



NSF

**ARCTIC  
Data  
Center**

# Overview and NSF Standards and Policies

Matthew B. Jones



**DataONE**



**<https://arcticdata.io>**  
**NSF Award #: 1546024**





Photographer: Jeremy Potter NOAA/OAR/OER <https://flic.kr/p/8F3Q6M>



Troms Fylke



Rama



Detroit Publishing Co.



# Key Deliverables

- **Data Archive**
- **Portal** for data discovery
- **Tools** for data and metadata submission
- **Repository** cyberinfrastructure
  - *Provenance features*
  - *Replication features*
  - *Metadata quality check*
- **Support** services
- **Training**
- **Outreach**
- **Overview:** <http://bit.ly/arctica-summary>



# Key Deliverables

- **Data Archive**
- **Portal** for data discovery
- **Tools** for data and metadata submission
- **Repository** cyberinfrastructure
  - *Provenance features*
  - *Replication features*





# Team



M.Jones



Arzayus



Baker-Yeboah



Budden



Casey



Dozier



Schildhauer



Walker\*



C.Jones\*



Mecum\*



Clark



Goldstein\*



Lortie

- Student Interns
  - J.Graybiel, A.Prescott, I.Su, J.S.Raquel, K.McGill, S.Halperin, S.Lee, A.Gordee, D.Mullen, H.Kim



# Milestones: Year 1

Cyberinfrastructure	✓	<b><i>CI-1: Launch initial system</i></b>
	✓	<b><i>CI-2: Prototype content replication</i></b>
	✓	<b><i>CI-3: Improve data portal</i></b>
	✓	<b><i>CI-4: Customize web submission</i></b>
	✓	<b><i>CI-5: Deploy client tools</i></b>
Support Systems	✓	<b><i>SS-1: Deploy support infrastructure</i></b>
	✓	<b><i>SS-2: Repository operation</i></b>
	✓	<b><i>SS-3: Support Data Management Planning</i></b>
Community Engagement	✓	<b><i>CE-1: Build public website</i></b>
	⟳	<b><i>CE-2: Create Data Science fellowship</i></b>
	⟳	<b><i>CE-3: Conduct training</i></b>
	⟳	<b><i>CE-4: Conduct synthesis project</i></b>



