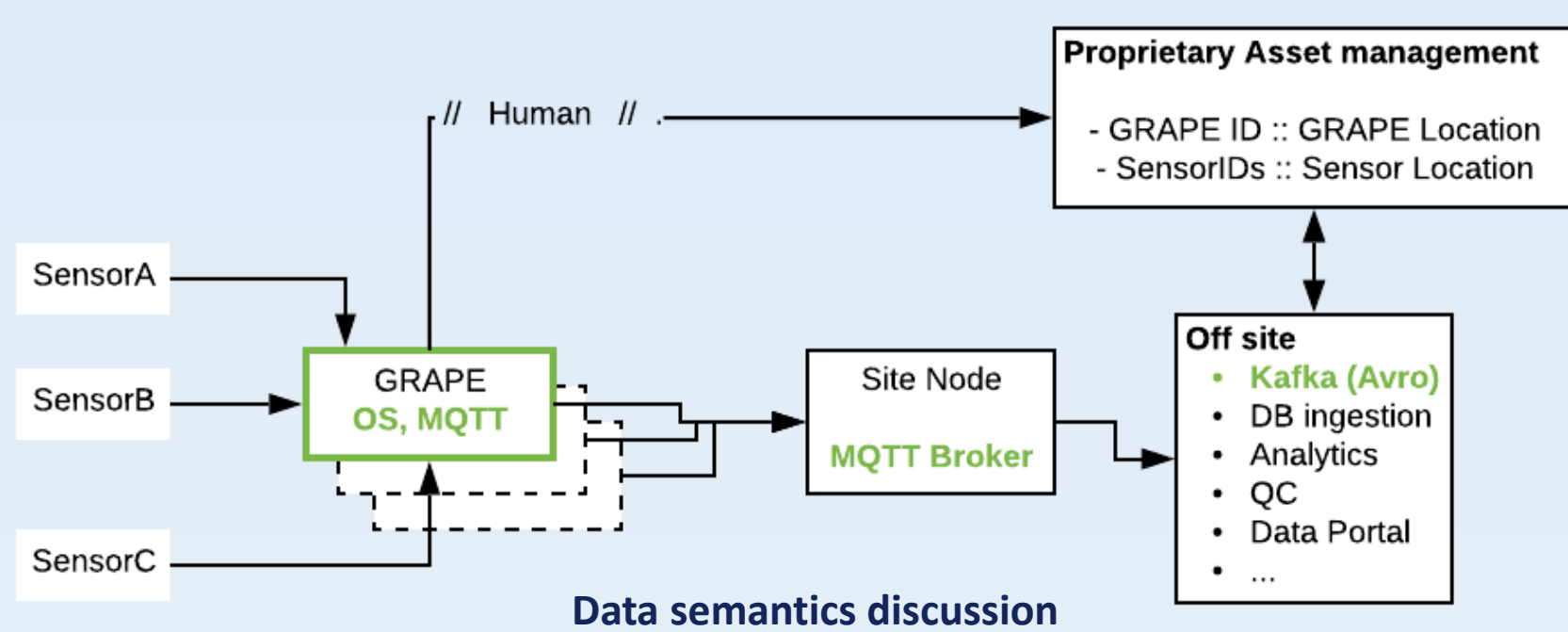
Broader Impacts
- **Cultivate and enhance** the CI workforce
- **Establish** a process to advance the CI and thus the scientific effectiveness of NSF Large Facilities
- **Develop community resources** that benefit other projects that need to discover and deploy CI software, utilize CI services, and leverage CI platforms
- **Improve the CI expertise** of the broad CI community

Evaluate approach and adjust engagement process

CI CoE Pilot
NSF Large Facilities

1. Engage with Large Facility
2. Learn
3. Provide expertise
4. Distill best practices
5. Disseminate
6. Foster a CI community

Working Groups

| Data Capture | Data Processing | Disaster Recovery | Data Storage, Curation and Preservation |
| Data Visualization and Dissemination | Identity Management | Engagement with Large Facilities |

Data Capture Working Group: NEON GRAPE

Proprietary Asset management
- GRAPE ID :: GRAPE Location
- SensorIDs :: Sensor Location

