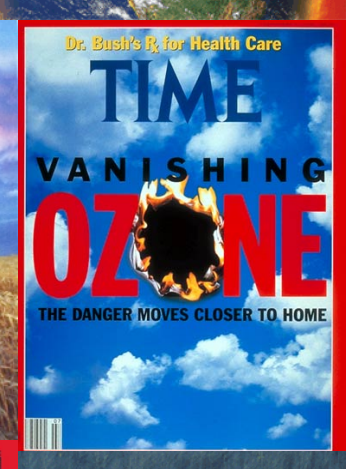


DataONE Overview

(Data Observation Network for Earth)

IDCC14
San Francisco, CA
February 2014





February 22, 2013

FROM: John P. Holdren
Director

1. Policy Principles

Scientific research supported by the Federal Government catalyzes innovative breakthroughs that drive our economy. The results of that research become the grist for new insights and are assets for progress in areas such as health, energy, the environment, agriculture, and national security.

To that end, I have issued a memorandum today ([pdf](#)) to Federal agencies that directs those with more than \$100 million in research and development expenditures to develop plans to make the results of federally-funded research publicly available free of charge within 12 months after original publication.

...the memorandum requires that agencies start to address the need to improve upon the management and sharing of scientific data produced with Federal funding.

The screenshot shows the top portion of the White House website. At the top, the text "the WHITE HOUSE" is followed by "PRESIDENT BARACK OBAMA" and five stars. To the right is the White House seal. Below this is a dark blue navigation bar with links: "BLOG" (with a US flag icon), "PHOTOS & VIDEO", "BRIEFING ROOM", "ISSUES", and "the ADMINISTRATION" (with an eagle icon). The main content area has a green background with the large text "WE the PEOPLE" and the subtitle "YOUR VOICE IN OUR GOVERNMENT". At the bottom is a dark green bar with a home icon and four links: "CREATE A PETITION", "OPEN PETITIONS", "RESPONSES", and "HOW & WHY".

Require free access over the Internet to scientific journal articles arising from taxpayer-funded research.

Increasing Public Access to the Results of Scientific Research

By Dr. John Holdren

Thank you for your participation in the We the People platform. The Obama Administration agrees that citizens deserve easy access to the results of research their tax dollars have paid for. As you may





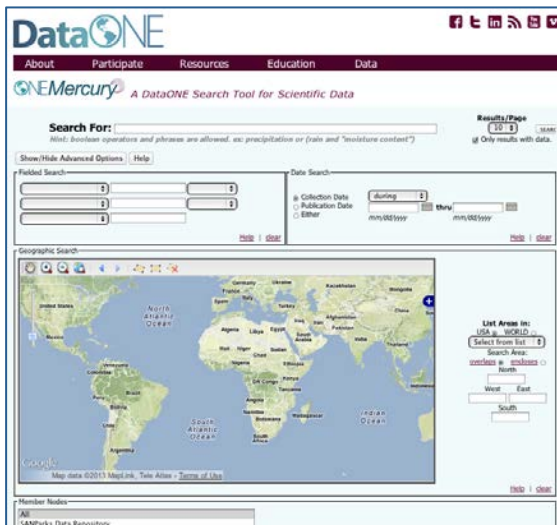
The DataONE Vision and Approach:

Providing universal access to data about life on earth and the environment that sustains it, as well as the tools needed by researchers

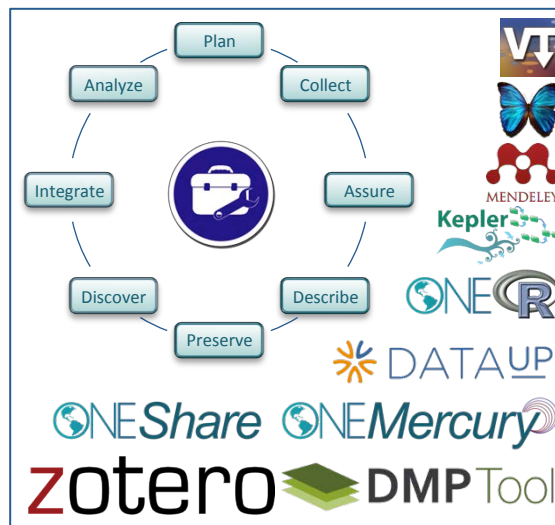
1. Building community



2. Developing sustainable data discovery and interoperability solutions



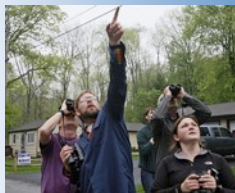
3. Enabling science through tools and services



