

ESS-DIVE Developed WITH the DOE ESS Community

Deb Agarwal

Charuleka Varadharajan, Shreyas Cholia, Cory Snavely,
Valerie Hendrix, Fianna O'Brien, Abdelrahman
Elbashandy, Yeongshnn Ong, William Riley, Chris
Jones, Matt Jones, and Karen Whitenack



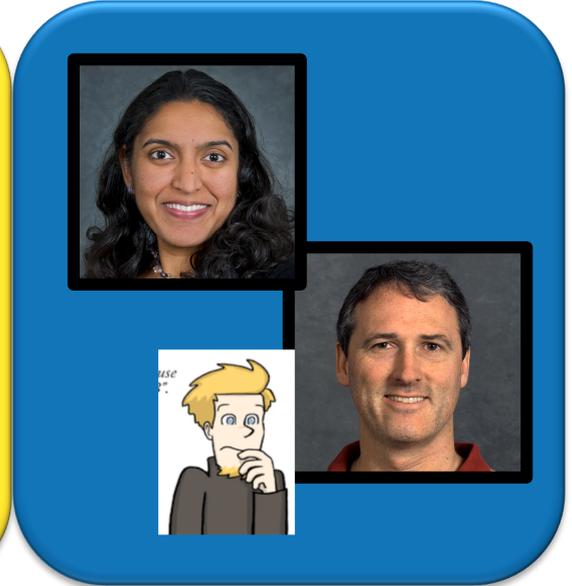
The ESS-DIVE Team



**Data Scientists and
Software engineers**

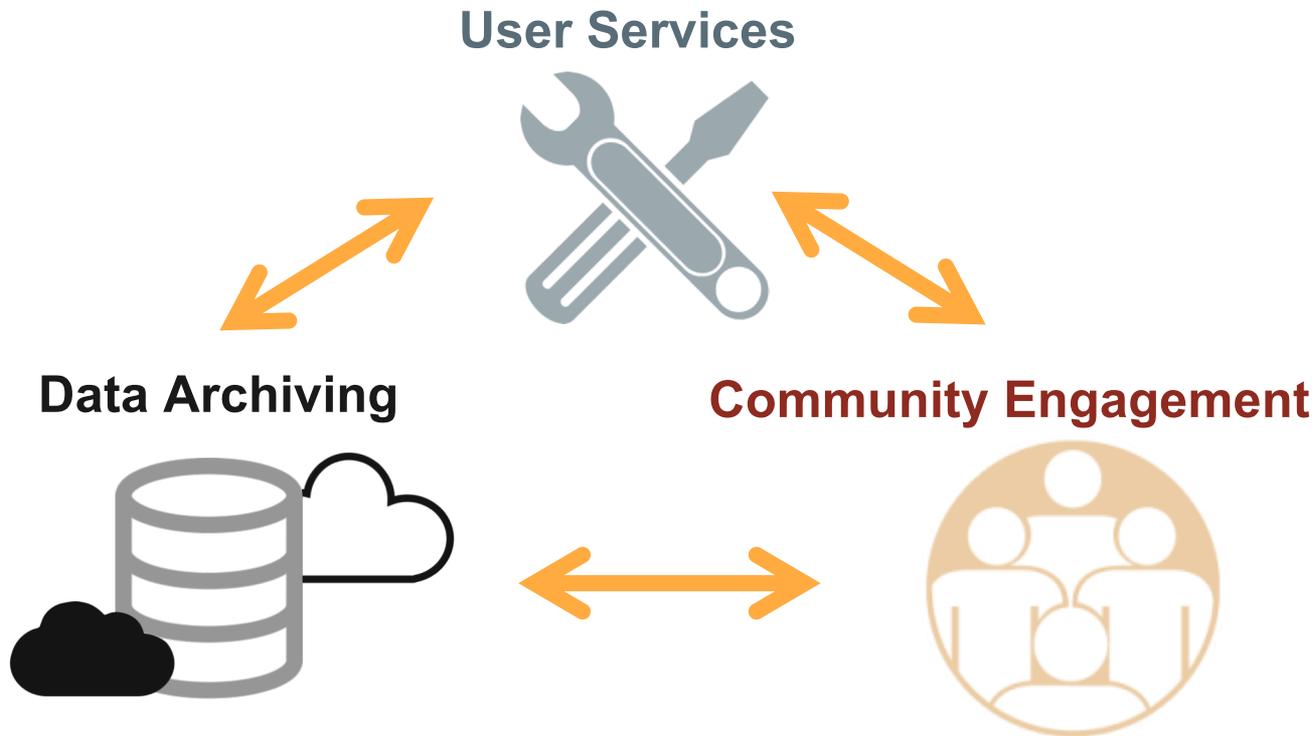


Digital Librarians