// Human //

SensorA
SensorB
SensorC

GRAPE OS, MQTT

Site Node
MQTT Broker

Off site
- Kafka (Avro)
- DB ingestion
- Analytics
- QC
- Data Portal
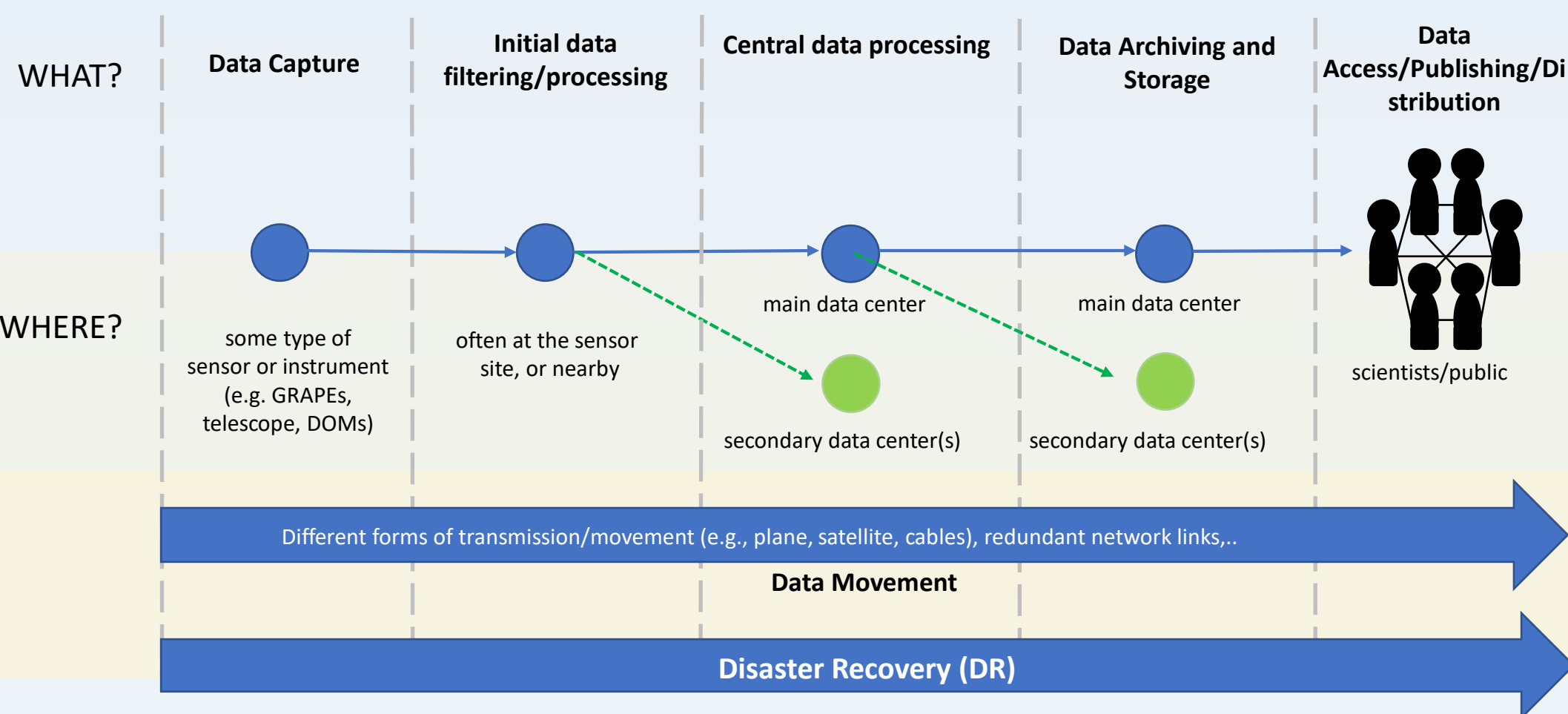- ...

