



A global federation of interoperable data repositories

Matthew B. Jones

National Center for Ecological Analysis and Synthesis
University of California Santa Barbara

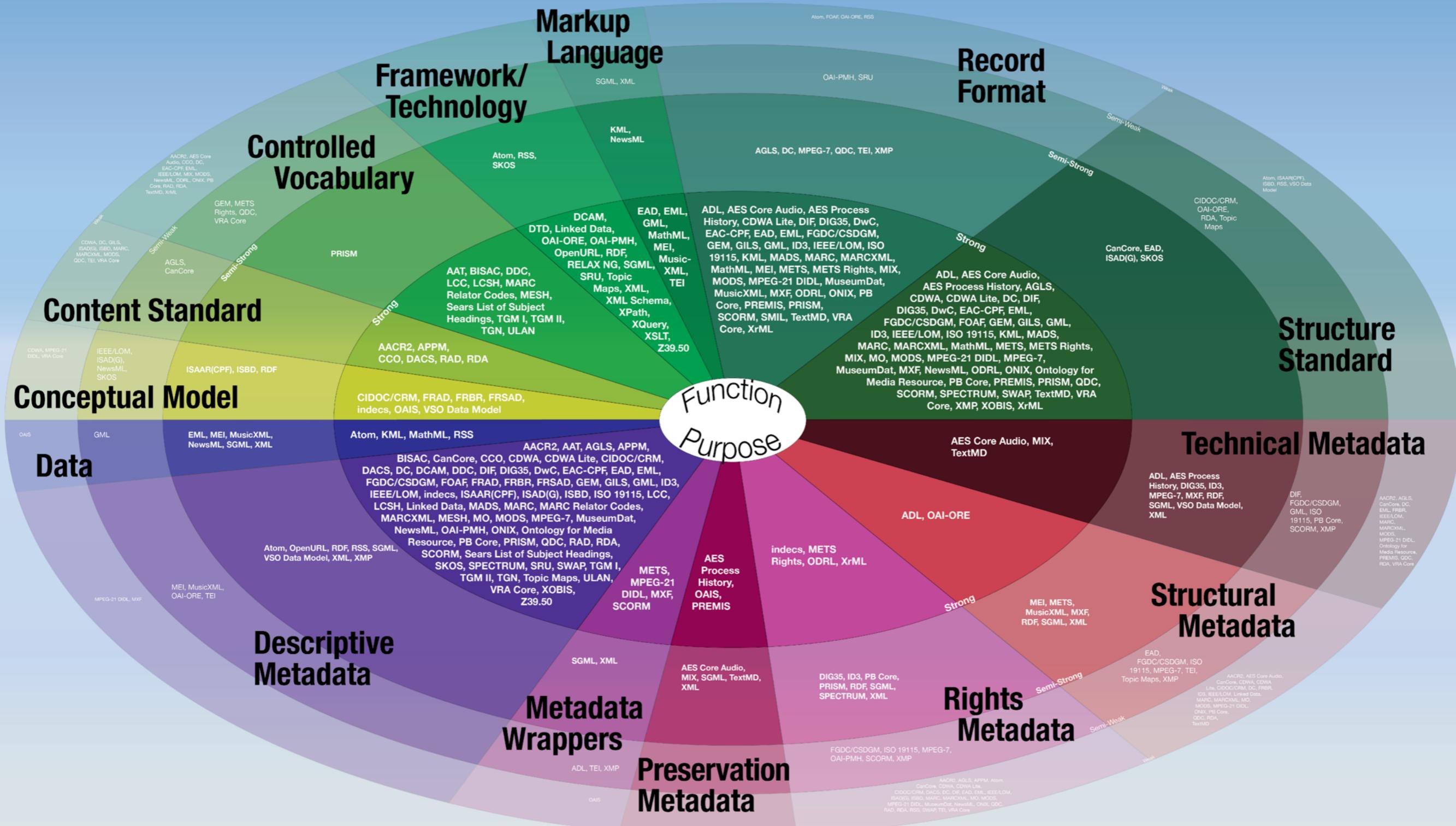


 0000-0003-0077-4738

@metamattj

jones@nceas.ucsb.edu

Diverse data and standards

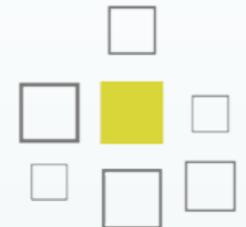


Diverse institutions



- Over 1000 repositories registered
 - Life sciences: 562
 - 85 for “Ecology”
 - Natural Sciences: 566
 - Humanities and Social Science: 329
 - Engineering Sciences: 134

Diverse software



D SPACE



The DataONE Vision and Approach

Facilitating long term access to earth sciences data through infrastructure, tools, and education.

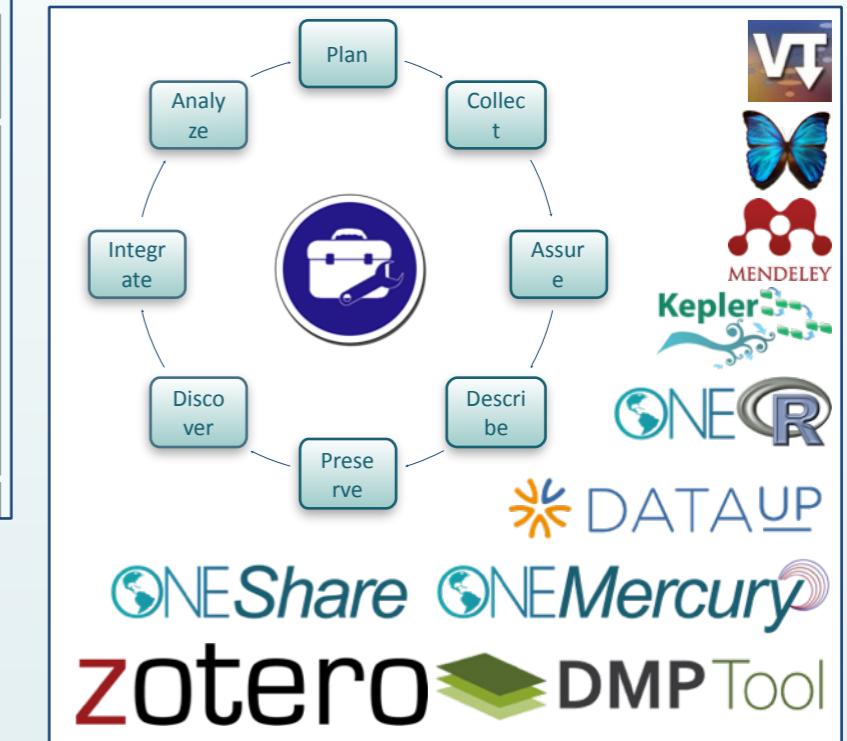
1. Build community



2. Develop sustainable data discovery and interoperability solutions

A screenshot of the DataONE ONEMercury search tool. The interface includes a header with links for About, Participate, Resources, Education, and Data. Below the header is a search bar with fields for "Search For:" and "Results/Page". It features sections for "Fielded Search", "Date Search", and "Geographic Search" which includes a world map. There are also filters for "List Areas in:" and "Search Area:". At the bottom, there are links for "Member Nodes" and "All SANParks Data Repository".