Enabling Scientific Discovery



eBird



Land Cover



Meteorology

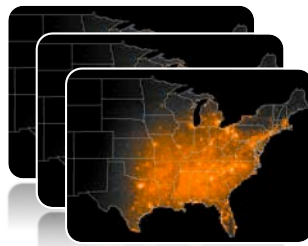


MODIS –
Remote
sensing data



The Cornell Lab
of Ornithology

Diverse bird observations and
environmental data from
300,00 locations in the US
integrated and analyzed using
High Performance Computing
Resources



$$F(X, s, t) = \frac{1}{n(s, t)} \sum_{i=1}^n f_i(X, s, t) I(s, t \in \theta_i)$$

Spatio-Temporal Exploratory
Model identifies factors
affecting patterns of migration

DataONE

Model results

Occurrence of Indigo Bunting (2008)

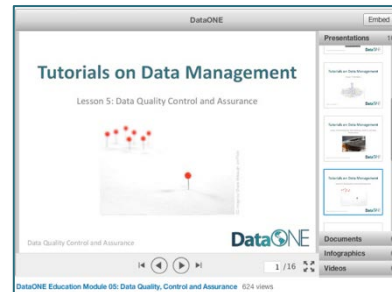
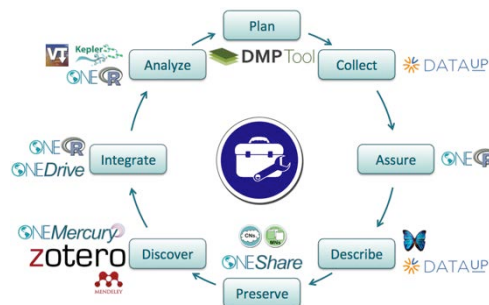
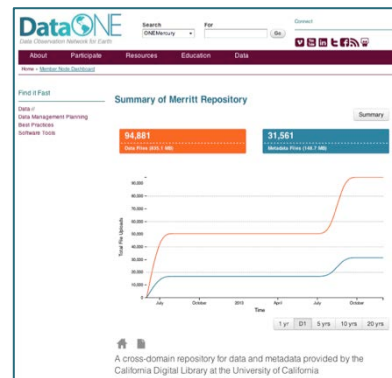


- Examine patterns of migration
- Infer how climate change may affect bird migration



DataONE Approach

- Understand and engage the community
- Provide interoperability solutions that benefit repositories and scientists
- Support the data life cycle





Researcher Challenges



Listening to the Community

stakeholder surveys



scientists



library's & librarians

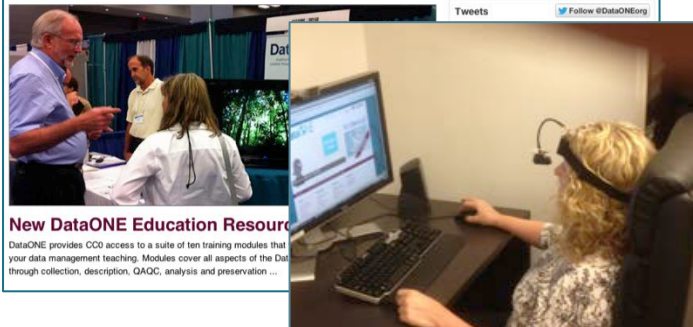
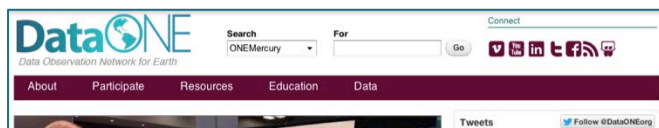