Data semantics discussion

- GRAPES requirements review
- GRAPEs to CI architecture requirements review
- Potential pipeline prototype

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

Disaster Recovery Working Group: Data Lifecycle

WHAT?	Data Capture	Initial data filtering/processing	Central data processing	Data Archiving and Storage	Data Access/Publishing/Distribution
WHERE?	some type of sensor or instrument (e.g. GRAPEs, telescope, DOMs)	often at the sensor site, or nearby	main data center / secondary data center(s)	main data center / secondary data center(s)	scientists/public

Different forms of transmission/movement (e.g., plane, satellite, cables); redundant network links,...
Data Movement

Disaster Recovery (DR)

Data Visualization and Dissemination Working Group

Prototype an interactive web portal for image data visualization/exploration

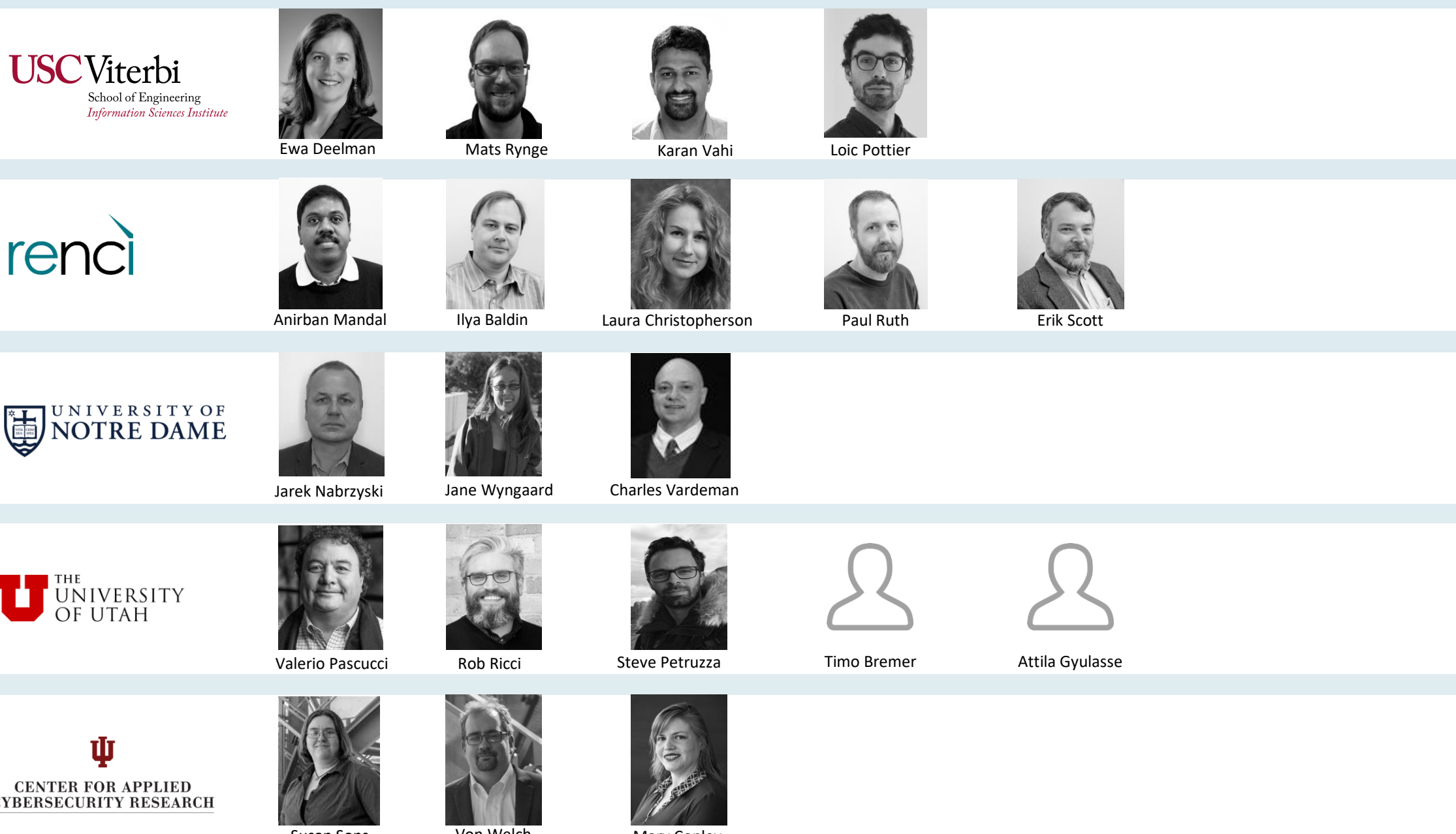Interactive exploration   Multiresolution download   View sharing
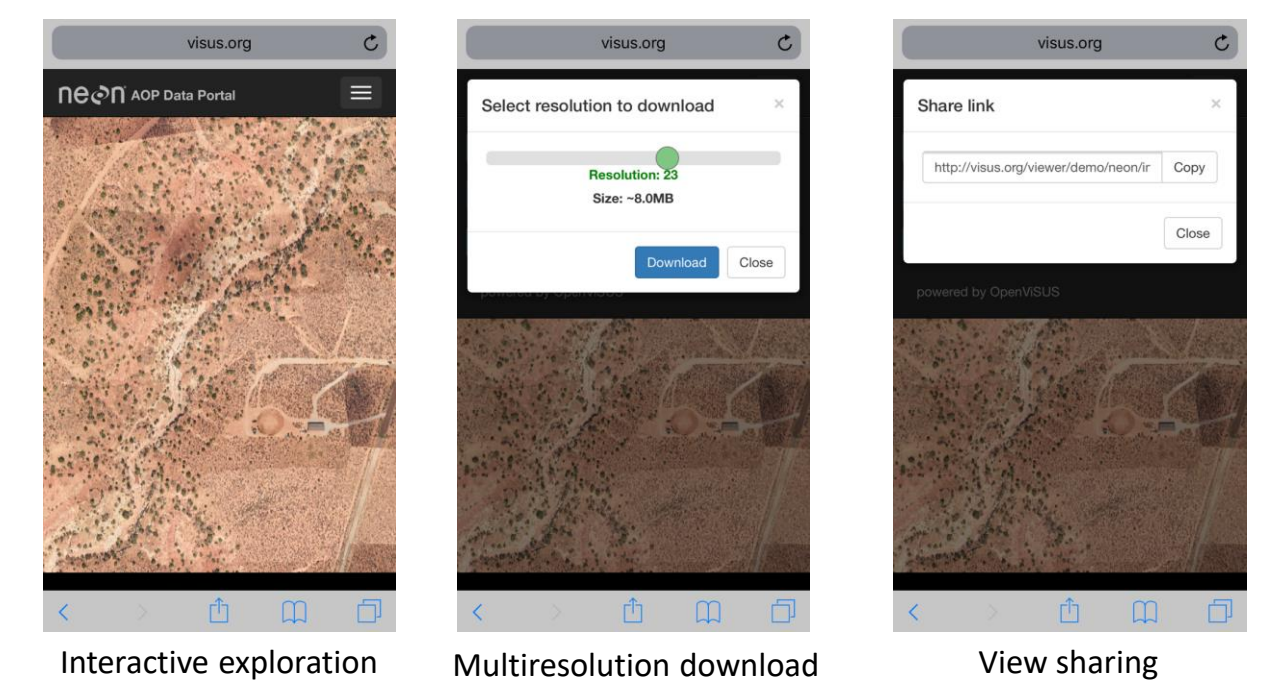
Data Processing Group: Workflow System Comparison

