

# Data Observation Network for Earth

## Demonstration of Tools and Services

Dave Vieglais

*University of Kansas*

*PRAGMA-24, March 2013*



# Open Science Movement

**EXECUTIVE OFFICE OF THE PRESIDENT  
OFFICE OF SCIENCE AND TECHNOLOGY POLICY  
WASHINGTON, D.C. 20502**

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren  
Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

## **1. Policy Principles**

The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.

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.

Access to digital data sets resulting from federally funded research allows companies to focus resources and efforts on understanding and exploiting discoveries. For example, open weather

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.



**OFFICIAL OFFICE OF SCIENCE AND TECHNOLOGY POLICY RESPONSE TO  
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*

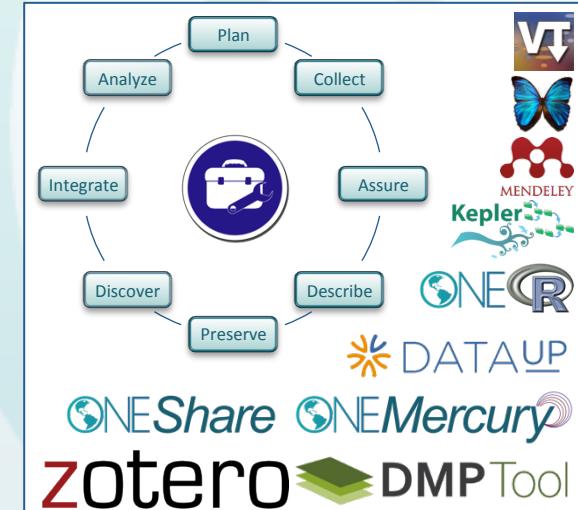
## 1. Building community



## 2. Developing sustainable data discovery and interoperability solutions

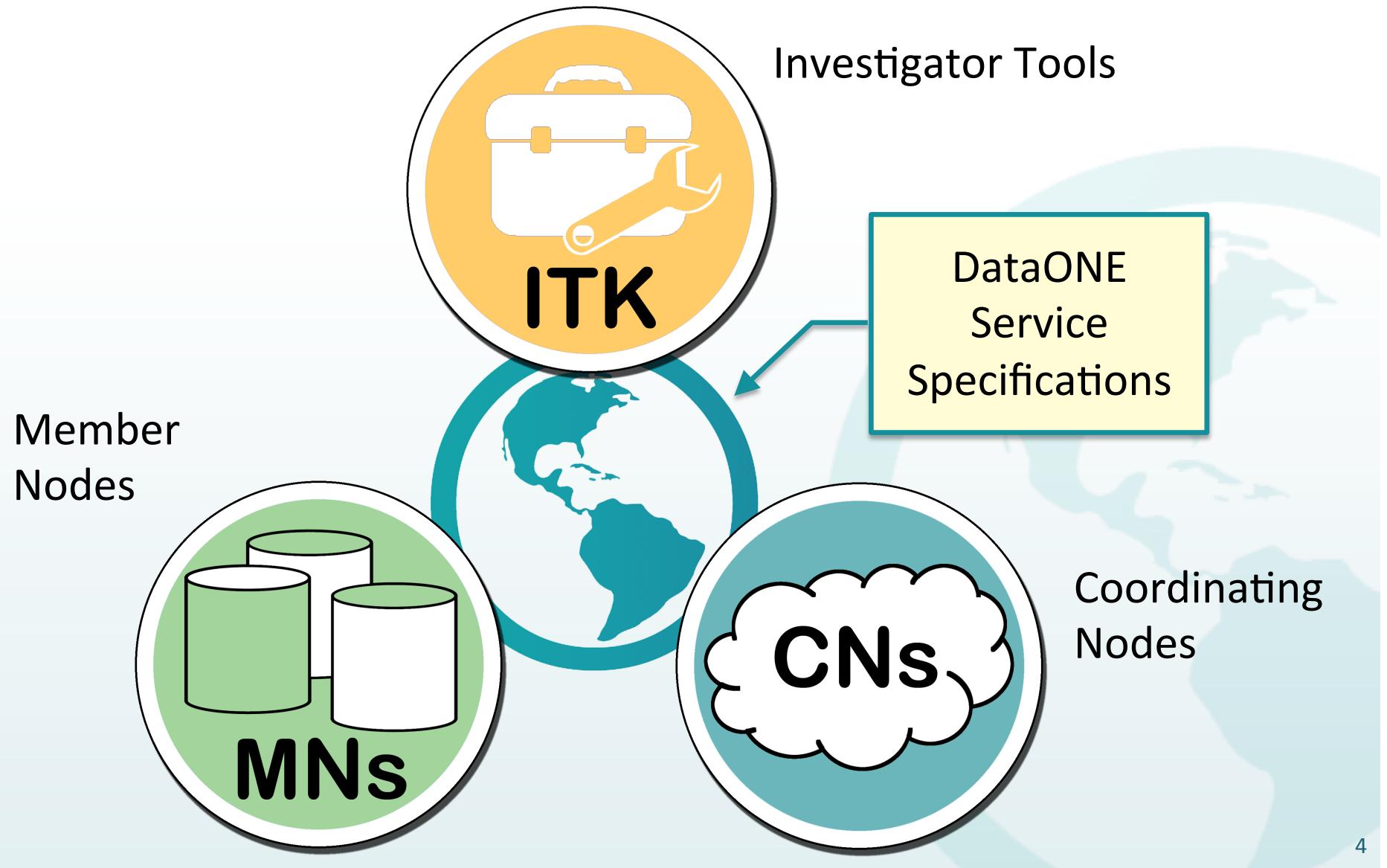
A screenshot of the DataONE ONEMercury search tool. The interface includes a top navigation bar with links for About, Participate, Resources, Education, and Data. Below the navigation is a search bar labeled "Search For:" with a hint about boolean operators. It features a "Fielded Search" section with dropdown menus for "Collection Data" and "Publication Date". A "Geographic Search" section contains a world map with country names and search filters for "List Areas" (USA, WORLD) and "Search Areas" (North, South, West, East). At the bottom, there are sections for "Member Nodes" and "SANParks Data Repository".