data managers



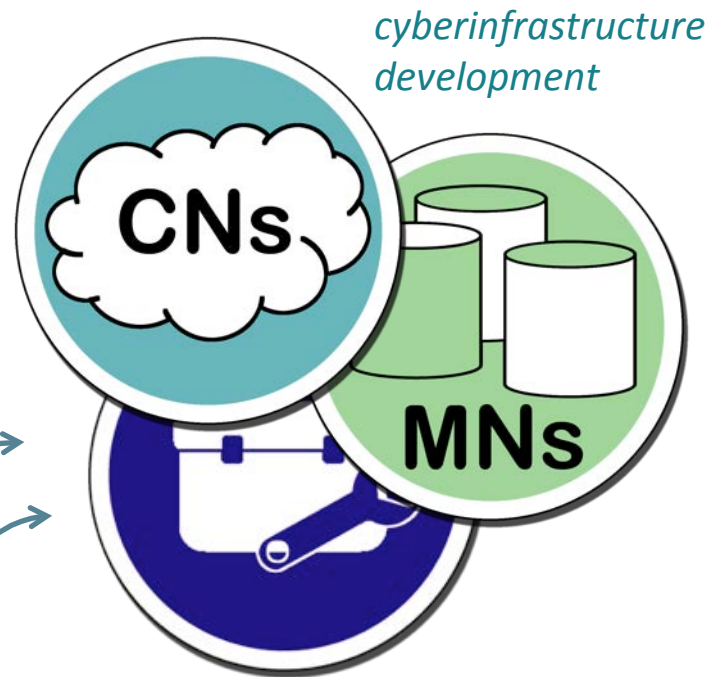
persona and scenario development

usability testing



external assessments / surveys

PLOS ONE		Articles	For Authors	About Us	Search
OPEN ACCESS		advanced search			
PEER-REVIEWED		10,244	21	189	40
		VIEWS	CITATIONS	ACADEMIC BOOKMARKS	SOCIAL SHARES
RESEARCH ARTICLE					
Data Sharing by Scientists: Practices and Perceptions					
Carol Tenopir, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, Mike Frame					





Understanding the Community

OPEN ACCESS Freely available online



Data Sharing by Scientists: Practices and Perceptions

Carol Tenopir^{1*}, Suzie Allard¹, Kimberly Douglass¹, Arsev Umur Aydinoglu¹, Lei Wu¹, Eleanor Read², Maribeth Manoff², Mike Frame³

¹ School of Information Sciences, University of Tennessee, Knoxville, Tennessee, United States of America, ² University of Tennessee Libraries, University of Tennessee, Knoxville, Tennessee, United States of America, ³ Center for Biological Informatics, United States Geological Survey, Oak Ridge, Tennessee, United States of America

Abstract

Background: Scientific research in the 21st century is more data intensive and collaborative than in the past. It is important to study the data practices of researchers – data accessibility, discovery, re-use, preservation and, particularly, data sharing. Data sharing is a valuable part of the scientific method allowing for verification of results and extending research from prior results.

Methodology/Principal Findings: A total of 1329 scientists participated in this survey exploring current data sharing practices and perceptions of the barriers and enablers of data sharing. Scientists do not make their data electronically available to others for various reasons, including insufficient time and lack of funding. Most respondents are satisfied with their current processes for the initial and short-term parts of the data or research lifecycle (collecting their research data; searching for, describing or cataloging, analyzing, and short-term storage of their data) but are not satisfied with long-term data preservation. Many organizations do not provide support to their researchers for data management both in the short-

- “More than half of the respondents (**56%**) reported that they did not use any metadata standard and about **22%** of respondents indicated they used their own lab metadata standard.”
- **Less than 6%** of scientists are making “All” of their data available via some mechanism.
- **2/3rd** report that organizational help and support is lacking



Engaging the Community

Building Community

- Working Groups
- All Hands' Meetings
- External Advisory Board
- DataONE Users Group
- MN "office hours"