# Operation Metrics



**4,700**  
DATA SETS



**631K**  
DATA FILES



**56K**  
FILE DOWNLOADS



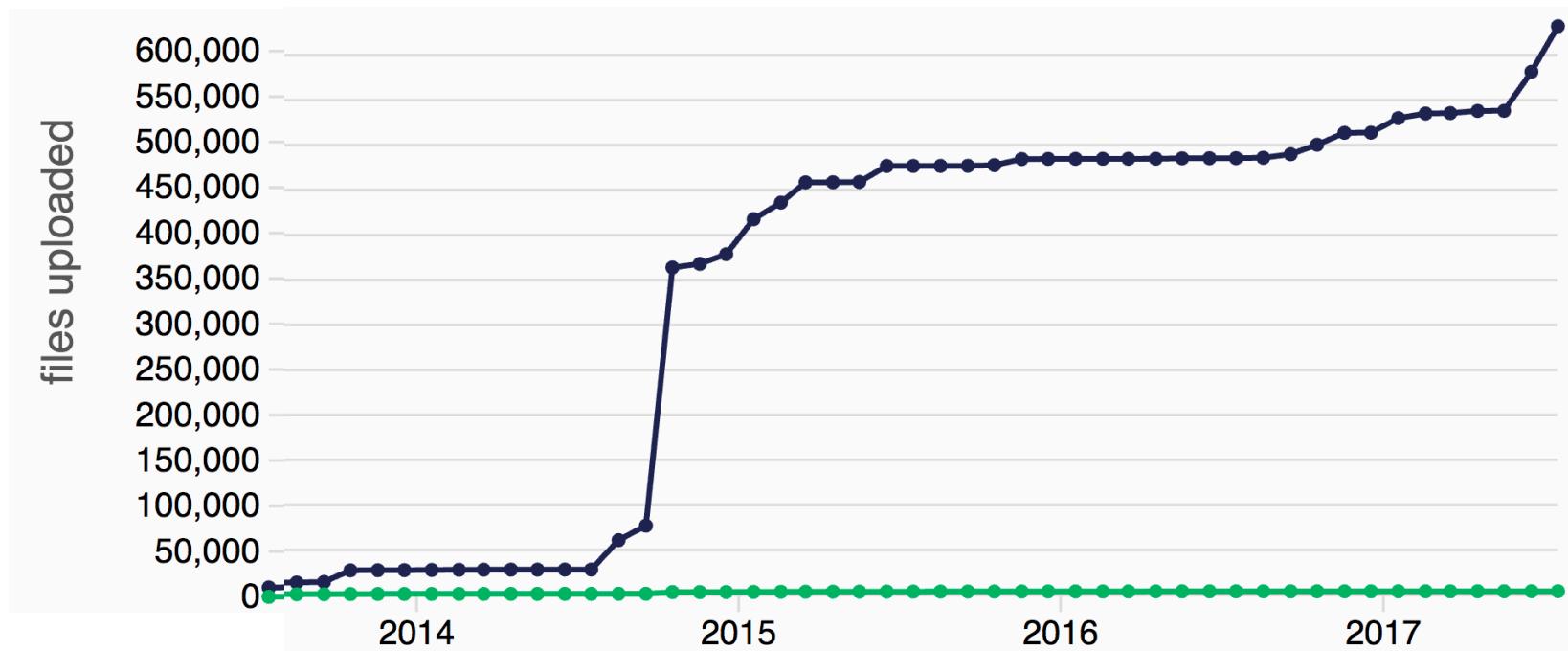
**1,900**  
CREATORS



**3,800**  
USERS

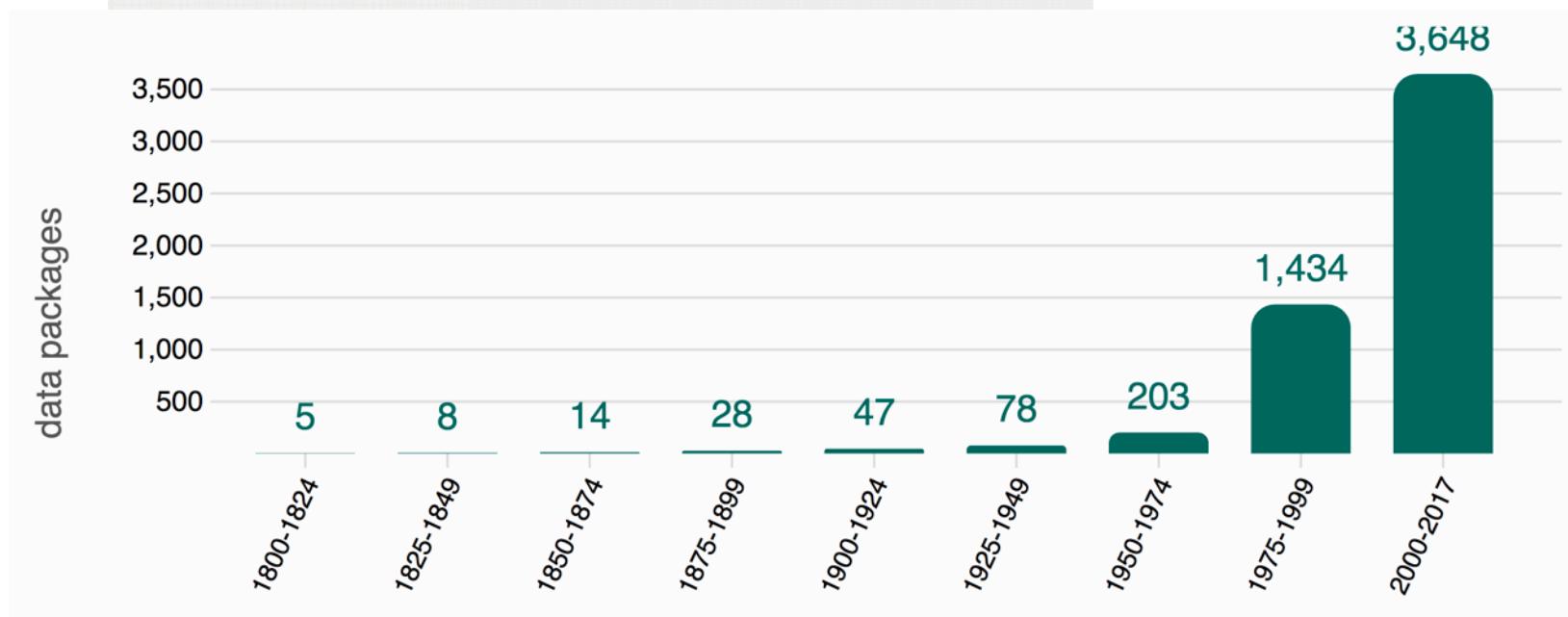
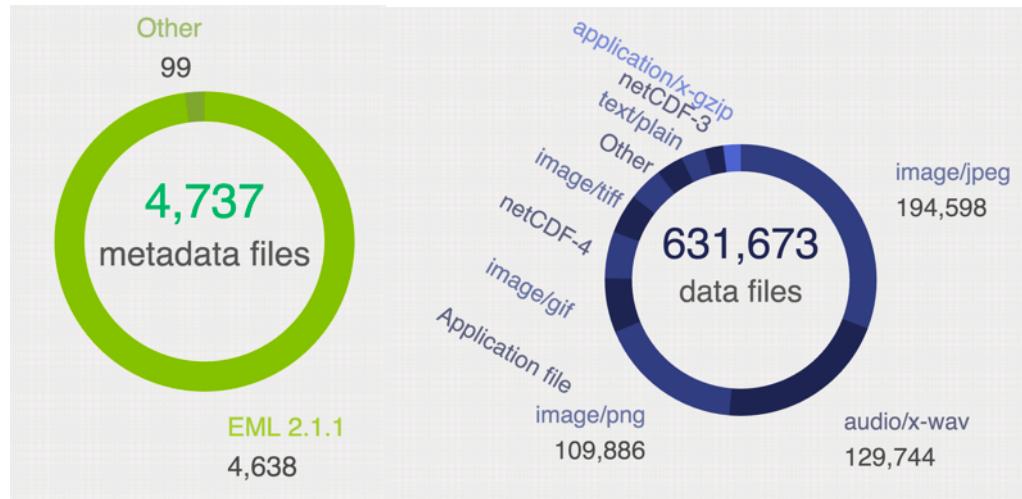