## 3. Enabling science through tools and services



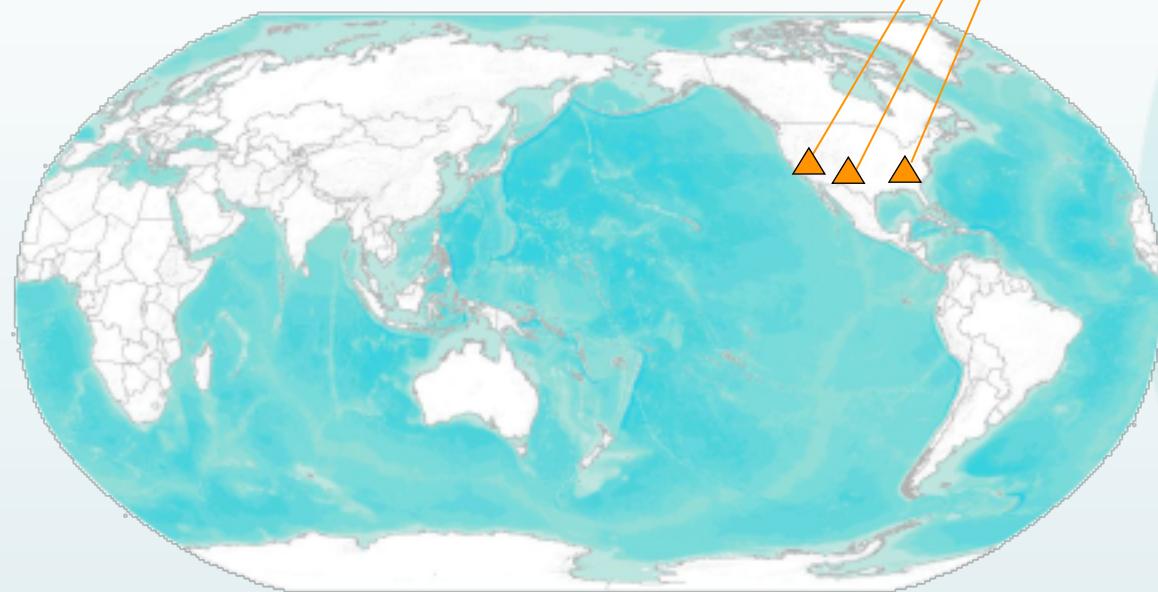
# Three Infrastructure Components

---



# Coordinating Nodes

Three major components for a flexible, scalable, sustainable network



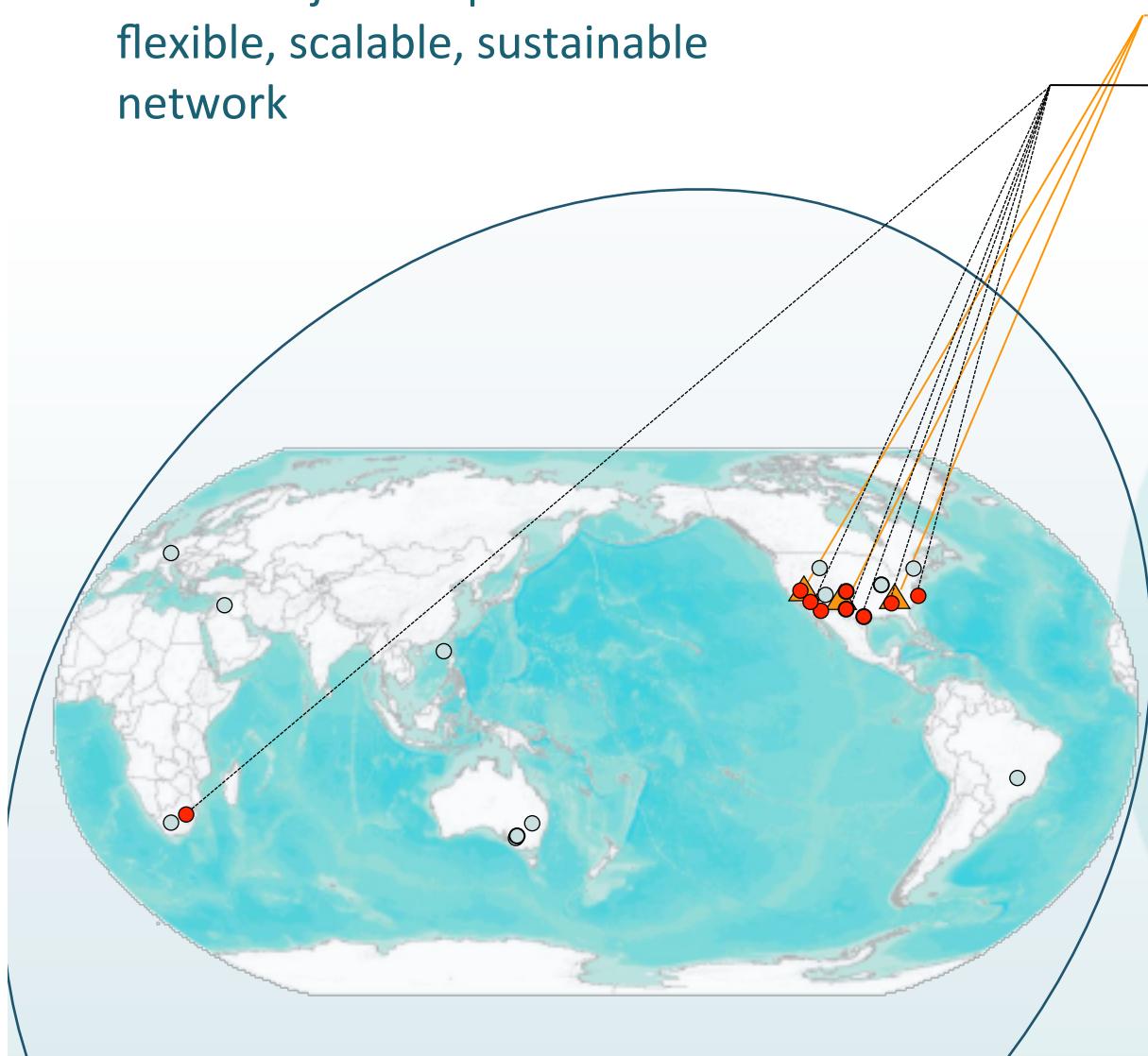
## Coordinating Nodes

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



# Member Nodes

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



The Cornell Lab



UC3Merritt



PISCO



esa

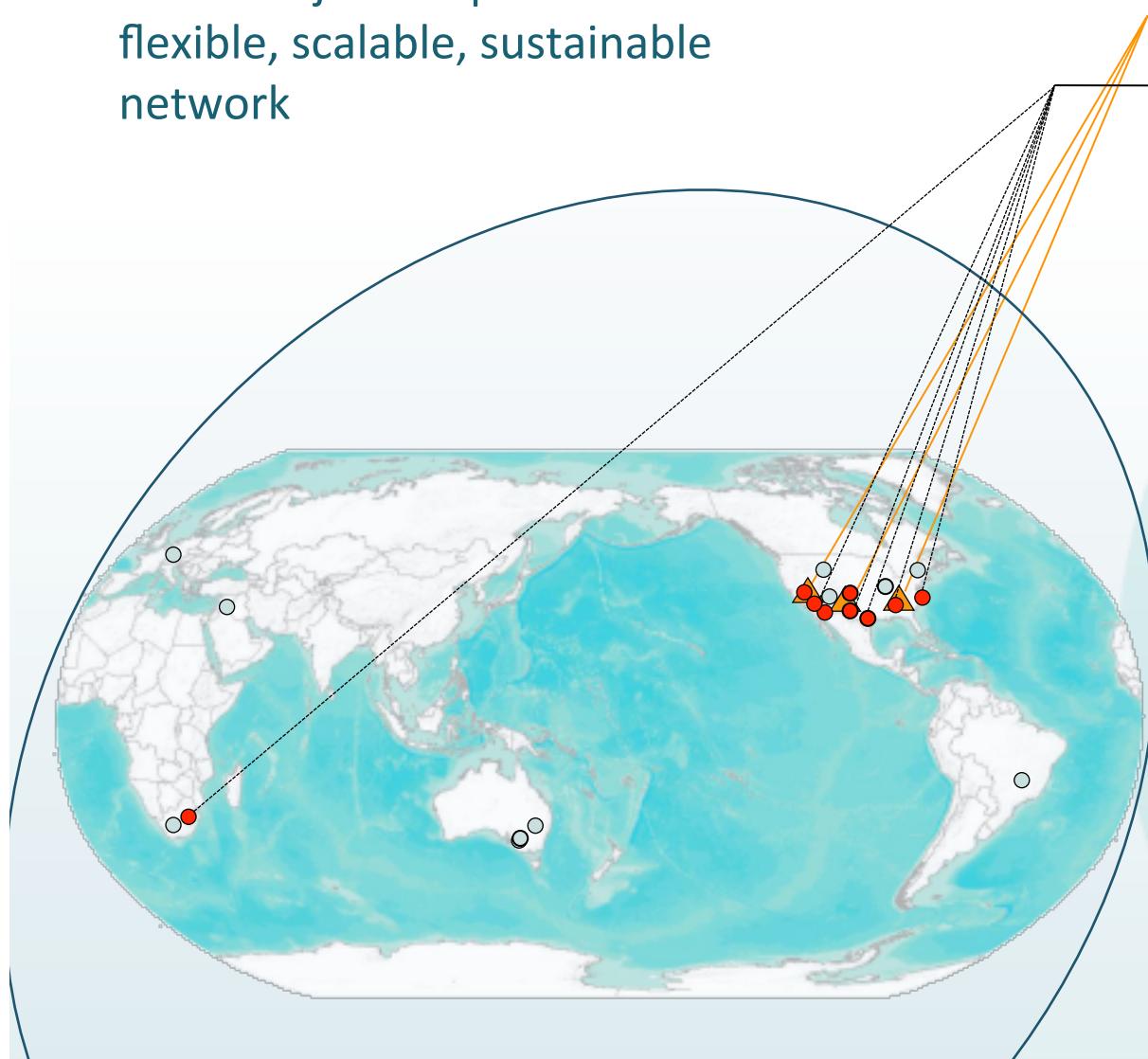


KNB

ONEShare

# Investigator Tools

Three major components for a flexible, scalable, sustainable network



## Coordinating Nodes

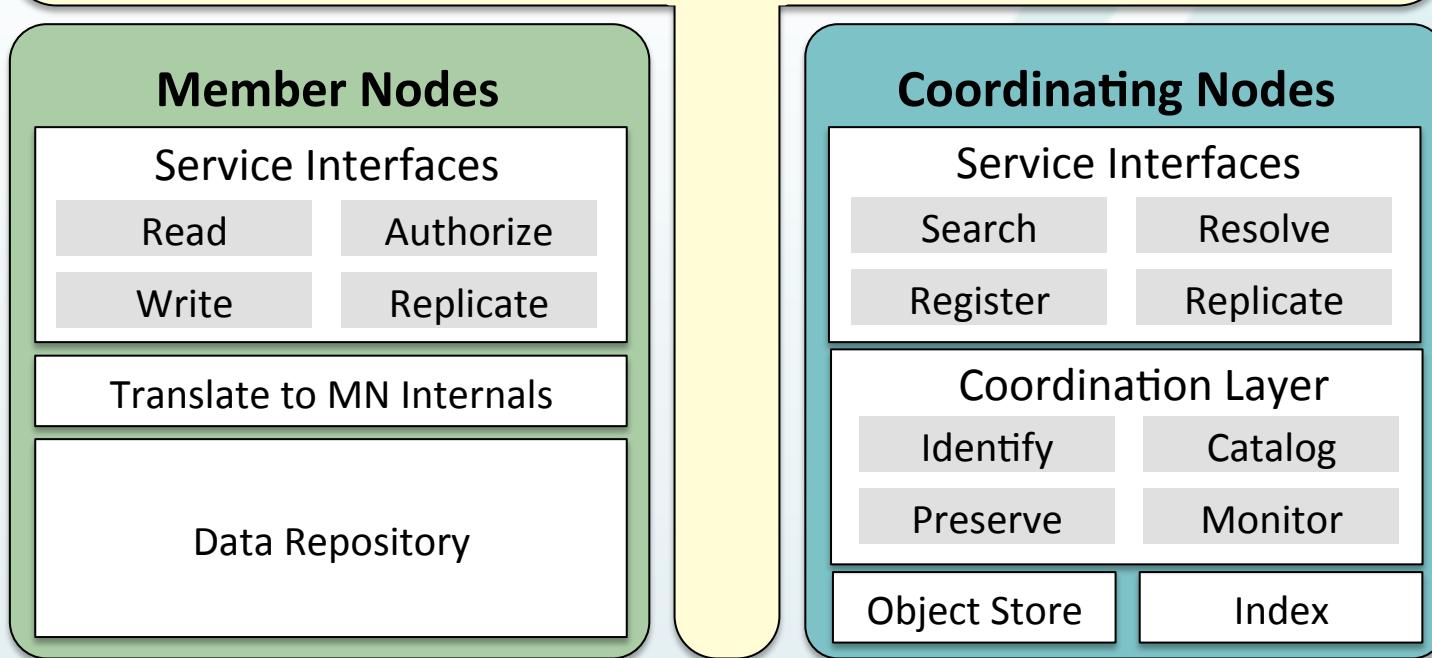
## Member Nodes

## Investigator Toolkit

>> command line interface



# Leveraging APIs – Investigator Tools





# Investigator Toolkit



- Bring DataONE to the tools scientists use
- Strengthen data management practices
- Enable community development



# Informing Priorities: Assessments

stakeholder surveys



Data Mgr.

usability testing



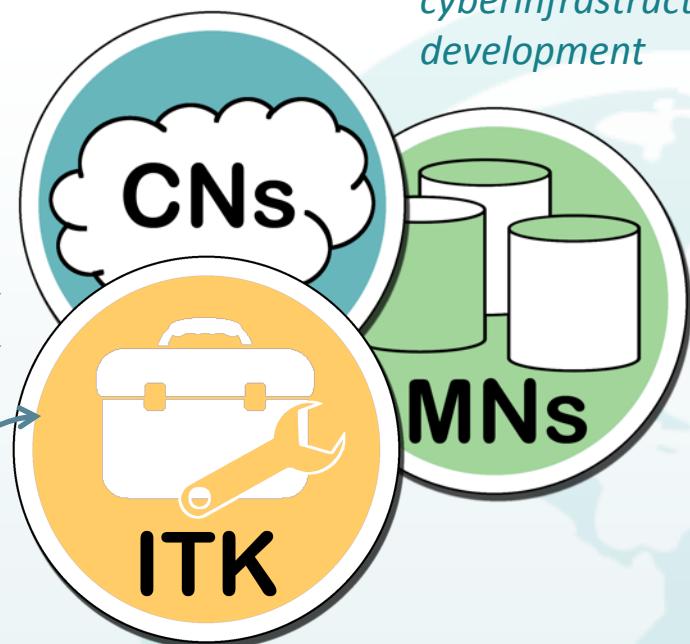
external assessments / surveys

A screenshot of a PLOS ONE article titled "Data Sharing by Scientists: Practices and Perceptions". The article has 10,244 views, 21 citations, 189 academic bookmarks, and 40 social shares. The authors listed are Carol Tenopir, Suzie Allard, Kimberly Douglass, Arsev Umar Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame.

persona and scenario development



cyberinfrastructure development



# DUG Surveys: 86 Tools Identified

Access	Fortran library	LiveAccessServer	Primer
ArcGIS	Fragstats	MapServer	Python library
Brahms	GBIF IPT	Mathematica	QuantumGIS
C/C++ library	Geoportal Server	Matlab	R
Command Line	Giovanni	Mendeley	RDV
Cyberintegrator	Google Forms	Metavist	SAS
DataONE Drive	Google Fusion Tables	Metawin	SciFlow
DataTurbine	Google Spreadsheet	ModelCenter	Specify
Density	Grads	MODIS subset tool	Spyder
Distance	GRASS	Morpho	SQLServer
DMP Tool	GridFTP	MP Parser	Stata
Drupal	HydroDesktop	MycoDB	Systat
Earth System Curator	HydroExcel	myExperiment	Taverna
Earth Sys. Mod Fr.	HydroR	Nature's Notebook	Thredds
ENVI/IDL	HydroTagger	OPeNDAP	Thredds Server
ERDAS/Imagine	IDRISI	Oxygen XML Editor	UDig
ERDDAP	IDV/MclDAS	Panoply	VisTrails
EstimateS	Java library	PC-ORD	Visual Fox Pro
Excel	JDBC/ODBC	Perl library	Web search
Fdiversity	JMP	PHP	Zotero
Ferret	Kepler	PlantList/TRY	
FishR	KPHP	Presence	