Key Collaborations

Education

Communication

- Newsletter
- Web presence
- Conferences
- Social Media



DataONE: Cyberinfrastructure

Three major components for a flexible, scalable, sustainable network



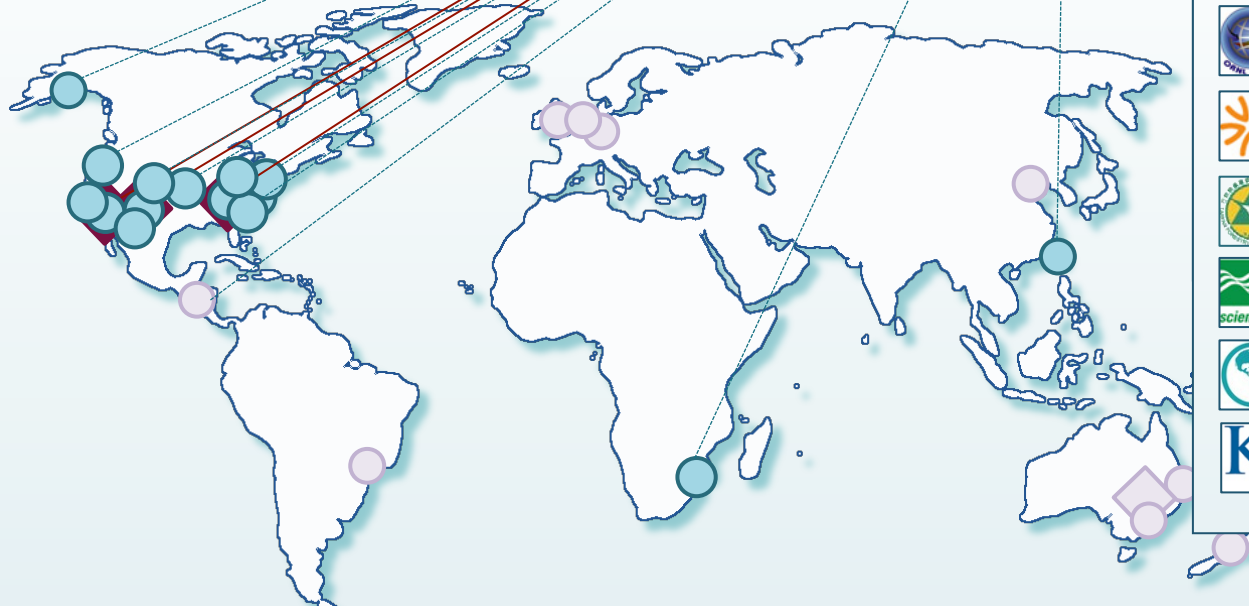
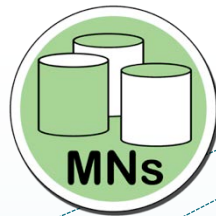
Coordinating Nodes

- retain complete metadata catalog
- indexing for search
- network-wide services
- ensure content availability (preservation)
- replication services



DataONE: Cyberinfrastructure

Three major components for a flexible, scalable, sustainable network

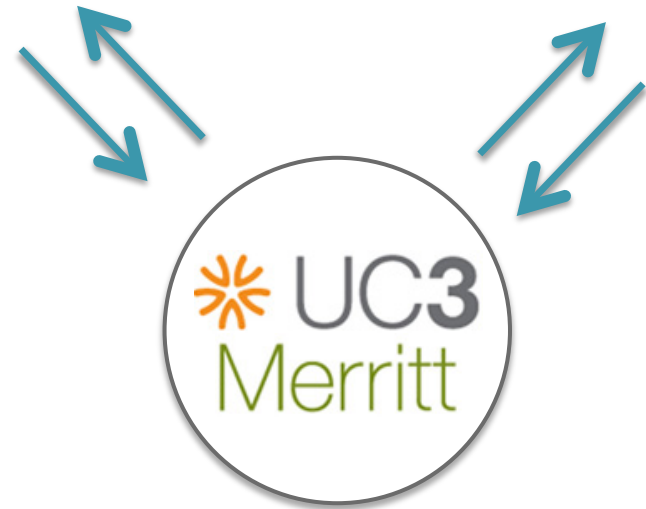
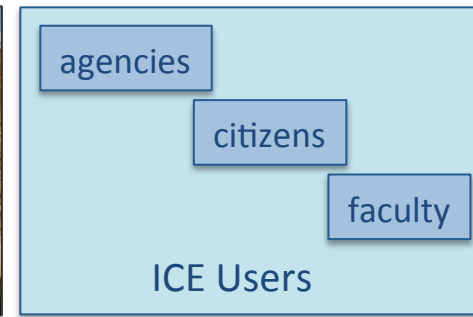
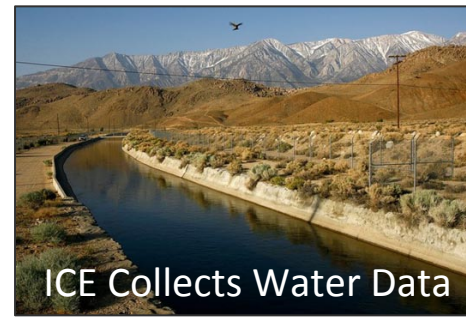
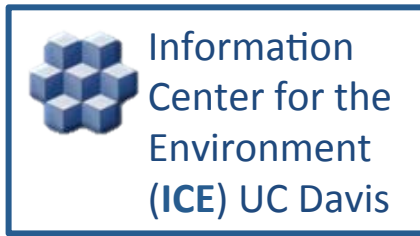



