

# DataONE Webinar Series

---

## What it Means to be a Member Node

Member Nodes Share Their Views



Dave Viegialis, Monica Ihli, Amy Forrester



Kenneth Casey



James Duncan

Mark Servilla

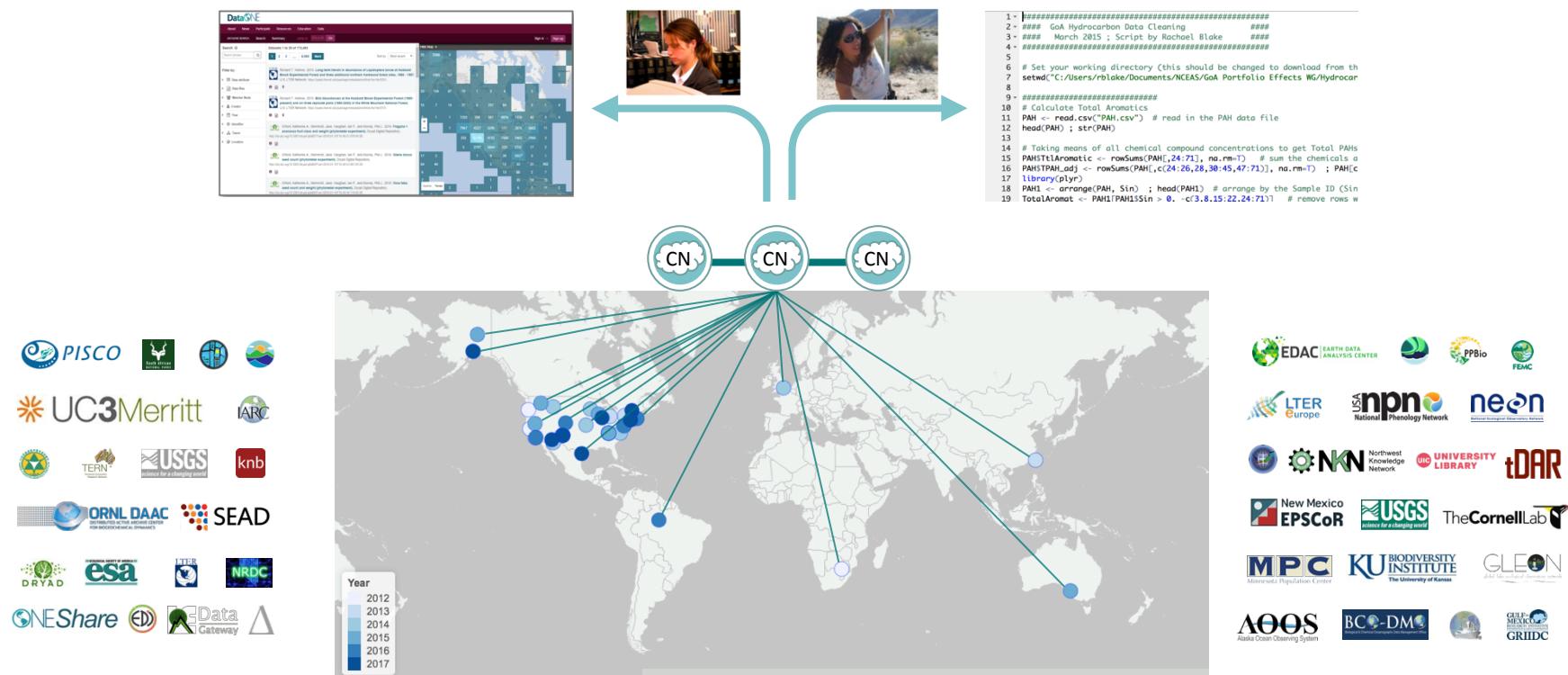


Grant Numbers 0830944 and 1430508

May 8, 2018

# Federation of Earth Science Repositories

## Member Nodes



# Federated Search Across Repositories

# Federated Search Across Repositories

The screenshot shows the DataONE search interface. A red box highlights the search results for 'Creator: Peterson'. The search bar at the top contains 'Search phrase' and a magnifying glass icon. Below it, the title 'My Search' is displayed, followed by 'Creator: Peterson' with a close button. The search results are filtered by 'Creator' and show 10 entries:

- Fabes, Jez, Seligman, William, Barrett, Carolyn, McKeon, Dayad. Intensive care discharge risk score and nurse-led interventions in the United Kingdom. Dryad Digital Repository. <https://doi.org/10.5061/dryad.1107>
- Park, Hyeree, Cho, Sooyoung, Woo, Hyeongtaek, Park, Dayad. In the Korean Multi-center Cancer Cohort. Dryad Digital Repository. <https://doi.org/10.5061/dryad.1108>
- Jason Schatz, Carly Ziter, and Christopher Kucharski. 2016. Madison, Wisconsin from 2012–2016. LTER Network Data Archive. <https://doi.org/10.5061/zenodo.1000000>
- Heather Bateman, Dan Childers, Madhusudan Katti, Eyer, Bird abundance and diversity in central Arizona-Phoenix. <https://pasta.lternet.edu/package/metadata/em/knb-lter-cap/40>
- Peter Bergholz. 2017. Escherichia coli isolated from a. knb. <https://doi.org/10.5061/zenodo.1000001>
- Márcia Cristina Mendes Marques, Anamaria Cequelin, B. naturaldo estrato arbóreo em um gradiente edáfico-Biodiversidade (PPBio). PPBioMA.31.3. <https://doi.org/10.5061/zenodo.1000002>
- Melinda Niewonger, Kristal Verhulst, Murat Aydin, and Antarctic ice cores. Arctic Data Center. [um.umd.746cc00](https://doi.org/10.5061/zenodo.1000003)
- Carlo, Michael. 2017. Colorectal\_EcoLett\_labandfield. Dayad. ver=2017-11-21T17:02:49.517+00:00. <https://doi.org/10.5061/dryad.pr1h0?ver=2017-11-21T17:02:49.517+00:00>
- Carlo, Michael A., Riddell, Eric A., Levy, Ofir, and Sears, Dayad. Carlo, Michael A., Riddell, Eric A., Levy, Ofir, and Sears inhibits juvenile growth, and alters species distribution. <https://doi.org/10.5061/dryad.pr1h0?ver=2017-11-21T17:02:49.517+00:00>

A red arrow points from the left side of the search results towards the search bar at the top.