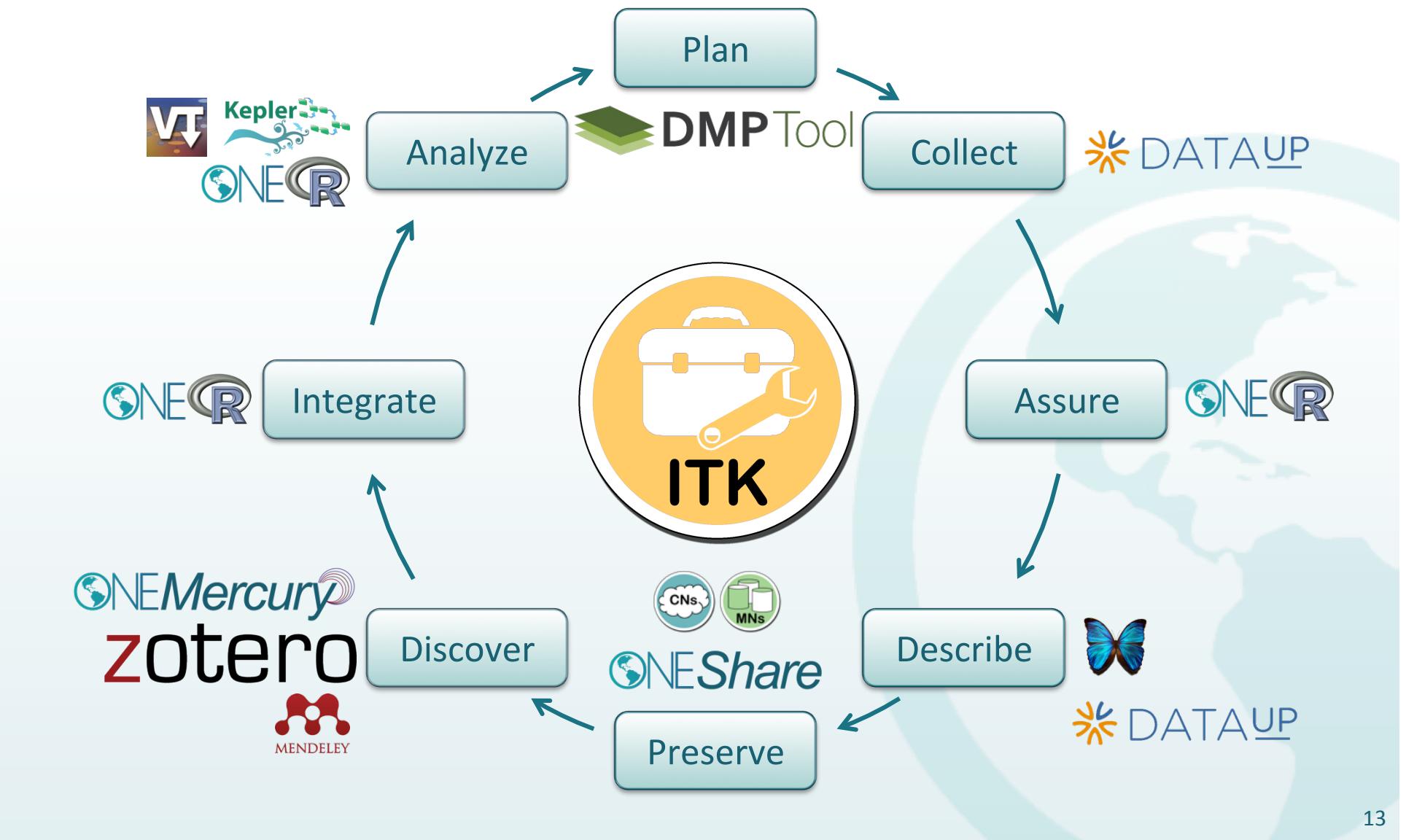
# DUG Surveys: ITK Priorities

---

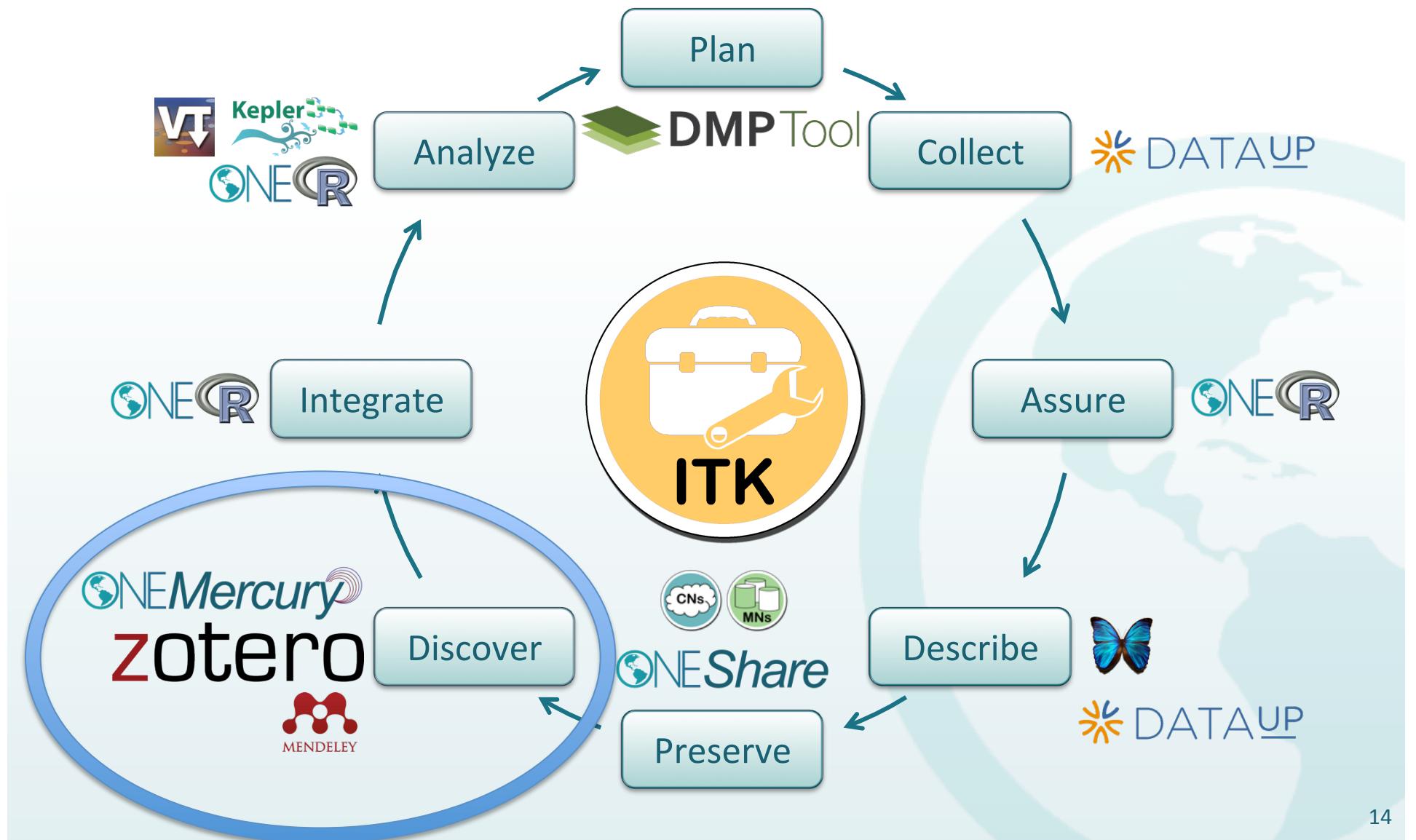
9	<b>DataONE Drive</b>	9	<b>R</b>	7	<b>ArcGIS</b>
5	<b>R</b>	2	<b>Excel</b>	3	<b>R</b>
4	<b>Python library</b>	2	<b>Java library</b>	2	<b>Geoportal</b>
4	<b>Java library</b>	2	<b>PHP</b>	2	<b>OPeNDAP</b>
3	<b>IDL</b>	2	<b>Python library</b>	2	<b>Python library</b>
2	<b>ArcGIS</b>	1	<b>QuantumGIS</b>	1	<b>Java library</b>
1	<b>Geoportal Server</b>			1	<b>Matlab</b>
1	<b>Fortran library</b>			1	<b>Mendeley</b>
1	<b>DMP Tool</b>			1	<b>Metavist</b>
1	<b>Matlab</b>			1	<b>MP Parser</b>
1	<b>MetaVist</b>			1	<b>Spyder</b>
				1	<b>Thredds</b>
				1	<b>VisTrails</b>

86 Total Tools Identified

# Tools Across the Data Life Cycle



# Tools Across the Data Life Cycle



# Discover

- Data discovery portal
- Search and retrieval of content indexed by DataONE
- Collates metadata across all Member Nodes
- Operates on each Coordinating Node

**DataONE**  
Data Observation Network for Earth

Connect  
[YouTube](#) [LinkedIn](#) [Twitter](#) [Facebook](#) [RSS](#)

About   Participate   Resources   Education   Data

**ONEMercury** A DataONE Search Tool for Scientific Data

Search For: spruce  
Hint: boolean operators and phrases are allowed. ex: precipitation or (rain and "moisture content")

Results/Page: 10   Only results with data.

Show/Hide Advanced Options  Help

Fielded Search Date Search

Collection Date Publication Date Either during mm/dd/yyyy thru mm/dd/yyyy

Geographic Search

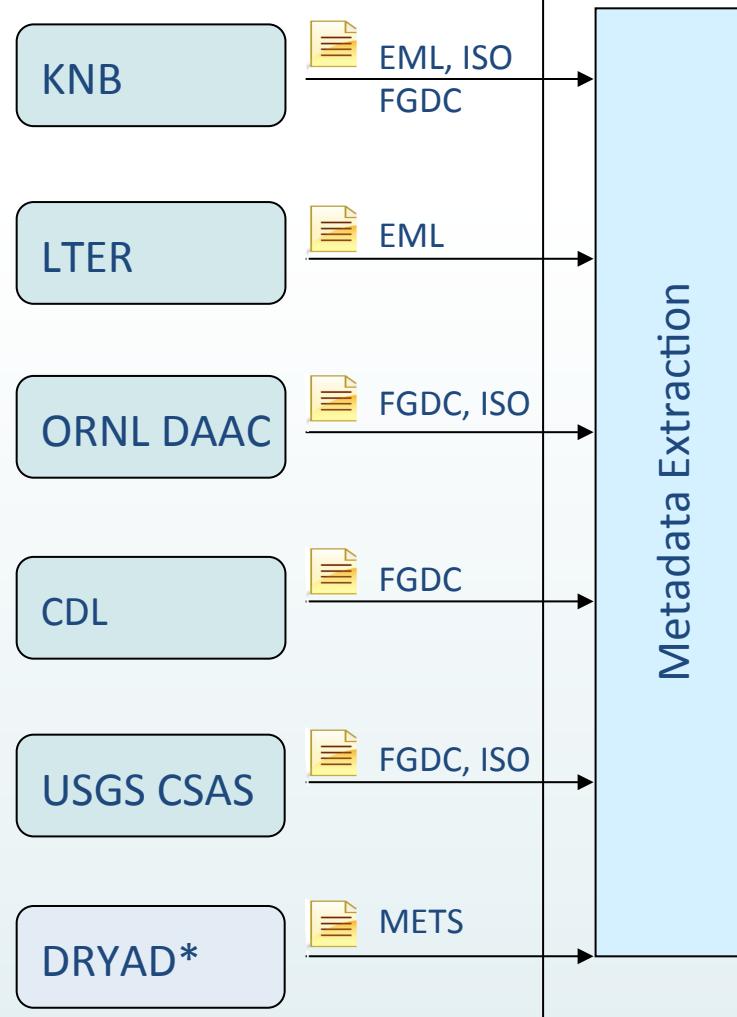
List Areas in: USA  WORLD   
Select from list  
Search Area: overlaps  encloses   
North: 49.809631 West: -87.97851 South: 35.710837 East: -59.15039

Member Nodes:

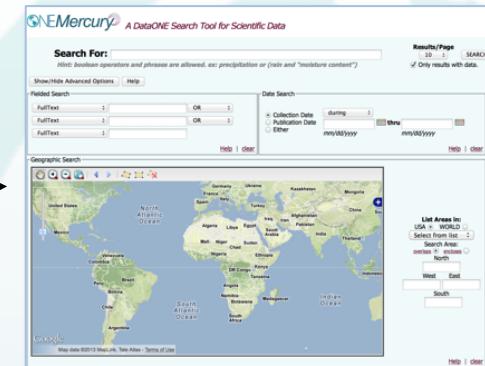
- All
- SANParks Data Repository
- PISCO MN
- LTER Network Member Node
- ORNL DAAC
- USGS Core Sciences Clearinghouse

# DataONE Metadata Interoperability

## Member Nodes



## Coordinating Nodes



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

\*Future

# ONEMercury - Demo



DataONE  
Data Observation Network for Earth

Connect

About

Participate

Resources

Education

Data



A DataONE Search Tool for Scientific Data

Search For: spruce

Hint: boolean operators and phrases are allowed. ex: precipitation or (rain and "moisture content")

Results/Page

10

SEARCH

Only results with data.

