

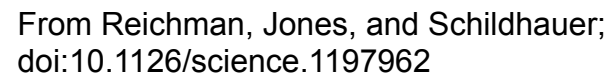
Data Observation Network for Earth Hands On

IDCC

February, 2014

Data **ONE**







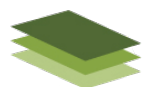
Investigator Toolkit



>> command line interface



python




DMPTool



DATAUP





Data Observation Network for Earth








Search

ONEMercury

For

Go

Connect

About

Participate

Resources

Education

Data

Home » Resources » [Investigator Toolkit](#)

Investigator Toolkit

Data Management Planning

Best Practices

Software Tools Catalog

Publications

Resources

Investigator Toolkit

Data Management Planning

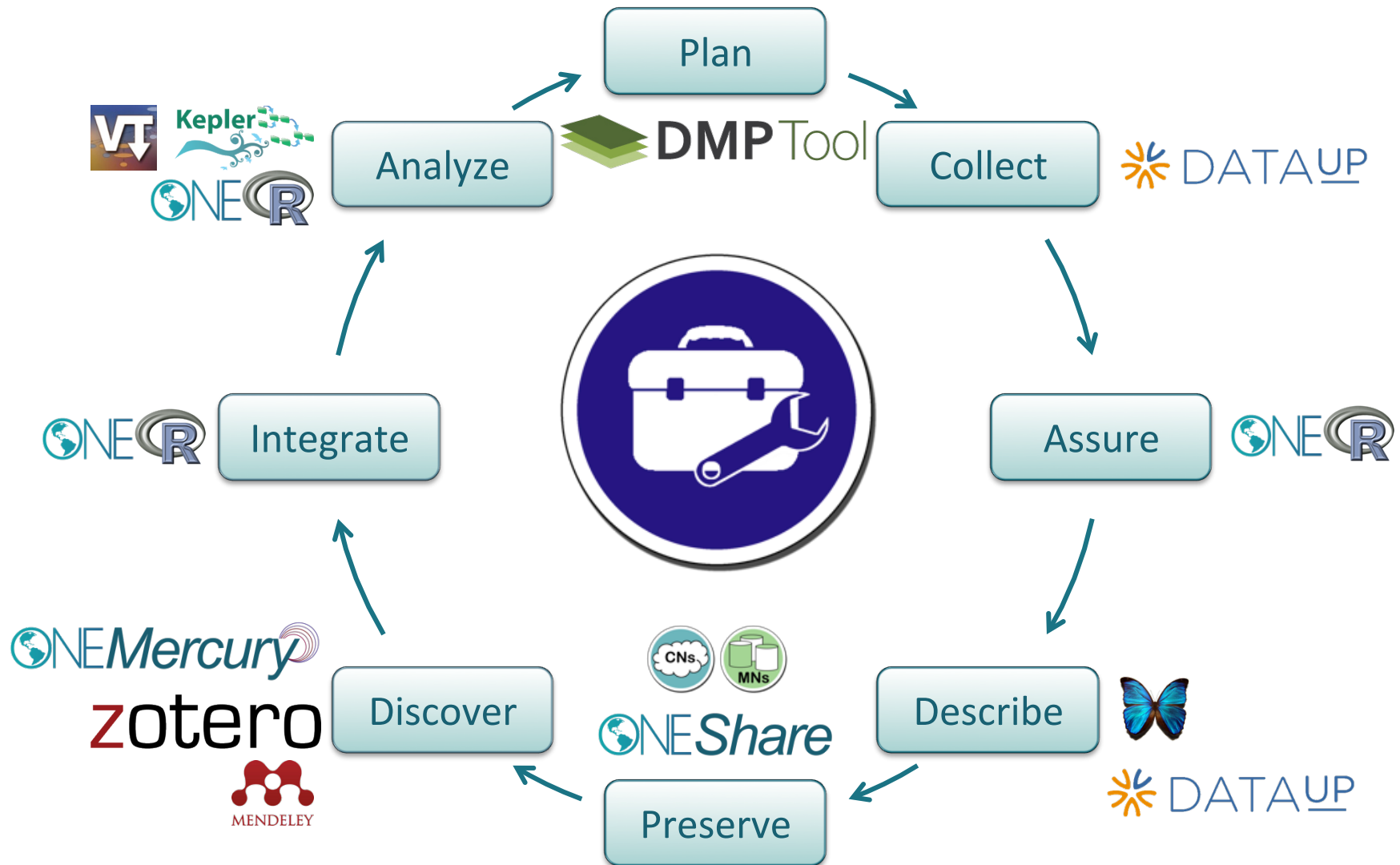
Best Practices

Software Tools Catalog

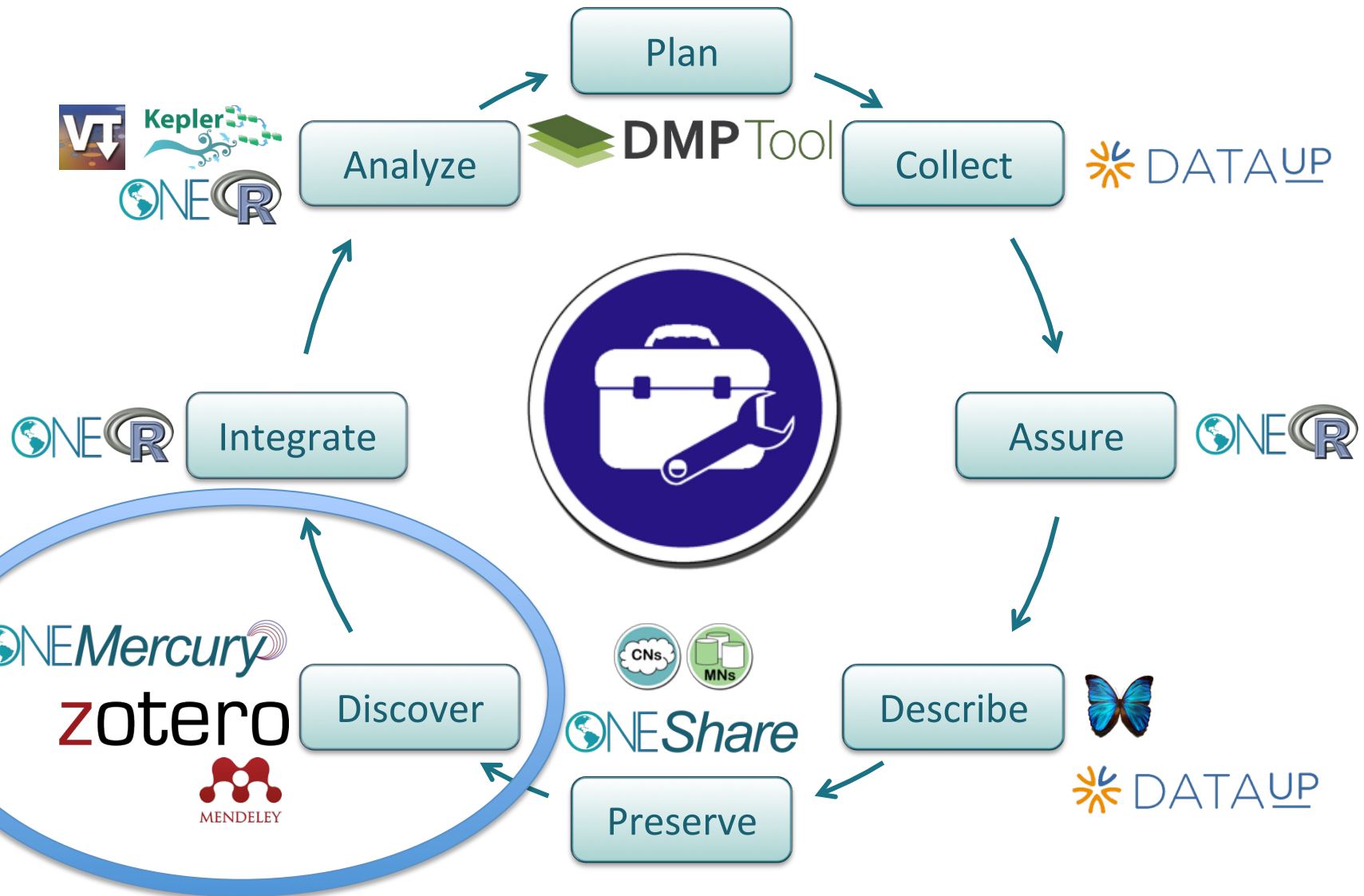
Publications

software tools for finding, using, and contributing data in DataONE. Some of DataONE, some are existing tools that have been modified to use the DataONE API, and some are tools that have well defined interfaces of their own. The current Investigator Toolkit includes:

Supporting the Full Data Life Cycle

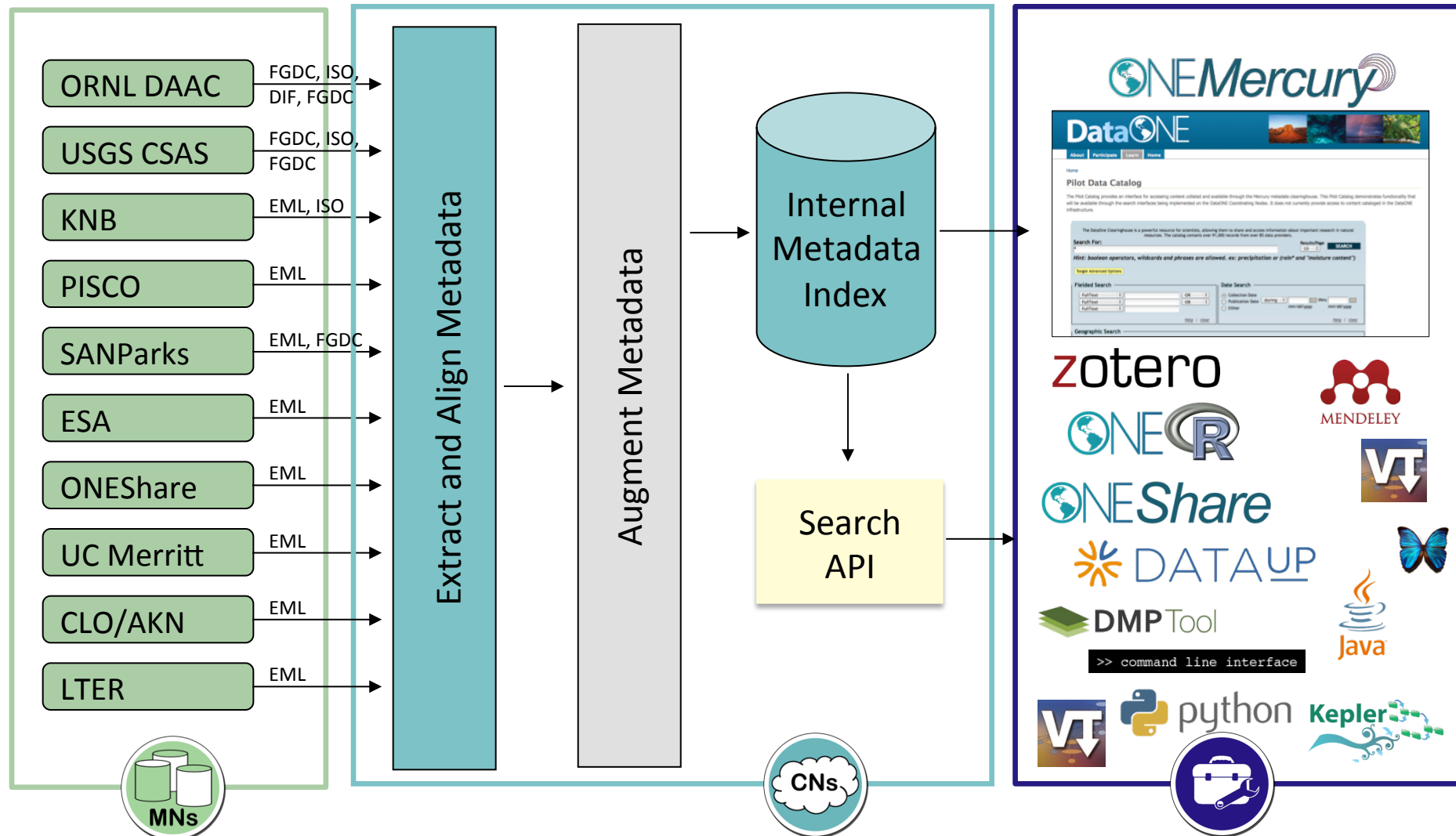


Supporting the Full Data Life Cycle



Enable Data Discovery

CNs



Discover

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

Search For:

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

Help | clear

Date Search

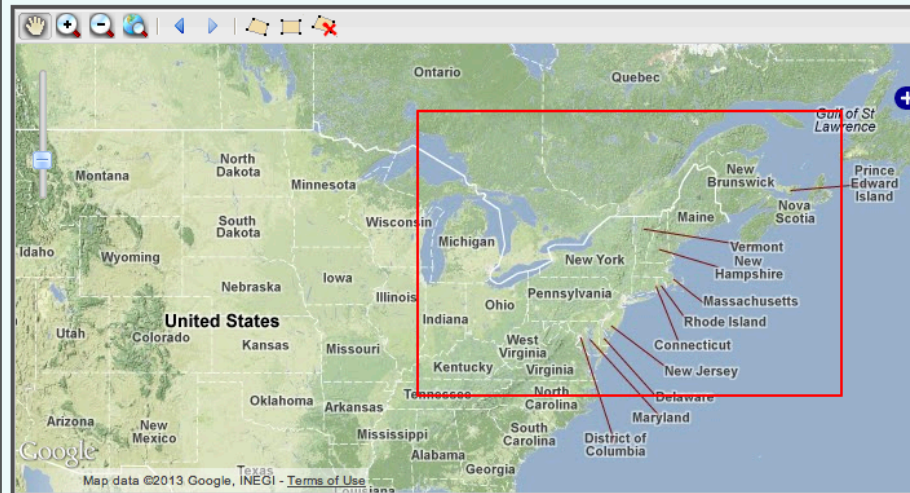
☒ Collection Date
☐ Publication Date
☐ Either

during thru

mm/dd/yyyy mm/dd/yyyy

Help | clear

Geographic Search



List Areas in:
USA ☒ WORLD ☐
Select from list
Search Area:
overlaps ☒ encloses ☐
North
49.809631
West East
-87.97851 -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

ONEMercury - Demo



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

Fielded Search interface with three input fields and dropdown menus.

