U.S. DEPARTMENT OF ENERGY | Office of Science

Federating DOE/SC Facilities

Phase 1: Federated Identity Management

Richard Carlson ASCR Program Manager

Richard.Carlson@science.doe.gov

ASCAC 14 January, 2020

Executive Summary

- ASCR has a long history of conducting research and supporting operations in Middleware, Grid, and higher level Services to form Distributed Science Infrastructures
- Operation of these infrastructures has been historically been performed by an individual Science domain (i.e., ESG - Climate, LHC – High Energy Particle Physics)
- A Pilot project built upon the success of the Future Lab Computing – Working Group to pilot the use of laboratory resources using a federated Identity service to access those resources
- Federating DOE/SC facilities as they continue to generate, process, analyze, and archive more data will significantly increase the value and usability of those facilities

Outline

- **Future Lab Computing background**
- **DCDE Pilot**
- **SC19 Demo**
- **Federated IdM across the lab**
- **Building for the future**
  - Past federations
  - Supporting 100x more users
  - Expected scientist skill set
- **Distributed facility Challenges**
  - The E2E mantra
  - Multiple Organizational Domains
  - Performance Tuning now and in the future
  - Operational Debugging now and in the future

From Grids to Clouds to Today

- **1999: MICS (ASCR) funded 3 pilot Grid computing projects**
  - Earth Systems Grid (ESG) – Distribution of Climate Models and Simulations
  - Particle Physics Data Grid – Distribution of LHC data to U.S. physicists
  - Fusion Grid – Remote control room for tokamak scientists
- **2009: ASCR funded a Software as a Service project**
  - Globus on-line – Hide complexity of GridFTP file transfers with a SaaS model
- **2017: ASCR pushes Federation of Lab Computing Resources**
  - Future Lab Computing – Working Group – Work with Lab staff to understand how to federate lab computers
  - Distributed Computing and Data Ecosystem (DCDE) – pilot project to demonstrate federation of lab computers

FLC – WG Report

- DOE/SC Laboratories provide computing/storage resources to lab staff, researchers, and visiting scientists
- Demands on these resources are increasing
- Labs have the capability to leverage decades of research to create modern Distributed Computing and Data Ecosystems (DCDE) to meet the current and future demands of DOE scientists
- ASCR constituted Future Laboratory Computing Working Group (FLC-WG). Met through 2018 and delivered report with findings.
- DCDE pilot established for FY2019 fleshes out the key components and documents procedures to establish the infrastructure.

FLC Working group report (2018): *Background and Roadmap for a Distributed Computing and Data Ecosystem*, https://doi.org/10.2172/1528707

DCDE Pilot – The Art of the Possible

- **Funded staff at ANL, BNL, LBNL, ORNL, and EMSL**
  - Goal is to deploy, not develop, existing tools and services
- **Services used:**
  - AuthN/AuthZ: InCommon, CILogon and COManage
  - Globus and auth-ssh
  - Application and Containers
  - Jupyter notebook and Parsl workflow
- **Issues, Challenges and Lessons**

- ***These Slides were taken from the DCDE team, particularly David Cowley (PNNL/EMSL)***

