



Overview

- Introduction (Deb Agarwal)
- Infrastructure (Cory Snavely)
- Demo (Val Hendrix)
- Discussion (Deb Agarwal)



Introduction

ESS-DIVE is live! A discussion of future scope and opportunities

ESS-DIVE: A New Data Archive for Earth and Environmental Science Data



Challenge

Preserve, expand access to, and improve usability of critical data generated or used by ESS projects

Approach and Results

- ESS-DIVE available for archiving of data as of April 1, 2018
- Leverage emerging digital library technologies and data standards
- Data citation using Digital Object Identifiers (DOI)
- Incentivize data providers to contribute well-structured, standardized, high-quality data
- Transition legacy DOE data (CDIAC) into the new archive

Significance and Impact

- Enables publishing of critical environmental science data
- Users able to find and obtain ESS data in usable formats

http://ess-dive.lbl.gov





ESS Community Engagement



- Worked with ESS Data Management Working Group in ESS-DIVE package metadata design
- Site visits to ORNL, CDIAC, OSTI, and SLAC/Stanford
- 2017 ESS CI and AGU presentation
- Formed Archive Partnership Board and had first meeting



Archive Partnership Board



Stan Wullschleger (ORNL) John Bargar (SLAC) Susan Hubbard (LBNL) Margaret Torn (LBNL) Paul Hanson (ORNL) James Stegen (PNNL) Eric Pierce (ORNL) Annie Kersting (LLNL) Ken Kemner (ANL) Kate Maher (Stanford) Julie Jastrow (ANL) David Moulton (LANL) Kristen DeAngelis (U Mass) Jennifer Power (U Minn) Forrest Hoffman (ORNL) Dean Williams (LLNL) (not yet confirmed)

ESS PI Meeting activities



- Posters (back of Ben Franklin Hall)
 - ESS-DIVE Overview
 - Community Engagement and Publishing Lifecycle
- 1:1 Help desk (besides posters)
- Tue 9:25-9:40 AM plenary talk (D. Agarwal)
- Wed 10-11:30 AM Tutorial Breakout Session (C. Varadharajan, Room 4)







Seeking input from ESS working groups on archive next steps.

Post hoc vs Post-Doc

We're hiring a postdoc to work on metadata standards, data processing and synthesis.

Job will be posted at jobs.lbl.gov

The Post hoc Fallacy To incorrectly assume "A" is the cause of "B" just because "A" preceded "B".

> e.g. "All Professors have Ph.D.'s, therefore getting a Ph.D. means you'll get a Professor job (right?)"

JORGE CHAM @ 2009

The Post-Doc Fallacy To incorrectly assume you'll have <u>a</u> job just because you have a PhD.

e.g. "Now what??"



Infrastructure

An overview of ESS-DIVE infrastructure, highlighting its primary features and the benefits brought by the LBNL / NERSC / NCEAS partnership.

ESS-DIVE Archiving Features





ESS-DIVE

Feature: User Services with Metacat

- Developed at NCEAS for earth sciences community
- Robust support for relevant metadata standards, data deposit, and data auditing / management
- Integrates with DataONE
- Web UI and REST API





Feature: Microservice Architecture at NERSC

- Built on supported *Spin* platform for web and network services
- Integrated with NERSC storage systems, HPC, and network
- Docker enables rigorous version control over software and underlying systems
- Multiple instances easy to "spin up" using isolated containers





Feature: DataONE network membership

- Robust protocol for data integrity and replication
- Geographic distribution ensures continuity of data preservation and access
- Federated search enhances discovery



Features and Benefits Summary

 \Rightarrow

 \Rightarrow



User Services based on Metacat archive management software and data tools

Microservice architecture on NERSC's new Docker-based platform, *Spin*

DataONE network membership

- Quick start-up; robust functionality; metadata and tools grounded in earth sciences community
- Integrated with petascale file systems, HPC, and fast network; rigorous version-controlled software stack and system architecture
- ⇒ Reliable data integrity and replication; geographic distribution protects against localized risks; broadened discovery through federation



ESS-DIVE Demo https://data.ess-dive.lbl.gov

Demonstration of ESS-DIVE Data Archive 1.0.0 capabilities launched on April 1, 2018.



Initial Capabilities

ESS-DIVE Data Archive 1.0 initially supports most common and important ESS community use cases.

- Upload data and metadata bundled together as a data package.
- Release data packages publicly with a digital object identifier (DOI).
- Public search and download of ESS data packages.







ESS-DIVE Main Site

https://data.ess-dive.lbl.gov

ESS-DIVE Playground

https://data-sandbox.ess-dive.lbl.gov



Discussion

Discussion Items



- Input on Metadata fields and standards
 - Project/Publisher
 - Variables
 - Keywords
- DOI Publisher and data versioning
- Data usage rights (CC)
- Management of project data and access rights
- Use-cases and priorities for future development
 - Access: How do you want to access
 - Upload: How do you want to upload
 - Do you want automated uploads?