3. Enable science through tools and services



Working with our Community

stakeholder surveys



scientists



libraries & librarians



data managers

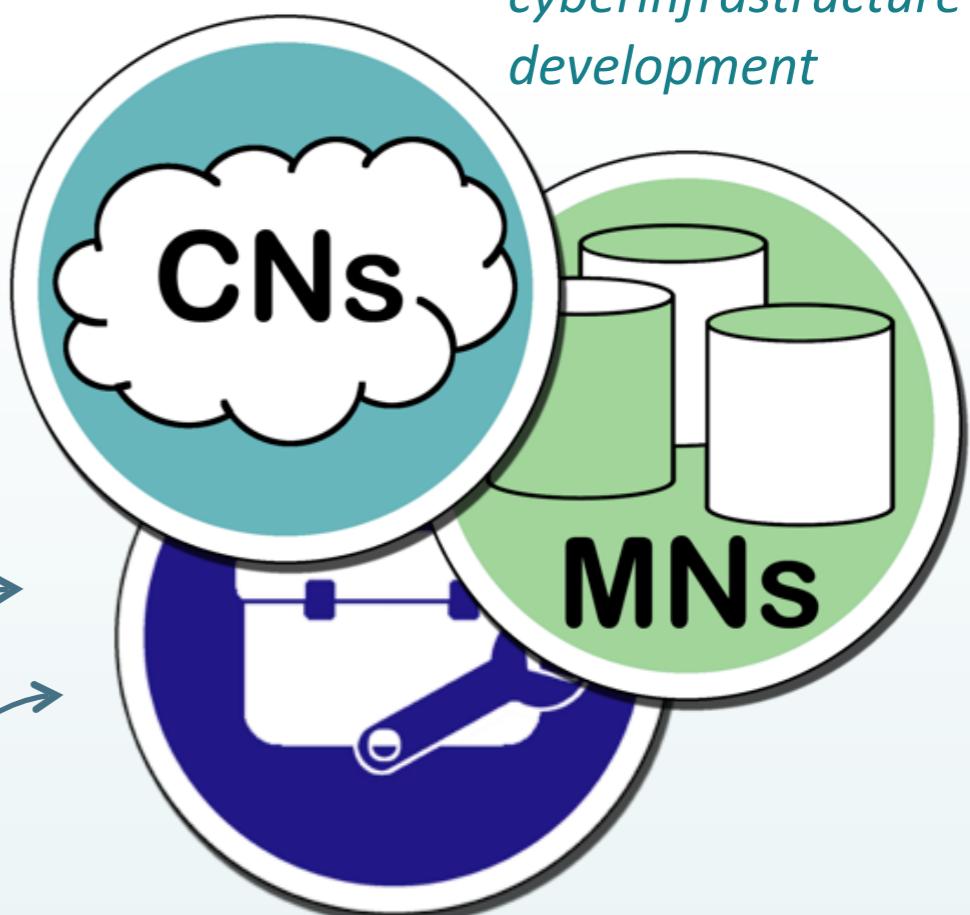
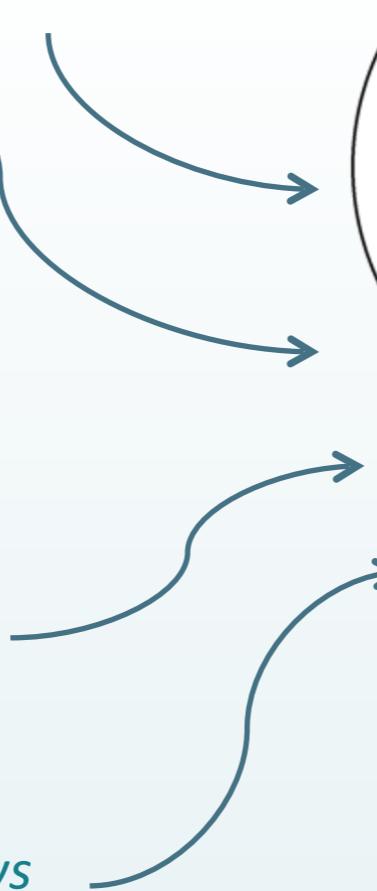


persona and scenario development

cyberinfrastructure development

usability testing

New DataONE Education Resources
DataONE provides CCO access to a suite of ten training modules that support your data management teaching. Modules cover all aspects of the DataONE system, from collection, description, QAQC, analysis and preservation ...



external assessments / surveys

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

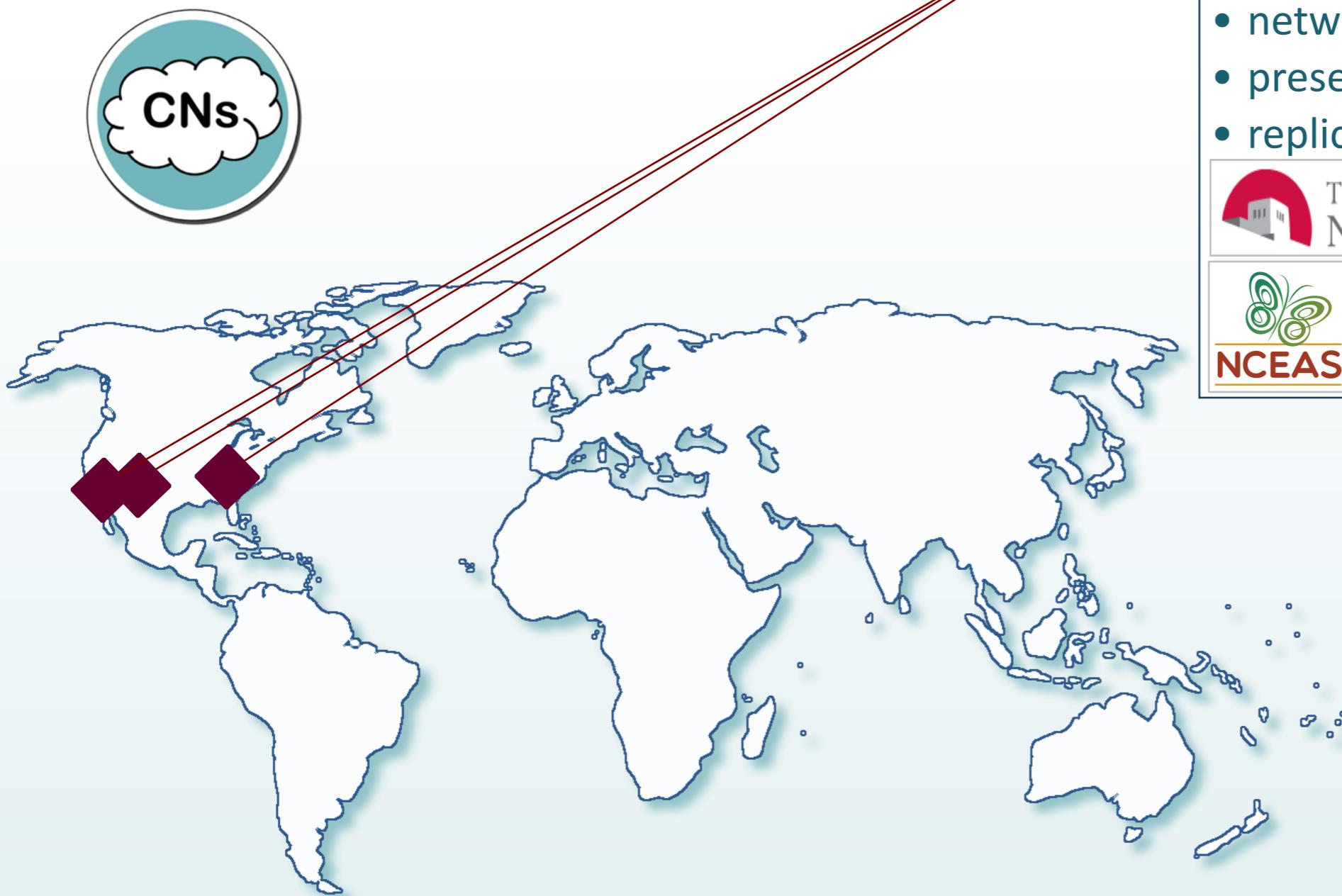
Data Sharing by Scientists: Practices and Perceptions

Carol Tenopir, Suzie Allard, Kimberly Douglass, Arsev Umar Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, Mike Frame