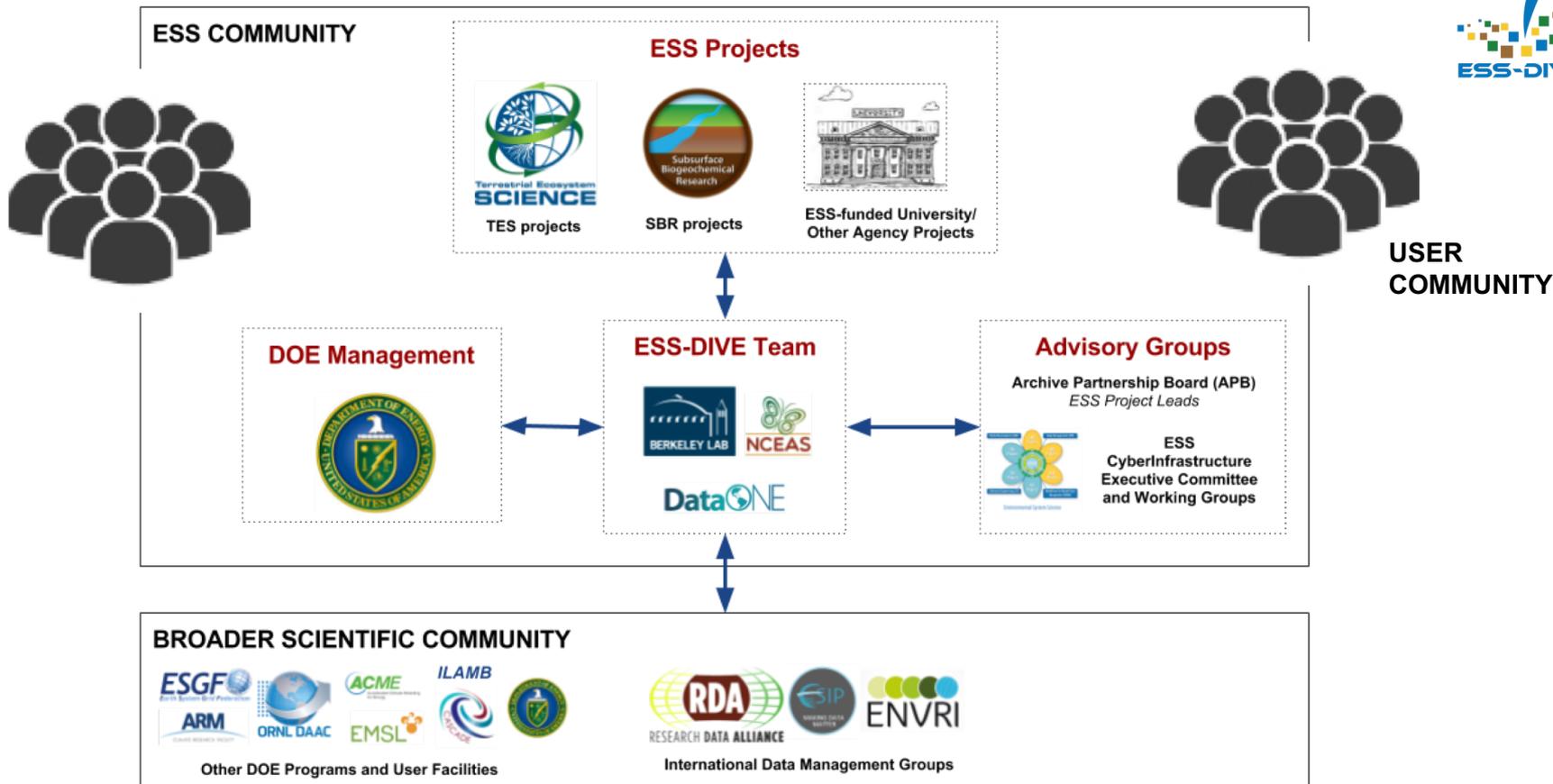


**Environmental
Scientists**

Three-Pronged Approach to Developing and Running ESS-DIVE

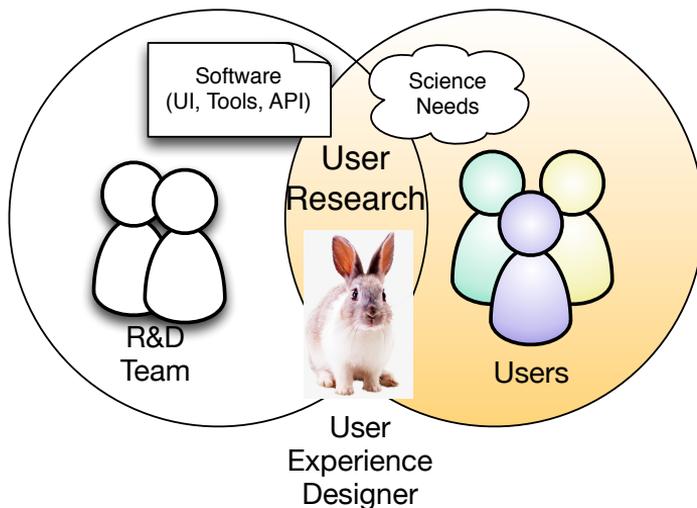


ESS-DIVE is a Partnership with the ESS Community



User Research – Based Approach

User research gives us a **process to verify/validate our “intuition (hypothesis) about what the user needs” and convert into action and build sense of ownership**