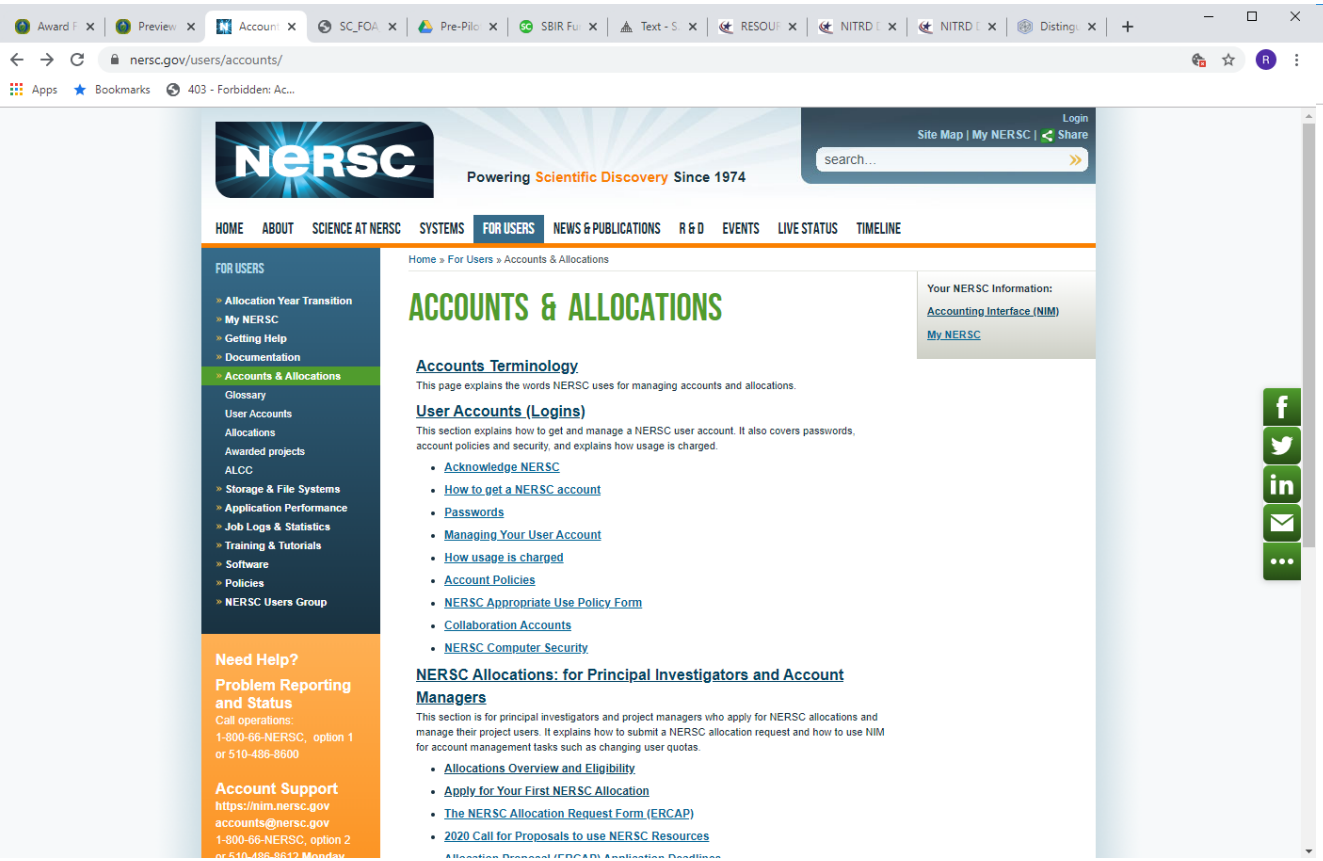
Identity Management Fundamentals

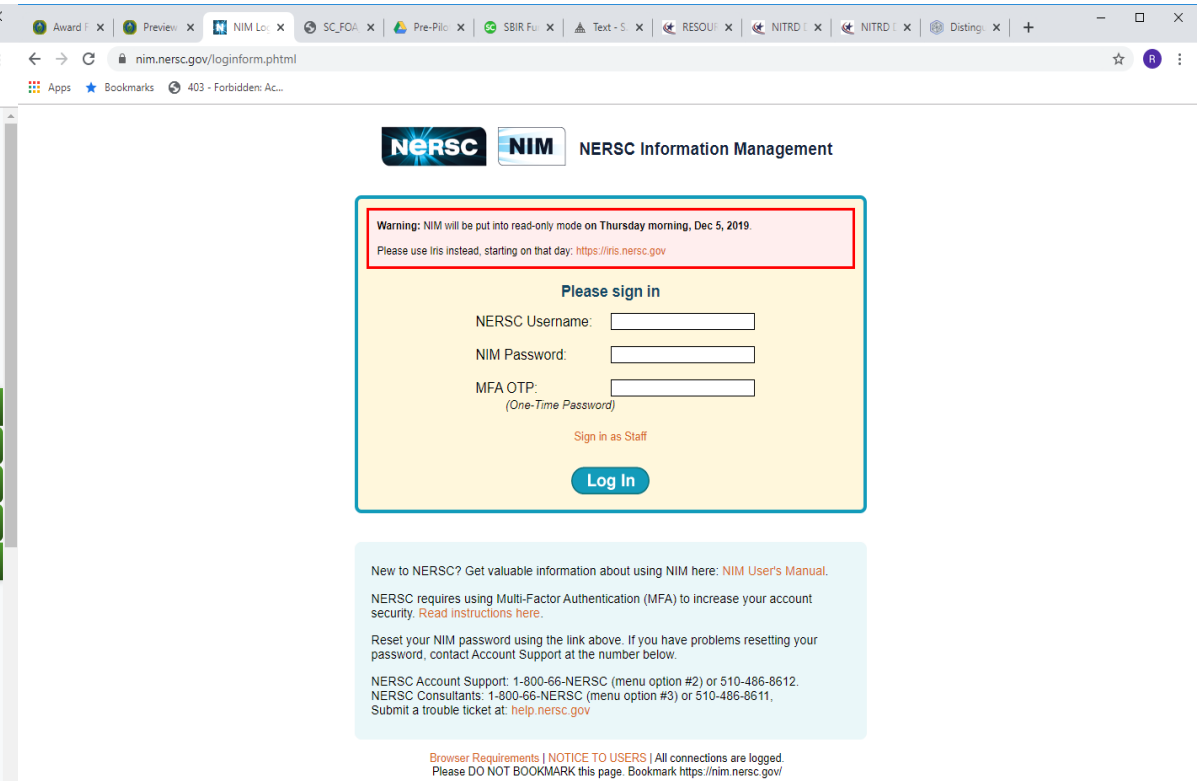
- **Authentication (AuthN): The Identity of an individual as defined by a username and some kind of password. An Identity Provider (IDP) server contains the login information and returns an identity token after a successful login**
- **Authorization (AuthZ): A Service Provider (SP) server controls access to a specific resource (i.e., computer, instrument) by accepting an identity token and a possible set of attributes.**