10,244 VIEWS 21 CITATIONS 189 ACADEMIC BOOKMARKS 40 SOCIAL SHARES

DataONE: Cyberinfrastructure

Three major components for a flexible, scalable, sustainable network



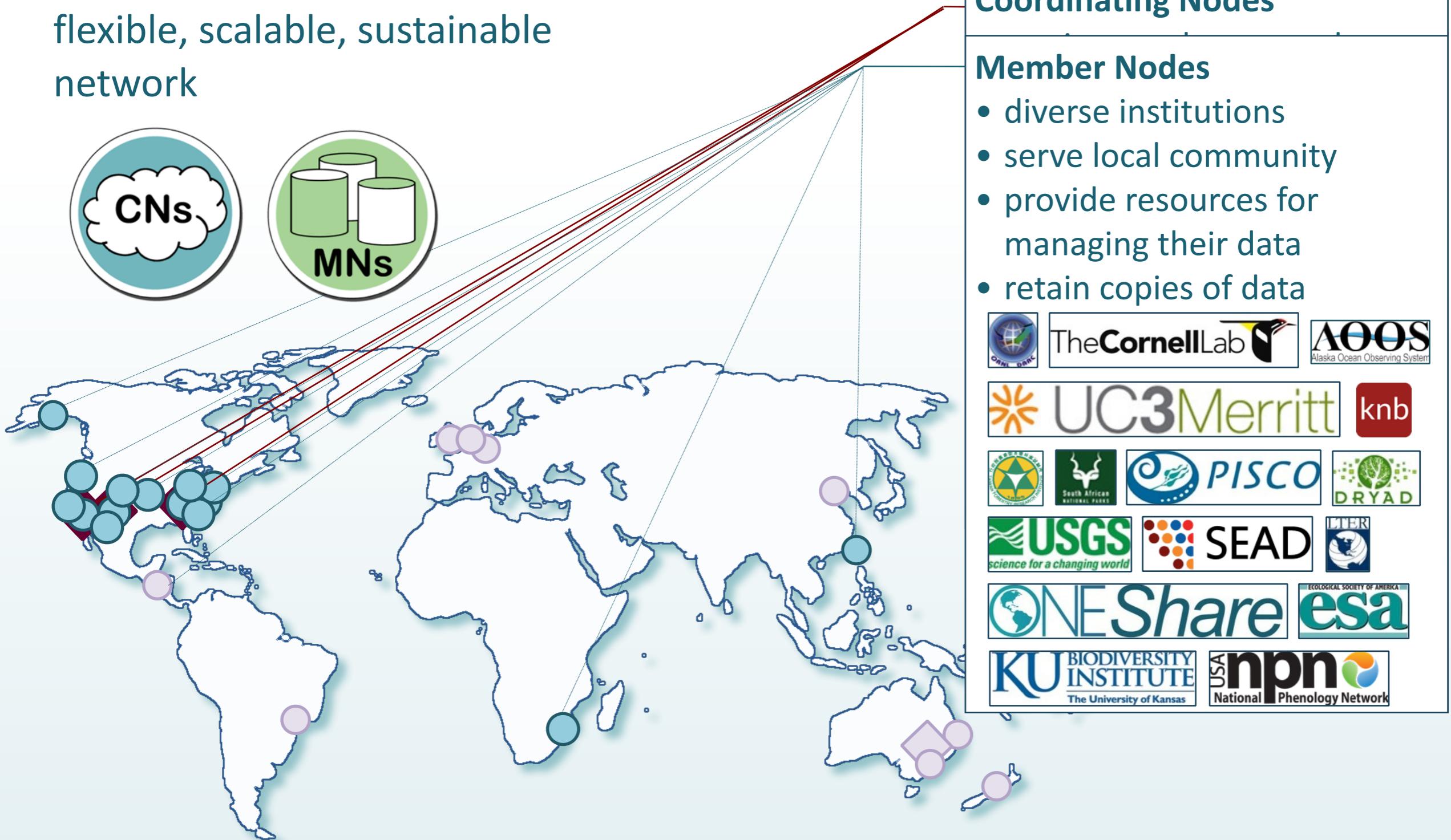
Coordinating Nodes

- metadata catalog
- search and discovery
- network-wide services
- preservation services
- replication services



