Closing thoughts and next steps: moving forward together

Ewa Deelman, University of Southern California
Thank you for participating

Big Thanks to the Organizers!

Wendy Whitcup
CI CoE Pilot

Todd Stone
ResearchSOC
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
- Currently fostering working group discussions around the science data lifecycle and identity management
- Partnering with other community efforts (TrustedCI, ResearchSOC, SGCI, OSG,..) to support science

http://cicoe-pilot.org/

Award #1842042
10/2018 – 9/2021
USC
Ewa Deelman (PI)
Mats Rynge
Karan Vahi
Loïc Pottier
Rafael Ferreira da Silva
Wendy Whitcup

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

Automation, Resource Management, Workflows, Project Management

Resource Management, Networking, Clouds, Social Science
**University of Notre Dame**
Jarek Nabrzyski (Co-PI)
Jane Wyngaard
Charles Vardeman
Mary Gohsman

**University of Utah**
Valerio Pascucci, Rob Ricci (Co-Pis)
Marina Kogan
Steve Petruzza

**Indiana University**
Angela Murillo

**Trusted CI**
Susan Sons
Josh Drake

**Project Team**

**Workforce development, Sensors, Semantic technologies**

**Data management, visualization, clouds, large-scale CI deployment, Crisis Informatics, Social Computing**

**Cybersecurity**

---

**Data Archiving**
• **CI CoE Mission Statement:**

CI CoE provides leadership, expertise, and active support to cyberinfrastructure practitioners at NSF Large Facilities and throughout the research ecosystem in order to enable ongoing evolution of our technologies, our practices, and our field, ensuring the integrity and effectiveness of the cyberinfrastructure upon which research and discovery depend.
Deep engagement:
- Identify a topic that is important and not-yet fully solved by the LF,
- Conduct focused discussions, mix of virtual and in-person presence, hands-on work
- Includes an engagement template that defines scope, sets expectations, identifies products
- Work products: documents/papers, prototypes, schema implementations, demos

Topical discussions:
- Identify a topic that is important to a number of LFs
- Facilitate virtual discussions, sessions at conferences, collect and share experiences, distill best practices
- Discover opportunities for shared infrastructure

Community building:
- Identify related efforts
- Collect information and disseminate information about the broad community activities
- Maintain a living resource for community information
- Host community activities

Each engagement has a working group with a leader and a set of work products.
<table>
<thead>
<tr>
<th>Working group</th>
<th>Goals</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Capture</td>
<td>Develop demonstrators and comparisons of the multiple architectures for data capture at the sensor to data deposition in a repository</td>
<td>• <strong>Prototype</strong>: architecture demo on github: <a href="https://github.com/cicoe/SensorThingsGost-Balena">https://github.com/cicoe/SensorThingsGost-Balena</a></td>
</tr>
<tr>
<td>Data Life Cycle &amp; Disaster Recovery</td>
<td>Develop a general set of DR requirements and policies that can inform the LFs about best practices for DR and how those can be adapted for specific facilities.</td>
<td>• <strong>Document</strong>: Disaster recovery template</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <strong>Document</strong>: Filled out template example (IceCube)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <strong>Webinar</strong>: Best Practices for NSF Large Facilities: Data Life Cycle and Disaster Recovery Planning</td>
</tr>
<tr>
<td>Data Processing</td>
<td>Provide support and distill best practices for workflows and services related to the processing of data.</td>
<td>• <strong>Paper</strong>: “Exploration of Workflow Management Systems Emerging Features from Users Perspectives”</td>
</tr>
<tr>
<td>Data Storage, Curation, &amp; Preservation</td>
<td>Compare and be able to consult on different data storage, curation and preservation technologies.</td>
<td>• <strong>Document</strong>: Competency questions based on scenarios that domain experts may use Google dataset search for NEON dataset discovery</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <strong>Presentation</strong>: at ESIP on schema.org</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <strong>Small containerized prototype</strong> of publishing neon vocabularies as linked data and linked data connection</td>
</tr>
<tr>
<td>Working group</td>
<td>Goals</td>
<td>Products</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Data Visualization & Dissemination   | Understand the access, visualization and user interaction workflows in large facilities. Distill best practices and provide solutions to improve the access and usability of the available data.                                                                 | • **Document** describing AOP data visualization cyberinfrastructure  
| Identity Management                  | Understand current practice in authentication and authorization and help mature practice across the NSF Large Facilities.                                                                                                                                         | • **Production deployment**: Connection to CI Logon NEON data download (using existing university / organization credentials) [https://cert-data.neonscience.org/home](https://cert-data.neonscience.org/home)  
• **Paper**: [NEON IdM Experiences](https://cert-data.neonscience.org/home) (NSF Cybersecurity Summit) |
| Engagement with Large Facilities     | Engage with Large Facilities and other large cyberinfrastructure projects to foster knowledge and effective practice sharing; 2) define avenues of engagement, modes of engagement, and plan community activities.                                                   | • **Document**: LF engagement template  
• **Presentations**: SCIMMA project meeting, 2019 LF meeting, PEARC’19, LF CI Workshop, Cybersecurity Summit’19, AGU’19  
• **Paper**: Invited e-Science 2019 paper |