# An Arctic Research Data Archive





# Data Holdings





# Data Discovery Portal

NSF  
**Arctic  
Data  
Center**

Data Support About Submit Data Sign in with Orcid

Search ?  
Search phrase

Filter by:  
Data attribute  
Creator  
Year  
Identifier  
Taxon  
Location

DATASETS 1 TO 25 OF 4,149  
1 2 3 ... 166 Next Sort by Most recent

Paul Bierman. 2016. **Greenland 10Be cosmogenic date in marginal sediment.** NSF Arctic Data Center. urn:uuid:1c821d0c-926d-4efc-b88f-7b99b4875425.  
 33

NSF Arctic Data Center. 2015. **1Learning about environmental research in the context of climate change: an international scholastic interchange.** NSF Arctic Data Center. doi:10.18739/A27H41.  
 7

NSF Arctic Data Center. 2015. **1Learning about environmental research in the context of climate change: an international scholastic interchange.** NSF Arctic Data Center. doi:10.18739/A2H08V.  
 6

Barbara Bodenhorn. 2015. **Learning about environmental research in the context of climate change: an international scholastic interchange.** NSF Arctic Data Center. urn:uuid:8eb64350-1e39-4b82-a2c7-2f62d84e8b5d.  
 5

Joseph McConnell. 2016. **Tunu, Greenland 2013 ice core chemistry.** NSF Arctic Data Center. urn:uuid:39ec61f1-0627-4ebc-ae3c-8b6261dce661.  
 5

Yasushi Fukamachi, Kay I. Ohshima, Daisuke Simizu, T. Takatsuka, K. Iwamoto, A.

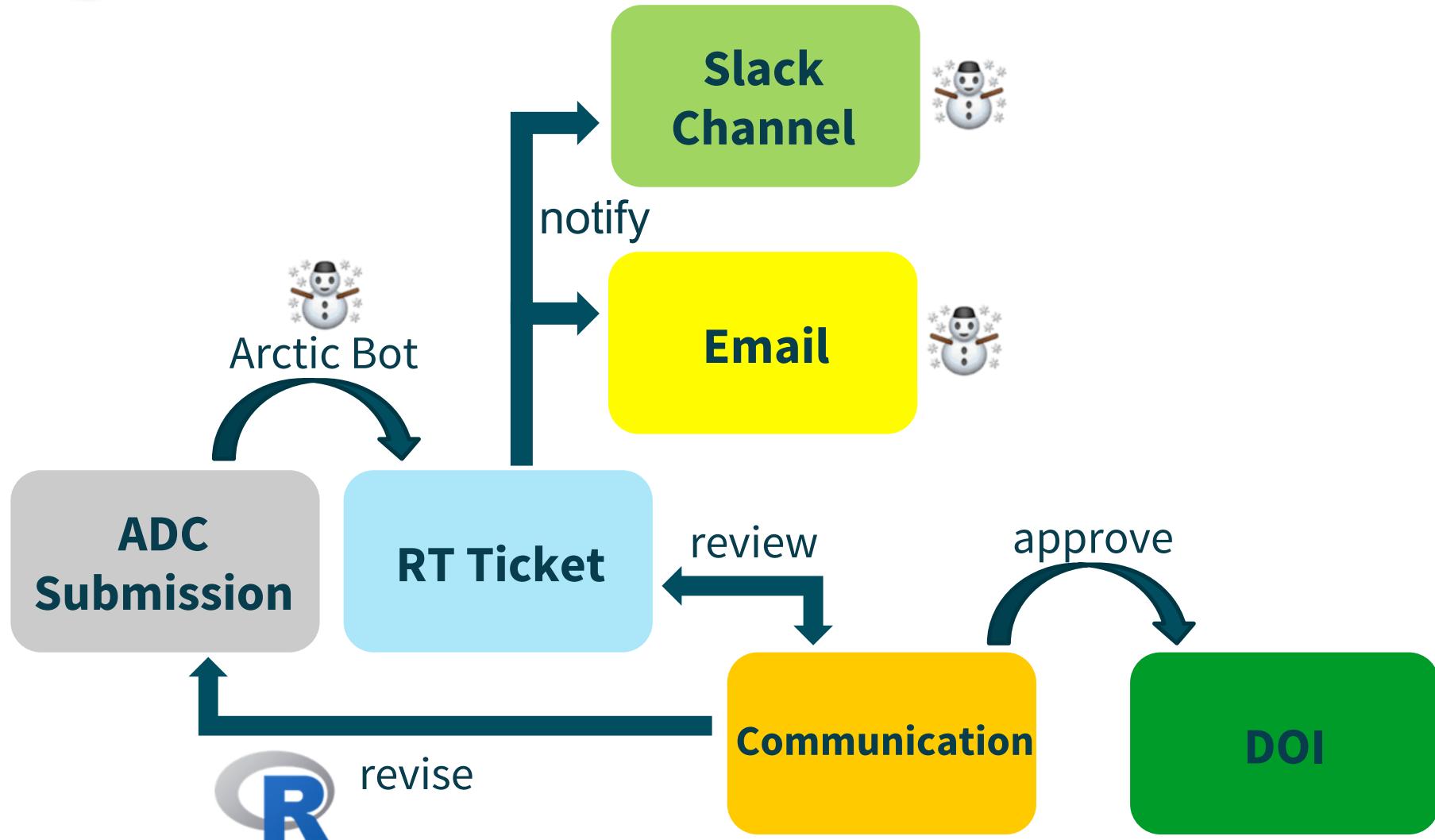
Hide Map »