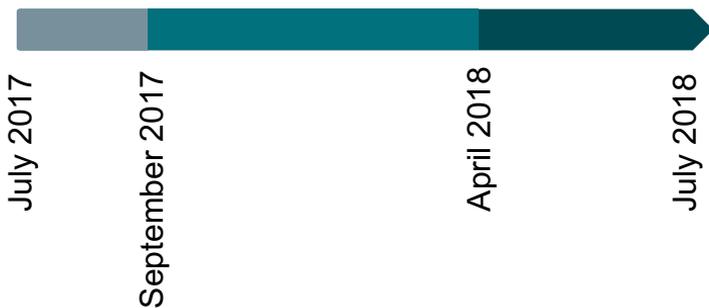




Project Community Engagement and Implementation

● Implementation

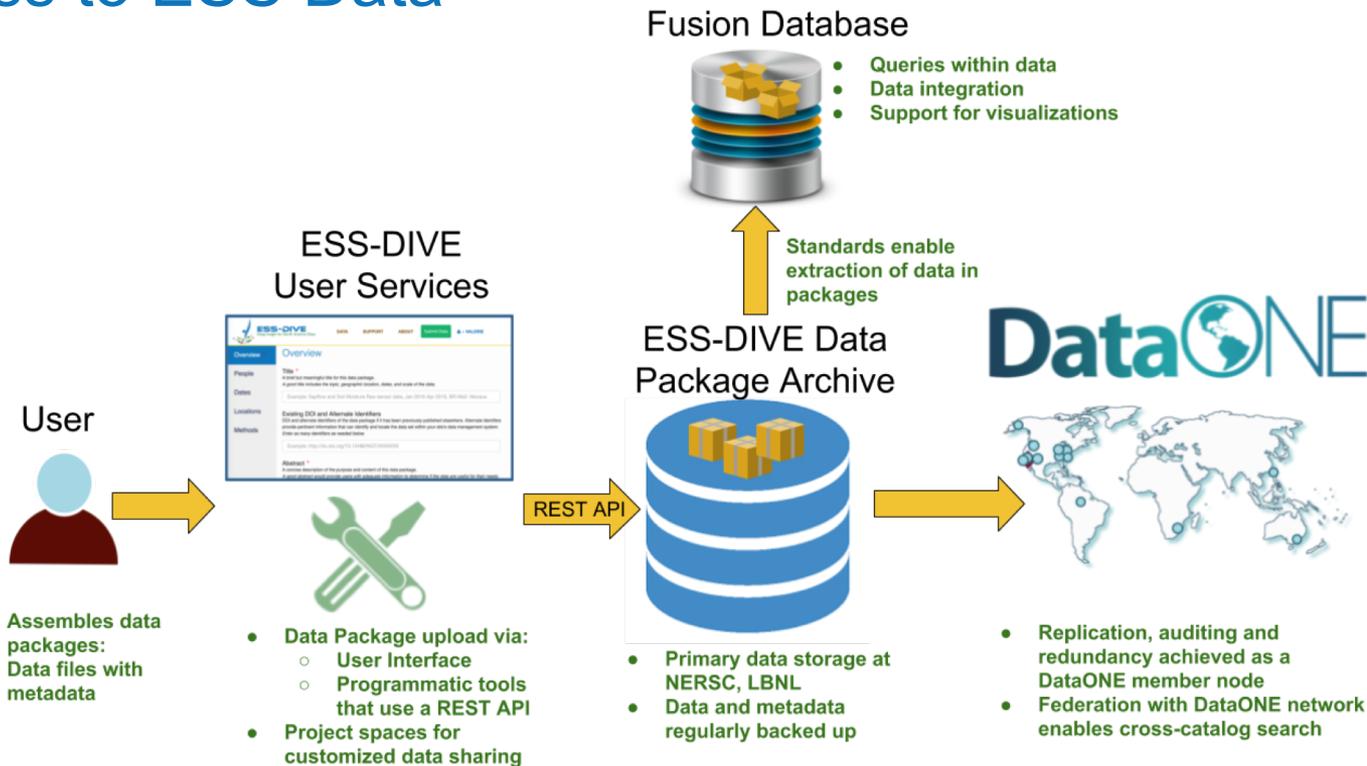
- 2017 July – Project start
- 2017 Sept. – Old archive transferred
- 2018 April – ESS-DIVE live
- 2018 August – Join **DataONE**



● Community engagement

- 2017 May – ESS CI and PI Meeting