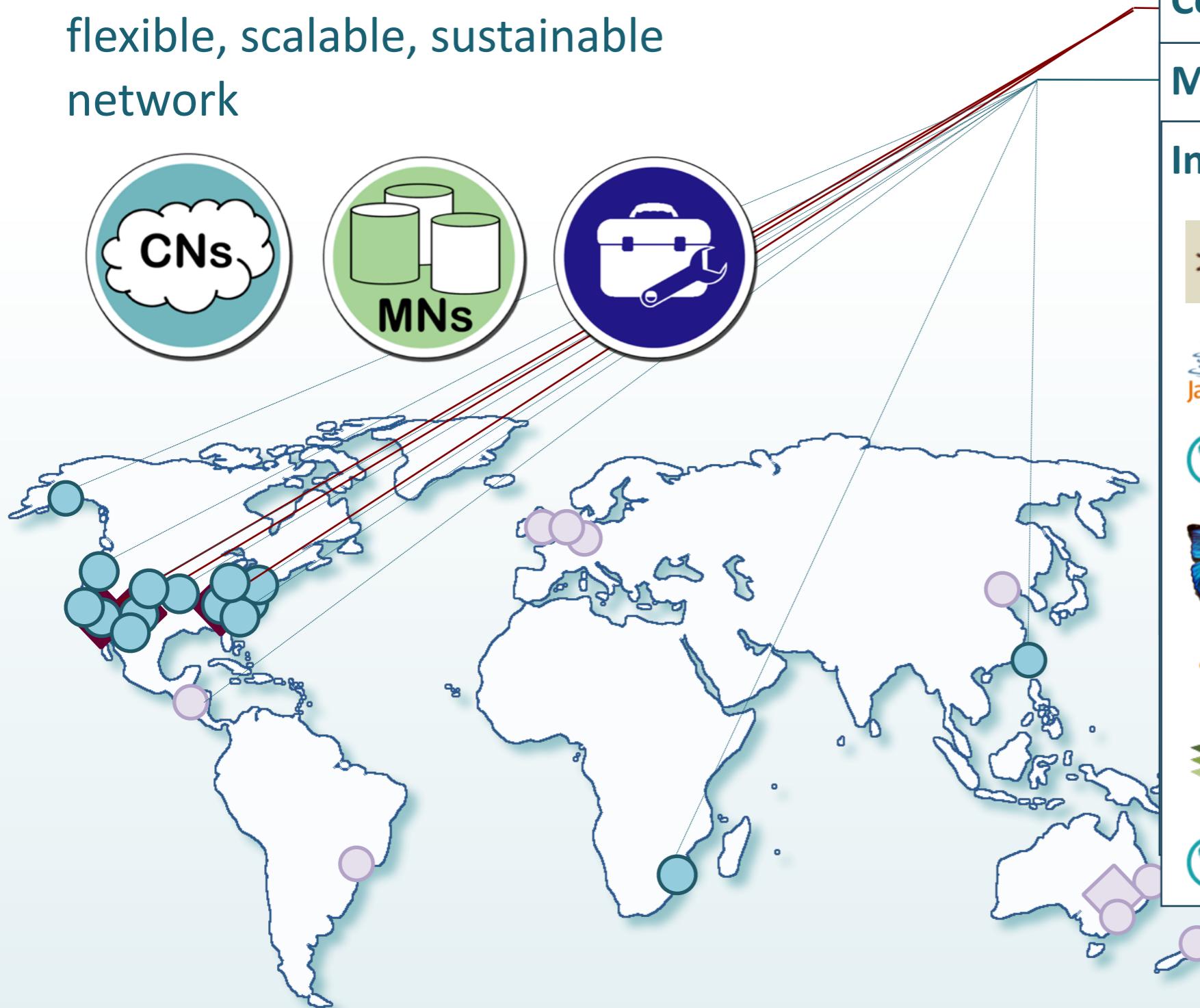
DataONE: Cyberinfrastructure

Three major components for a flexible, scalable, sustainable network



DataONE: Infrastructure and Tools

Three major components for a flexible, scalable, sustainable network



Coordinating Nodes

Member Nodes

Investigator Toolkit

>> command line interface



ONEDrive zotero



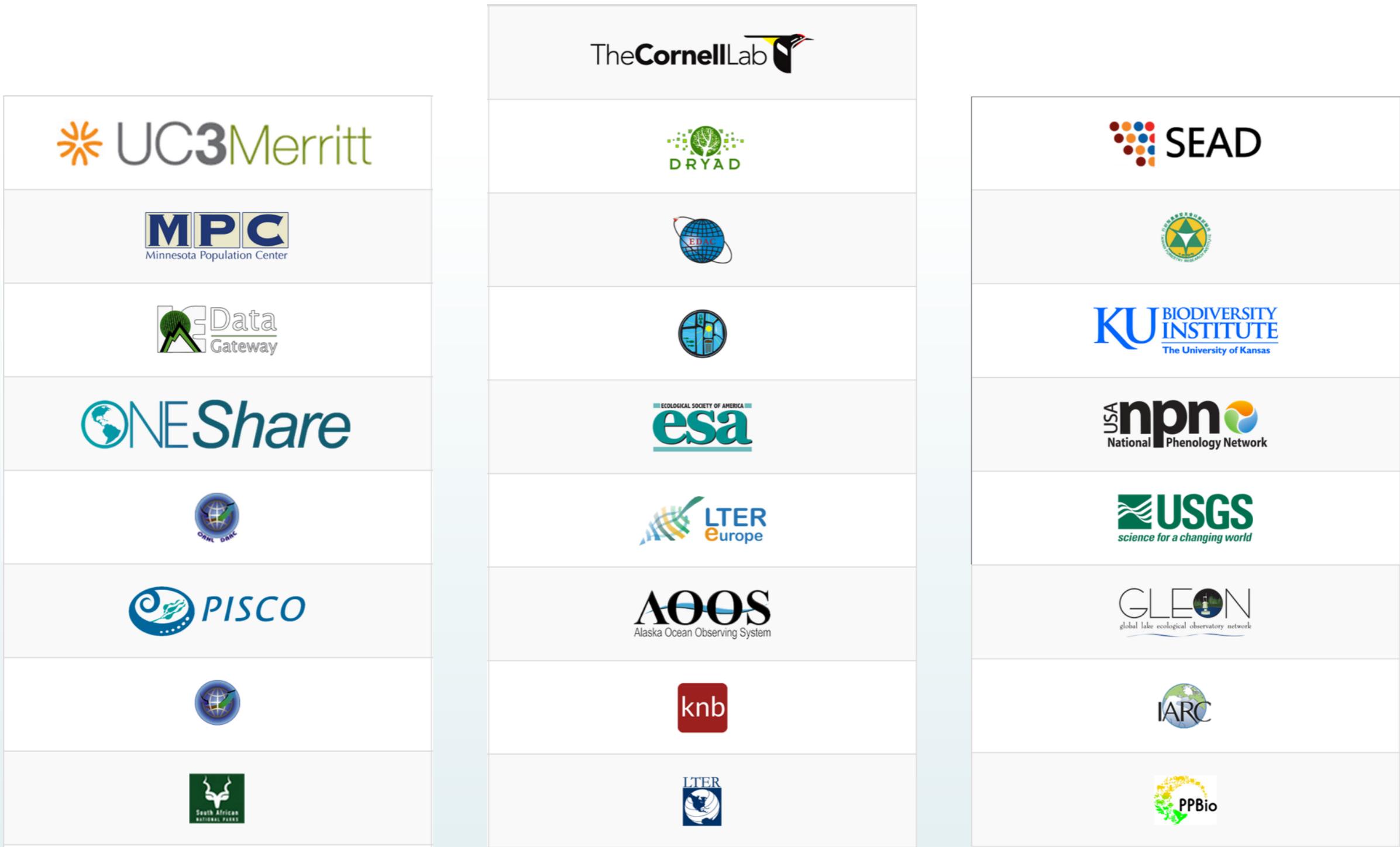
DATAUP



ONEMercury VT

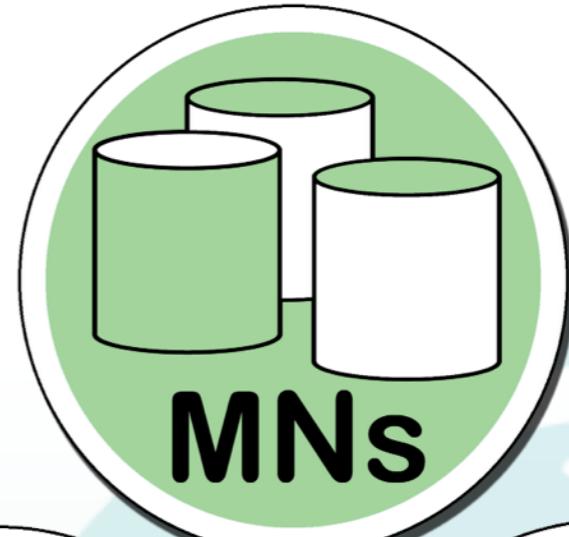
25 DataONE Repository Members

North America (21), Europe (1), Africa (1), Asia (1), South America (1)



- **Member Nodes (MNs)**

- Authoritative data stores
- Harness the power of local curation



- **Coordinating Nodes (CNs)**

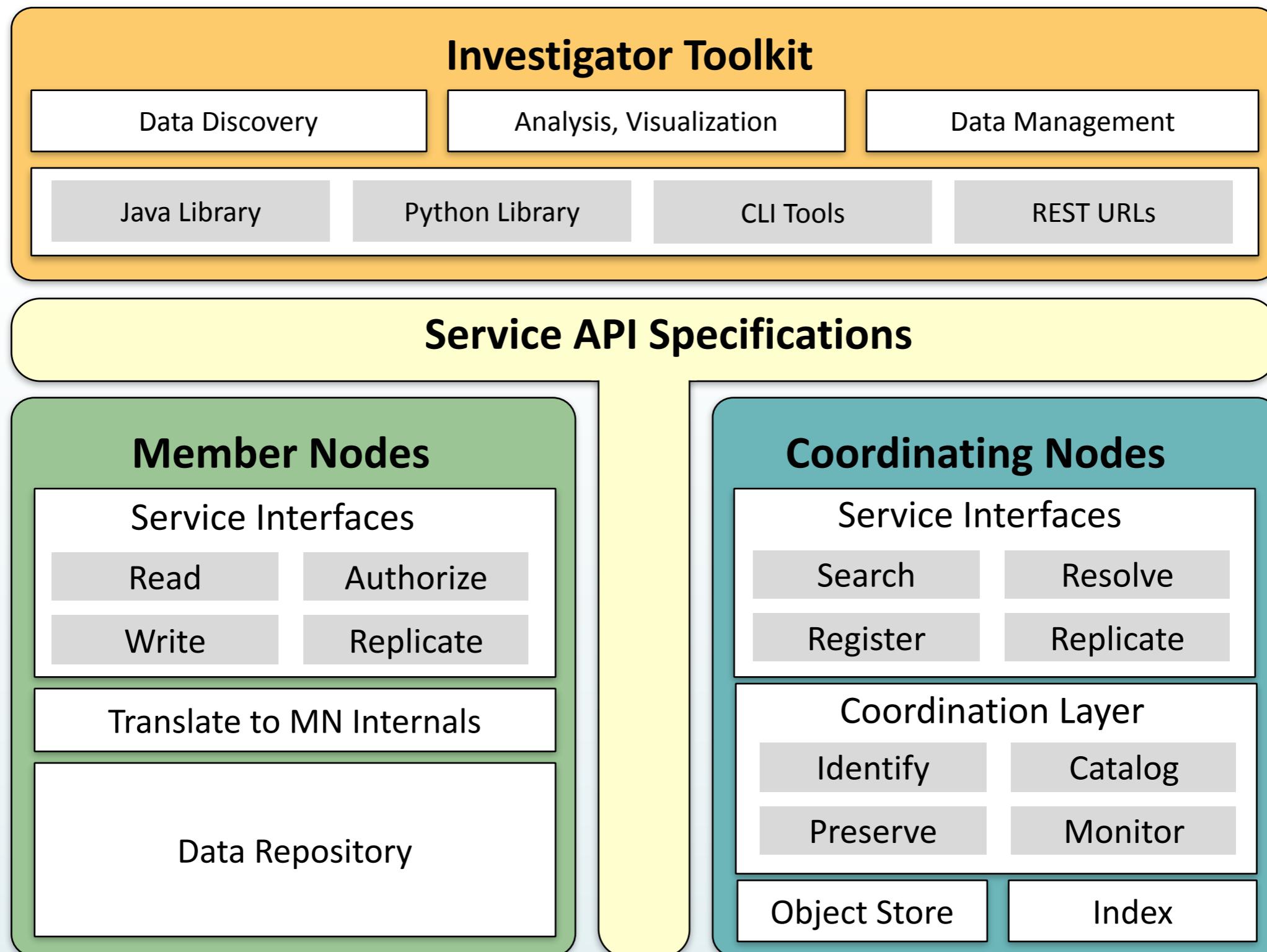
- Services to link Member Nodes
- REST APIs