Sample list of Account Attributes

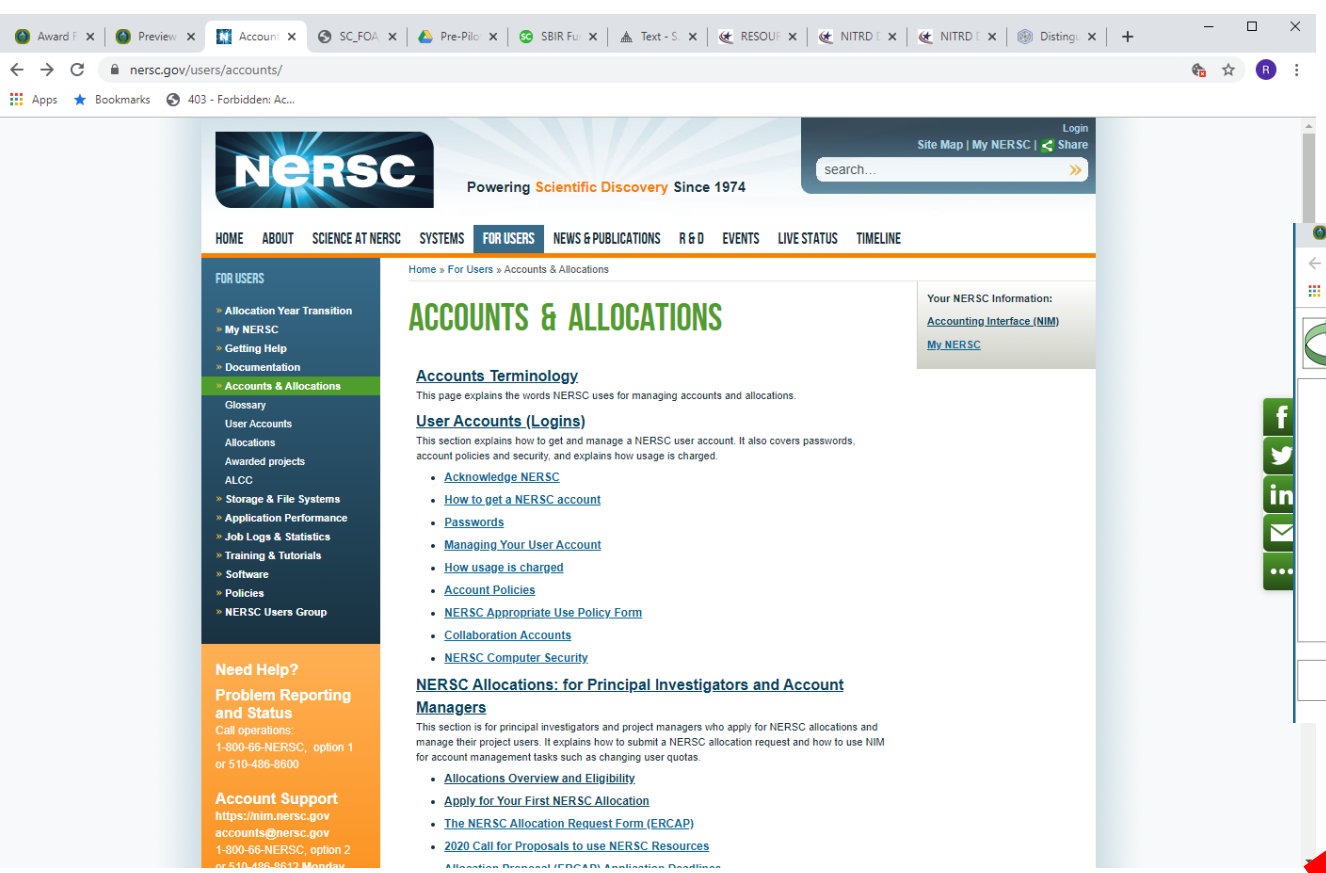
Information required to get a computing account for non lab employees	ANL	BNL	LBNL	ORNL	PNNL
Site ID number		TRUE			TRUE
Online Cybersecurity training	TRUE	TRUE			TRUE
Online Computer use agreement	TRUE	TRUE		TRUE	TRUE
First Name	TRUE	TRUE		TRUE	TRUE
Last Name	TRUE	TRUE		TRUE	TRUE
DoB	TRUE	TRUE			TRUE
Citizenship	TRUE	TRUE		TRUE	TRUE
SSN					TRUE
email	TRUE	TRUE		TRUE	TRUE

Example: Accessing NERSC's NIM





Example: Accessing NERSC's NIM



Bridged with CILogon + COManage

Also usable for SSO to in-campus (in-lab for us) resources, without obvious redirect to external login page.

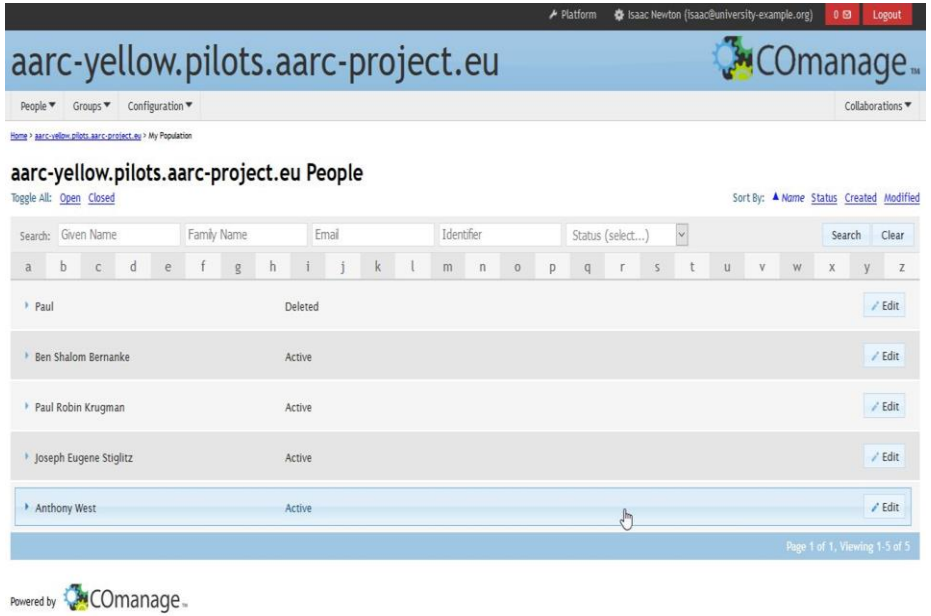
Allows usage of non-InCommon auth sources (Google, Github, any OAUTH2).