Show/Hide Advanced Options

Clear All

Help

Fielded Search


Date Search

Collection Date      during      mm/dd/yyyy

Publication Date      thru      mm/dd/yyyy

Either

Geographic Search



A map of the Eastern United States and parts of Canada. A rectangular search area is highlighted in red, covering a large portion of the Great Lakes region and extending into the northern Midwest and New England. State and provincial boundaries are shown, along with major rivers and lakes.

List Areas in:

USA  WORLD

Select from list

Search Area:

overlaps  encloses

North

49.809631

West

-87.97851

East

-59.15039

South

35.710837

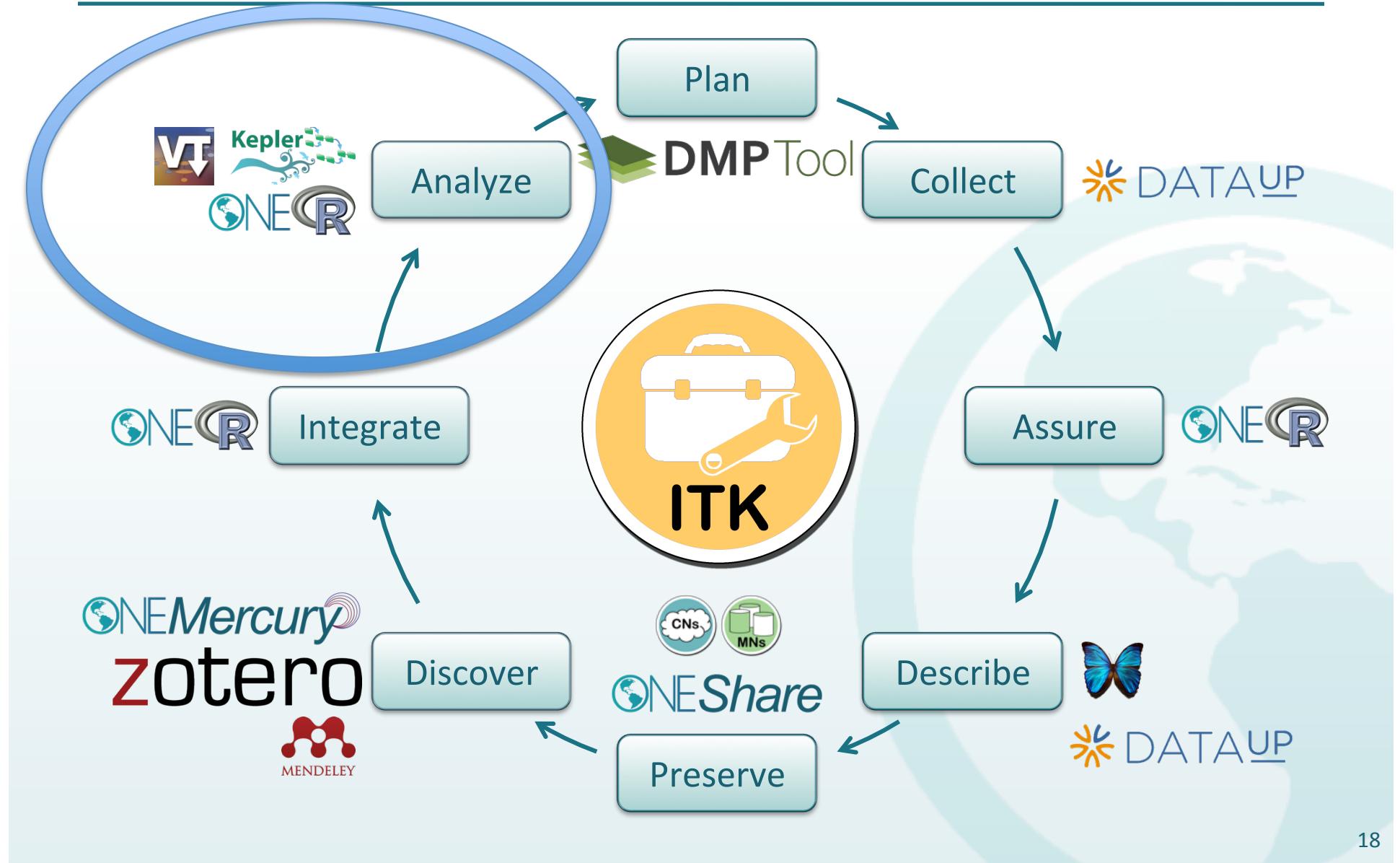
Help | clear

Member Nodes

All  
SANparks Data Repository  
PISCO MN  
LTER Network Member Node  
ORNL DAAC  
USGS Core Sciences Clearinghouse

17

# Tools Across the Data Life Cycle



# DataONE R Client



- Goals
  - Reproducible science
  - Citable analysis
  - Published results
- Data access integrated directly in R
  - Access via global identifiers
  - Enables R Script portability and preservation
- Data submission with system metadata generation
  - Versioning of objects as submitted



# R Expectations

---

- R is command driven
  - It's not for everybody
  - But it is heavily used by our target user groups
- Many people work from scripts with R



# Using the DataONE R Client

---

## Initialize client object

```
d1 <- D1Client(env, mn_nodeid)
```

## Resolve, download, and convert data

```
d1Object <- getD1Object(d1, "erd.362.1")
getBytes <- getData(d1Object)
dataPackage <- getPackage(d1, packageId)
data <- asDataFrame(dataPackage, 1)
```

## Store data on Member Node

```
d1Object <- new(Class="D1Object", dataId,
                 csvdata, format, mn_nodeid)
setPublicAccess(d1Object)
createD1Object(d1, d1Object)
```

RStudio File Edit Code View Project Workspace Plots Tools Window Help

~/development/dataone/allsoftware/demos/d1\_client\_r\_demo\_3 – RStudio

d1\_r\_demo.R × Source on Save Run Source Workspace History Load Save Import Dataset Clear All

```
1 library(dataone)
2
3 # Main loop -- call the functions in proper order
4 # dataId <- buildAndUploadDataPackage()
5 # print(paste("ID of data object:", dataId))
6 # downloadDataAndPlot(dataId)
7
8 ##
9 ## retrieves a data object and plots the data
10 ## contained therein, assuming that it would have
11 ## variables for count and species
12 ##
13 downloadDataAndPlot <- function(id) {
14   print(" ")
15   print(paste("Plotting data from: ", id))
16   d1 <- D1Client("STAGING2")
17   dlo <- getD1Object(d1, id)
18   mydf <- asDataFrame(dlo)
19
20   str(mydf)
21   boxplot(count ~ species, data=mydf)
22   print(summary(mydf))
23 }
24
25 ##
26 ## an over-simplified method to upload a package
27 ## with minimal metadata in a package, creates
28 ## style identifier
29 ##
30 ## (returns the identifier of the data file in the package)
31 ##
32 buildAndUploadDataPackage <- function() {
33
34   formatid.csv <- "text/csv"
35   formatid.eml <- "eml://ecoinformatics.org/eml-2.1.0"
```

About RStudio

RStudio Version 0.96.331 © 2009–2011 RStudio, Inc. This program is licensed to you under the terms of version 3 of the GNU Affero General Public License.

RStudio includes other open source software components. The following is a list of these components (full copies of the license agreements used by these components are included below):

- The R Project
- Qt (LGPL v2.1)
- QtSingleApplication (LGPL v2.1)
- Ace (LGPL v2.1)
- Boost
- RapidXml
- JSON Spirit
- Google Web Toolkit
- Guice
- GIN
- AOP Alliance
- RSA-JS
- FindPAM.cmake
- CPackRPM.cmake
- tree.hn

OK

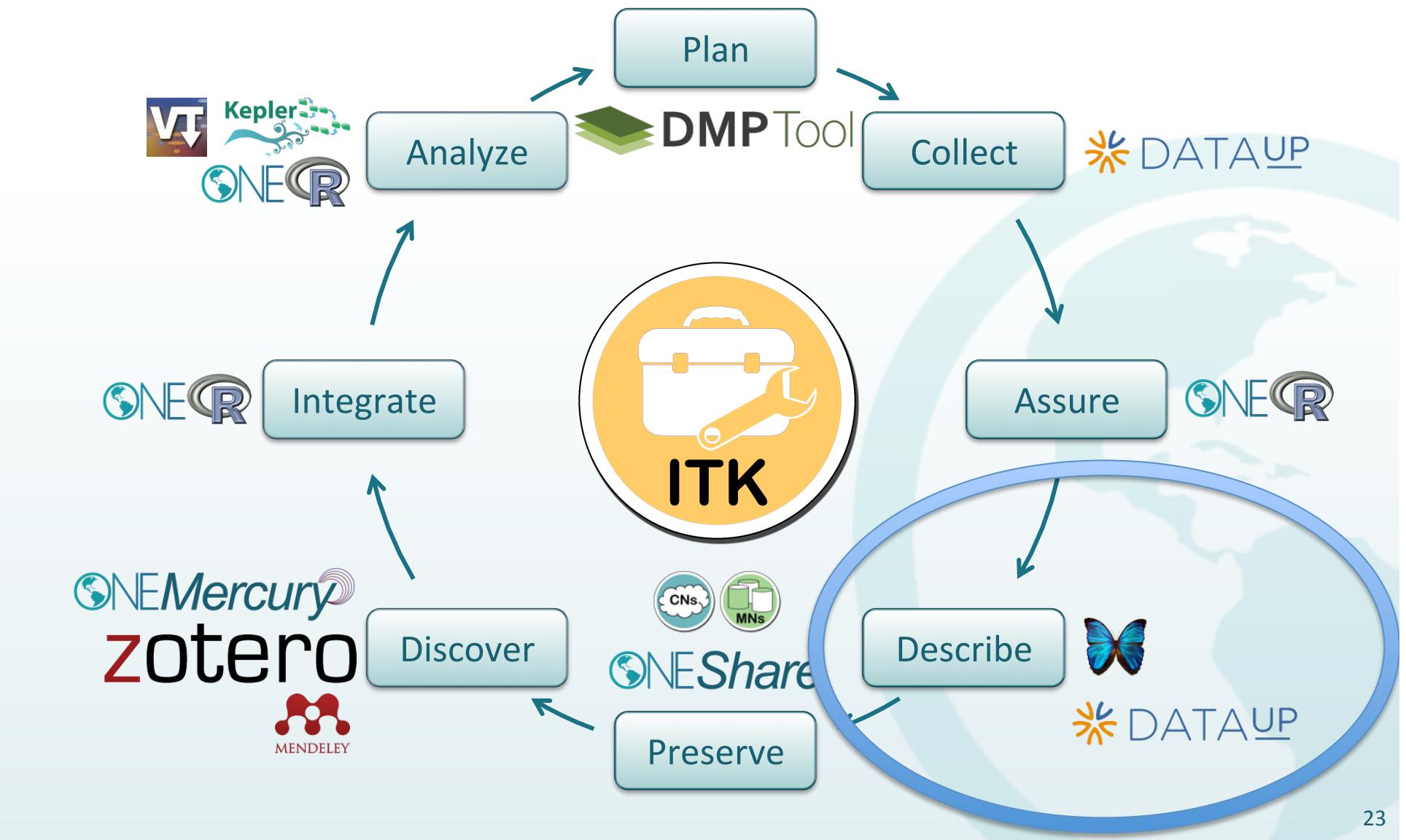
Show HD 00.00.00 5730 Tue Feb 2

Packages Help Export Clear All

Console

22

# Tools Across the Data Life Cycle



# Describe: DataUp



- Extension for Microsoft Excel
- Widely used for management of simple data
- Promote best practices for spreadsheet data
- 65% of data in DataONE is readable by Excel