Coordinating Nodes

Member Nodes

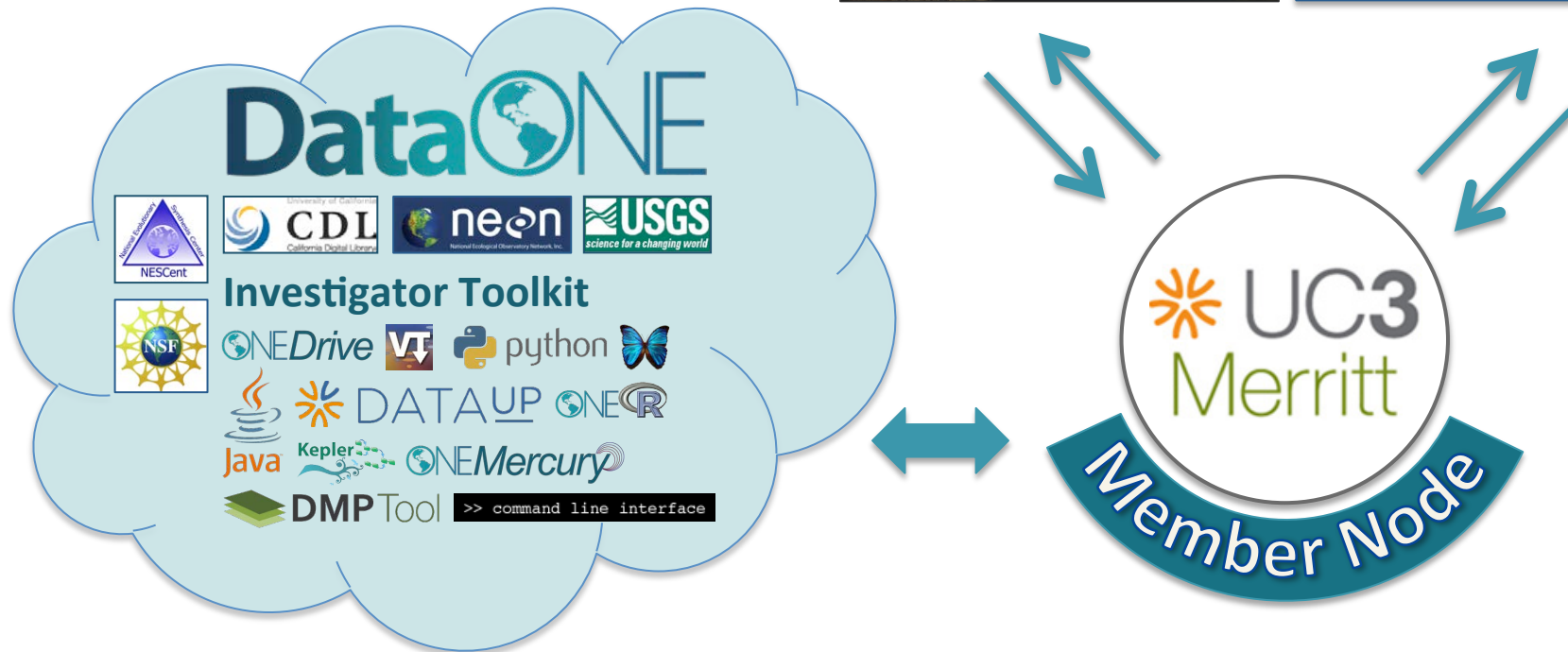
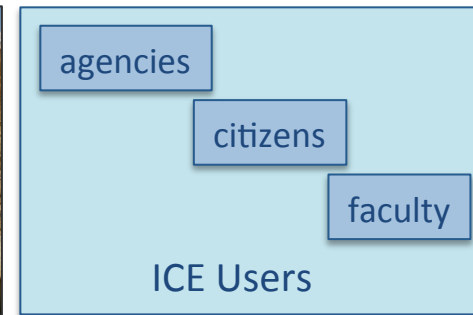
- diverse institutions
- serve local community
- provide resources for managing their data
- retain copies of data