Help | clear

Date Search

☒ Collection Date
☐ Publication Date
☐ Either

during

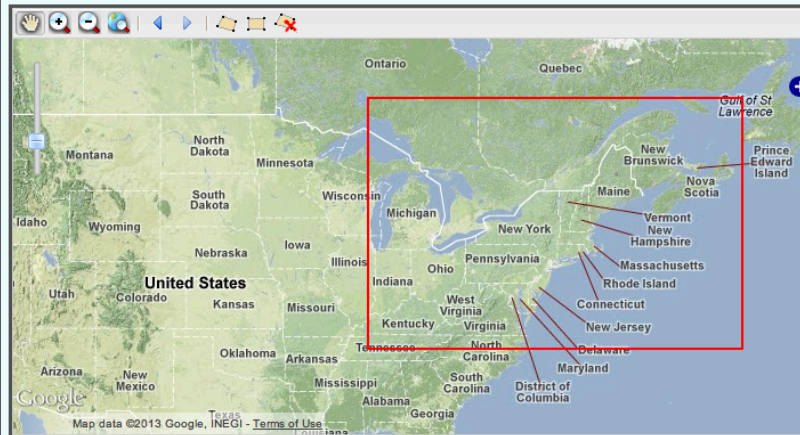
mm/dd/yyyy

thru

mm/dd/yyyy

Help | clear

Geographic Search



List Areas in:

USA ☒ WORLD

Select from list

Search Area:

☒ overlaps ☐ encloses

North

49.809631

West East

-87.97851 -59.15039

South

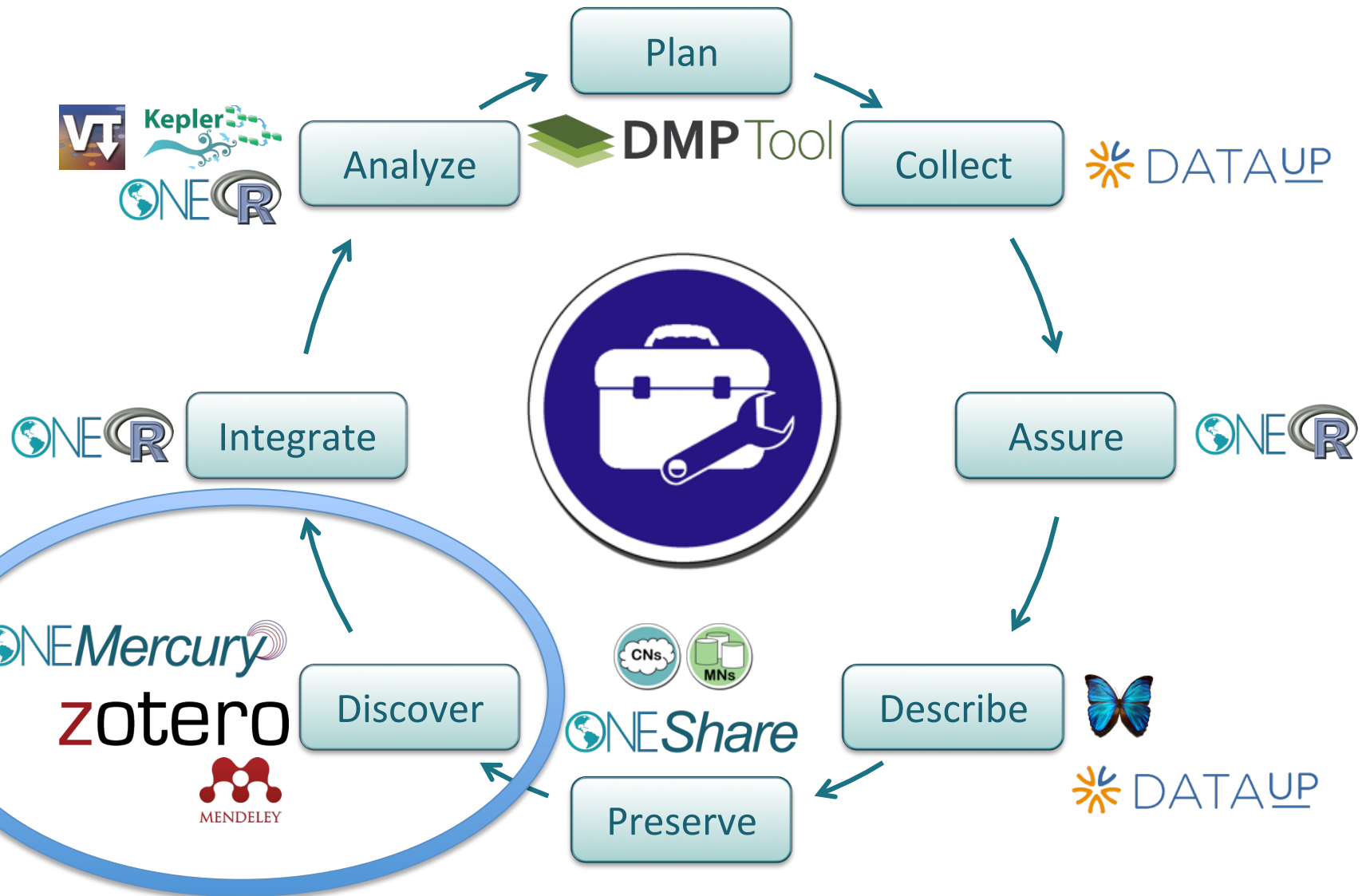
35.710837

Help | clear

Member Nodes

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

Supporting the Full Data Life Cycle



- Online citation manager support
- e.g. Zotero, Mendeley

DataONE
Data Observation Network for Earth

About Participate Resources Education Data

ONEmercury A DataONE Search Tool for Scientific Data

Return to Search

Query: text : water AND has data

Hide Filters

Filter by Member Node	Filter by author	Filter by project	Filter by keywords
Merritt Repository (31446)	Bruce Menge (5716)	Partnership for Interdisciplinary (17279)	California (42882)
PISCO MN (18029)	Libe Washburn (5196)	health (31447)	contaminants (31444)
ORNL DAAC (400)	Margaret McManus (4263)	Olympic Coast National (612)	drinking water (31444)
Knowledge Network for Biocomplexity (379)	Jack Barth (1885)	Gulf of the (84)	marsh Ecology Research (40)
Dryad Digital Repository (296)	CDPH Merced District (1523)	No project title (20)	continental shelf (17978)
TRFRI Data Catalog (103)	CDPH San Bernardino District (1365)	Partnership for Interdisciplinary	Oceanographic Sensor Data (17975)

Sort By: Relevance Start Date Most Recent

Viewing Documents 1 - 10 out of 50842
Prev 1 2 3 4 5 6 7 8 9 10 Next

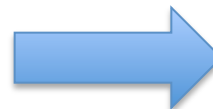
DAVIDSON, E.A.; FIGUEIREDO, R. DE; MARKIEWITZ, D. 08/12/1999. LBA-ECO ND-02 STREAM WATER CHEMISTRY, PARAGOMINAS, PARA, BRAZIL: 1999-2005.
Identifier: scmetri_1067.xml Data source: ORNL DAAC
This data set includes measurements of dissolved nutrient and organic carbon concentrations, as well as dissolved oxygen, alkalinity, conductivity, turbidity, pH, and discharge from three streams located in mixed land use (crop fields, pastures, secondary vegetation, and forest) and two streams in entirely forested landscapes near Paragominas in the state of Para, Brazil. Stream water samples were collected during two different periods: 1) weekly from August 1999 to July 2001 at location Igarape 54, Station 5 and 2) monthly from April 2003 through October 2005 at all of the stations. The exact...

View full metadata Download

Cal Water Service. 04/05/2002. DWSAP Assessment for 4900546-001: HAWKINS WATER CO-CAL WATER SERVICE (PUC) / WELL 02.
Identifier: ark:/13030/m5mk8wv1/cadwsp-4900546-001.xml Data source: Merritt Repository
A source water assessment identifies the vulnerability of the drinking water supply to contamination from typical human activities. The assessments are intended to facilitate and provide the basic information necessary for a local community to develop a program to protect the drinking water supply...

View full metadata Download

Cal Water Service. 04/05/2002. DWSAP Assessment for 4900546-002: HAWKINS WATER CO-CAL WATER SERVICE (PUC) / WELL 01 - STANDBY.
Identifier: ark:/13030/m5gb2vz2/cadwsp-4900546-002.xml Data source: Merritt Repository