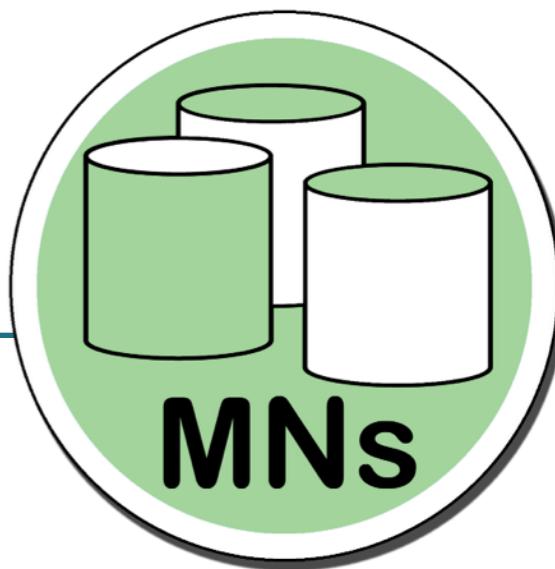


- **Investigator Toolkit (ITK)**

- Tools for the whole data lifecycle

Interoperability
API





- **Authoritative members of the Federation**
- **Curate data holdings**
 - *Provide unique identifiers for each object*
 - *Ensure availability, quality, and reliability*
- **Replicate holdings for other MNs**
- **Provide access and access control**
- **Log and report accesses to objects**
- **Engage with DataONE community**
- **Deploy a DataONE-compatible software system**



Citation and usage tracking



knb

ABOUT

DATA

SHARE

TOOLS

Search for data



SIGN IN

NCEAS 2058: Costanza: Value of the world's ecosystem services and natural capital: Toward a dynamic, integrated approach , National Center for Ecological Analysis and Synthesis , and Costanza R. **Value of the world's ecosystem services and natural capital datasets** ([doi:10.5063/AA/nceas.183.13](https://doi.org/10.5063/AA/nceas.183.13))

Files

	Package	resourceMa...	.zip	85.38 KB	
	metadata	doi:10.5063/A...	.xml (EML	17.32 KB	7154 views
	data	doi:10.5063/A...	application/octet-stream	180 B	850 downloads
	data	doi:10.5063/A...	application/octet-stream	67.88 KB	785 downloads

[Download all](#)

General

Title

[Value of the world's ecosystem services and natural capital datasets](#)

Identifier

[nceas.183.13](#)

Abstract

The services of ecological systems and the natural capital stocks that produce them are critical to the functioning of the Earth's life-support system. They contribute to human welfare, both directly and indirectly, and therefore represent part of the total

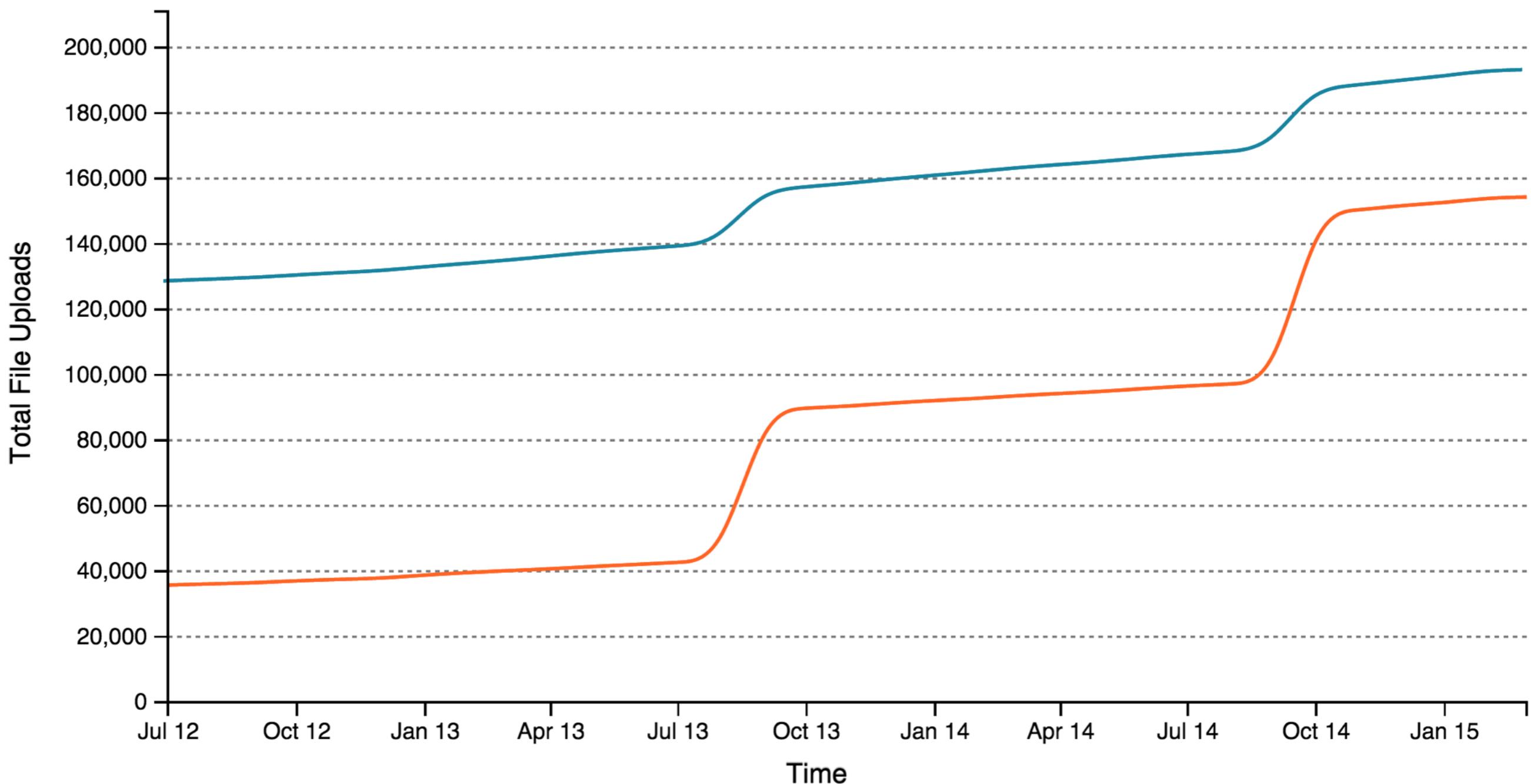
Current Member Nodes Summary

154,341

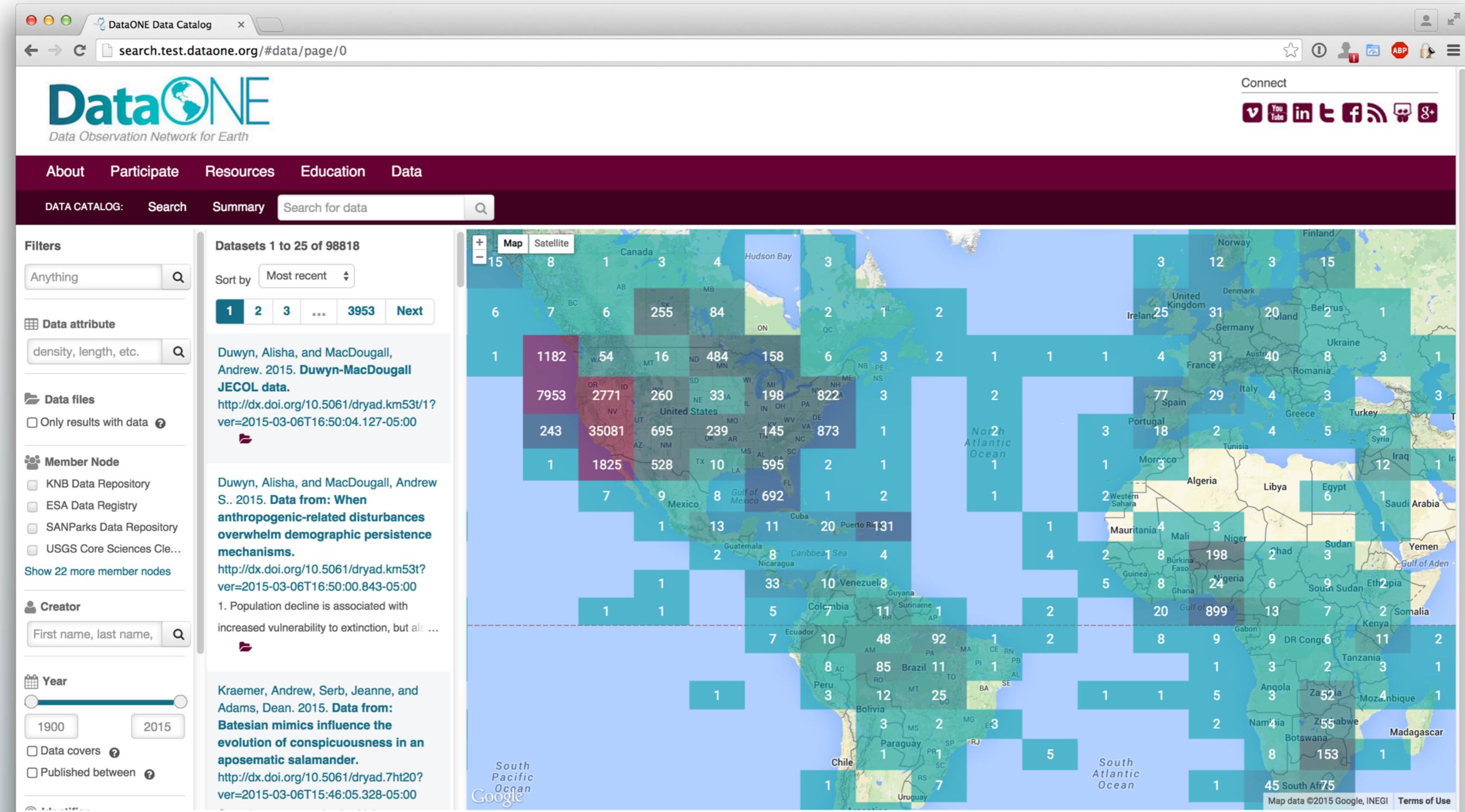
Data Files (1.5 TB)