Describe, Manage, & Share Your Data  
A Service of the University of California Curation Center (UC3)



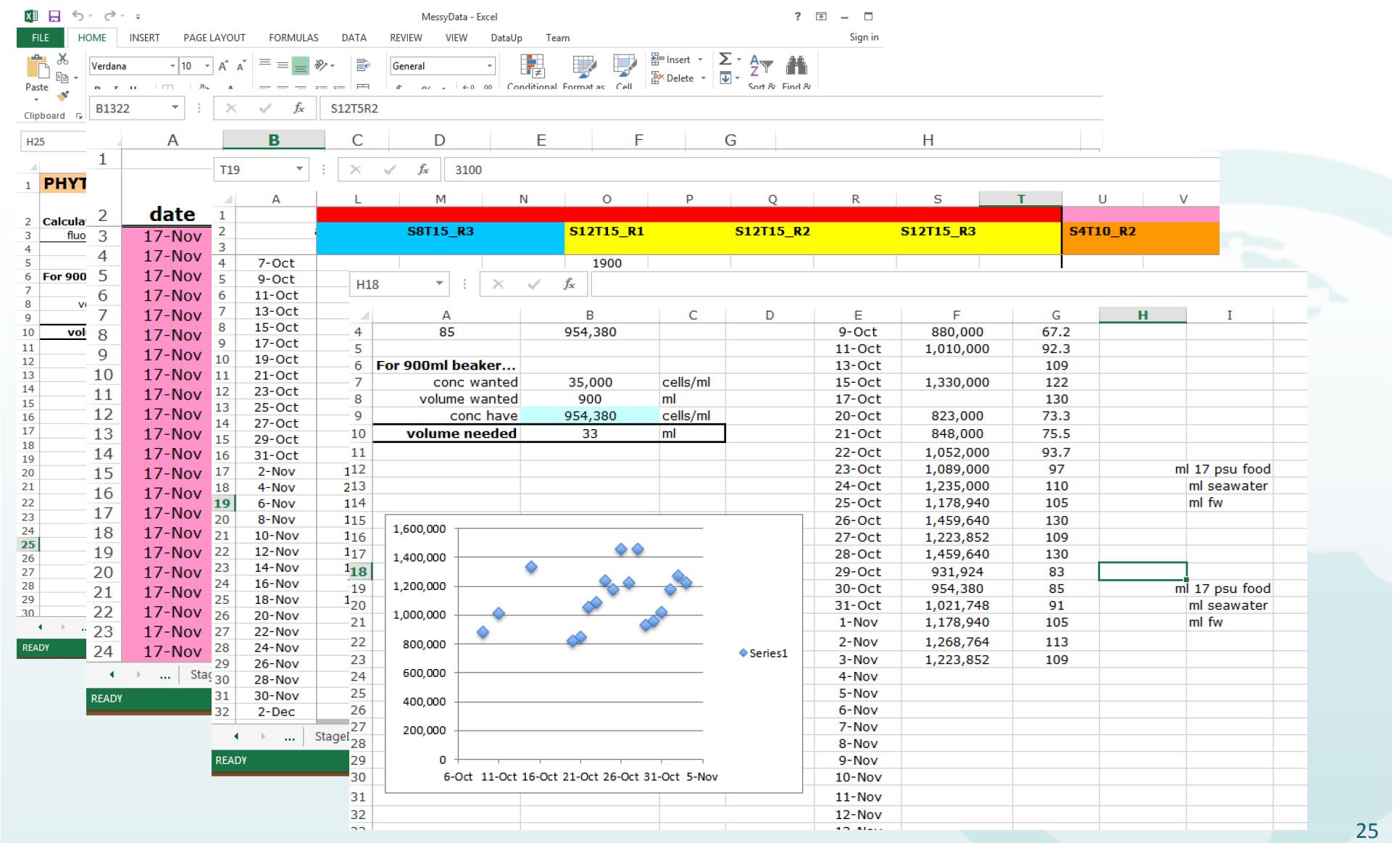
University of California  
**CDL**  
California Digital Library

**ONEShare**

Microsoft®  
**Research**

GORDON AND BETTY  
**MOORE**  
FOUNDATION

# Flexibility Detracts from Reusability



# DataUp Solution

---



1. Guide user to clean spreadsheet
  - Remove spurious formatting
  - Promote naming conventions
  - Promote consistency in spreadsheet layout
2. Gather metadata describing spreadsheet
  - Stored in a new worksheet
  - Expressed as EML
3. Apply a unique identifier to the spreadsheet
4. Upload the content to DataONE

# DataUp Add-In

Screenshot of Microsoft Excel showing the DataUp Add-In ribbon tab and various data analysis tools.

The ribbon tabs include:

- FILE
- HOME
- INSERT
- PAGE LAYOUT
- FORMULAS
- DATA
- REVIEW
- VIEW
- DataUp (selected)
- Team

The DataUp tab contains the following icons:

- Check
- Data Description
- Column Description
- Generate
- Post to Repository
- Demo

The ribbon also includes tabs for Best Practices, Describe Data, Citation, Post File, and About.

The main worksheet displays a table and a scatter plot.

**Table Data:**

	Calculating conc based on fluorescence...		Daily phytoplankton concs			
	fluorescence	cells/ml	date	cells/ml	fluorescence	
2	85	954,380	9-Oct	880,000	67.2	
3			11-Oct	1,010,000	92.3	
4			13-Oct		109	
5			15-Oct	1,330,000	122	
6	For 900ml beaker...		17-Oct		130	
7	conc wanted	35,000	cells/ml	20-Oct	823,000	73.3
8	volume wanted	900	ml	21-Oct	848,000	75.5
9	conc have	954,380	cells/ml	22-Oct	1,052,000	93.7
10	volume needed	33	ml	23-Oct	1,089,000	97
11			24-Oct	1,235,000	110	ml 17 psu food
12			25-Oct	1,178,940	105	ml seawater
13			26-Oct	1,459,640	130	ml fw
14			27-Oct	1,223,852	109	
15			28-Oct	1,459,640	130	
16			29-Oct	931,924	83	
17			30-Oct	954,380	85	ml 17 psu food
18			31-Oct	1,021,748	91	ml seawater
19			1-Nov	1,178,940	105	ml fw
20			2-Nov	1,268,764	113	
21			3-Nov	1,223,852	109	
22			4-Nov			
23			5-Nov			

**Scatter Plot Data:**

Row	X	Y
15	Oct_7to15	850,000
16	scope_data	1,050,000
17	pop_size	1,100,000
18	algae	1,150,000
19	Metadata	1,200,000
20		1,250,000
21		1,300,000
22		1,350,000
23		1,400,000
24		1,450,000
25		1,500,000

The chart is titled "Series1" and shows a positive correlation between the variables.

# DataUp: Check Data

Screenshot of the DataUp Excel add-in interface showing data validation results and error details.

The ribbon tabs include FILE, HOME, INSERT, PAGE LAYOUT, FORMULAS, DATA, REVIEW, VIEW, DataUp, and Team.

The DataUp tab is selected, showing the following options:

- Check
- Description
- Column Description
- Generate Citation
- Post to Repository
- Post File
- Demo
- About

The main worksheet displays a table of phytoplankton data and a scatter plot. The table includes columns for date, cell count, and various concentrations. A chart titled "Series1" plots "pop\_size" against dates from October 6 to November 5.

A sidebar titled "Error Details" lists validation errors:

- pop\_size (4) algae (6)
- Remove Selected

Category	Count
pop_size	4
algae	6

Details for each error type:

- Embedded charts, pictures: Remove, Advice, Info
- Merged cells: Remove, Advice, Info
- Color coded text or cell shading: Remove, Advice, Info
- Non-contiguous data: Remove, Advice, Info
- Blank Cells: Remove, Advice, Info
- Special Characters: Remove, Advice, Info

# DataUp: Add Metadata

The screenshot shows a Microsoft Excel spreadsheet titled "Metadata". The spreadsheet contains data about environmental parameters and their descriptions, along with geographic coverage information.

**Table Structure:**

	A	B				
1	Name	Value				
2	Title					
3	Create					
4	EntityName	EntityDescription	Name	Description	Type	Units
5	Station1		Sample	Sampling station	Text	
6	Station1		Depth		Numeric	m
7	Station1		Temperature		Numeric	°C
8	Station1		Salinity		Numeric	psu
9	Station1		Sigma-T		Numeric	kg/m^3
10	Station1		Chlorophyll-a		Numeric	mg/m^3
11	Station1		DO	Dissolved Oxygen	Numeric	ml/l
12	Station1		pH		Numeric	
13	keyw	Station1	log10(PAR)	Log(10) of Photosynthetically Active Radiation	Numeric	uM/m^2s
14	keyw	Station1	PAR	Photosynthetically Active Radiation	Numeric	uM/m^2s
15	Inte					
16	url f					
19	Geographic coverage: Description					
20	Geographic coverage: West bounding coordinate					
21	Geographic coverage: East bounding coordinate					
22	Geographic coverage: North bounding coordinate					
23	Geographic coverage: South bounding coordinate					

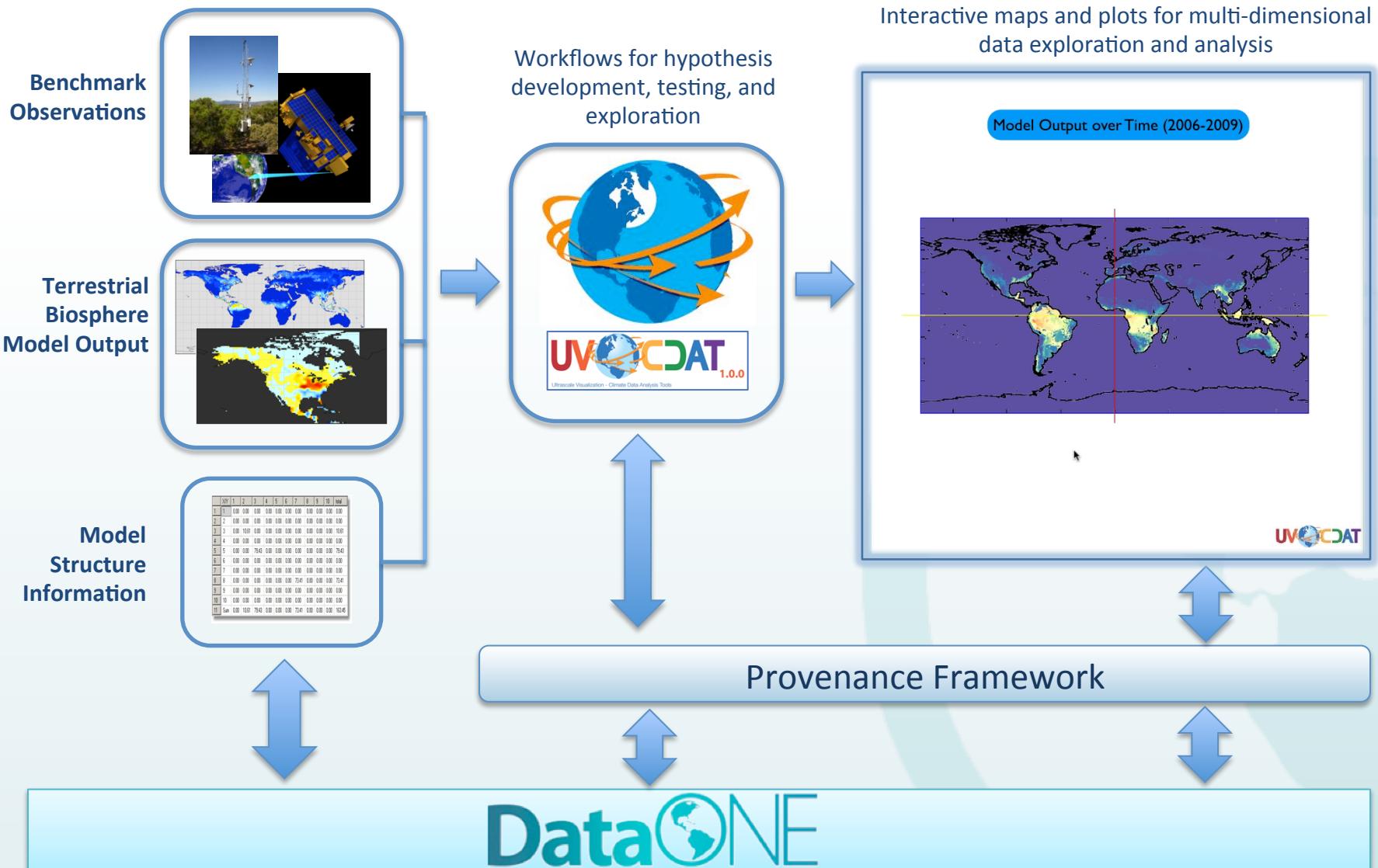
**Toolbars and Status Bar:**

- Paste, Clipboard, Font, Alignment, Number, Styles, Cells, Sort & Find & Filter, Editing.
- Cells A14, Formula Bar: Abstract.
- Status Bar: Station1, Sheet1, Metadata, 100%.