Map data ©2016 1000 km Google Terms of Use

The map displays a global distribution of datasets. A prominent red dashed box highlights a cluster of datasets in the Arctic region, specifically around Greenland and the North Atlantic. The numbers in the boxes represent the count of datasets for each location. The map includes labels for countries like Canada, United States, Mexico, Venezuela, Colombia, Brazil, Argentina, Chile, Bolivia, Peru, and others. It also shows the Great Lakes and the St. Lawrence River area. The legend at the bottom indicates 'Satellite' and 'Terrain' modes, and the 'Google' logo is present.



# Curation process





# Policies

- **Data producers**
  - Who, what, and how
- **Data consumers**



# Who Must Submit

<https://arcticdata.io/submit/#who-must-submit>

- **Arctic Research Opportunities (ARC):**
  - Complete metadata and all appropriate data and derived products
  - Within 2 years of collection or before end of award, whichever comes first
- **ARC Arctic Observing Network:**
  - Complete metadata and all data
  - Real-time data made public immediately
  - Within 6 months of collection

ACADIS	Standards	Systems	Metrics	Discussion
--------	-----------	---------	---------	------------



# Who Must Submit: Social Sciences

<https://arcticdata.io/submit/#who-must-submit>

- Arctic Social Sciences Program (ASSP):
  - NSF policies include special exceptions for ASSP and other awards that contain sensitive data
  - Human subjects, governed by an Institutional Review Board, ethically or legally sensitive, at risk of decontextualization
  - Metadata record that documents non-sensitive aspects of the project and data
    - Title
    - Contact information
    - Abstract
    - Methods

ACADIS	Standards	Systems	Metrics	Discussion
--------	-----------	---------	---------	------------



# Terms of Use: Licensing and Distribution

<https://arcticdata.io/submit/#license>

- All metadata and (non-sensitive) data will be released under either:
  - CC-0 Public Domain Dedication:  
  
*“... can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.”*
  - Creative Commons Attribution 4.0 International License:  
  
*“... free to... copy,... redistribute,... remix, transform, and build upon the material for any purpose, even commercially,... [but] **must give appropriate credit**, provide a link to the license, and indicate if changes were made.”*



# Data Citation

- We assign a DOI to each published data set
- Researchers should cite data they use

Nina J. Karnovsky and Ann M. A. Harding. 2016. At-sea density of foraging little auks (*Alle alle*) near Hornsund Fjord. Arctic Data Center. doi:10.5065/D6MK6B17.

- We are working with DataCite to track the citations to data
  - To be displayed next year

 **NOTE:** A newer version of this dataset exists

[Home](#) / [Search](#) / [Metadata](#)

Nina J. Karnovsky, Pomona College, Ann M. A. Harding, Environmental Science Department, Alaska Pacific University, and UCAR/NCAR - Earth Observing Laboratory. 2016. **At-sea density of foraging little auks (*Alle alle*) near Hornsund Fjord.** Arctic Data Center. urn:uuid:849a7036-8dc4-400e-a584-9d1aaafacca63.