191,858

Metadata Files (2.6 GB)



Data discovery



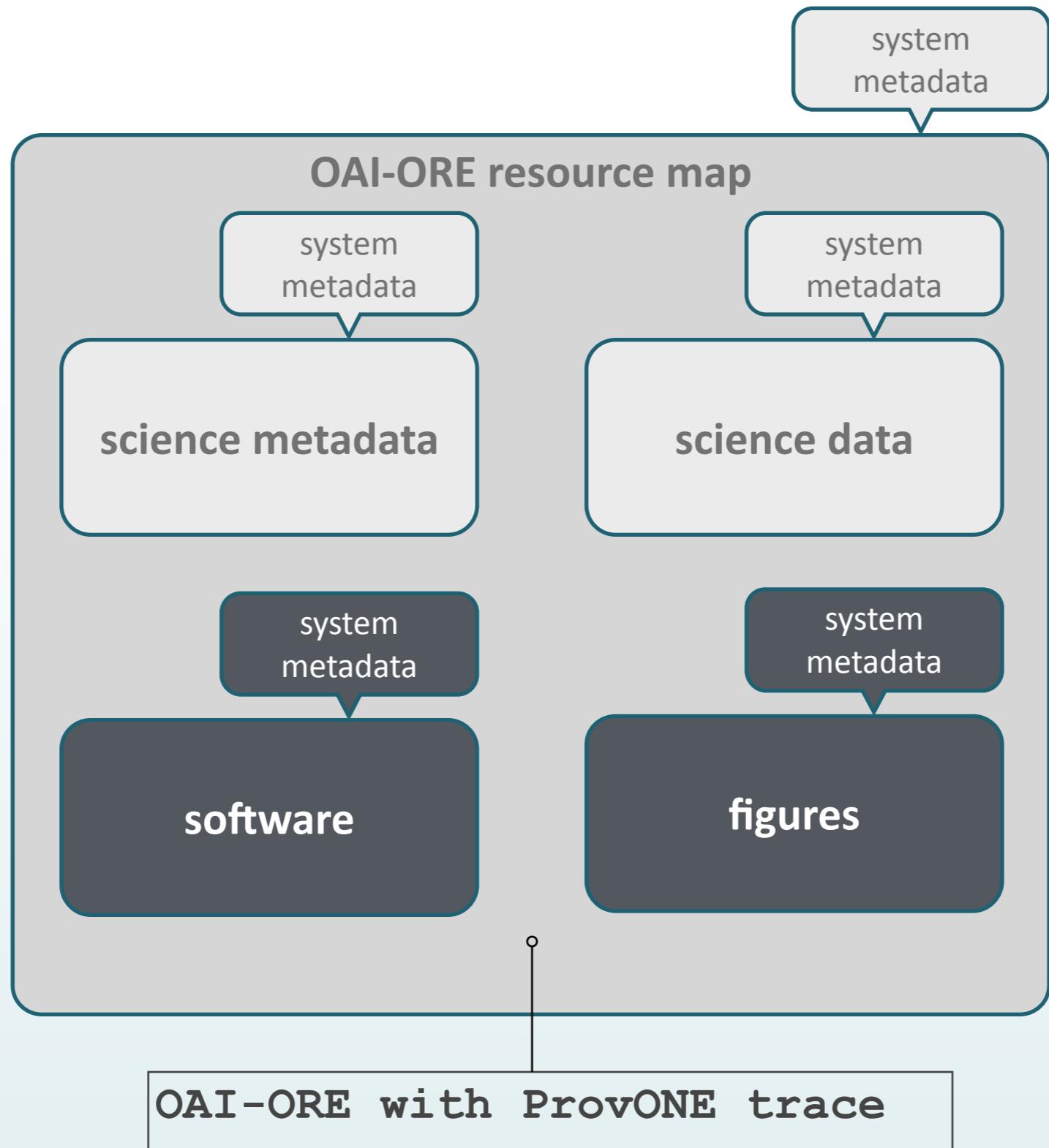
Provenance: time travel in DataONE

- Phase II Goal: Facilitate reproducible science
 - Track **data derivation** history
 - Track data **inputs** and **outputs** of analyses
 - Track analysis and model **executions**
 - Preserve and document **software**
 - Link all of these to **publications**