Cite

Zotero

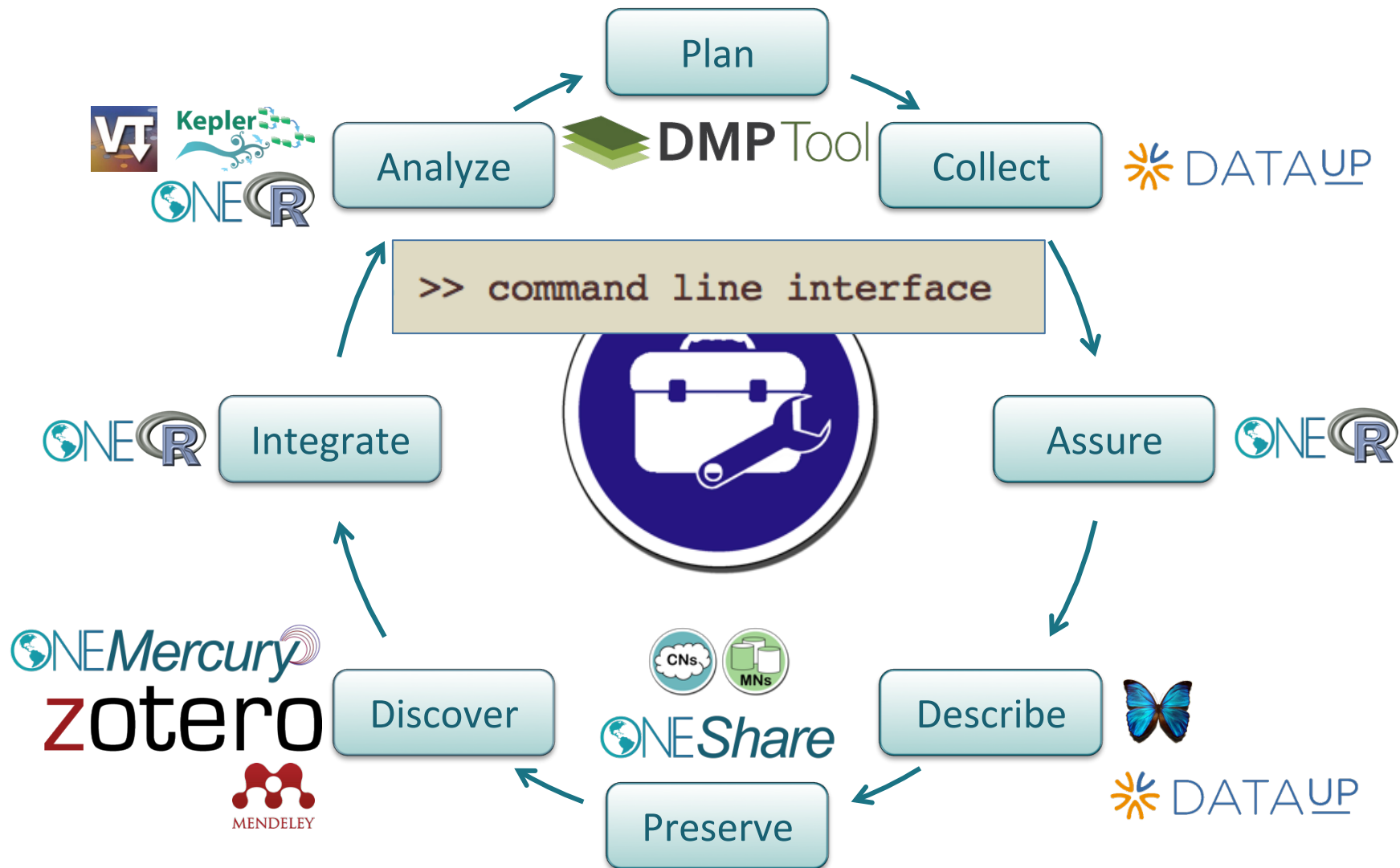
Search: All Fields & Tags

Title	Creator	Info	Notes	Tags
A global synthesis reveals biodiversity loss as a major driver of ecosystem change	Hooper et al.			
Analysis of Carbon Sequestration and Carbon Dioxide Reduction by Post and Beam Woo...	TU			
A Search Tool for Scientific Data				
BOREAS T28-08 PHOTOSYNTHETIC RATE DATA OVER THE SSA-OBS AND THE SSA-OJP	LEIRDAU			
CANOPY PHOTOSYNTHESIS RATES (PFE)	ASRAR			
Cost-benefit and mechanism design of the forest carbon sink in Taiwan (1/2)	Lin			
Data Sharing by Scientists: Practices and Perceptions	Tenopir et al.			
Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the ha...	Möller-Levet et al.			
Elevated CO2 stimulates net accumulations of carbon and nitrogen in land ecosystems: ...	NCIAS 3880 : Luo: Nitrogen...			
Environment Canada - Climate Change - Quality Assurance/Quality Control and Data Ma...	Government of Canada			
LEAF PHOTOSYNTHESIS RATES (PFE)	NORMAN			
Science and Decisions: Advancing Risk Assessment				

Item Type: Journal Article
Title: A global synthesis reveals biodiversity loss as a major driver of ecosystem change
Author: Hooper, David U.
Adair, E. Carol
Cardinale, Bradley J.
Byrnes, Jarrett E.
Hungate, Bruce A.
Matalich, Kristin L.
Gonzalez, Andrew
Duffy, J. Emmett
Camfield, Lars
O'Connor, Mary L.
Abstract: Nature
Publication: Nature
Volume:
Pages: 2012-5-2
Series:
Series Title:
Series Text:
Journal Abb: Journal Abbr:
Language:
DOI: 10.1038/nature.1...
ISSN: 0028-0836, 1476...
Short Title:
URL: http://en.da...
Accessed: Wed Mar 20 21:41...
Archive:
Loc. in Archive:
Library Catalog: CrossRef
Call Number:
Rights:

[to be added] AIR TEMPERATURE
AIR TEMPERATURE/GAS EXCHANGE SYSTE...
bloodomics BOREAL FORESTS/CANADA
CANOPY CHARACTERISTICS/GAS EXCHAN...
carbon dioxide reduction
carbon sequestration chronobiology

Tools Across the Data Life Cycle





Programmatic Search

`https://cn.dataone.org/cn/v1/query/solr/q=abstract:water`

Diagram illustrating the components of the URL:

- CN REST endpoint
- Method
- Query Engine
- Query

- Result is a list of identifiers (and more)
- Use `resolve()` to determine locations
- Use `get()` to retrieve bytes



GETting data: the main target

Where can I get doi:10.5063/AA/mbauer.75.1?