- Each update has a unique identifier
- Cite the exact version used
  
- Newer versions are clearly indicated



# Data usage counts

Files in this dataset   Package: resource\_map\_urn:uuid:6cf078d8-9466-4ce

Name	File type	Downloads
Metadata: iso19139.xml	http://www.isotc211.org/2005/gmd	3 views
dispatches_imnavait_apr2012.pdf	PDF	852 downloads
depth_happyvalleylines_apr2012.xlsx	Microsoft Excel OpenXML	274 downloads
depth_imnav_apr2012_1by1grid.xlsx	Microsoft Excel OpenXML	209 downloads
<a href="#">▶ Show 4 more items in this data set</a>		

[Download All](#)

[Download](#)

[Download](#)

[Download](#)

[Download](#)



# Preparing Data and Metadata

<https://arcticdata.io/submit/#preparing-data>



- **ADC supports upload of any data file format**
- **Sharing greatly enhanced if using open-source formats**
  - Microsoft Excel files are commonplace, but better to export to CSV text files (do not require Microsoft products)
  - Export GIS data to ESRI shapefiles
  - Export data created in MATLAB or other matrix-based programs to NetCDF (an open binary format)
- **Describe data with complete metadata**
  - What, who, when, where, how, and why



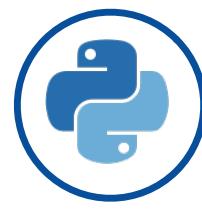
MATLAB



R



WEB PORTAL



PYTHON



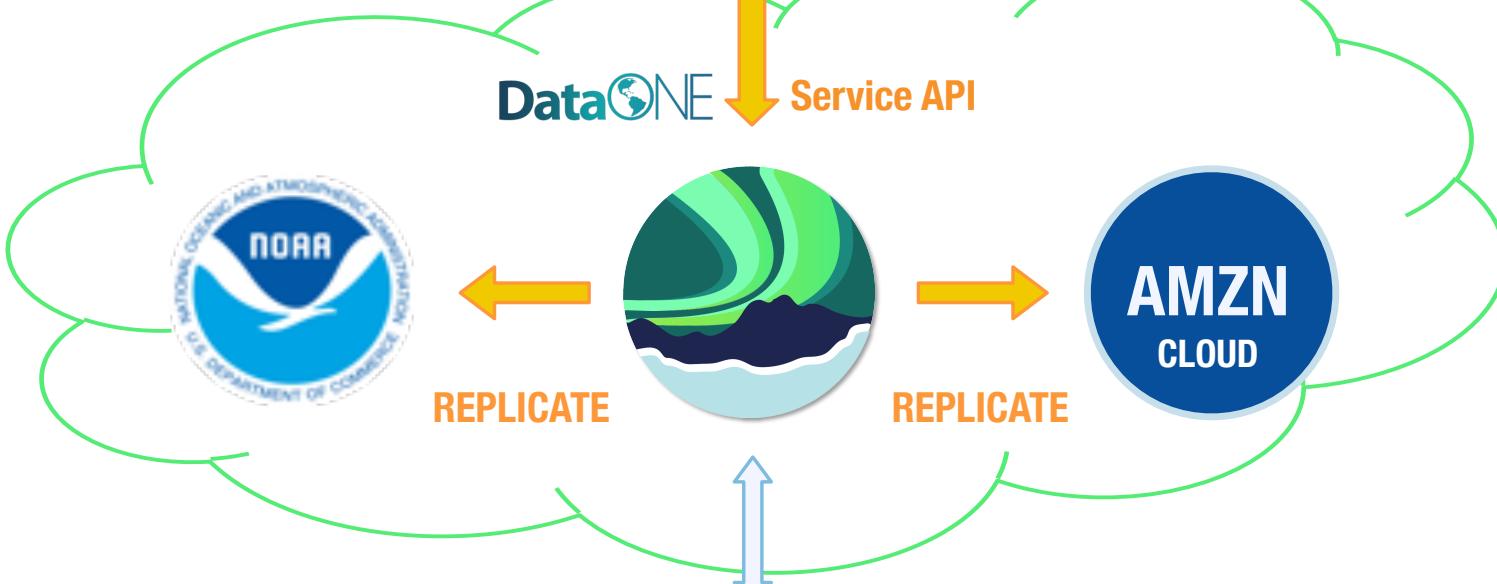
METACATUI



SEARCH / SUBMIT / DOWNLOAD

DATA / METADATA / CODE

DataONE Service API



REPLICATE

REPLICATE



Complementary Services



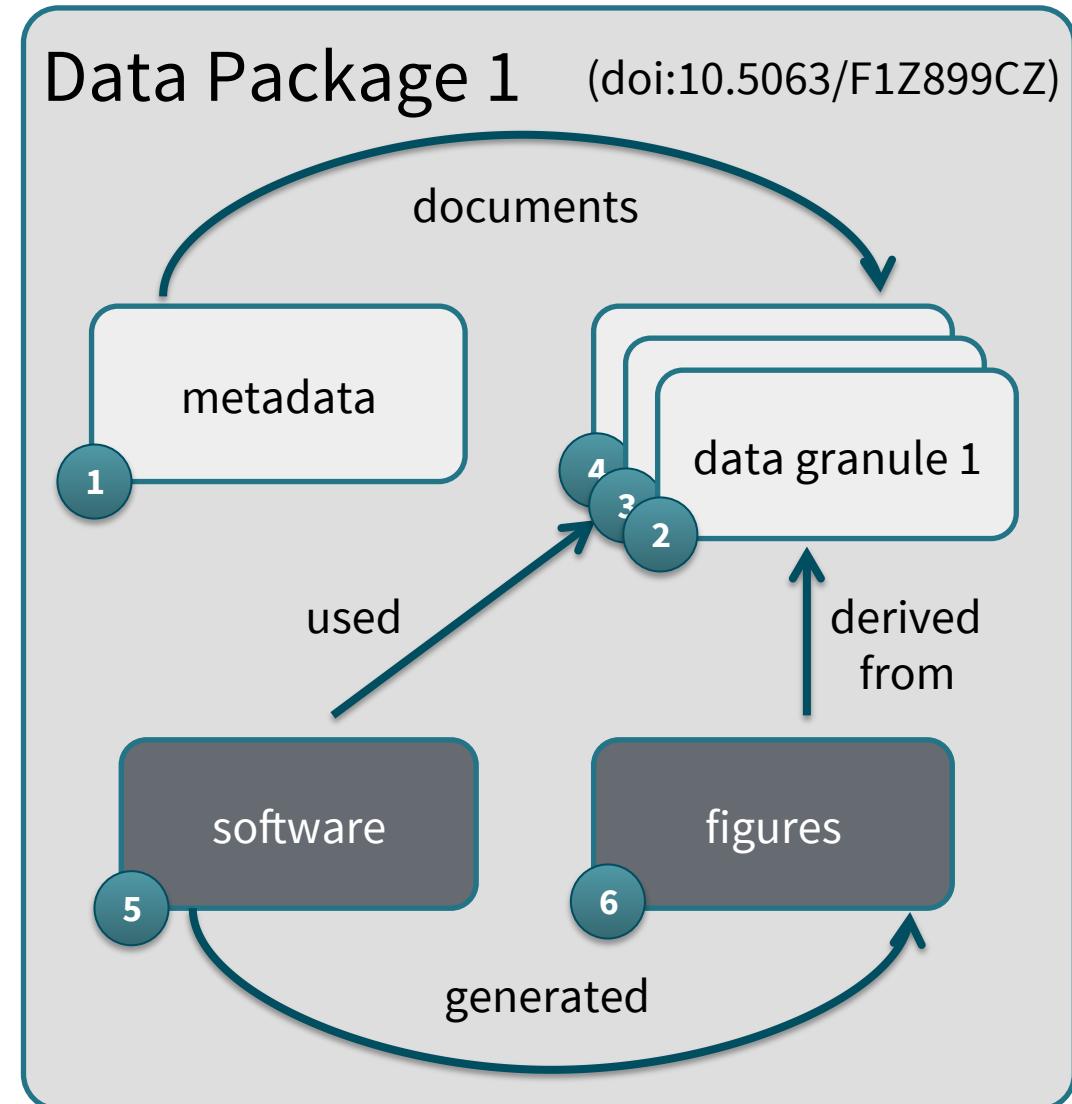
# Key Features

- Archival formats
    - Data, Code, Figures/Images
  - Identification
    - DOIs assigned on publication