**Search ?**

Search phrase

## My Search

**Creator: Peterson**

### Filter by:

▼  **Data attribute**

►  **Data files**

►  **Member Node**

▼  **Creator**

►  **Year**

►  **Identifier**

►  **Taxon**

►  **Location**

**Hide Map >**

Map showing data distribution across the world. The map is overlaid with a grid of numbers representing data values. The legend indicates 'Satellite' and 'Terrain' data types. A scale bar shows 0 to 1000 km.

Map Foundation (NSF) under a Cooperative Agreement. Acknowledgments: This material is based upon work supported by the National Science Foundation (NSF) under a Cooperative Agreement. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

# Data Package

## OAI-ORE Binding Graph

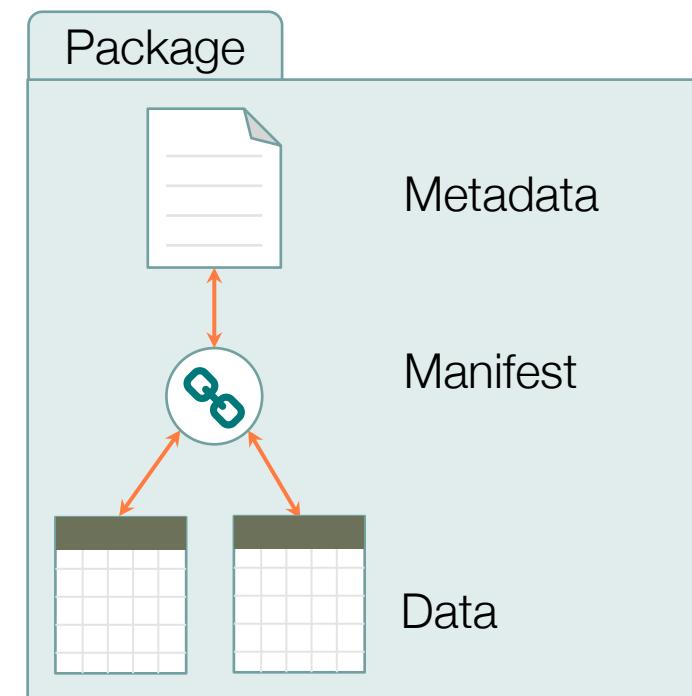


### Data Package components

- Metadata
- Data
- Manifest

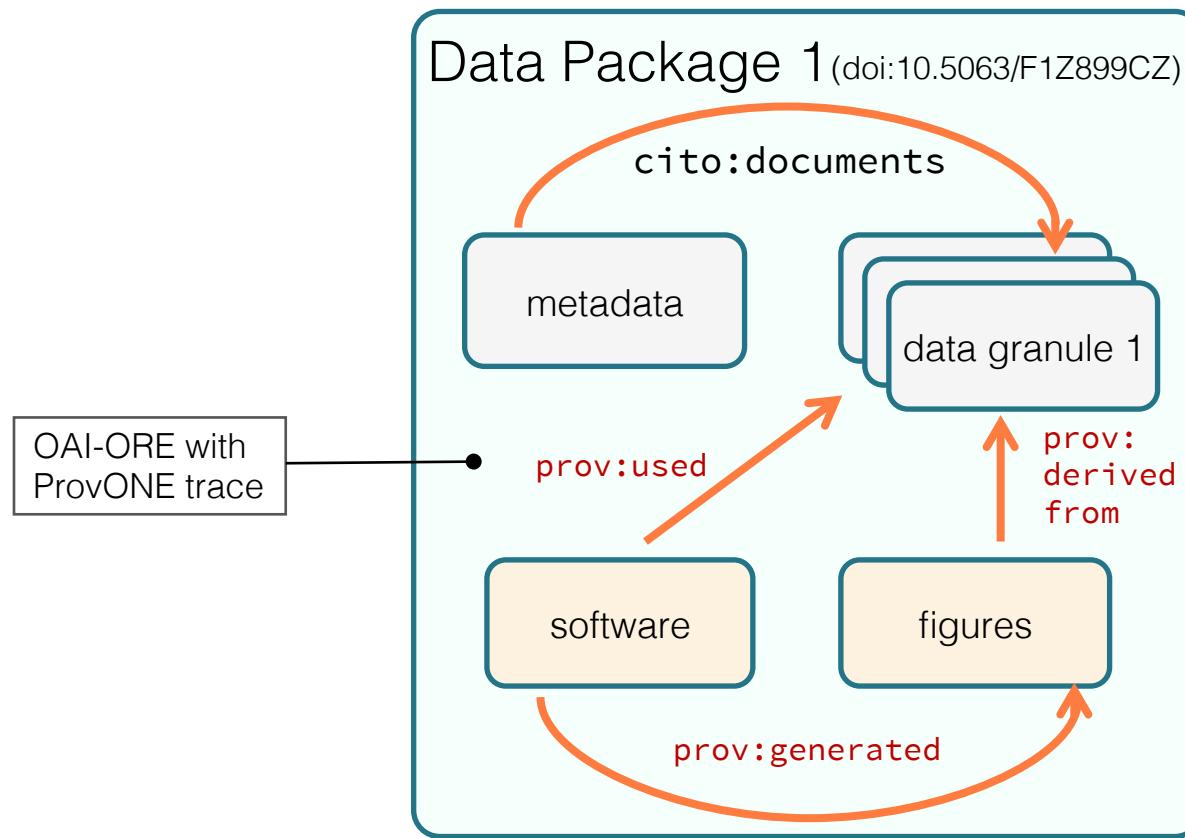
### All components

- Immutable
- Uniquely identified
- Resolvable
- Retrievable



## Data Package with Provenance

---



# Provenance Display

## DataONE Search

About News Participate Resources Education Data

DATAONE SEARCH: [Search](#) [Summary](#) Jump to: [DOI or ID](#) [Go](#)

[Sign in](#) or [Sign up](#)

[◀ Back to search](#) | Search / Metadata

Mark Carls. 2017. Analysis of hydrocarbons following the Exxon Valdez oil spill, Gulf of Alaska, 1989 - 2014. Gulf of Alaska Data Portal. urn:uuid:3249ada0-afe3-4dd6-875e-0f7928a4c171.