- 2017 July - Visit to ORNL and OSTI
- 2017 Dec – Visit to Stanford/SLAC
- 2018 March – Archive Partnership Board Meeting
- 2018 May – ESS PI Meeting
- 2018 July – Visit to PNNL
- 2018 July – Archive Partnership Board Meeting

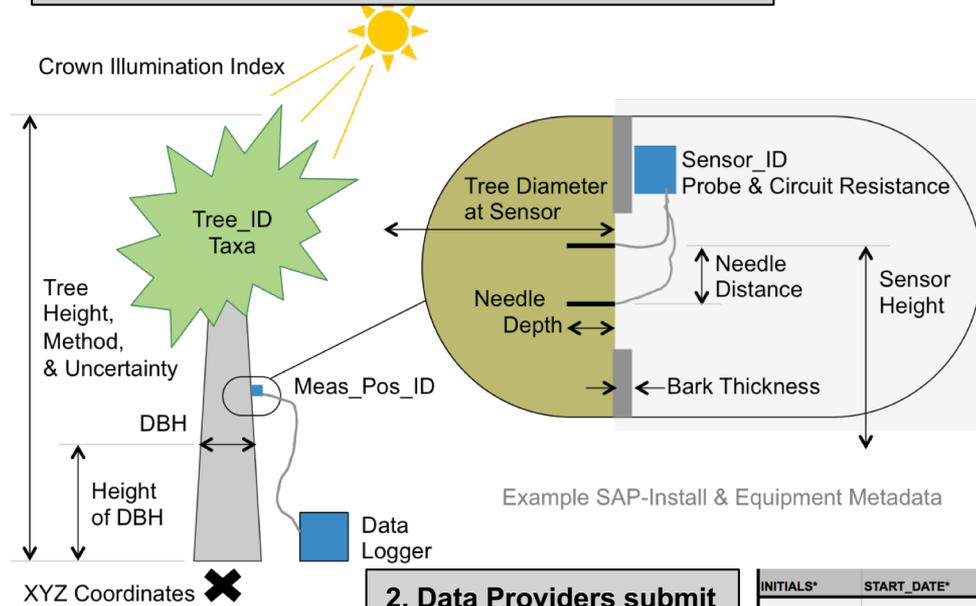
2018: A Data Archive that Stores and Provides Easy Access to ESS Data