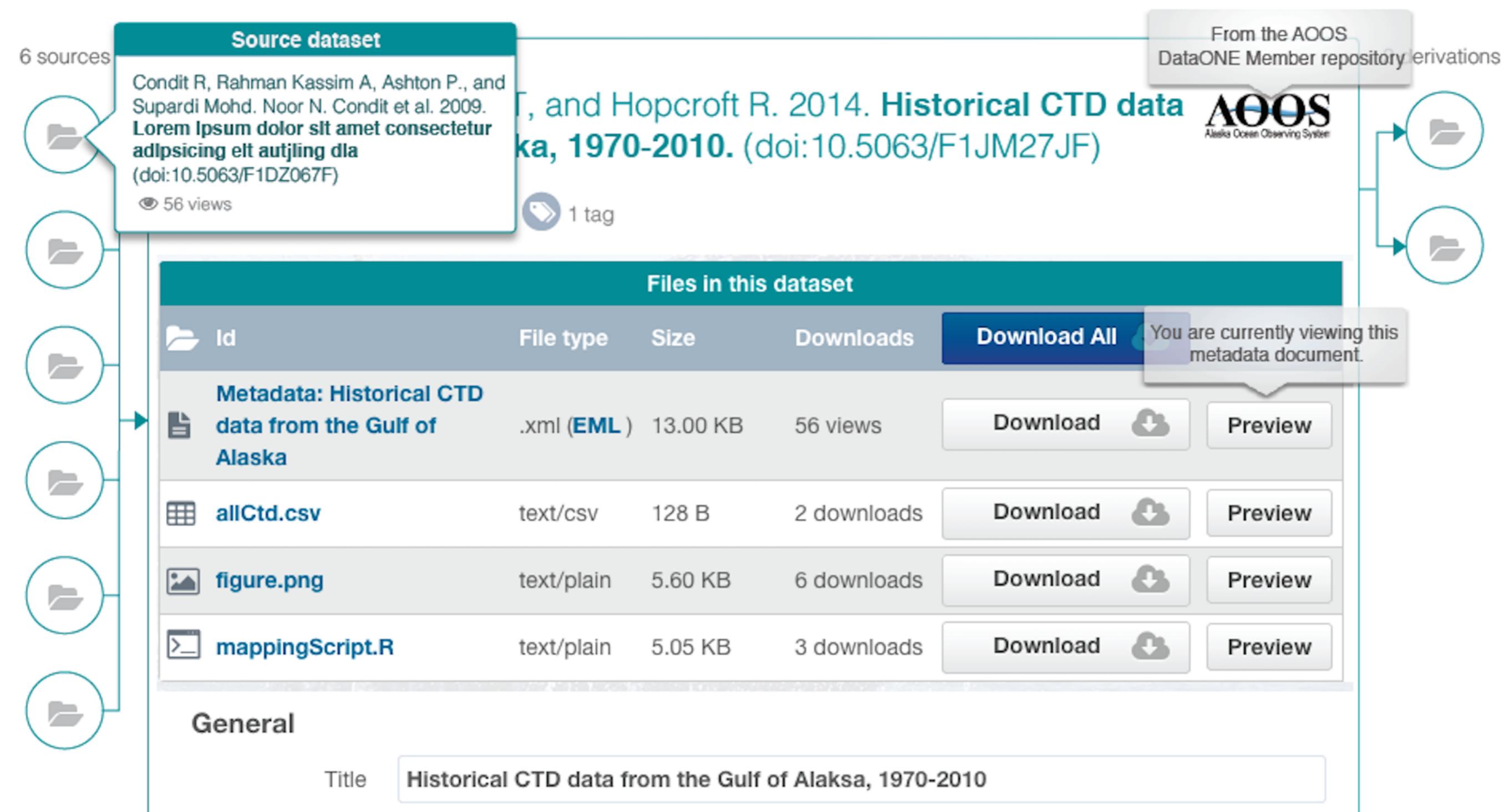
Provenance at-a-glance

- **ProvONE extensions to W3C PROV**
- **New first-class objects in DataONE**
 - Figures
 - Software
- **Extended Data Package with Provenance**
- **New Tools**
 - Web user interface for browsing provenance
 - Matlab tool for generating provenance
 - R tool for generating provenance

New objects, ProvONE trace



Provenance: Data derivation



Provenance: Code execution

This data was used by the program
[mappingScript.R](#) to generate the image
you are currently viewing, figure.png.

This data is in the dataset you are currently viewing.

Weingartner T, Royer T, Hopcroft R. 2014.
Historical CTD Data from the Gulf of Alaska,
1970-2010. (doi:10.5063/FJM27JF)

Accuracy
Assessment

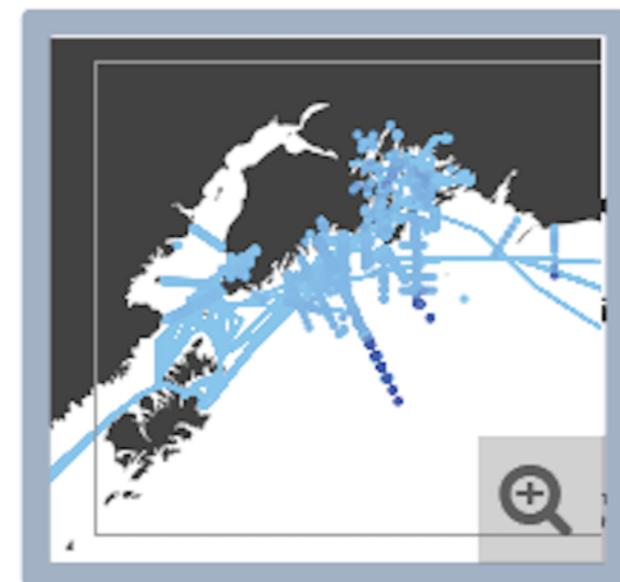
Coverage

Method



[figure.png](#) (ctdImage20141015371)

56 views



[Download](#)