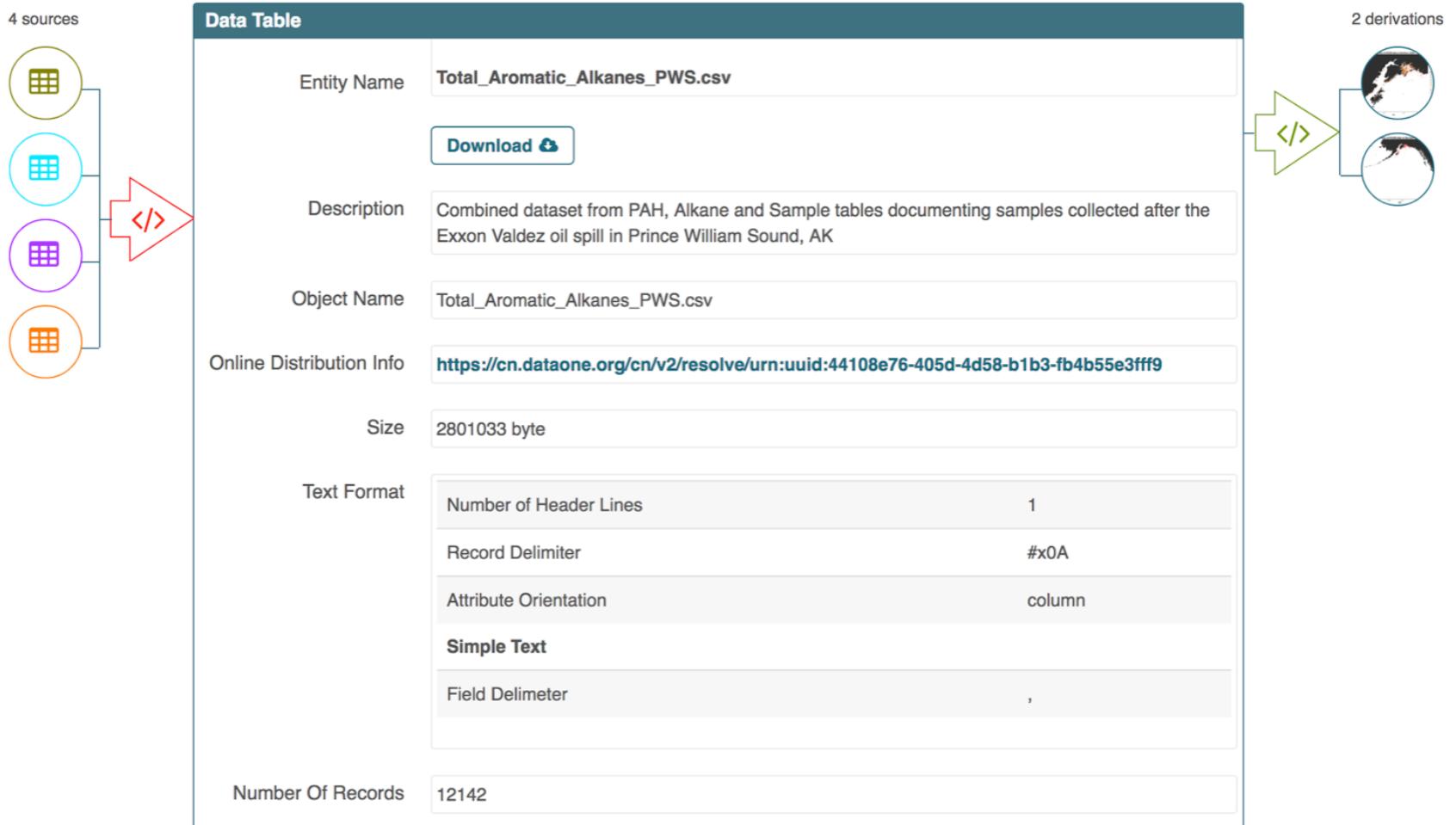
[Copy Citation](#)