Membership lists, attributes, and enrollment and lifecycle management via (hosted) COManage.

- **NSF/Internet2 project**
- **SAML attribute provider**
- **Handles registration workflow**

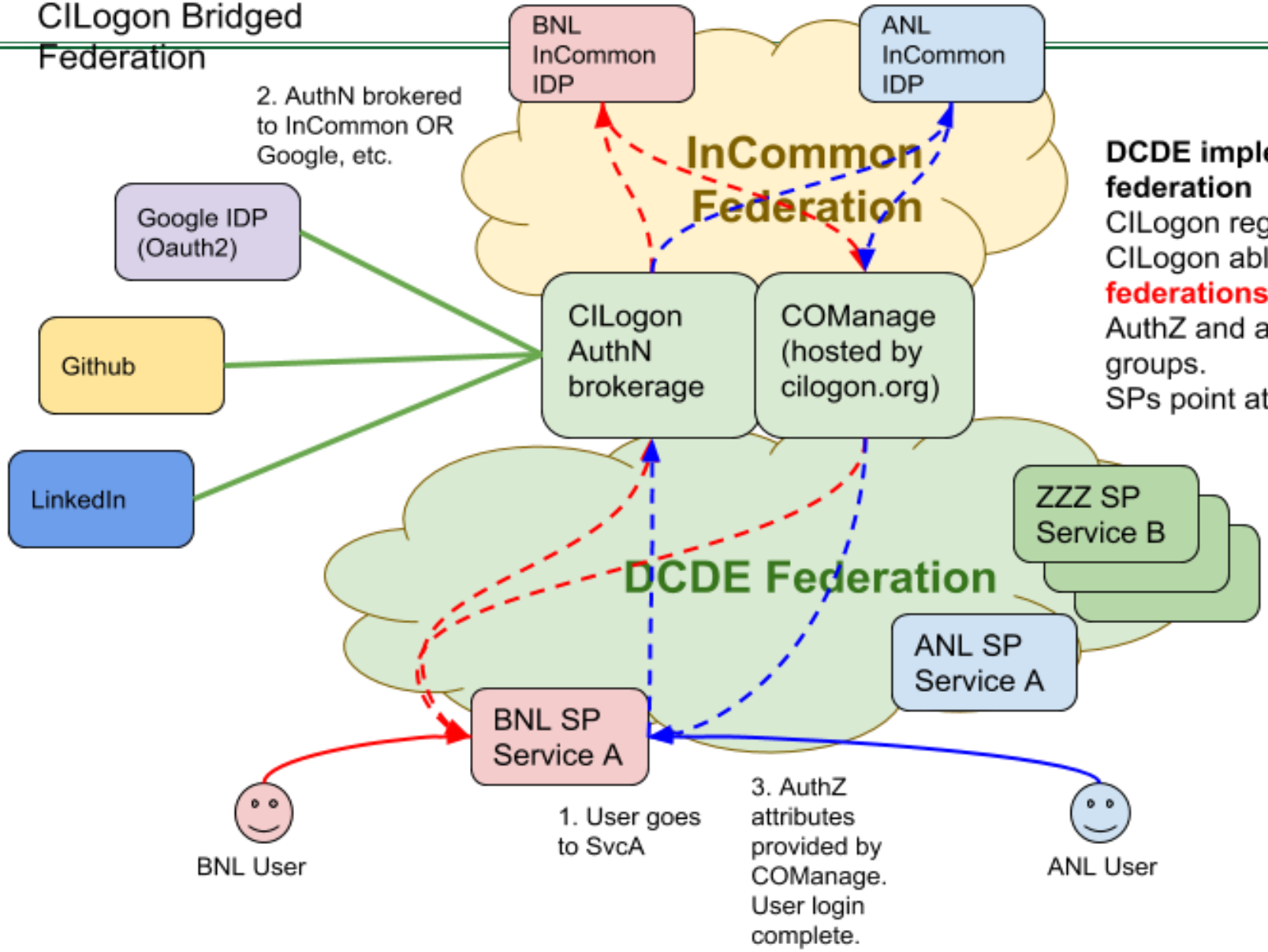
https://incommon.org/docs/iamonline/20180117_IAMOnline.pdf