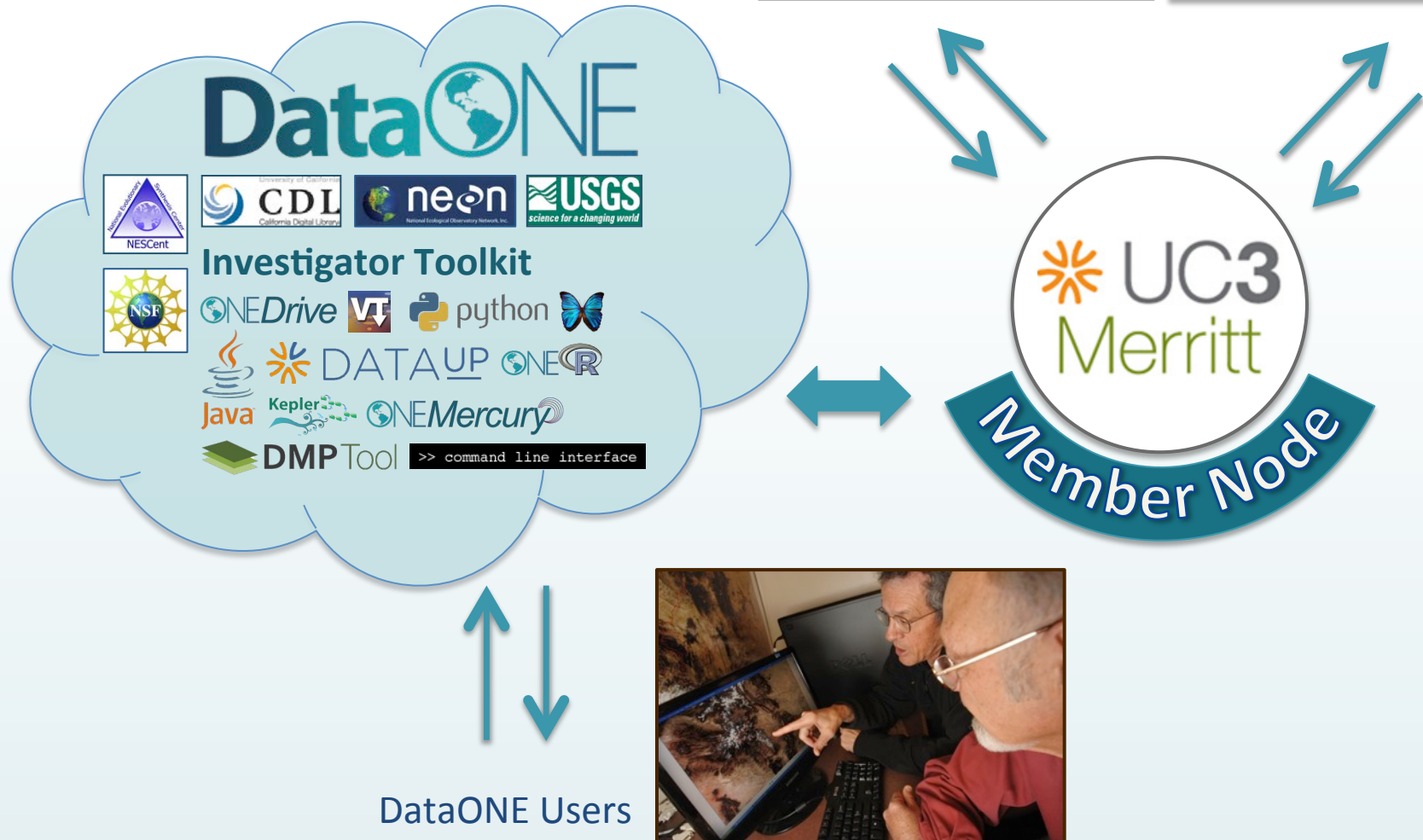
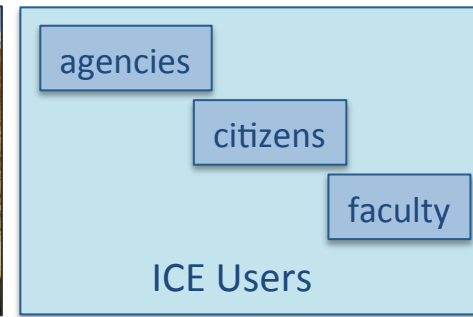
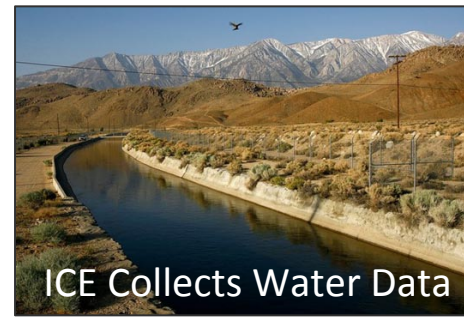