Files in this dataset Package: urn:uuid:1d23e155-3ef5-47c6-9612-027c80855e8d

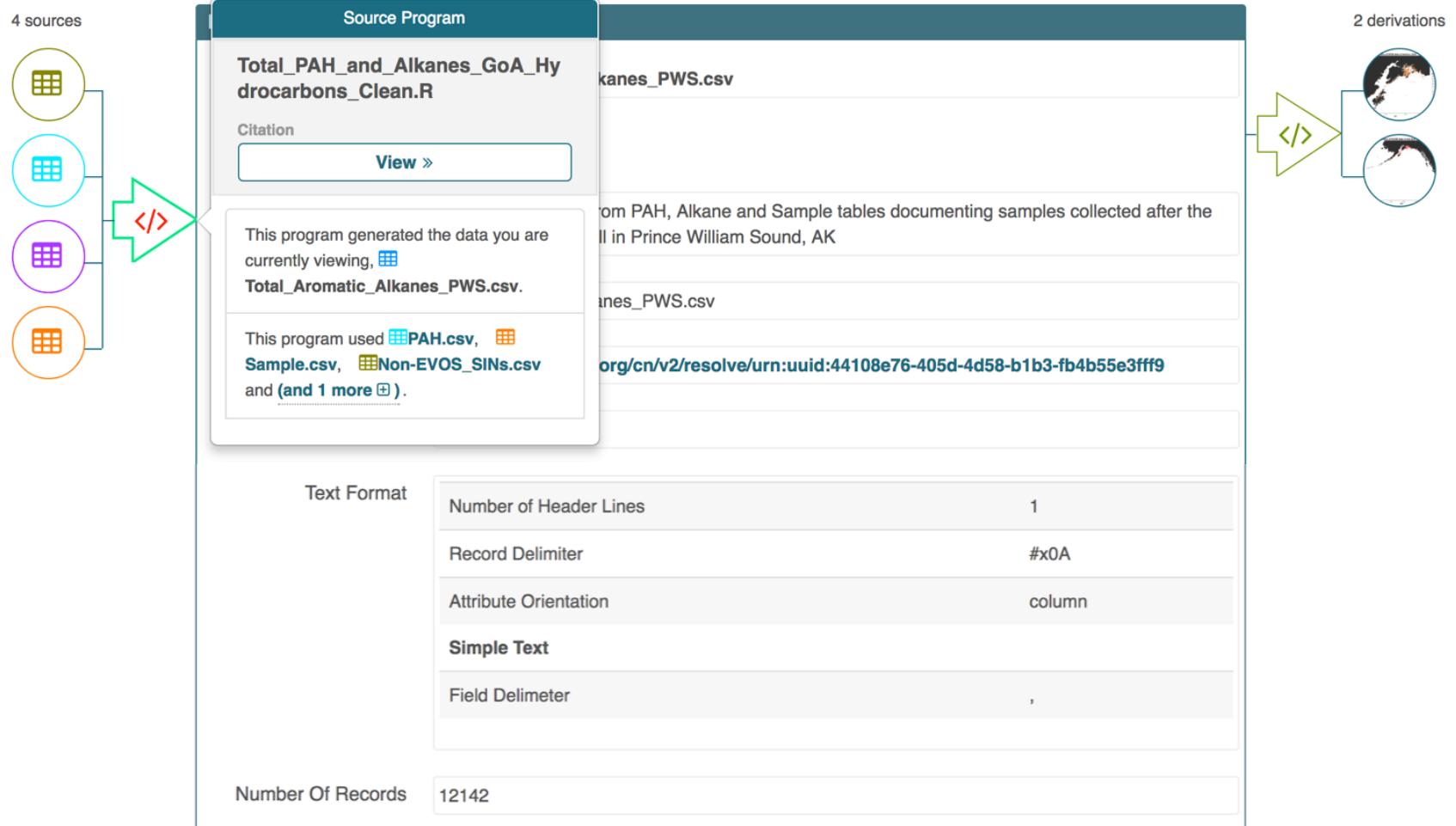
Name	File type	Size	<a href="#">Download all</a>		
Metadata: metadata.xml	EML v2.1.1	140 KB	<a href="#">Download</a>		
Total_Aromatic_Alkanes_PWS.csv	More info	text/csv	3 MB	3 downloads	<a href="#">Download</a>
CollectionMethods.csv	More info	text/csv	793 B	2 downloads	<a href="#">Download</a>
Non-EVOS_SINs.csv	More info	text/csv	3 KB		<a href="#">Download</a>

[▶ Show 8 more items in this data set](#)

## Data Table, Image, and Other Data Details



## Data Table, Image, and Other Data Details



# Data Use Metrics

DATAONE SEARCH: [Search](#) [Summary](#) Jump to: [DOI or ID](#) [Go](#) [Sign in](#) or [Sign up](#)



## Timothy Assal

Username: assal



National Center for Ecological Analysis and Synthesis



Contributer for 7 months  
since Nov. 17, 2015

14 contributions

166 downloads

304 views

6 citations

9 saves

10 mentions

Datasets 1 to 3 of 3

Sort by [Most recent](#)

United States Geological Survey and Timothy Assal. 2015. [Data from: Mapping forest functional type in a forest-shrubland ecotone using SPOT imagery and predictive habitat distribution modelling](#). KNB Data Repository. doi:10.5063/F1639MP5.

94 202 3 4 5

Timothy Assal. 2015. [Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium](#). KNB Data Repository. doi:10:5063/F183MK

30 46 1 2

Assal et al. 2015. [Ut enim ad minim veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur](#). KNB Data Repository. doi:10:5063/F187FR6

42 58 2 4 3

# Data Use Metrics

DATAONE SEARCH: [Search](#) [Summary](#) Jump to: [DOI or ID](#) [Go](#) [Sign in](#) or [Sign up](#)



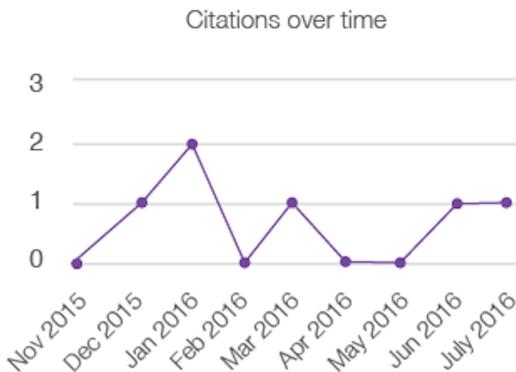
## Timothy Assal

Username: assal



National Center for Ecological Analysis and Synthesis

6 Citations



UC Curation Center



Datasets 1 to 3 of 3

Sort by [Most recent](#)

### Works that cited Timothy Assal's datasets

Assal T., Anderson, P. and Sibold J. 2016. [Spatial and temporal trends of drought effects in a heterogeneous semi-arid forest](#). Forest Ecology and Management. doi:10.1016/j.foreco.2016.01.017

