



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

# **CESD Cyberinfrastructure Working Groups**

**ESS-PI**

**April 24, 2017**

## ***Data Management***

**Leads: Kirsten Hofmockel**

**Danielle Christianson** (acting lead for **Charuleka Varadharajan**)

**Team Members: Deb Agarwal, Tom Boden, Paul Hanson, Trevor Keenan, Terri Killeffer, Lara Kueppers, Dave Millard, Eric Pierce, Gilberto Pastorello, Daniel Ricciuto, James Stegen, Margaret Torn, Roelof Versteeg, Anthony Walker, Dali Wang, Chongang Xu**

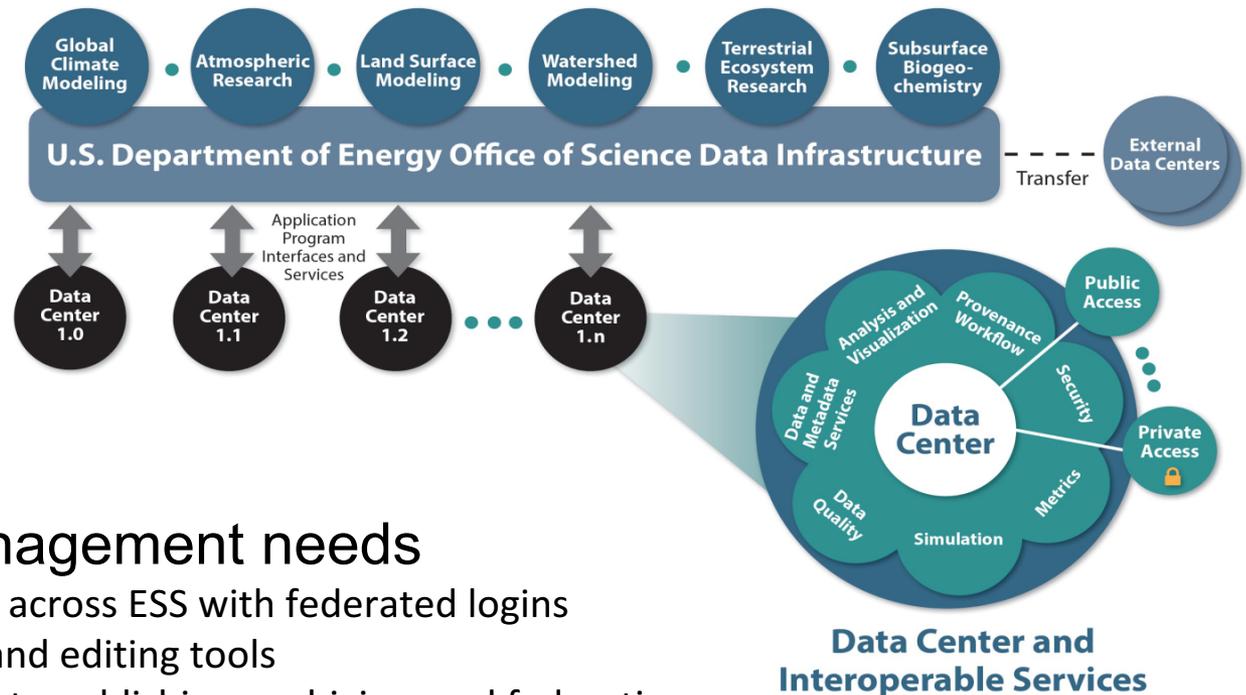
# Enable data practices for science outcomes

---

- **Management and Archival** of DOE climate and environmental datasets
  - Initial focus on large Earth System Science (ESS) projects  
e.g. NGEES, SFAs, AmeriFlux, SPRUCE
- Common Data and Metadata **Standards**
- Data **Sharing, Distribution, and Publication**
- Data **Citation and Attribution**
- Data **Federation** across various ESS datasets
- Development of common **Tools** for QA/QC, processing, analysis, mining and visualization data to prepare them for use in new research projects.