Information
Center for the
Environment
(ICE) UC Davis



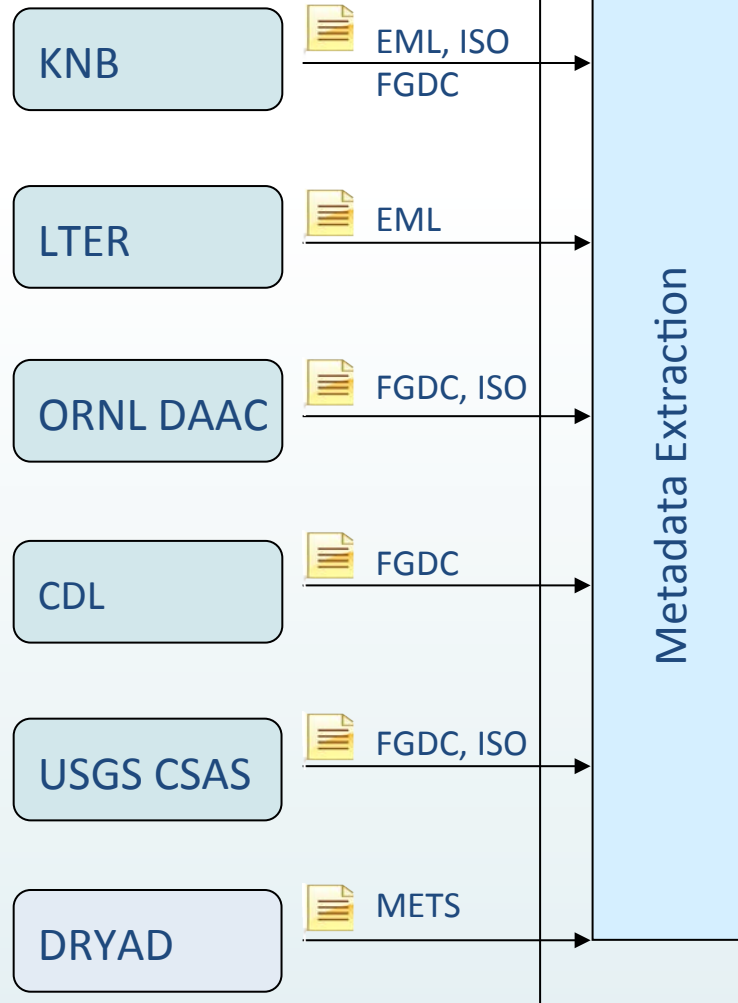


Information
Center for the
Environment
(ICE) UC Davis

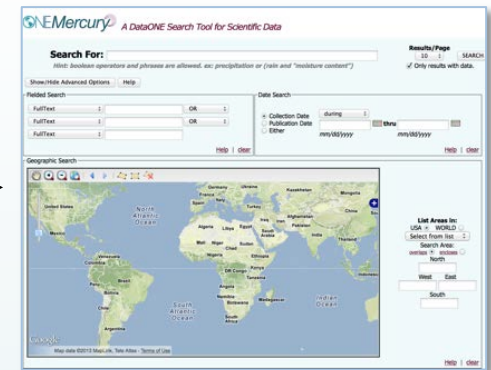
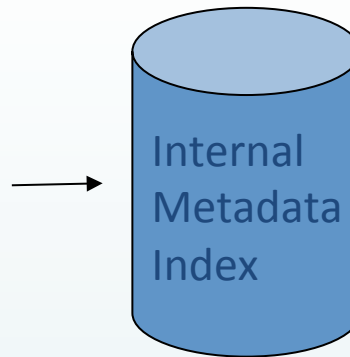