Example: User Research Applied to Metadata Standards and Tools



1. Sapflow metadata reporting templates developed with community input



3. Climate modeler writes R script to read metadata templates and data files

```

1 # RSD 2015-2016
2 # Read in R object generated from the "RSD_data_collecting.R" script
3 # - connecting data and metadata pertaining to meteorology, sap flux, and soil moisture for Menus site
4 # Makes plots of sap flux data
5 # R: Circadianrhythm (DroeflyWan1.gov)
6 # December 2016
7
8 library(cron)
9 library(foreign)
10 mc(114-143)
11
12 # Paths and filenames
13 rootDir <- "C:/Users/296342/Documents/"
14 path_ensd <- paste(rootDir, "data/NSD-Tropics/NSD", sep="")
15 path_man <- paste(path_ensd, "data-20151121T194232/2015-06-NSD_menus", sep="")
16 path_man_sapgr <- paste(path_man, "Menus_20150720/Counter (sap velocity)", sep="")
17 path_man_sapICT <- paste(path_man, "Menus_20150720/ICT (sap velocity)", sep="")
18
19 # Load the R object containing the definitions and lists containing data and metadata
20 load(file=paste(path_ensd, "NSD_Robj", sep=""))
21
  
```

2. Data Providers submit data to archive with standard templates

INITIALS*	START_DATE*	END_DATE*	EVENT_TYPE*	EVENT_DESCRIP*
Initials of person entering info	Start date of event	End date of event; if single day event, report same day as start date	Type of event being reported	Short description of event
Abbrev. from General_Info	YYYYMMDD	YYYYMMDD	LIST	free text

Enhancing Value of Public Archive to ESS Community



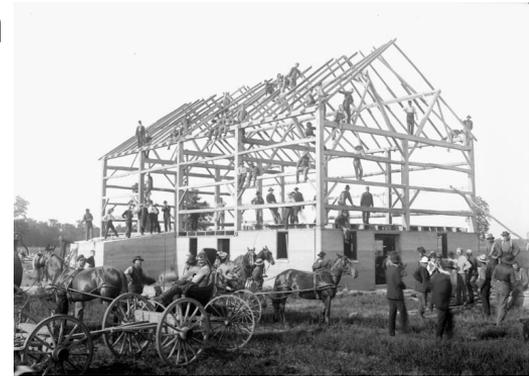
- Understand **user needs** with User Research
- Adopt **community standards** for data and metadata
- Provide archiving of **private/project** data packages to support cradle-to-grave data archiving
- Enable new data tools and capabilities through **community development**



ESS Community Initial Requests



- **Publish** data with **DOIs** for citation and provenance
- Easy **web services** for uploading data and metadata
- **Bulk** data **upload**
- Provide archiving of **private/project** data packages
- Support for **project-level control/curation** of data packages
- Enable new data tools and capabilities through **community development**
- Expand **metadata fields**



How do we move forward and build upon this discussion with our community?



- Integrate data management into the science process
- Project-level administration of contributed datasets – credit, curation, tools, and tracking
- Fix the citation and credit problem
 - Large numbers of datasets used in a paper
 - Data curators/processors/archivists (particularly domain)
- Continue to develop the carrots – DOIs, download stats, usage tracking, ...
- Data versioning and change tracking
- Methods to track the more nuanced data relationships
- Methods for dealing with model and experiment data



Acknowledgements



- DOE BER Data Management program within the Climate and Environmental Science Division - Funding
- National Center for Ecological Analysis and Synthesis (NCEAS) – Help getting up and running quickly
- DOE Office of Scientific and Technical Information (OSTI) – Transition DOIs from prior archive
- Datacite – Consultations and transfer of DOIs
- National Energy Research Scientific Computing facility (NERSC) – Hosting archive