# DataUp: Publish to DataONE

A screenshot of a Microsoft Excel spreadsheet titled "CleanData-3 - Excel". The ribbon menu is visible with tabs like FILE, HOME, INSERT, PAGE LAYOUT, FORMULAS, DATA, REVIEW, and VIEW. The "DataUp" tab is selected, highlighted in yellow. A modal dialog box titled "FILE POST" is displayed in the center of the screen. The dialog contains a green checkmark icon and the text "FILE PUBLISHED SUCCESSFULLY". Below this, a message states: "Your data file was submitted successfully. You will receive an email shortly with your unique identifier and submission information." A "CLOSE" button is at the bottom of the dialog. The main Excel interface shows a grid of cells with column headers A through K and row numbers 1 through 23. Cell A14 is currently selected. The status bar at the bottom shows "Station1 | Sheet1 | Metadata" and "READY". The page number "30" is located in the bottom right corner.

# Leveraging DataONE: UV-CDAT



# The Future



## Four areas of activity:

1. Expand content through Member Nodes
2. Increase use through Investigator Tools
3. Grow capability with new features
4. Maintain and refine the infrastructure

Priority #1: Stable infrastructure

# DataONE Resources

---

## General Information

<http://dataone.org>

## Data

<https://cn.dataone.org>

## Project Documents

<http://docs.dataone.org>

## Architecture Documentation

<http://mule1.dataone.org/ArchitectureDocs-current/>

## Software Repository

<https://repository.dataone.org>

## FAQ (New!)

<https://ask.dataone.org>



# Developers and Core CI Team

---

Andy Pippin

Ben Leinfelder

Bob Sandusky

Bruce Wilson

Chad Berkeley

Chris Brumgard

Chris Jones

Dave Vieglais

Giri Palanisamy

Jeff Horsburgh

Jim Green

Jing Tao

John Kunze

Mark Servilla

Matt Jones

Nick Dexter

Paul Allen

Ranjeet Devarakonda

Rob Nahf

Robert Waltz

Roger Dahl

Ryan Scherle

Skye Roseboom



# Community Driven Development

---

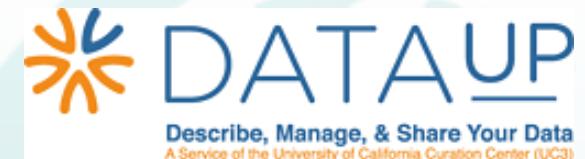
- ITK tools developed and sustained by the community
- DataONE provides support
  - Interface documentation
  - Development consultation
  - Infrastructure
- Examples:
  - DMP Tool: CDL & Andrew Sallans
  - DataUp: Strasser, Kunze, and Cruse; Microsoft Research
  - VisTrails: Dave Koop
  - Kepler: Barseghian for AOOS
  - R-Provenance: K. Ram for ROpenSci



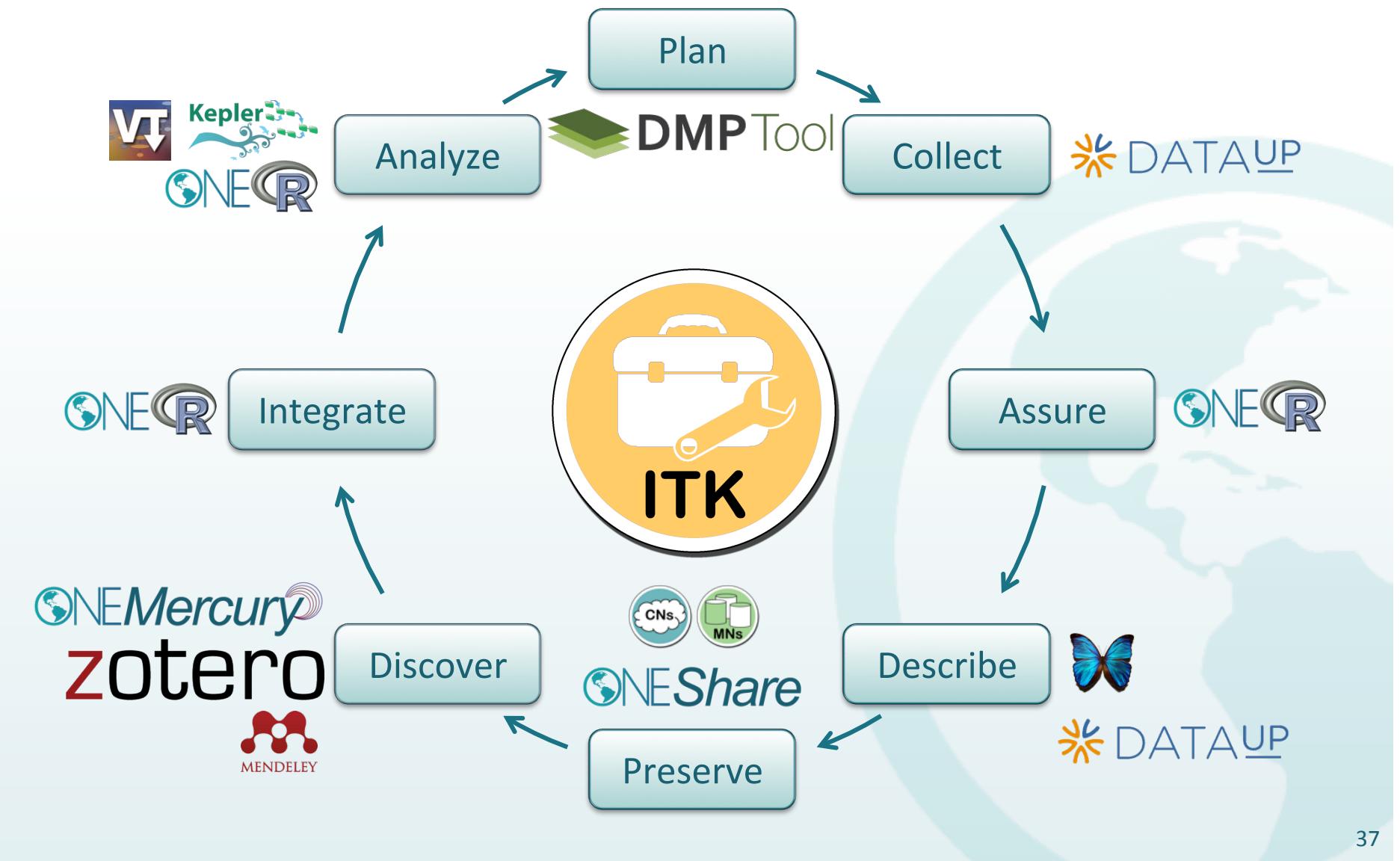
# Next Steps

---

- Finalize and release initial versions
- Gather feedback
- Iterate on next versions designs
- Develop and release
- Emerging features:
  - Enhanced discovery through semantics
  - Tracking provenance



# Questions?



---

[[ Break for questions: 20min

Next session is research demos: 40 + 15 min

Followed by CI Overview: 20 + 20 min

]]

# Command Line Client

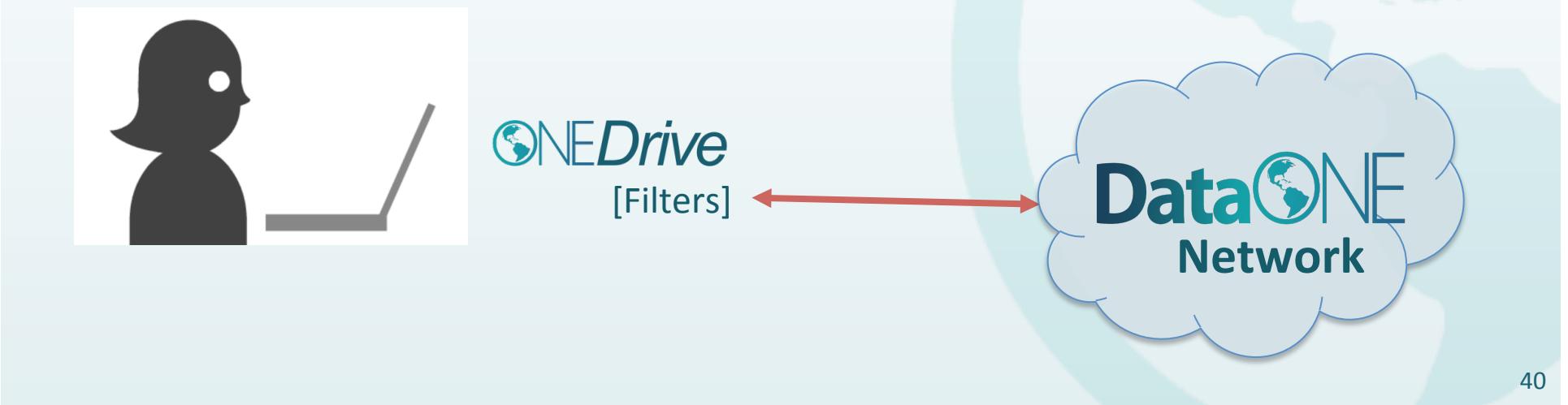


- \$ dataone
- --mn-url=https://knb.ecoinformatics.org/knb/d1/mn
- --rights-holder="CN=urn:node:D1USGSXXX,DC=dataone,DC=org"
- --allow-public **save**
  
- \$ dataone --format-id text/csv **create** doi:10.xxxx/5003 data-sites.csv
- \$ dataone **get** doi:10.xxxx/5003 mydata.csv
  
- \$ dataone --query "id:doi\:10.xxxx/5\*" **search**
- \$ dataone --query "datasource:urn\:node\:D1USGSXXX" **search**

# ONEDrive



- DataONE appears as network drive
- Live view – changes appear immediately
- Focus view using filters
  - keyword="solar radiation" AND project="Canopy Arthropods"
  - surname="Fink"
- Access content with unmodified applications