Metadata Interoperability

Member Nodes



Coordinating Nodes



- Virtual Portals
- Numerous search capabilities
- Metadata has link to data, which reside at Member Nodes

DataONE: Cyberinfrastructure

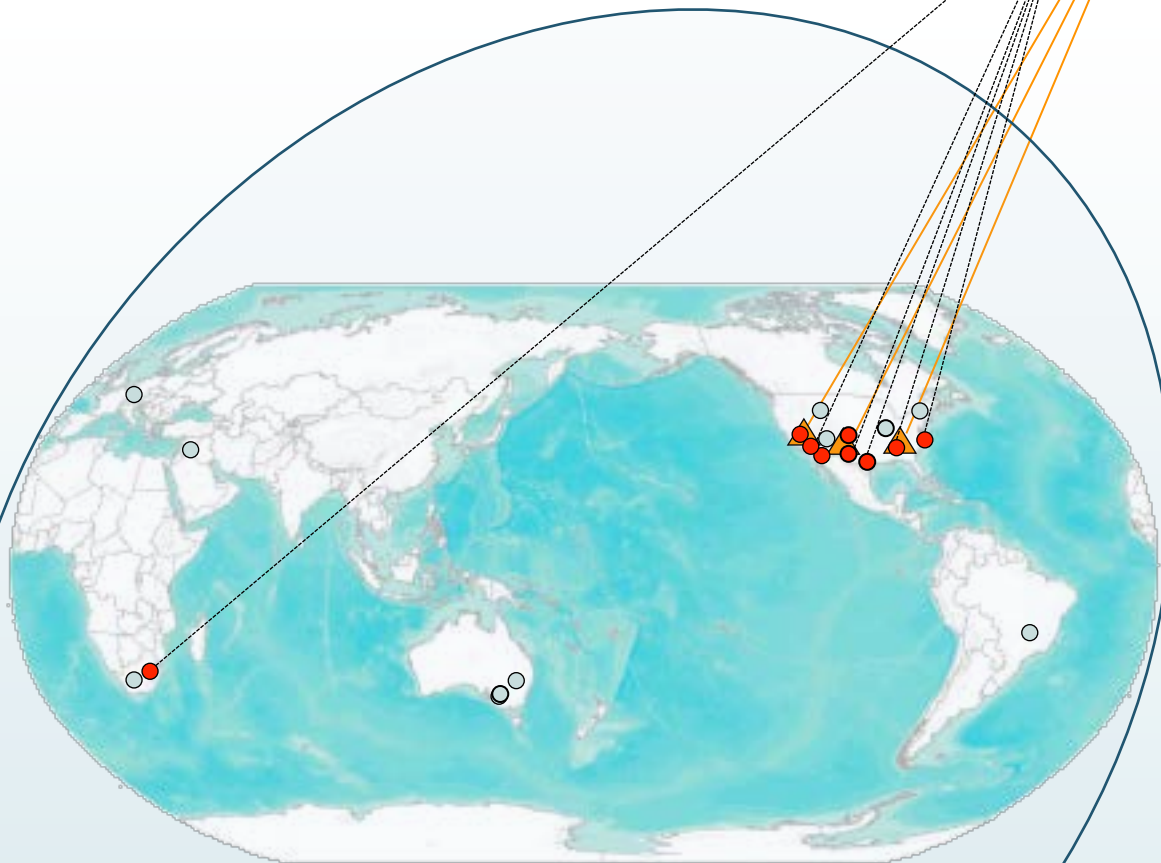
Three major components for a flexible, scalable, sustainable network

Coordinating Nodes

Member Nodes

Investigator Toolkit

>> command line interface





DataONE Tool Interoperability

Member Nodes

KNB

EML, ISO
FGDC

LTER

EML

ORNL DAAC

FGDC, ISO

CDL

FGDC

USGS CSAS

FGDC, ISO

DRYAD

METS

Metadata Extraction

Coordinating Node

Internal
Metadata
Index

Query Service Interface

VisTrails



DataONE
Library

DataONE Client
Python Library
wrapped in
VisTrails Modules

Data Life Cycle Tool Support