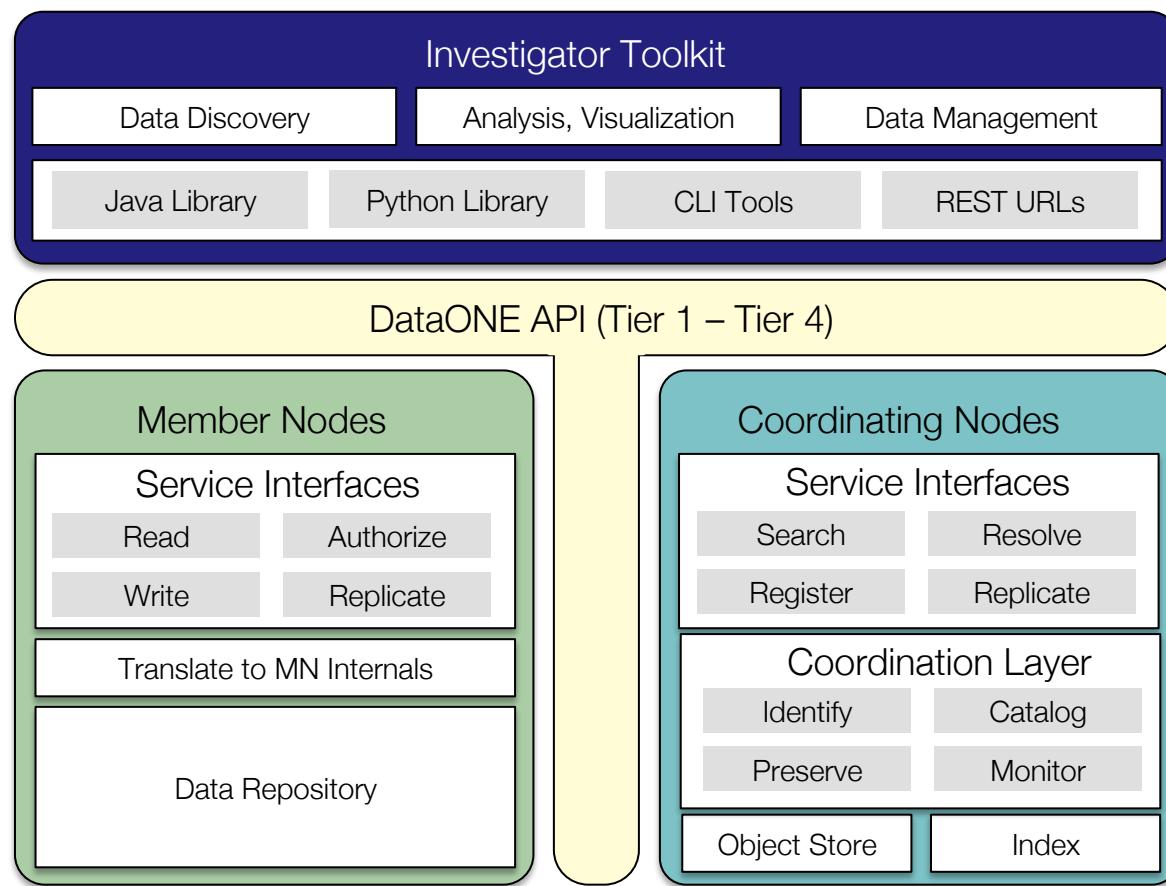
Tim Assal. 2016. [Documenting Spatial and Temporal Trends of Drought](#). NASA.gov

Smith, Joe. 2016. [Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque audantium](#). Ecology. doi:1234/123JK

Douglas et al. 2016. [At vero eos et accusamus et iusto odio dignissimos ducimus qui blanditiis praesentium voluptatum deleniti atque corrupti](#). Nature. doi:1J78/397HF9

[See 2 more](#)

# DataONE Core Architecture



## Member Nodes

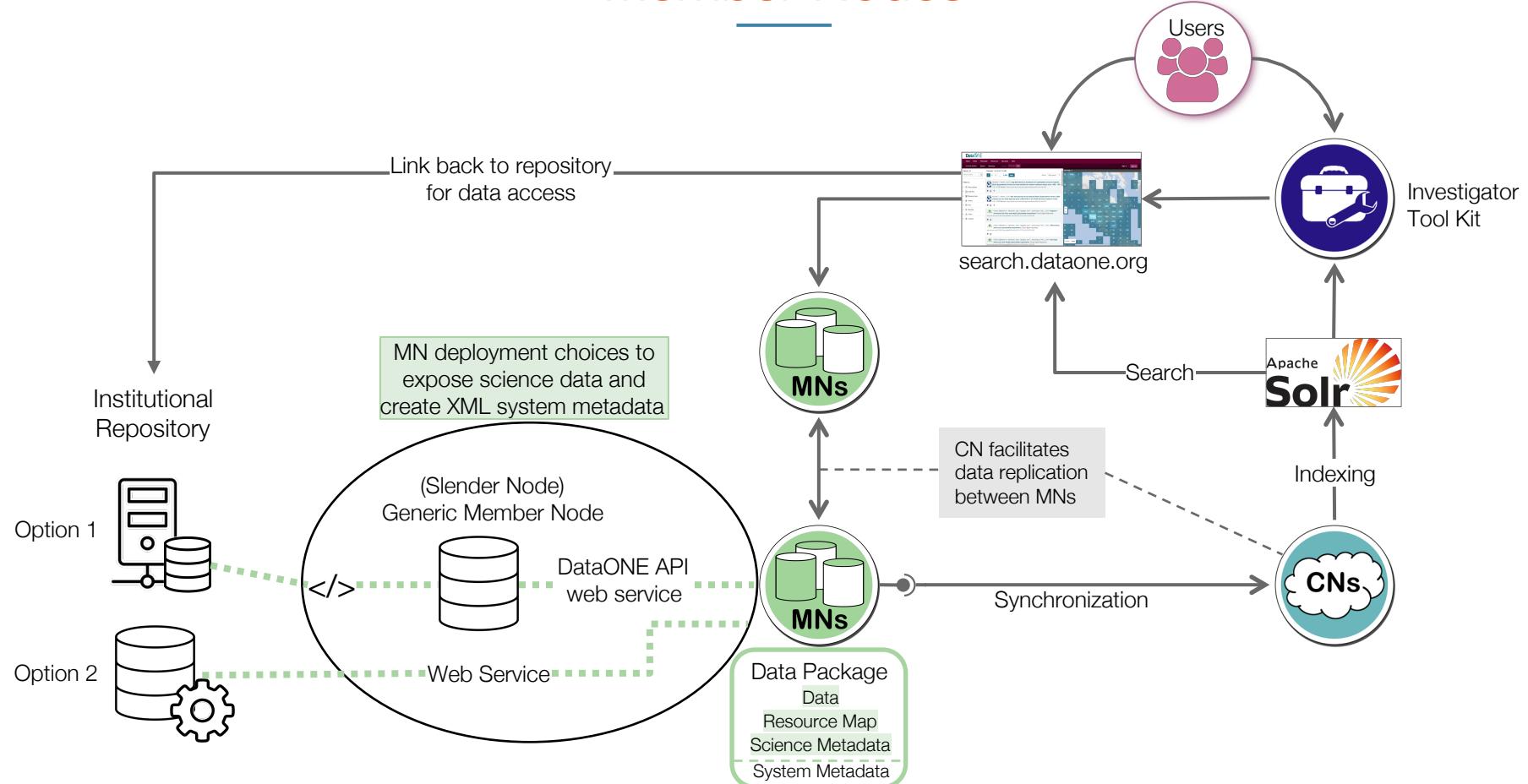
---

Factors which can influence MN implementation choices:

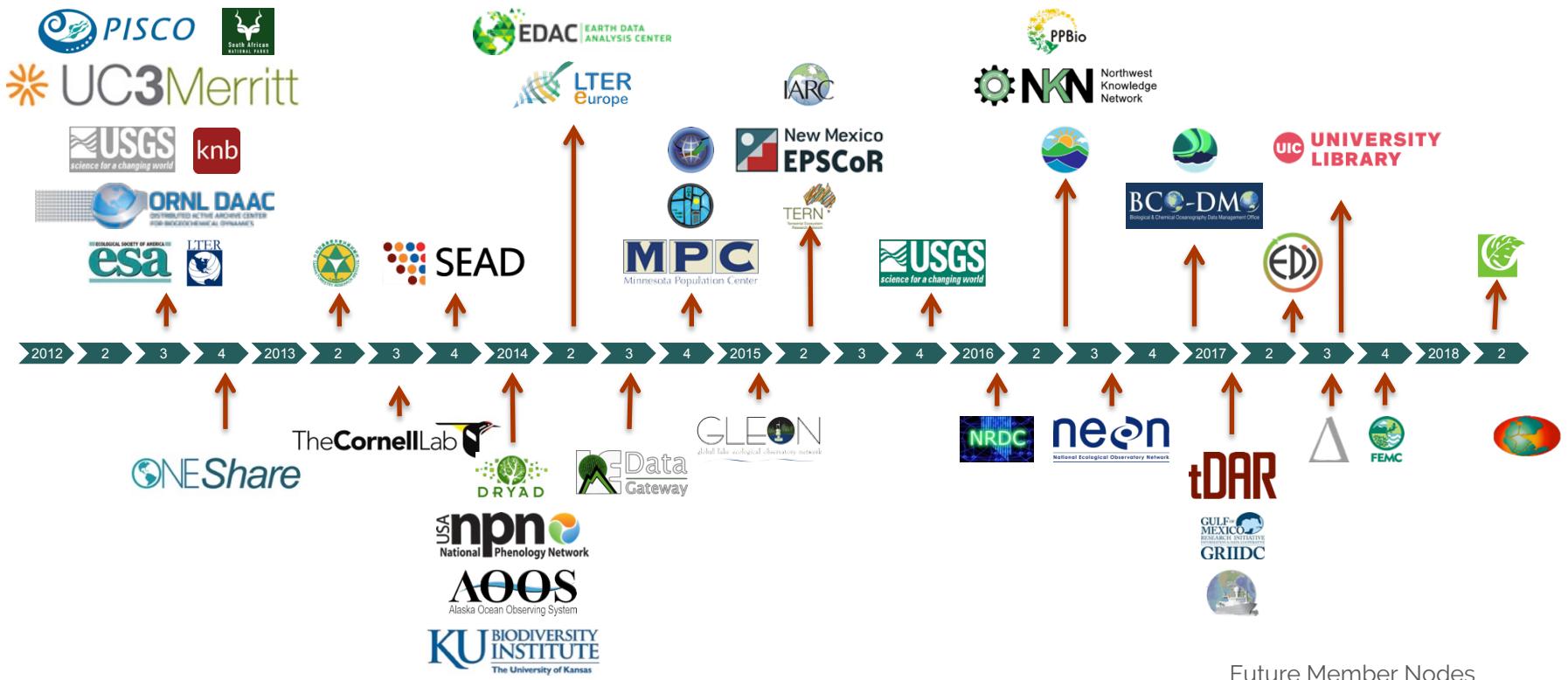
- Direct communication with CN through DataONE API, or via proxy?
- Existence of other services for exposing data.
- Satisfaction with current repository solution.
- Leveraging or contributing resources for replication of science data.



## Member Nodes

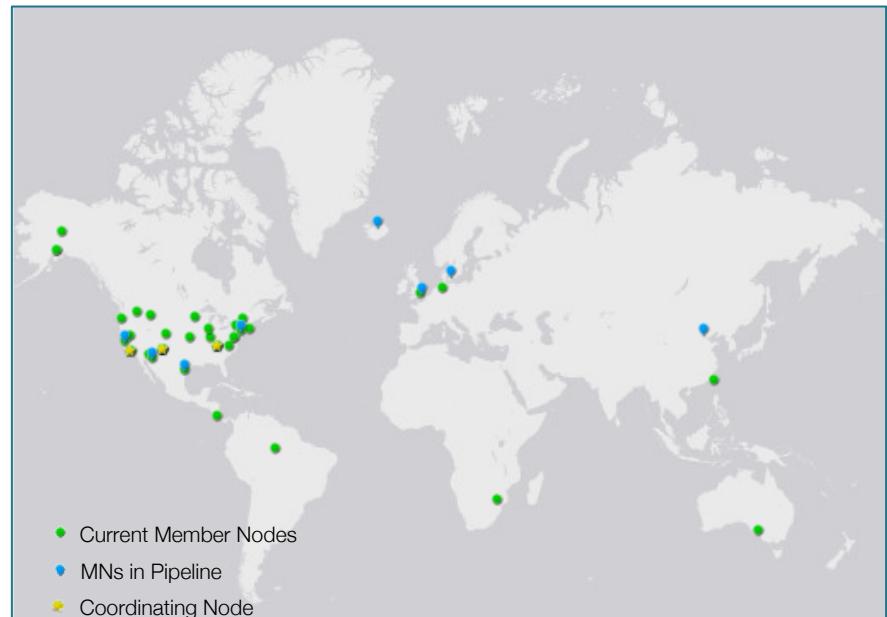
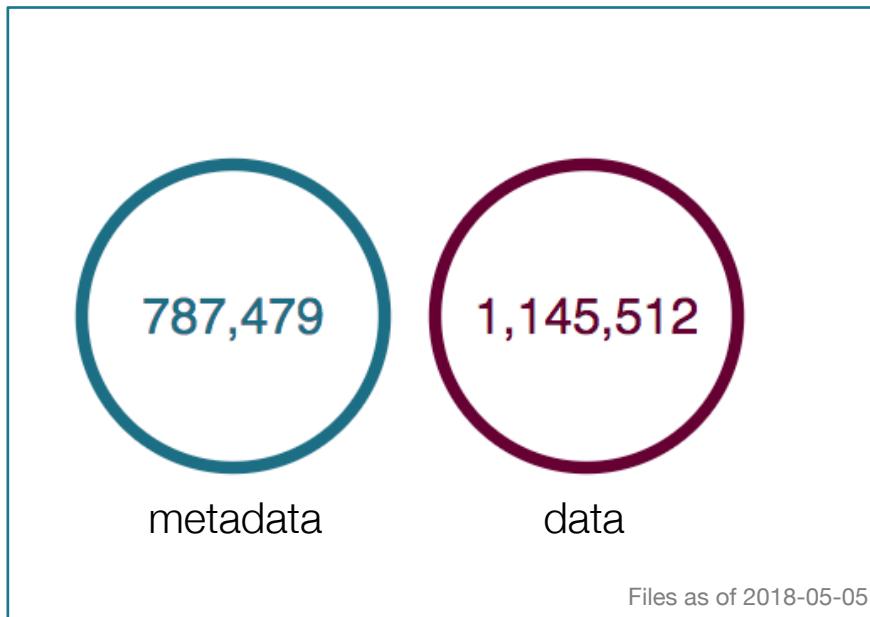