CILogon provides interface for registering/managing services.



This approach doesn't require any DCDE-specific infrastructure--all hosted by cilogon.org.

CILogon Bridged Federation



DCDE implemented as CILogon bridged federation

CILogon registered as InCommon ***SP***

CILogon able to broker AuthN to **other ID federations** like Google, Github, Microsoft, etc.

AuthZ and allowed AuthN managed in KeyCloak groups.

SPs point at CILogon for AuthN.

InCommon

- **InCommon Federation chosen as identity platform**
- **Participation from multiple DOE labs**
  - ANL, BNL, LBNL, JLab, ORNL
- **Users from participating labs can authenticate themselves using their lab credentials**
- **InCommon provides only SAML standard but no Oauth (which is a bit of a problem)**
  - DOE HPC and General Purpose compute systems have limited support for SAML

CILogon

- **Authentication hub for DCDE**
- **Serves as a proxy/broker service linked to Incommon**
  - Able to translate SAML to Oauth
- **Also provides X509 certificate service that is useful for integration with some services (eg. gsissh, gridftp)**

COManage

- **COManage service is integrated with CILogon**

- **It provides a web-portal like platform**
  - for users to self-register (optional)
  - to manage and federate user attributes from multiple sources
  - admins to create groups and sub-groups
  - admins to manage project and user account lifecycle

AuthN/AuthZ: Participating Site Roles

- A site admin for each site
- Provide access to resources for DCDE project
- Provide a site entry point gateway "host" and a batch scheduler
- Obtain registered users' DN from COManage admin
- Create appropriate "mapfiles" mapping the user DNs to local site accounts
- Create local accounts and groups as appropriate to the local site policies

Globus and oauth-ssh

- **Globus ssh provides the ssh over oauth**
  - users can ssh to sites via oauth-ssh
- **Globus transfer is used for data transfer purposes between participating sites**
- **Technical requirements for oauth-ssh**
  - Choose a port (we chose 2222) for each site to be opened
  - Update DNS at site to add a text metadata record for the site gateway host for oauth-ssh to authenticate site using nslookup

Application and Containers

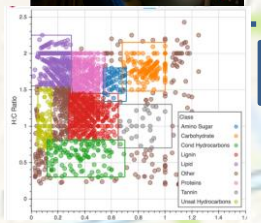
- **We chose a microscopy application called Relion for prototyping purposes**
- **The application is containerized using Singularity**
- **Each site runs the same container and same version of Singularity for ease of portability and troubleshooting**

Jupyter and Parsl

- **Jupyter was integrated into the DCDE project whereby users can sign into the Jupyter web-interface using their DCDE credentials**
- **Jupyter's capability of custom authentication was linked with the CILogon interface**
- **CILogin OAuth tokens are available within the Jupyter notebooks to provide seamless authentication across the DCDE resources**
- **Parsl is chosen as the workflow platform**
  - Written in Python -- a python package
  - Natural for Jupyter
  - Well integrated with Globus and oauth

Distributed Computing and Data Ecosystem (DCDE) Demo Overview

1. Cowley@PNNL



Jupyter Hub

Parsl Invocation

2. Dong@BNL

CoManage Federated ID Mapping

Jupyter Hub

Parsl Invocation

3. Murphy-Olson@ANL

Jupyter Hub

Parsl Invocation

globus

5. Chawla@LBNL

Jupyter Hub

Parsl Invocation

4. Maheshwari@ORNL

Jupyter Hub

Parsl Invocation

Demonstration Components

- **Science Driver**
- **Federated Identity Management**
- **Portability across laboratories**
- **Workflow through analytic notebooks**
- **Data Transfer**

Try very hard to not reinvent anything: use available technologies and capabilities!