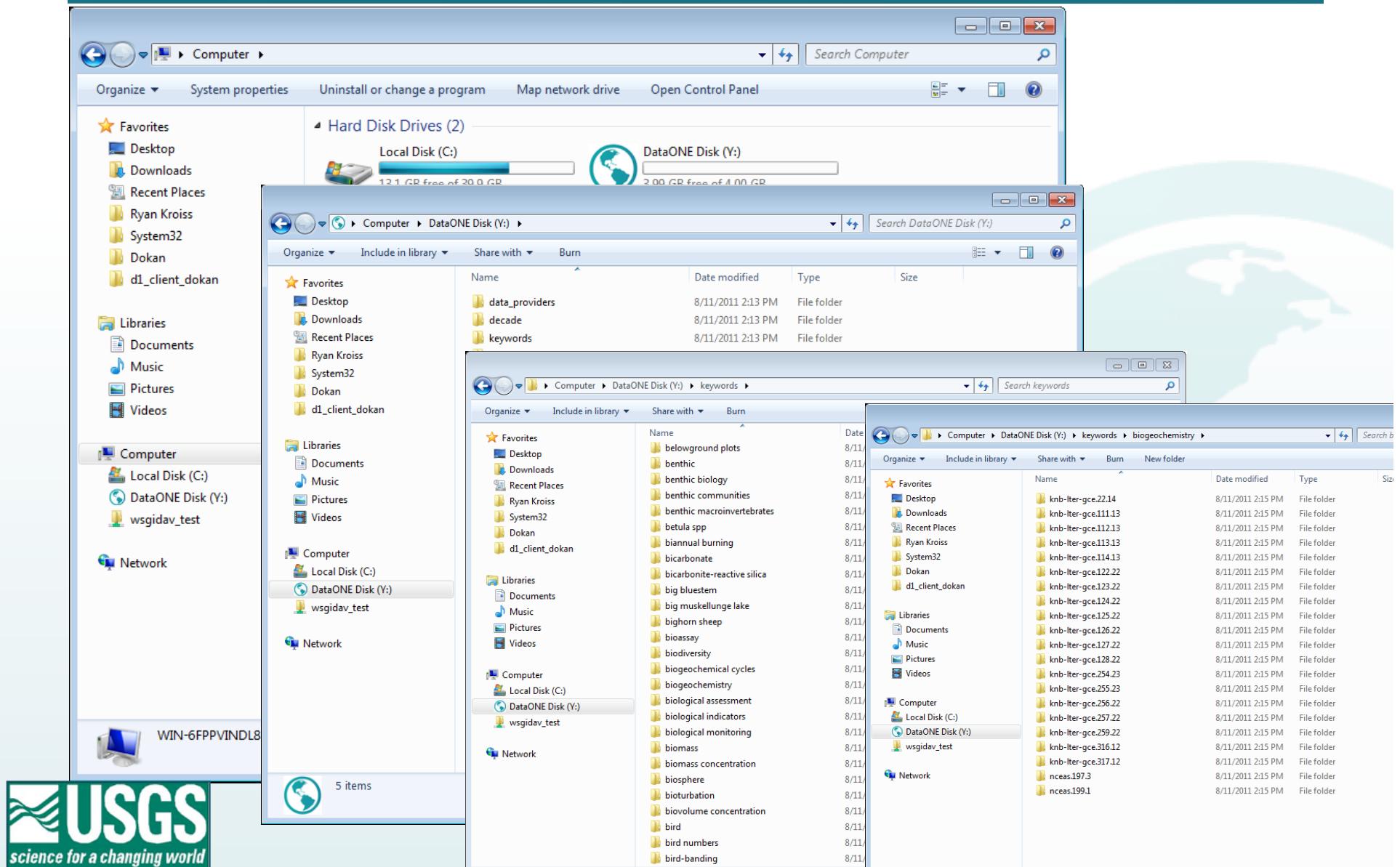


# ONE Drive

- Concept: Virtual drive integrated into the OS which allows for scientists to access & deposit data
  - DataONE has developed a MAC/Linux OS implementation
  - USGS is working on a Windows OS implementation

Name	Date Modified	Size	Kind
▼ DataONE	Today, 15:39	--	Volume
► data_providers	Today, 15:39	--	Folder
► decade	Today, 15:39	--	Folder
► keywords	Today, 15:39	--	Folder
► projects	Today, 15:39	--	Folder
▼ title	Today, 15:39	--	Folder
► Axial hydraulic segmentation in shrubs.	Today, 15:39	--	Folder
► Butterfly surveys of B...e, Boulder, Colorado	Today, 15:39	--	Folder
► California Ecological Condition 2000	Today, 15:39	--	Folder
▼ eBird Reference Data... Demonstration Data	Today, 15:39	--	Folder
▼ erd.368.1	Today, 15:39	--	Folder
► Effects of land-use o...ans %28fish data%29.	Today, 15:39	--	Folder
► Effects of land-use o... %28habitat data%29.	Today, 15:39	--	Folder
► esn039_supp.doc	Today, 15:39	--	Folder
► Spacing behavior an...tral coastal California	Today, 15:39	--	Folder
► Spatial and temporal...pper Coweeta Creek.	Today, 15:39	--	Folder

# ONE Drive – Windows implementation



# DUG Surveys: 86 Tools Identified

Access	Fortran library	LiveAccessServer	Primer
ArcGIS	Fragstats	MapServer	Python library
Brahms	GBIF IPT	Mathematica	QuantumGIS
C/C++ library	Geoportal Server	Matlab	R
Command Line	Giovanni	Mendeley	RDV
Cyberintegrator	Google Forms	Metavist	SAS
DataONE Drive	Google Fusion Tables	Metawin	SciFlow
DataTurbine	Google Spreadsheet	ModelCenter	Specify
Density	Grads	MODIS subset tool	Spyder
Distance	GRASS	Morpho	SQLServer
DMP Tool	GridFTP	MP Parser	Stata
Drupal	HydroDesktop	MycoDB	Systat
Earth System Curator	HydroExcel	myExperiment	Taverna
Earth Sys. Mod Fr.	HydroR	Nature's Notebook	Thredds
ENVI/IDL	HydroTagger	OPeNDAP	Thredds Server
ERDAS/Imagine	IDRISI	Oxygen XML Editor	UDig
ERDDAP	IDV/MclDAS	Panoply	VisTrails
EstimateS	Java library	PC-ORD	Visual Fox Pro
Excel	JDBC/ODBC	Perl library	Web search
Fdiversity	JMP	PHP	Zotero
Ferret	Kepler	PlantList/TRY	
FishR	KPHP	Presence	

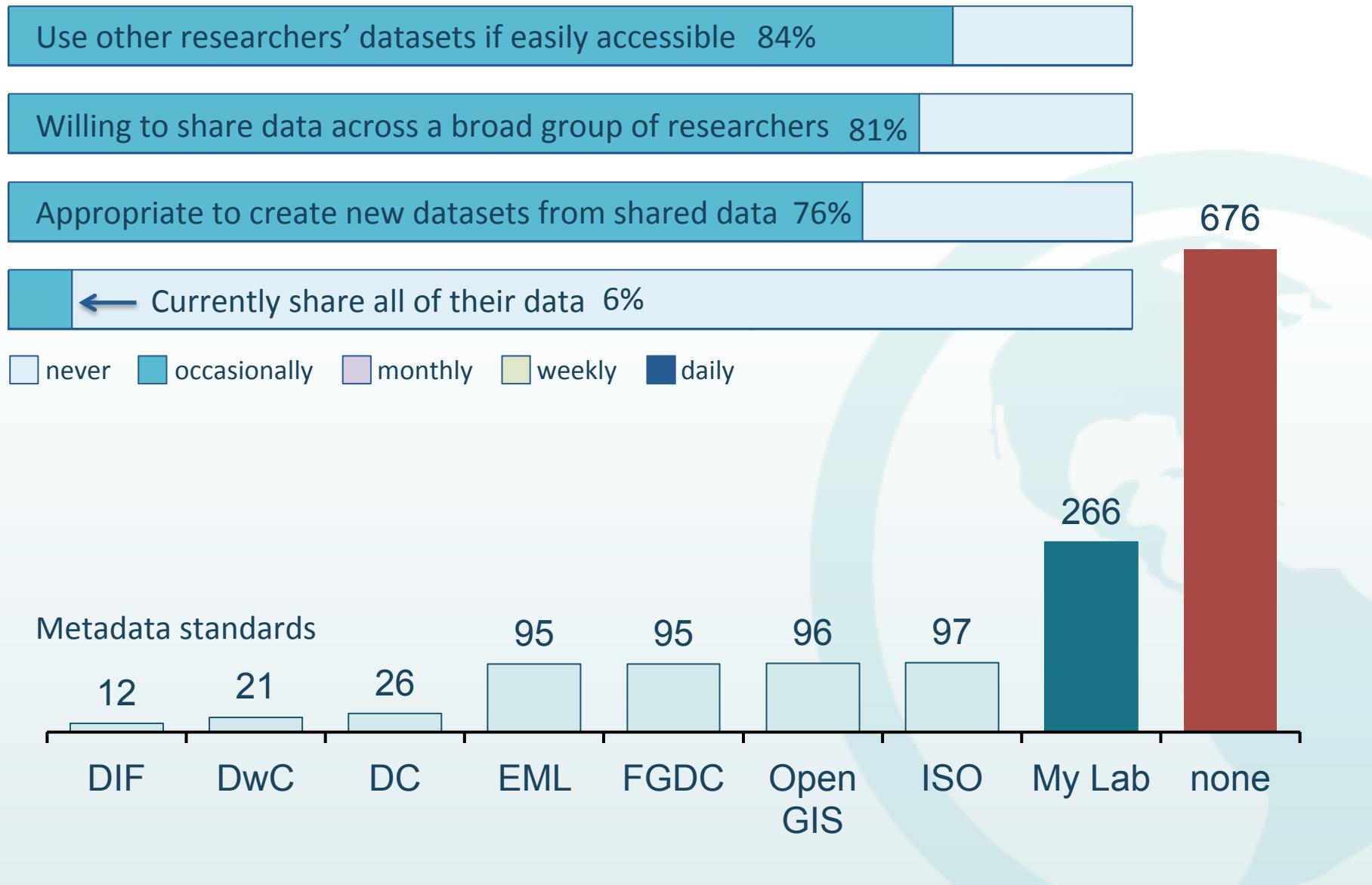
# DUG Surveys: ITK Priorities

---

9	<b>DataONE Drive</b>	9	<b>R</b>	7	<b>ArcGIS</b>
5	<b>R</b>	2	<b>Excel</b>	3	<b>R</b>
4	<b>Python library</b>	2	<b>Java library</b>	2	<b>Geoportal</b>
4	<b>Java library</b>	2	<b>PHP</b>	2	<b>OPeNDAP</b>
3	<b>IDL</b>	2	<b>Python library</b>	2	<b>Python library</b>
2	<b>ArcGIS</b>	1	<b>QuantumGIS</b>	1	<b>Java library</b>
1	<b>Geoportal Server</b>			1	<b>Matlab</b>
1	<b>Fortran library</b>			1	<b>Mendeley</b>
1	<b>DMP Tool</b>			1	<b>Metavist</b>
1	<b>Matlab</b>			1	<b>MP Parser</b>
1	<b>MetaVist</b>			1	<b>Spyder</b>
				1	<b>Thredds</b>
				1	<b>VisTrails</b>

86 Total Tools Identified

# Scientists not yet providing data services



# Investigator Toolkit



**DataONE**  
Data Observation Network for Earth

Search: ONEMercury For: Go Connect: YouTube LinkedIn Twitter Facebook RSS

About Participate Resources Education Data

Home » Resources » [Investigator Toolkit](#)