## Building the Federation



## Member Nodes are the Key

---





# The NCEI DataONE Member Node

Kenneth S. Casey, Ph.D.  
NOAA's National Centers for Environmental Information

For the DataONE Webinar Series: MN Showcase

National Oceanic and Atmospheric Administration | NOAA Satellite and Information Service



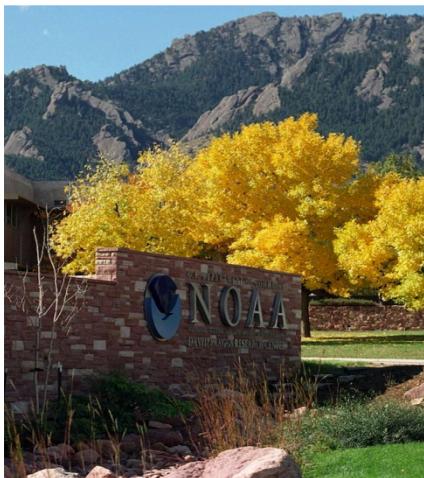
[ResearchGate Profile](#)  
[Google Scholar Profile](#)

ResearcherID: [D-4065-2013](#)  
ORCID: [0000-0002-6052-711](#)

## Deputy Director of NOAA's National Centers for Environmental Information (NCEI) Data Stewardship Division (DSD)

- Set technical direction for NCEI's archive and data stewardship activities
- Promote NCEI as a responsible citizen of the global environmental data management community
- Physical oceanographer, developed in situ and satellite-based datasets to document and understand interactions between climate change and marine ecosystems (past experience, informs actions today)

- NOAA's designated long term archive and data stewardship facility
- Responsible for maintaining one of the largest environmental archives on Earth, spanning oceanic, atmospheric, and geophysical data (over 17 PB, ~30K collections, ~billions of granules, ~200 routine data feeds)
- Four sites: Asheville NC, Boulder CO, Stennis MS, Silver Spring MD



- We also build value-added products and perform higher tiers of stewardship where the mission demands and resources allow
- We share data broadly: Data.gov, WMO Information System, GEOSS Portal, ICSU World Data Systems (four), CEOS, etc.
- We interact a lot with other repos - e.g, long term archival services for NSF-funded BCO-DMO, CCHDO, and soon, the Arctic Data Center
- Currently: ocean archive is Tier 1, working toward Tier 4 for ADC



- DataONE presents an interesting opportunity, to both leverage community data management developments and contribute to them
- If successful, this relationship could help us streamline how we bring in “long tail” data useful to the NOAA mission
- Currently operating only as a Tier-1, metadata-only member node for our ocean archive.
- Actual impacts are not yet well understood. DataONE providing another way for users to find and reuse our data but no hard metrics
- Working on Tier-4 as partner in the Arctic Data Center project will provide a much clearer picture of the value of the relationship
  - Will the interface be a stable and reliable source of data?
  - Will we be able to easily maintain it?
  - Will it help us bring data in at a lower overall cost?



# Kenneth S. Casey, PhD

Deputy Director, Data Stewardship Division

NOAA National Centers for Environmental

Information (NCEI)

[Kenneth.Casey@noaa.gov](mailto:Kenneth.Casey@noaa.gov)

301-713-4849 (Office)

# Who Am I?

Director of the FEMC  
Formerly the Data and Web Coordinator  
Research Staff at Rubenstein School of Environment and Natural Resources