<https://cn.dataone.org/cn/v1/resolve/doi%3A10.5063%2FAA%2Fmbauer.75.1>

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<d1:objectLocationList xmlns:d1="http://ns.dataone.org/service/types/v1">
```

```
<identifier>doi:10.5063/AA/mbauer.75.1</identifier>
```

```
<objectLocation>
```

```
<nodeIdentifier>urn:node:KNB</nodeIdentifier>
```

```
<baseURL>https://knb.ecoinformatics.org/knb/d1/mn</baseURL>
```

```
<version>v1</version>
```

```
<url>https://knb.ecoinformatics.org/knb/d1/mn/v1/object/doi:10.5063%2FAA%2Fmbauer.75.1</url>
```

```
</objectLocation>
```

```
</d1:objectLocationList>
```

CN Resolve
method call

MN Get call to
retrieve data

A DataONE GET (digital object) call

<https://knb.ecoinformatics.org/knb/d1/mn/v1/object/doi:10.5063%2FAA%2Fmbauer.75.1>

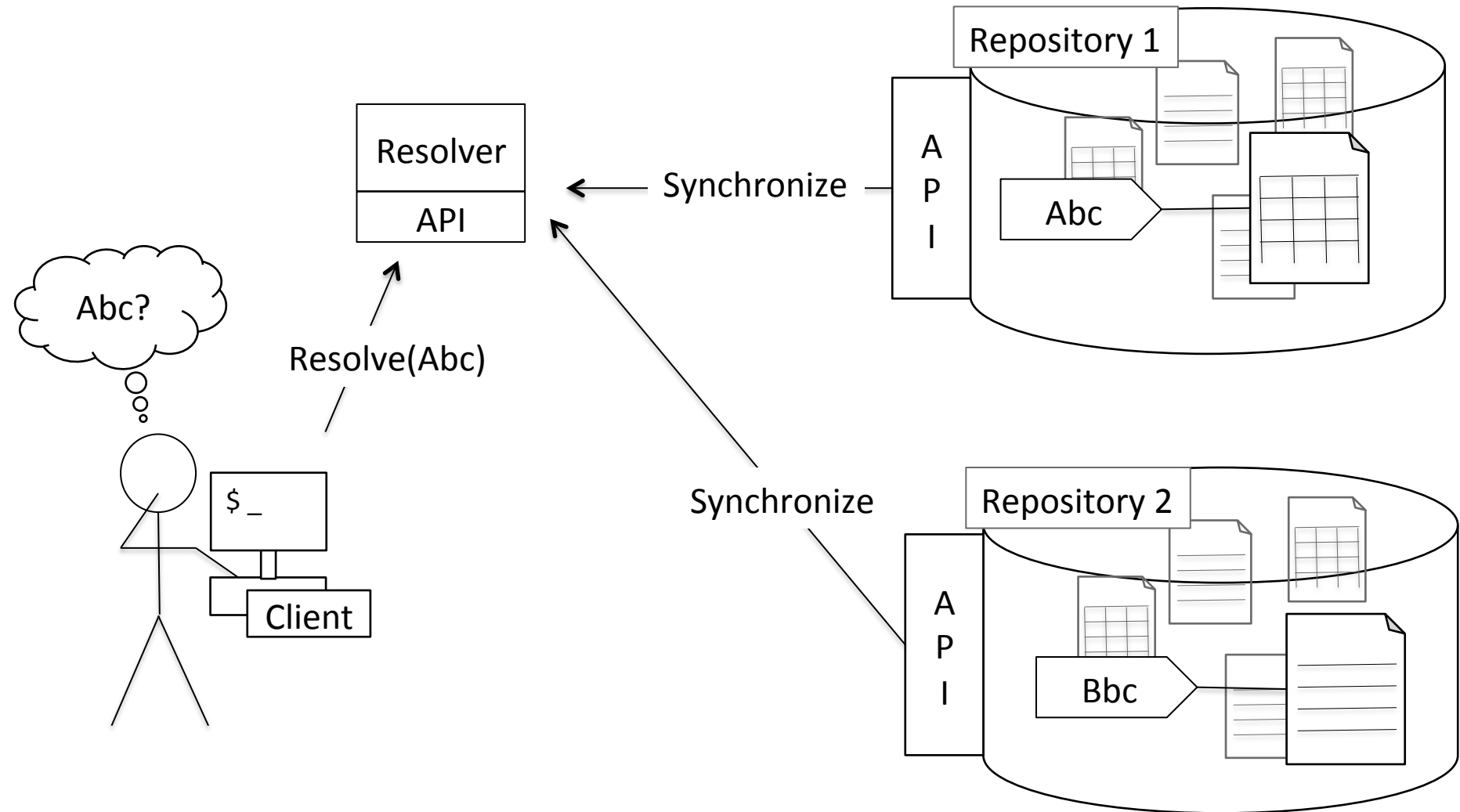
MN REST Endpoint

Method

Object Identifier



Data – Resolve?





Resolving Identifiers



Where can I get doi:10.5063/AA/mbauer.75.1?

<https://cn.dataone.org/cn/v1/resolve/doi%3A10.5063%2FAA%2Fmbauer.75.1>

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<d1:objectLocationList xmlns:d1="http://ns.dataone.org/service/types/v1">
```

```
<identifier>doi:10.5063/AA/mbauer.75.1</identifier>
```

```
<objectLocation>
```

```
<nodeIdentifier>urn:node:KNB</nodeIdentifier>
```

```
<baseUrl>https://knb.ecoinformatics.org/knb/d1/mn</baseUrl>
```

```
<version>v1</version>
```

```
<url>https://knb.ecoinformatics.org/knb/d1/mn/v1/object/doi:10.5063%2FAA%2Fmbauer.75.1</url>
```