  - Versioning
  - Replication and audit
- 
- Usage tracking
  - Provenance
  - Quality engine
  - R and Matlab packages



# Archival Packages: Provenance inside

- Globally unique identifiers
- Strong versioning
- OAI-ORE manifest
- BagIt serialization
- Hierarchical
- Provenance





# Data Discovery Portal

NSF  
**Arctic  
Data  
Center**

Data Support About Submit Data Sign in with Orcid

Search ?  
Search phrase

Filter by:  
Data attribute  
Creator  
Year  
Identifier  
Taxon  
Location

DATASETS 1 TO 25 OF 4,149  
1 2 3 ... 166 Next Sort by Most recent

Paul Bierman. 2016. **Greenland 10Be cosmogenic date in marginal sediment.** NSF Arctic Data Center. urn:uuid:1c821d0c-926d-4efc-b88f-7b99b4875425.  
 33

NSF Arctic Data Center. 2015. **1Learning about environmental research in the context of climate change: an international scholastic interchange.** NSF Arctic Data Center. doi:10.18739/A27H41.  
 7

NSF Arctic Data Center. 2015. **1Learning about environmental research in the context of climate change: an international scholastic interchange.** NSF Arctic Data Center. doi:10.18739/A2H08V.  
 6

Barbara Bodenhorn. 2015. **Learning about environmental research in the context of climate change: an international scholastic interchange.** NSF Arctic Data Center. urn:uuid:8eb64350-1e39-4b82-a2c7-2f62d84e8b5d.  
 5

Joseph McConnell. 2016. **Tunu, Greenland 2013 ice core chemistry.** NSF Arctic Data Center. urn:uuid:39ec61f1-0627-4ebc-ae3c-8b6261dce661.  
 5

Yasushi Fukamachi, Kay I. Ohshima, Daisuke Simizu, T. Takatsuka, K. Iwamoto, A.