Thank you!

Acknowledgment: DOE SC/ASCR

Laboratories: LCRC@ANL, SDCC@BNL, LRC@LBNL, CADES@ORNL, EMSL@PNNL

Issues, Challenges and Lessons

- **Some learning curve for users -- third-party auth, oauth-ssh, parsl, etc.**
  - An approach is to provide a templated solutions to common user issues
- **Site admin overhead eg. firewalls management, installation and configuration of oauth stack, jupyterhub etc.**
  - Scripted several install steps

Pilot to Production

- **Federated Identity Management across the SC lab complex**
  - Generate a production level Federated IdM service based on pilot labs
  - Integrate ASCR facilities into this federation
  - Integrate other SC labs into this federation
  - Integrate other SC facilities into this federation
- **Resolve open policy issues**
  - What attributes are required by a Resource Provider?
  - How will Federated IDs map to local accounts (multiple options)
- **Delay decision on implementing a workflow service**

The Future's so Bright

- **Associate Professor at HBCU**

The Future's so Bright

- **Send sample to Light Source Facility**

The Future's so Bright

- **Real-Time verification of data**

The Future's so Bright

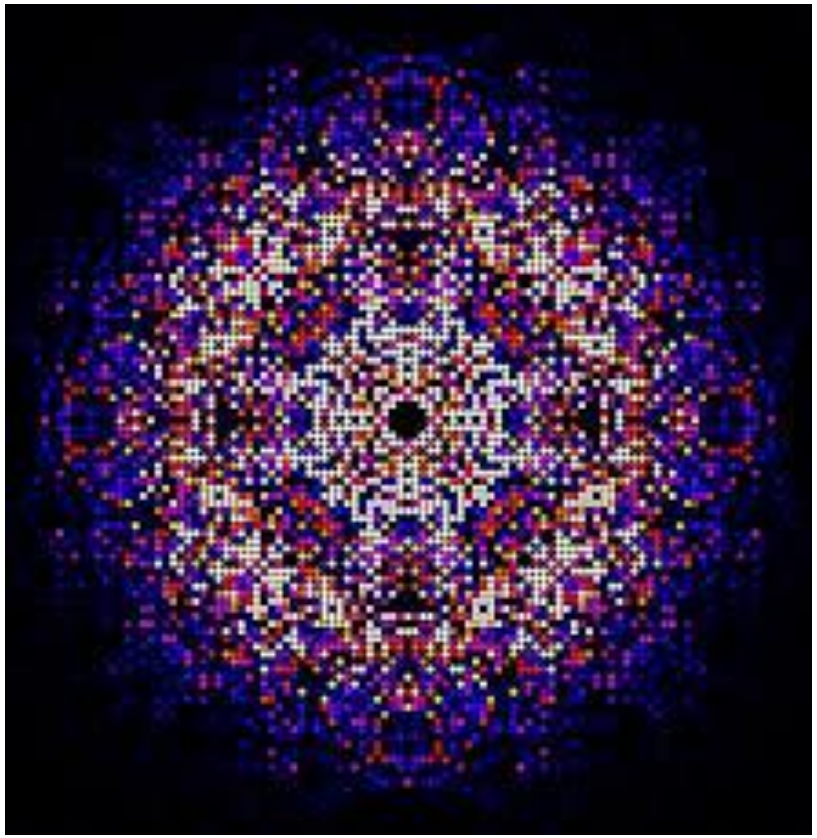
- **Data Collection**

The Future's so Bright

- **Data processing**

The Future's so Bright

- **Image Processing and Discovery**



The Future's so Bright

- **Discovery and Recognition**

Conclusion

- **Federated Identity Management is a key enabling service to foster scientific discovery**
- **The DCDE pilot project demonstrated that IdM services are ready for full scale deployment within the DOE/SC lab complex**
- **While some policy and trust issues need to be resolved, there are significant benefits to creating and using a federated IdM service**