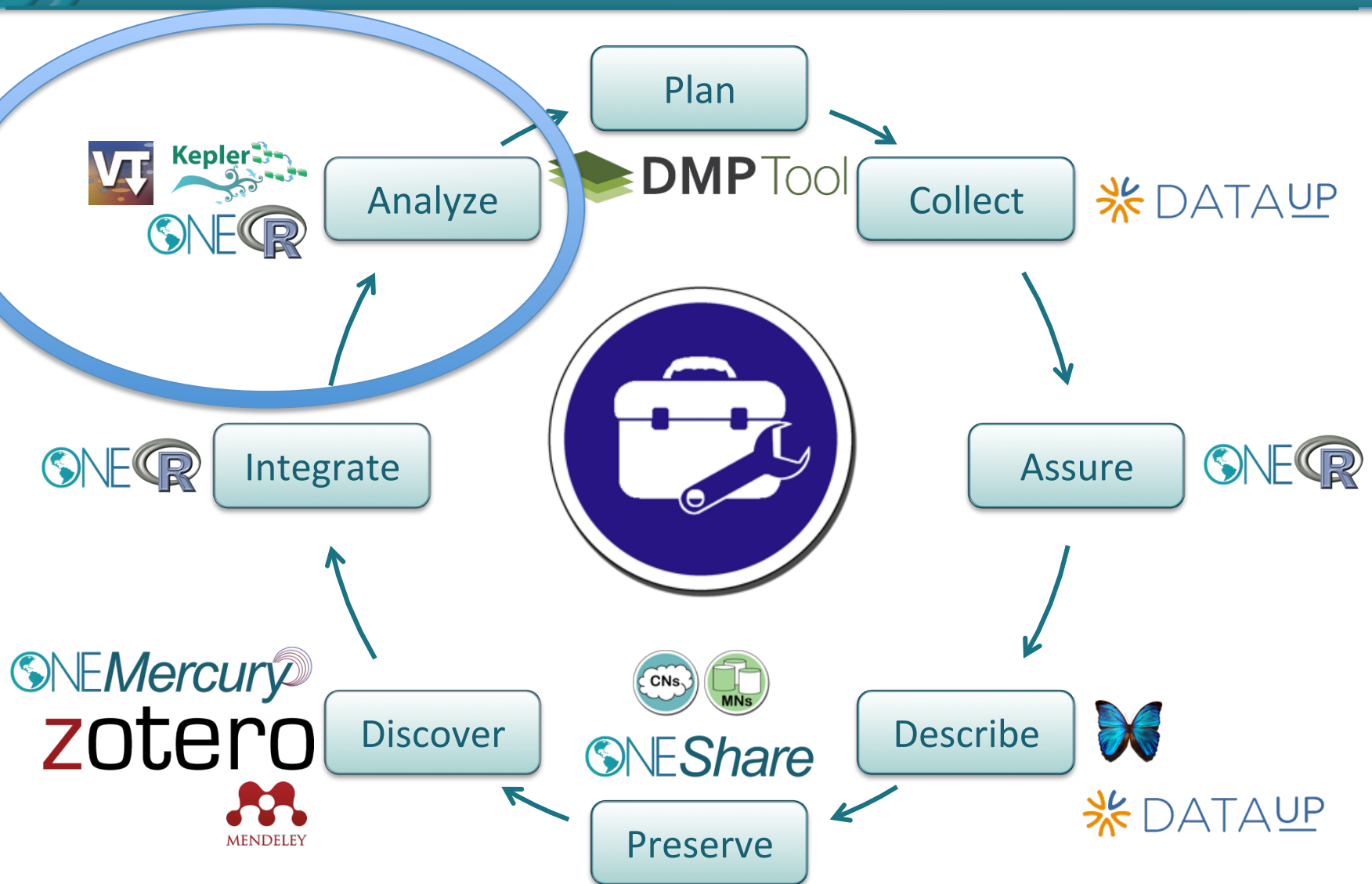
```
</objectLocation>
```

```
</d1:objectLocationList>
```

CN Resolve
method call

MN Get call to
retrieve data

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")  
databytes <- 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

```
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 be
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"
```


Workspace History
Load Save Import Dataset Clear All

Run Source

1:1 (Top Level) R Script

Console

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
- ADP Alliance
- RSA-JS
- FindPAM.cnake
- CPackRPM.cnake
- tree.h

OK

Packages Help
Export Clear All