Hide Map »

Map data ©2016 1000 km Google Terms of Use



# Precise Drill Down

NSF  
ARCTIC  
DATA  
CENTER

Data Support About Submit Data Sign in with Orcid

[Clear all filters](#)

Search ?  
Search phrase

Filter by:  
 Data attribute  
 Creator  
 Year  
 Identifier  
 Taxon  
 Location

DATASETS 1 TO 14 OF 14

Sort by

Jacqueline M. Grebmeier. 2016. [Regional Alaska Community Meetings](#). NSF Arctic Data Center. doi:10.5065/D6QR4V4P.  
 8

Arny L. Blanchard and Howard Feder. 2016. [Macrobenthic data from the Chukchi Sea \[Blanchard\]](#). NSF Arctic Data Center. doi:10.5065/D6X63K0J.  
 7

Patrick Sullivan. 2016. [Agashashok White Spruce Branch Growth 2005-2008](#). NSF Arctic Data Center. doi:10.5065/D6Q81B6D.  
 11

Patrick Sullivan. 2016. [Agashashok White Spruce Needle Gas Exchange 2007-2008](#). NSF Arctic Data Center. doi:10.5065/D6H1304V.  
 6

Patrick Sullivan. 2016. [Agashashok Microclimate](#). NSF Arctic Data Center. doi:10.5065/D6BP00X1.  
 8

Patrick Sullivan. 2013. [White Spruce Needle Gas Exchange, Nitrogen Concentration, 13C and 15N](#). NSF Arctic Data Center. doi:10.18739/A2TK8H.

[Hide Map](#)  Limit my search to the map area



# Metadata Record



NSF  
**ARCTIC  
Data  
Center**

Data Support About [Submit Data](#) [!\[\]\(8a482f7d703b2f3a21bcbafdaced4511\_img.jpg\) Sign in with Orcid](#)

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

Arny L. Blanchard and Howard Feder. 2016. Macrobenthic data from the Chukchi Sea [Blanchard]. NSF Arctic Data Center. doi:10.5065/D6X63K0J.

[!\[\]\(b8cfa9474de969a8f1ab00018de932d4\_img.jpg\) Copy Citation](#) [!\[\]\(c7f1392ae34784539f14d48ef82b6bc5\_img.jpg\) Quality report](#)

Files in this dataset Package: resource_map_doi:10.5065/D6X63K0J				
File type	Name	Size	Downloads	Download all
EML v2.1.1	Metadata: science_metadata.xml	7 KB	7 views	<a href="#">Download</a>
PDF	Feder_Chukchi_Sea_readme.pdf	34 KB	4 downloads	<a href="#">Download</a>
PDF	Feder_et_al-1991_NE_Chukchi_Sea.pdf	5 MB	4 downloads	<a href="#">Download</a>
text/csv	Feder_Chukchi_Sea_data.csv	1 MB	4 downloads	<a href="#">Download</a>
Microsoft Excel OpenXML	Feder_Chukchi_Sea_data.xlsx	1002 KB	4 downloads	<a href="#">Download</a>

[▲ Show less](#)



# Data Submission

**Sign in to submit data**



Sign in with ORCID



# Data Submission

[Data](#)[Support](#)[About](#)[Submit Data](#) [Amber Budden](#)

## Upload your data

Use this form to submit a new data package to the repository.

 [Show Help Guide](#)

\* Denotes a required field.

### Basic Information

\*Data Set Title

\*Award Number

[+ Add](#)

Enter an award number or search for an NSF award by keyword.

### People and Organizations

Role

First Name

Last Name



# Data Submission

[Data](#)[Support](#)[About](#)[Submit Data](#) [Amber Budden](#)

## Temporal Coverage

Start Date

\*Year (yyyy)

End Date (leave blank if your data set is open-ended)

Year (yyyy)

Month

MM

Month

MM

Day

DD

Day

DD

## Spatial Coverage

\*Geographic Description

General description of the geographic area in which the data were collected. It can be a simple place name (e.g., Santa Barbara) or a fuller description.

## Coordinates

Degrees

Minutes

Seconds

\*Latitude

- North
- South



# Other Submission Methods

## DataONE R and Matlab Tools



Scientists can submit data files and code programmatically through R or Matlab.

But...EML Metadata stills needs to be created separately or by using the limited EML R library.



# Metadata Quality Reports: MetaDIG

NSF ARCTIC DATA CENTER

Data Support About Submit Data Sign in with Orcid

< Back to search | Home / Search / Metadata

Arny L. Blanchard and Howard Feder. 2016. Map of the Chukchi Sea. [v1]. NSF Arctic Data Center. doi:10.5065/D6X63K0J.