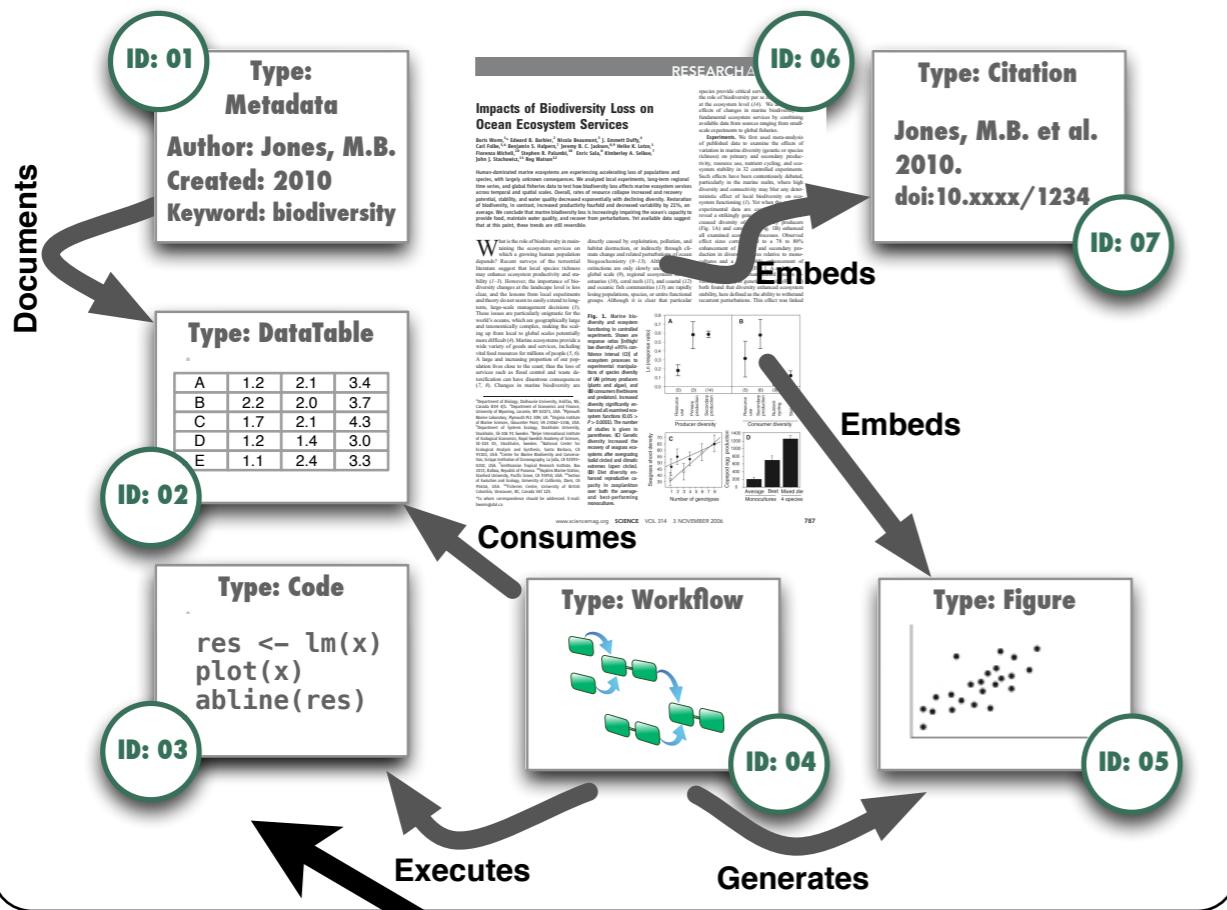
Entity Name

figure.png

Description

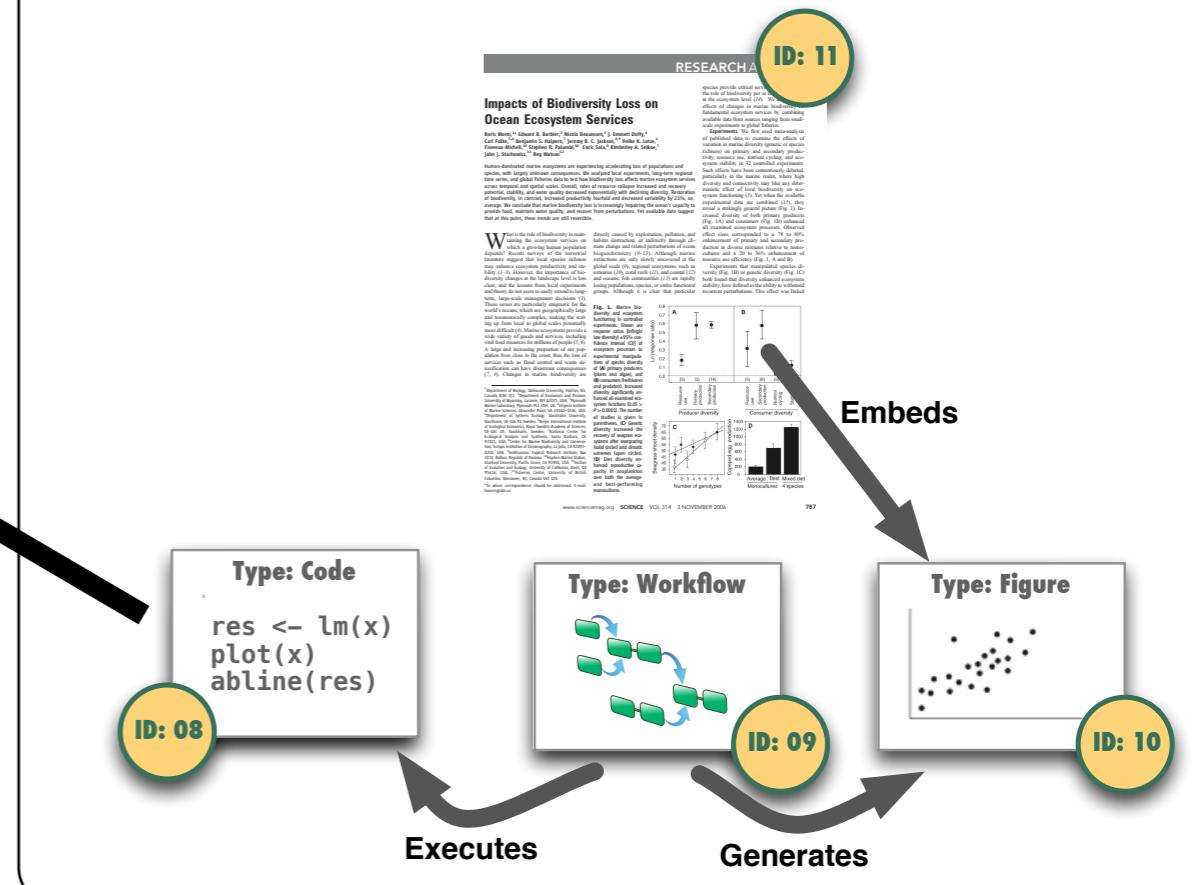
plot

Publication: version 1

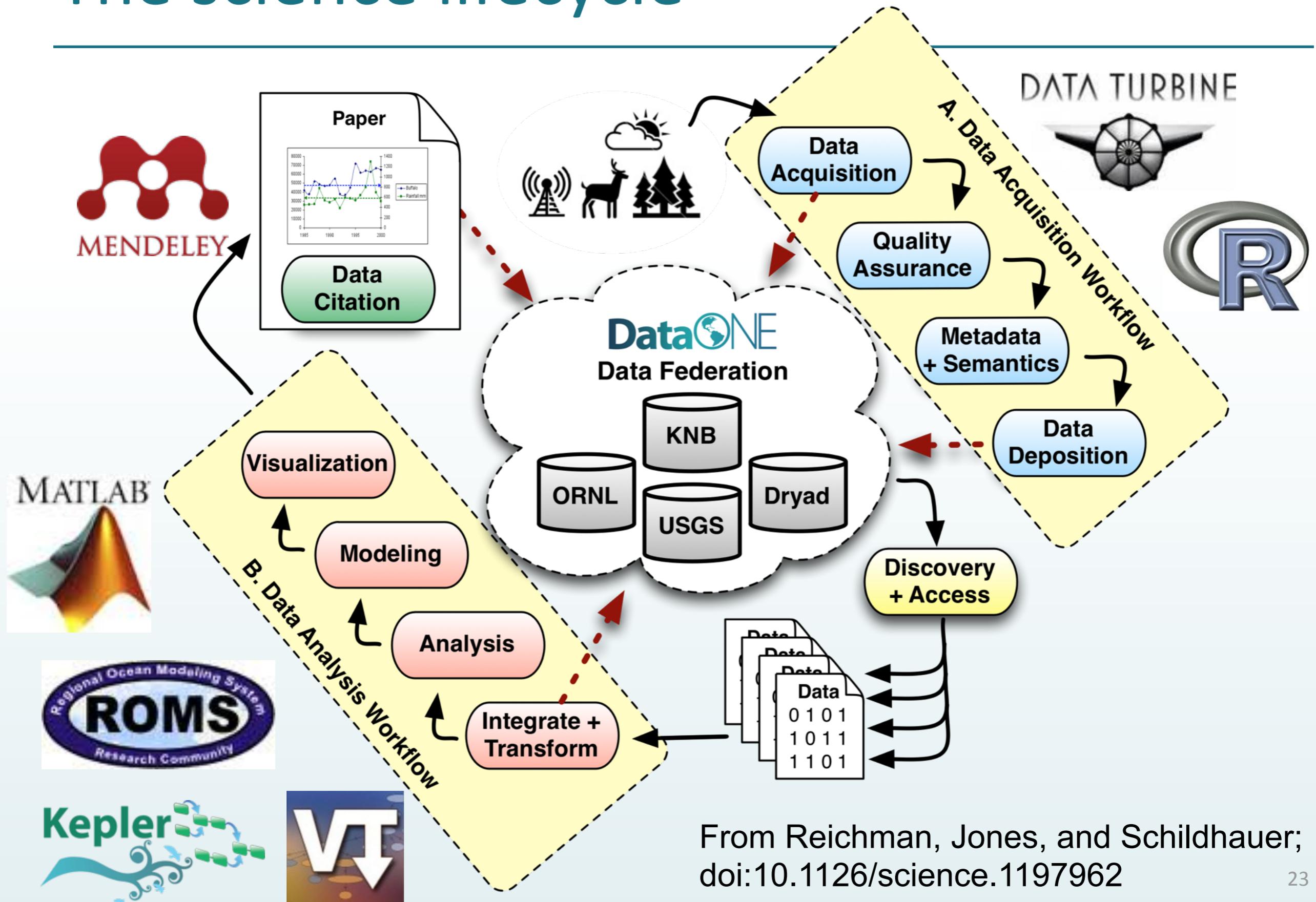


Replaces

Publication: version 2



The science lifecycle



Resources

- <http://www.dataone.org>
- [#dataone">irc.ecoinformatics.org #dataone](irc.ecoinformatics.org)
- developers@dataone.org
- <http://mule1.dataone.org/ArchitectureDocs-current/>

The screenshot shows the DataONE website. At the top is the DataONE logo with the tagline "Data Observation Network for Earth". To the right is a search bar with dropdown menus for "Search" (set to "ONEMercury") and "For", and a "Go" button. Further right is a "Connect" section with icons for YouTube, LinkedIn, Twitter, Facebook, and a feed symbol. Below the header is a dark red navigation bar with links for "About", "Participate", "Resources", "Education", and "Data". The "Participate" link is underlined, indicating the current page. A breadcrumb trail "Home » Participate » Developer Resources" is visible. The main content area has a title "Developer Resources" and a paragraph about contributions and collaborations. On the left sidebar, there are links for "DataONE Users Group", "Member Nodes", "Internships", "Developer Resources" (which is also the active page), and "Open Positions".

DataONE welcomes contributions and collaborations with all who are interested in helping the goals of meeting the needs of science and society for open, persistent, robust, and secure access to well-described and easily discovered Earth observational data. This page describes resources that are available to those interested in learning more about the cyberinfrastructure for DataONE, particularly for contributing to DataONE or using the products that DataONE has developed in other efforts. DataONE code is generally available under an Apache 2.0 license and our code repository.

DataONE Team and Sponsors



- Amber Budden, Roger Dahl, Rebecca Koskela, Bill Michener, Robert Nahf, Skye Roseboom, Mark Servilla



• Ewa Deelman



- Dave Vieglais



• Deborah McGuinness



- Suzie Allard, Kimberly Douglass, Laura Moyers, Carol Tenopir, Robert Waltz, Bruce Wilson



• Jeff Horsburgh



- John Cobb, Bob Cook, Ranjeet Devarakonda, Giri Palanisamy, Line Pouchard



• Robert Sandusky



- Patricia Cruse, John Kunze



• Bertram Ludaescher



- Sky Bristol, Mike Frame, Richard Huffine, Viv Hutchison, Jeff Morisette, Jake Weltzin, Lisa Zolly



• Peter Honeyman



- Chris Jones, Matt Jones, Ben Leinfelder, Mark Schildhauer, Jing Tao, Lauren Walker, Peter Slaughter, {Stephanie Hampton, Chad Berkley, Andrew Pippin}



• Cliff Duke



- Paul Allen, Rick Bonney, Steve Kelling



• Carole Goble



- Jane Greenberg, Ryan Scherle, Todd Vision



• Donald Hobern



- Randy Butler



• David DeRoure



- Paolo Missier