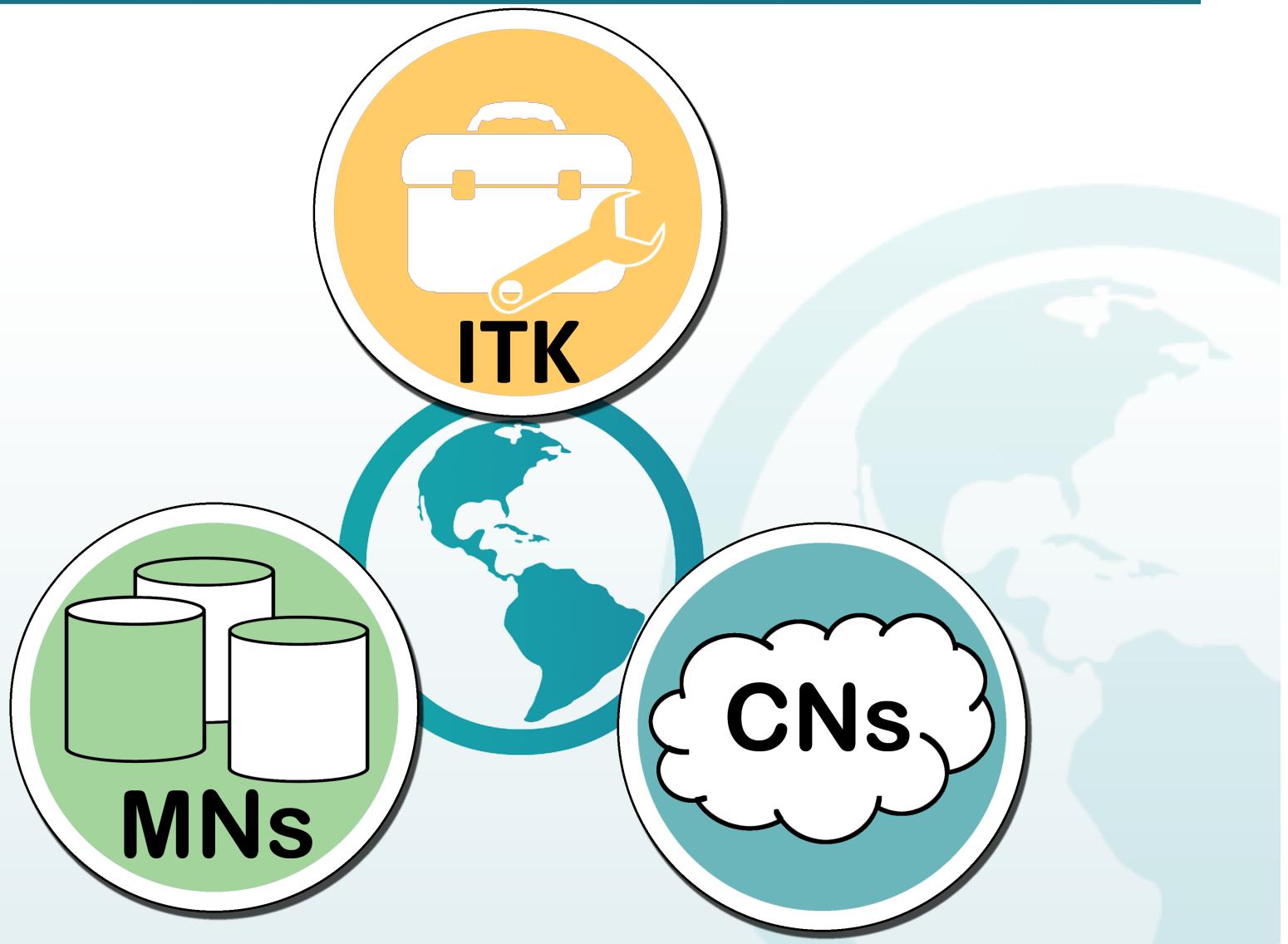
**Resources**

- Investigator Toolkit
- Data Management Planning
- Best Practices
- Software Tools Catalog
- Publications

The current Investigator Toolkit includes:

# Investigator Toolkit (ITK)

---



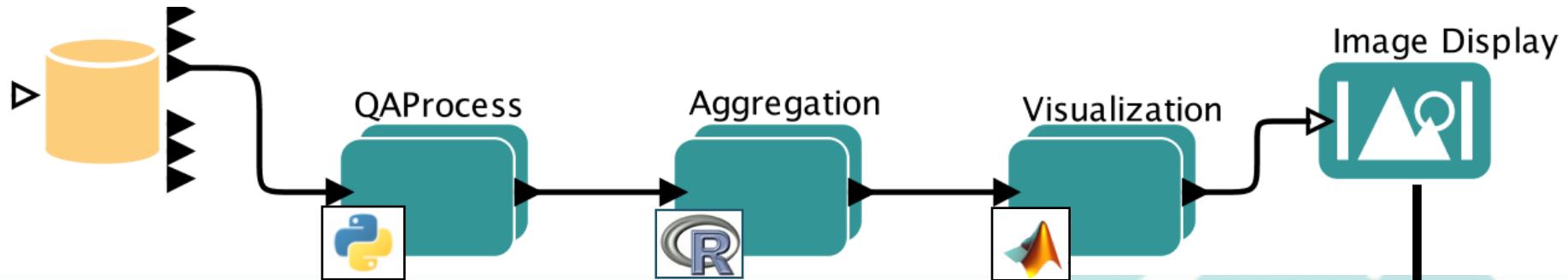
# Adapt Tools to DataONE Services



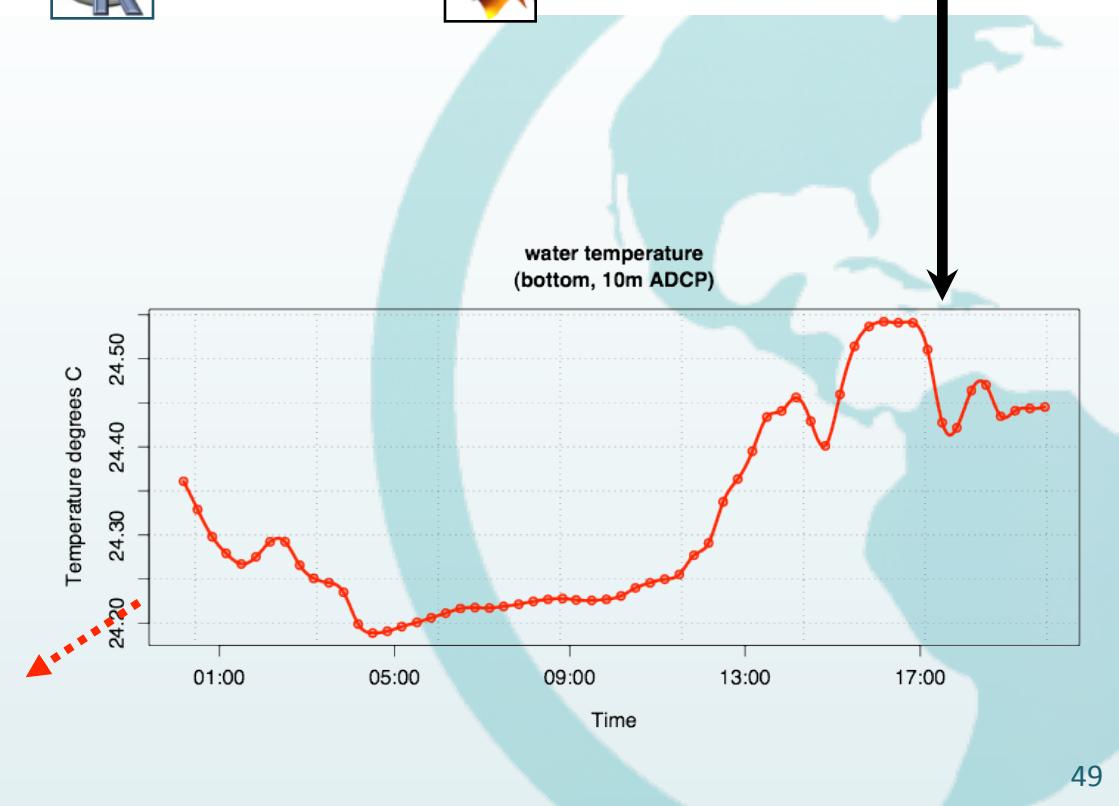
- Reduce stovepipe systems
  - Use tools with any Member Node
- Improve efficiency
  - Work together on Open Source
  - Reduce redundant efforts, better results possible
- Build community

# Integrate and Analyze: Kepler

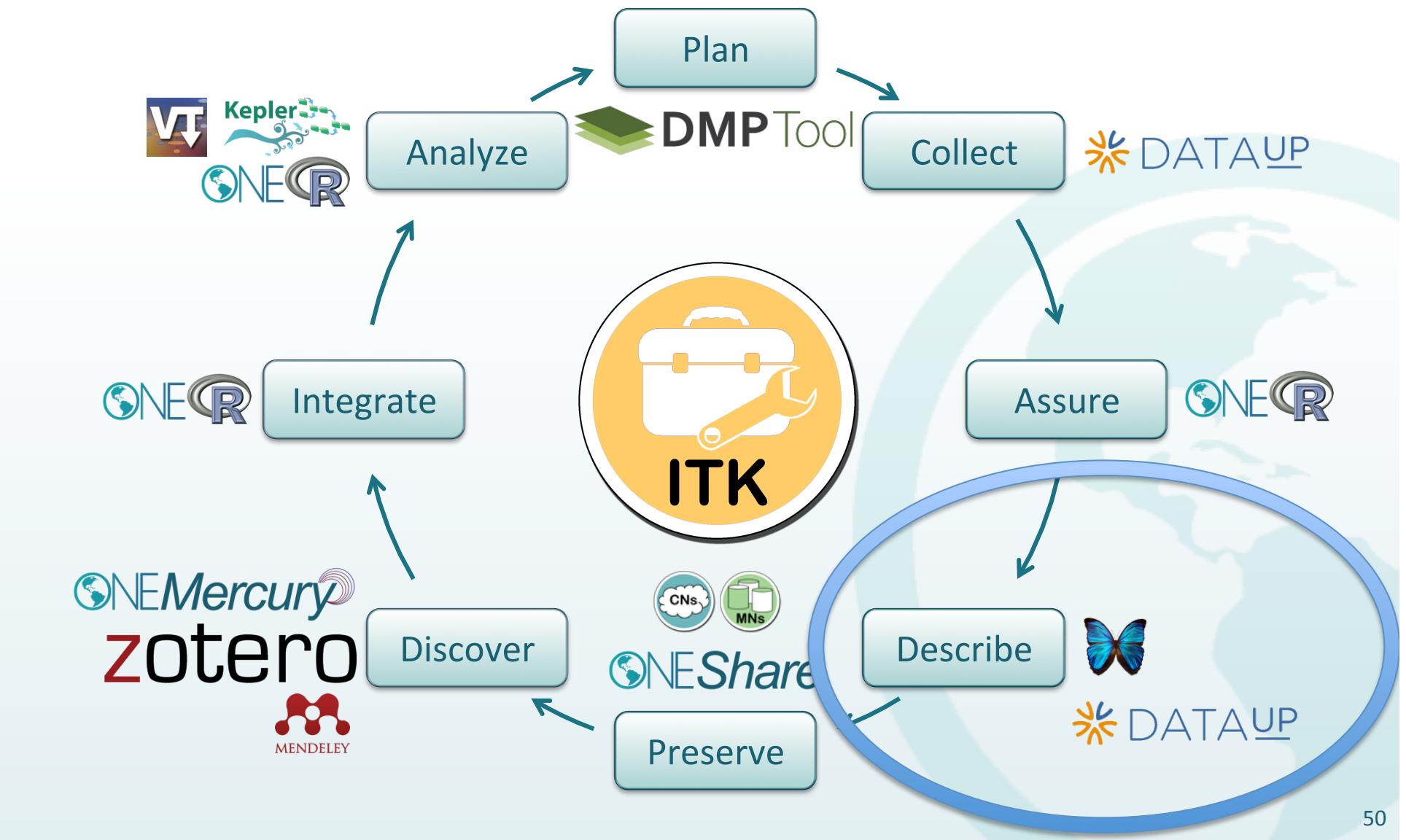
DataONE Data Source



Graphs and derived  
data can be  
archived in DataONE



# Tools Across the Data Life Cycle



# Describe: Morpho



- Data management for researchers
- Generates and captures metadata
- Organize and search your own data
- Version management
- Targets individual researchers
- Adaptation goals:
  - Work with any Member Node
  - Search DataONE
  - Support DOIs