Digital Object Identifier - DOI



- Unique DOI prefix for ESS-DIVE with OSTI https://doi.org/10.15485
- Not planning to assign an infix
- Citation format
 - Creator (PublicationYear). Title. Version. Publisher. https://doi.org/10.15485/xxxxxx
- How do you want to handle publisher?
- Versioning?

Usage Right Options - Creative Commons



CC0 - Public domain, "No rights reserved"

- Public usage of your work allowed without restriction
- User is provided DOI citation information and requested to cite the DOI

CCby4 - Public use with attribution of source

- Others who use your work in any way must cite the DOI
- Use of your work without citation requires author permission



Backup Slides





Initial Capabilities

ESS-DIVE Data Archive 1.0 initially supports most common and important ESS community use cases.

- Upload data and metadata bundled together as a data package.
- Release data packages publicly with a digital object identifier (DOI).
- Public search and download of ESS data packages.



Main Page - Public view



ESS-DIVE







	Deep Insight for Earth	Science Data	SUPPORT ABOUT	Submit Data
	Search @ fossil fuels Q	DATASETS 1 TO 25 OF 93 1 2 3 4 Next Sort by Most recent •	Hide Map »	
Keyword	Filter by:	J. H (2002): Throughfall Displacement Experiment (TDE) Ecosystem Model Intercomparison Project Data Archive. Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (USA). doi:10.3334/CDIAC/VRC.002 Gornitz V; White T (1994): A Coastal Hazards Data Base for the U.S. Gulf Coast (1993) (NDP-04bB). Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). doi:10.3334/CDIAC/SSR.NDP043B	-	Creeland 1 Canado 1 1 1 1 1 1 1 1 1 1 1 1 1
search for "fossil fuels	,1	 2 • • • • • • • • • • • • • • • • • • •	Satellite Terrain Google	Venezuela Colombia Pera Bolivia Chie? Argentina Map data ©2018 2000 km Terms of Use

ESS-DIVE



ESS-DIVE

Menu Options - Data



ESS-D Deep Insight for Ear	IVE h Science Data	DATA	SUPPORT	ABOUT	Submit Data	🛎 🕶 VALERIE
Search @	DATASETS 1 TO 9 OF 9	SEARCH	lide Ma	ap »		
Search phrase Q	Sort	SUMMARY				

Menu Options - Support



ESS-D Deep Insight for Eart	DATA DATA	SUPPORT	ABOUT	Submit Data	🛎 🕶 VALERIE
Search @	DATASETS 1 TO 9 OF 9	CONTACT US			
Search phrase Q	Sort by Most rece	NEW USER?			And the second second
Filter by:	Jones P; Limbert D (1989): A Databank of Antarctic Surface Temperature and Pressure Data (NDP-032)	DATA ACCESS HEL	P	Part and	
 Q Location 	Carbon Dioxide Information Analysis Center (CDIAC), Ridge National Laboratory (ORNL), Oak Ridge, TN (U States): University of East Anglia, Norwich, United Kin	UPLOAD HELP		Per V Ch.	
Creator	ess-dive-c4bec7fac06eabf-20180424T085437278159.	FAQ		A seal of	220 6
🕨 🋗 Year	🔒 📤 💿 👁 🚯 📄 💡	1.3	- Sec	× 11	



Data Package - Create





Data Package - Create





Data Package - Create









Data Package - Public view

		\ ``	Deep Insight for Earth Sci Back to search Ho	ence Data me / Search / Metadata	DATA	SUPPORT ABOUT	Sub	mit Data 🚺 s	ign in with Orcid	ESS-DIVE
	Paul Hanson, Jeffrey Riggs, C Dorrance, William Nettles, and Les Hook (2017): SPRUCE S1 Bog Environmental Monitoring Data: 2010-2016. SPRUCE doi:10.3334/CDIAC/spruce.001								object identifier (DOI)	
	a	Files in this dataset Package: ess-dive-f38/87b4/cf1bce-20180330T150358358173								
Metadata in		-	Name			File type	Size	Downloads	Login to Download	
format and		ß	Metadata: SPRUCE Monitoring Data: 20	S1 Bog Environmental 10-2016.xml		EML v2.1.1	10 KB	19 views	Download	Scroll down the
uploaded data		⊞	em_data_dictionary_	_public_20170620.xlsx	More info	Microsoft Excel OpenXML	23 KB	2 downloads	Download	page to see the
for download.		⊞	SPRUCE_EM_DAT	A_2010_2016_20170620.pdf	More info	PDF	1 MB	21 downloads	Download	data package metadata: Title.
		■	EM123_Combined_ csv	public_data_2010_20170620.	More info	text/csv	6 MB	2 downloads	Download	Abstract, Project
				•	Show 7 moi	re items in this data set				Name, DOE
										Contract Numbers
		General Identifier							etc	
			ess-dive-c611d6b5be792d3-2	20180330T2	02613132					
										. ↓