# Support DOE (CESD/ESS) Data Infrastructure

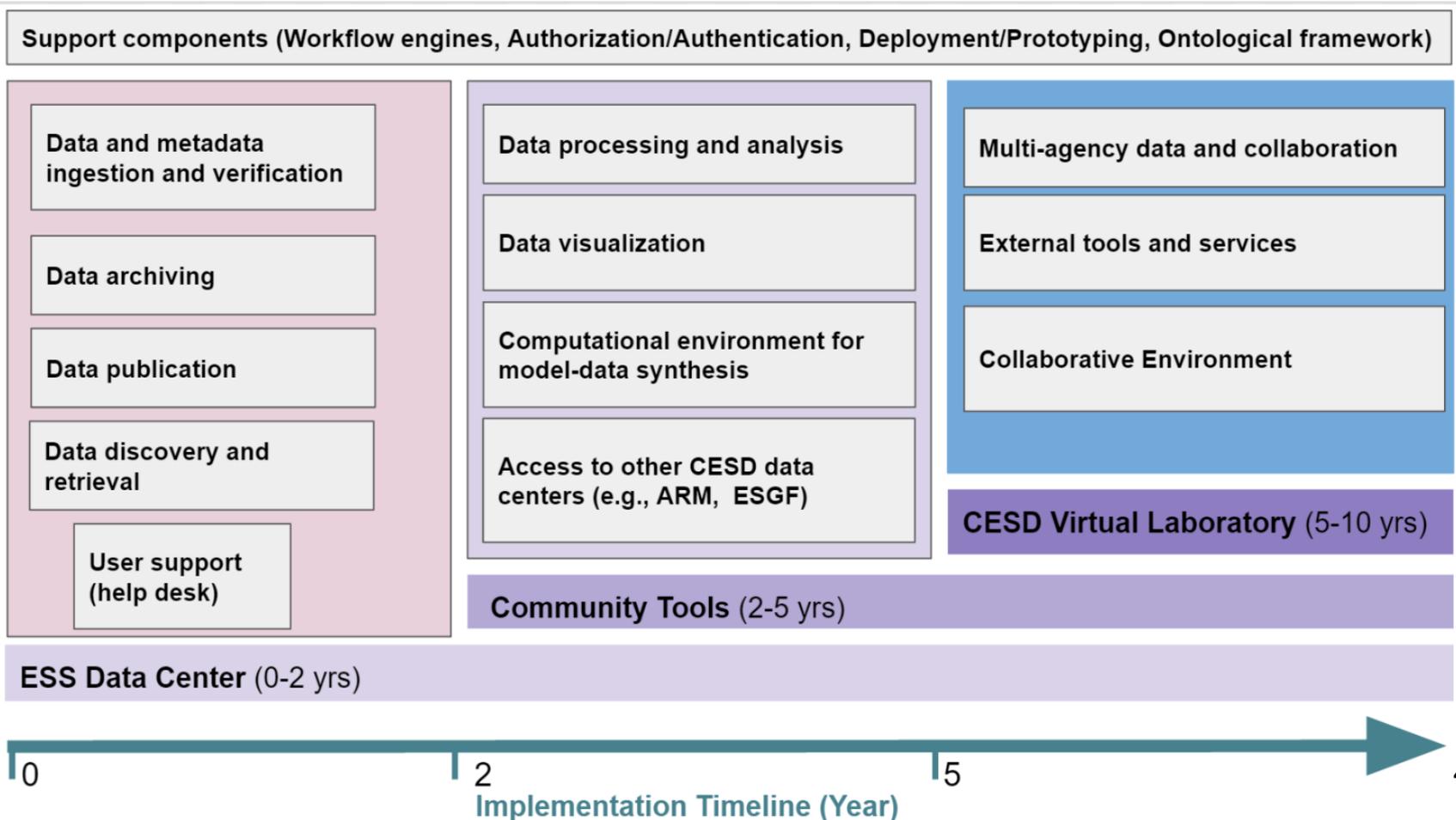
- Working group supports CESD/ESS data infrastructure with input from community via face-face meetings and conference calls
- “Data” = Observations, Analysis products, Model data, Publications



- Identified data management needs
  - Common data repository across ESS with federated logins
  - Metadata development and editing tools
  - Unified procedures for data publishing, archiving, and federation
  - Established data policies and data citation/attribution mechanisms

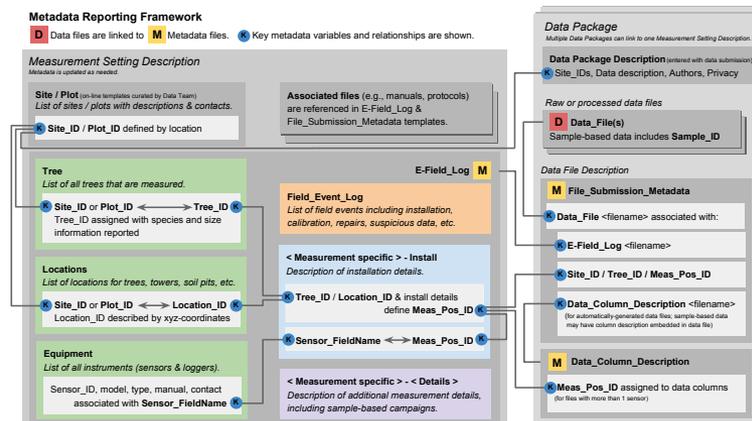
# Long-Term Goals

- Shared database to host DOE data (versioning and provenance)
- Enhanced search for DOE and external datasets
- Tools for analysis, modeling, and creating value added products
- Domain-specific metadata templates with web-based interfaces



# Short-Term Goals

- Develop workflow to capture and serve publication data that meets DOE digital data requirements
- Review of methods for uniquely identifying electronic data content (e.g. DOIs)
- Define publication packages
- Centralized DOE single sign-on
- Standardized sample tracking
- Identify best practices for metadata



# Breakout: ESS Data Center Interaction/Support

- What areas of **standardization** does the DM-WG want to develop with the Data Center?  
E.g., data policies, metadata, citation / attribution (DOI)
- How can the DM-WG **support the ESS community** data providers in complying with standards / policies of the Data Center? And data users in using data provided via these standards / policies?  
E.g., QA/QC resources, tools for analysis & visualization

Activities (1.5 hours total):

20 minutes: Brainstorm questions — augment / expand past areas of interest

10 minutes: Prioritize areas of interest — confirm / update past priorities

45 minutes (15 min each for top 3 interest areas):

-- Brainstorm resources, DOE project connections

-- Discuss how interaction with ESS / WG community might happen OR how to structure development

-- Identify leads, interested members

15 minutes: Wrap up to discuss next steps / action items / report