Comparing different workflow system and services for processing of data, for example transforming raw sensor data to more specific data products.

Pachyderm   **Parsl**   **Airflow**   **Makeflow**   **Pegasus**

Meet our Team

USC Viterbi School of Engineering Information Sciences Institute
Ewa Deelman   Mats Rynge   Karan Vahi   Loic Pottier

renci
Anirban Mandal   Ilya Baldin   Laura Christopherson   Paul Ruth   Erik Scott

UNIVERSITY OF NOTRE DAME
Jarek Nabrzyski   Jane Wyngaard   Charles Vardeman

THE UNIVERSITY OF UTAH
Valerio Pascucci   Rob Ricci   Steve Petruzza   Timo Bremer   Attila Gyulasse

CENTER FOR APPLIED CYBERSECURITY RESEARCH
Susan Sons   Von Welch   Mary Conley

TELL US ABOUT

LF workforce development

What is your biggest CI challenge?

What is your data lifecycle?

CI technology evaluation/selection

Your CI success story

Advisory Board
- **Stuart Anderson**, Caltech
- **Pete Beckman**, ANL, Northwestern University
- **Tom Gulbransen**, Battelle
- **Bonnie Hurwitz**, University of Arizona
- **Miron Livny**, University of Wisconsin, Madison
- **Ellen Rathje**, University of Texas at Austin
- **Von Welch**, Indiana University
- **Michael Zentner**, Purdue University