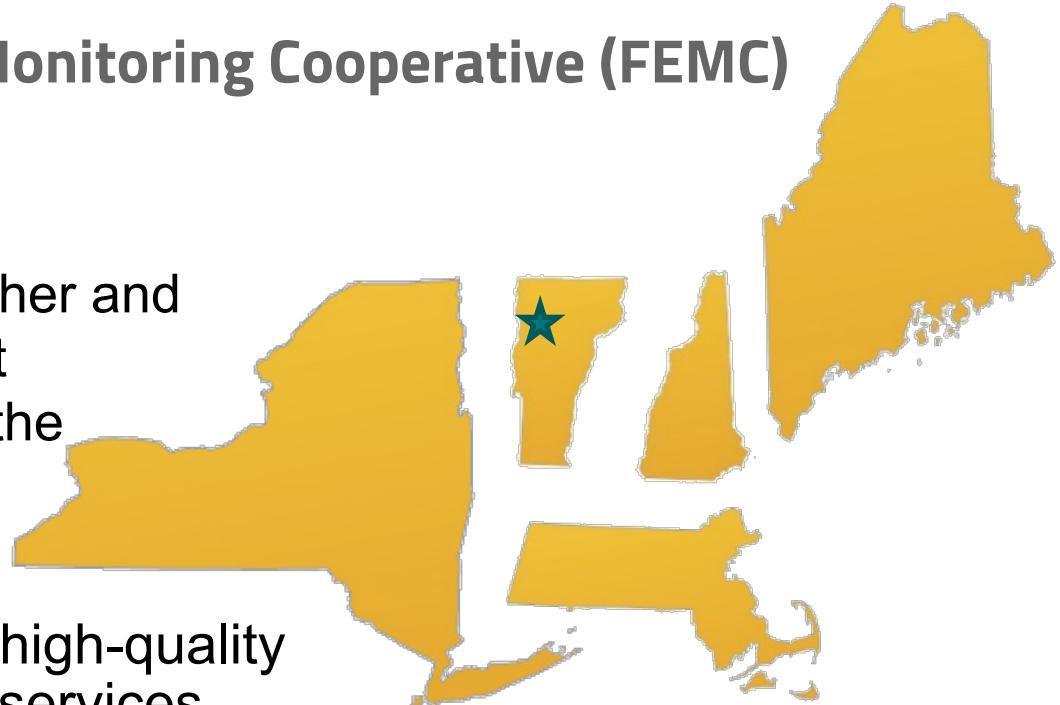


- GIS and Spatial Modeling
- Forest Ecology
- Socioecological Systems



# The Forest Ecosystem Monitoring Cooperative (FEMC)

A cooperative effort to gather and synthesize trends in forest ecosystem health across the region...



By providing unique, nimble, high-quality analytical and monitoring services



The University of Vermont  
**THE RUBENSTEIN SCHOOL**  
OF ENVIRONMENT AND NATURAL RESOURCES



**VERMONT**  
AGENCY OF NATURAL RESOURCES

# The Forest Ecosystem Monitoring Cooperative (FEMC)



Ecosystem  
Monitoring



Collaborator  
Network



Data Archive,  
Access,  
Integration



# Why Participate in DataONE?



Standards + Data + Online

## Key benefits

Discoverability  
Marketability  
Longevity



# The Environmental Data Initiative

Mark Servilla  
University of New Mexico  
8 May 2018



# Who we are...

An open data repository for environmental and ecological data

- Long Term Ecological Research (LTER) Network
- Long Term Research in Environmental Biology (LTREB)
- Macro-systems Biology (MSB)
- Organization of Biological Field Stations (OBFS)

In continuous operation since January 2013

- Expansion of the LTER Network PASTA data repository in July 2016
- 42,700 data packages, 8.5TB of data
  - 6,300 research contributed
  - 21,000 Landsat images
  - 15,400 Ecotrends project
- DataCite Digital Object Identifier allocator
- Recognized by Springer Nature, Scientific Data, ESA, r3data, PloS, and others

# About our data repository...

## Data Package structure

- Science metadata - Ecological Metadata Language
- Science data (1 .. more) - tabular, spatial, database, other BLOBs
- Data package quality report

## Data Package Quality Evaluation

- Metadata validation
- Data validation
- Metadata and data congruence

## Strong Versioning

Based on a Service Oriented Architecture Design Pattern

Accessible via REST web service API

## Our DataONE Member Node(s)...

EDI since 2017, LTER since 2012

Uses DataONE Generic Member Node (Python)

Is true slender node - proxies all resources to PASTA

Is replication home for orphaned Member Node data

- GLEON
- USA\_NPN

**Why we ❤️ DataONE...**



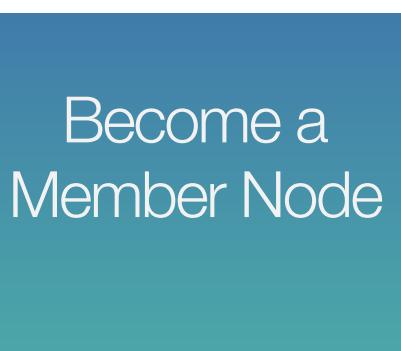
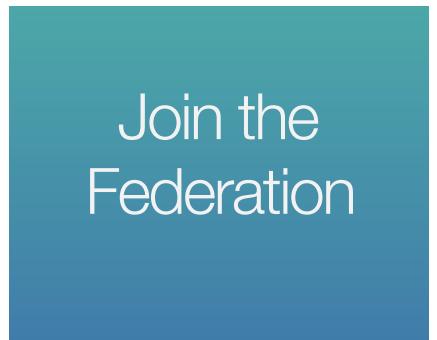
...exposes our customer's data products to the broadest possible audience of scientists, students, and policy makers.

# DataONE Member Nodes

Why participate?



- Increased visibility
- Broader dissemination
- Long-term data management
- Community engagement



Contact us: <https://www.dataone.org/contact>





## DataONE User Group (DUG)

More info:

[Bit.ly/DUG2018](https://bit.ly/DUG2018)

### 2018 DUG Meeting

Monday, July 16<sup>th</sup>  
Tucson, AZ

### DATAONE USERS GROUP MEETING

Building a Community of Scientific Data Repositories  
in an Open Science Landscape

July 16<sup>th</sup> 2018

Marriott University Park, Tucson AZ; co-located with the ESIP meeting

- *Bringing together repository managers and users in support of open science*
- *Community contributed talks and posters*
- *DataONE updates and visioning*
- *Topical breakout sessions and workshops*

Full information at [bit.ly/DUG2018](https://bit.ly/DUG2018)





Questions?

**DataONE**



**FEMC**  
Forest Ecosystem Monitoring Cooperative



Thank you...

Dave Vieglais ([dave.vieglais@gmail.com](mailto:dave.vieglais@gmail.com))

Monica Ihli ([email@monicaihli.com](mailto:email@monicaihli.com))

Amy Forrester ([aforres4@utk.edu](mailto:aforres4@utk.edu))

Kenneth Casey ([kenneth.casey@noaa.gov](mailto:kenneth.casey@noaa.gov))

James Duncan ([James.Duncan@uvm.edu](mailto:James.Duncan@uvm.edu))

Mark Servilla ([mark.servilla@gmail.com](mailto:mark.servilla@gmail.com))