Contact: Ewa Deelman, [deelman@isi.edu](mailto:deelman@isi.edu)
Engagement with NEON and NCAR

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

CI CoE Pilot Engagement

- Consult on cloud technologies, including data, containers, etc.
- Helped inform NEON/NCAR’s proposal to NSF
- Expect to engage further if/when funded

2020
**Goal:** Provide CI expertise on a Common Cloud Platform (CCP) to pilot the migration of their collective data resources and services from individual in-house solution to common Cloud

**Engagement**
- Embedded in a number of CCP working groups
- Provided high-level schematics and guidance for contents, organization and structuring of the CCP high-level requirements draft
- Presented existing best practices for designing a ConOps document
- Discussed CI best practices for various aspects of cloud architecture design for CCP
- Drafted a companion document to inform CPP platform design (cloud migration considerations for data, resource orchestration, cloud storage, messaging, supporting FAIR data principles, Identity Management (in collaboration with Trusted CI)

**SAGE:** Seismological Facilities for the Advancement of Geoscience
**GAGE:** Geodetic Facilities for the Advancement of Geoscience

2020
In collaboration with Trusted CI

- Monthly meetings with speakers and discussions on topics relevant to LFs: e.g. CILogon
- Engagements, primarily focusing on federated identity management
  - Issues of identifying data usage and enabling reporting
  - ARF and GAGE, exploring solutions and developing demonstrations
  - SCIMMA, contributing to their Identity Access Management prototype
Trustworthy Data Working Group

Aims to provide guidance on data security for open science, to improve scientific productivity and trust in scientific results.

Open science relies on data integrity, collaboration, high performance computing, and scalable tools to achieve results, but currently lacks effective cybersecurity programs that address the trustworthiness of scientific data.

Community Survey: Scientific Data Security Concerns and Practices
- 111 participants
- Report available: https://doi.org/10.5281/zenodo.3906865

PEARC’20 Workshop on Trustworthy Scientific Cyberinfrastructure

Next: creating a “Guidance for Science Projects and Cyberinfrastructure Developers” document
CI CoE Pilot: Connecting CI, Connecting Science, Connecting People

- CI Practitioners
- CI Facilitators
- Computer Scientists
- Domain Scientists

- Workshop on CI for Large Facilities
- We are thinking of hosting it every 2 years with smaller thematically focused events in the interim
  - Community building events with ResearchSOC, TrustedCI, SGCI

Understanding the impact of Covid-19 on NSF Large Facilities
Kerk Kee, Texas Tech University

http://cicoe-pilot.org

Supporting related projects
NSF workshops on CI for Large Facilities (2017, 2019)

http://facilitiesci.org

NSF workshop on Smart CI (2020)

http://smartci.sci.utah.edu/
Questions about Next Steps

- Very short survey (<5 mins), input for the future
- Is there some CS/CI training you would like to see in the future?
- Would it make sense to have periodic office hours/forums and if so what topics? Clouds
- Is this format useful, what other community activities, working groups, would you like to see?
- Did you like the 1 day on 1 day off mode?

[https://cics-workshop.org/](https://cics-workshop.org/) Slack channels will remain open