 Quality report

Copy Citation Quality report

Files in this dataset Package: resource\_map\_doi:10.5065/D6X63K0J

Name	File type	Size	Downloads	Download all
Metadata: science_metadata.xml	EML v2.1.1	7 KB	7 views	Download
Feder_Chukchi_Sea_readme.pdf	PDF	34 KB	4 downloads	Download
Feder_et_al-1991_NE_Chukchi_Sea.pdf	PDF	5 MB	4 downloads	Download
Feder_Chukchi_Sea_data.csv	text/csv	1 MB	4 downloads	Download
Feder_Chukchi_Sea_data.xlsx	Microsoft Excel OpenXML	1002 KB	4 downloads	Download

▲ Show less

# Metadata Quality Report

Mackenzie Grieman. 2016. Ice core vanillic acid and para-hydroxybenzoic acid concentrations, Akademii Nauk, Severnaya Zemlya archipelago, Russia, 2013-2015. Arctic Data Center. doi:10.18739/A23Q23.

After running your metadata against our standard set of metadata, data, and congruency checks, we have found the following potential issues. Please assist us in improving the discoverability and reusability of your research data by addressing the issues below.



**Identification:** 100% complete

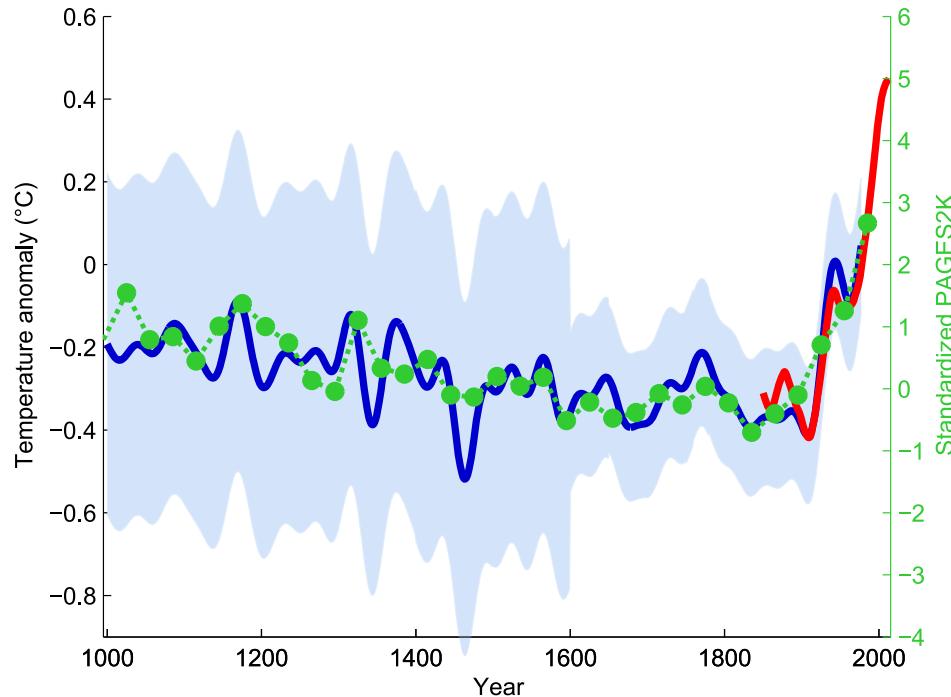
**Discovery:** 100% complete

**Interpretation:** 100% complete

- ▶ Passed 20 checks out of 22. Good job!
- ▶ Warning for 2 checks. Please review these warnings.
- ▼ Failed 0 checks.
- ▶ 4 informational checks. These may include skips, errors and failures.



# Provenance



What **input data** went into this study?

What **methods** were used?

... with what  
**parameter** settings,  
**calibrations**, ...?

Can we **trust** the data and methods?

- **Provenance** (*lineage*): track **origin** and **processing history** of data → trust, data quality ~ audit trail for attribution, credit

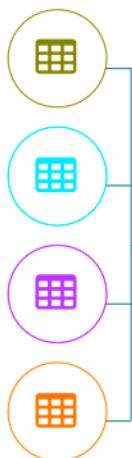


# Provenance Web UI

D

Data Table, Image, and Other Provenance

4 sources



## Source Program

Total\_PAH\_and\_Alkanes\_GoA\_Hydrocarbons\_Clean.R

### Citation

[View »](#)

This program generated the data you are currently viewing, [Total\\_Aromatic\\_Alkanes\\_PWS.csv](#).

This program used [PAH.csv](#), [Sample.csv](#), [Non-EVOS\\_SINs.csv](#) and [\(and 1 more\)](#).

This program used [Alkane.csv](#).

## Derivation

Total\_Aromatic\_Alkanes\_PWS.csv

from PAH, Alkane and Sample tables documenting samples from Exxon Valdez oil spill in Prince William Sound, AK

Alkane.csv

<http://org/cn/v2/resolve/urn:uuid:44108e76-405d-4d58-b1b3->

## Text Format

Number of Header Lines

1

Record Delimiter

#xA

Attribute Orientation

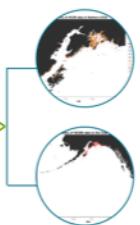
column

Simple Text

Field Delimiter

,

2 derivations

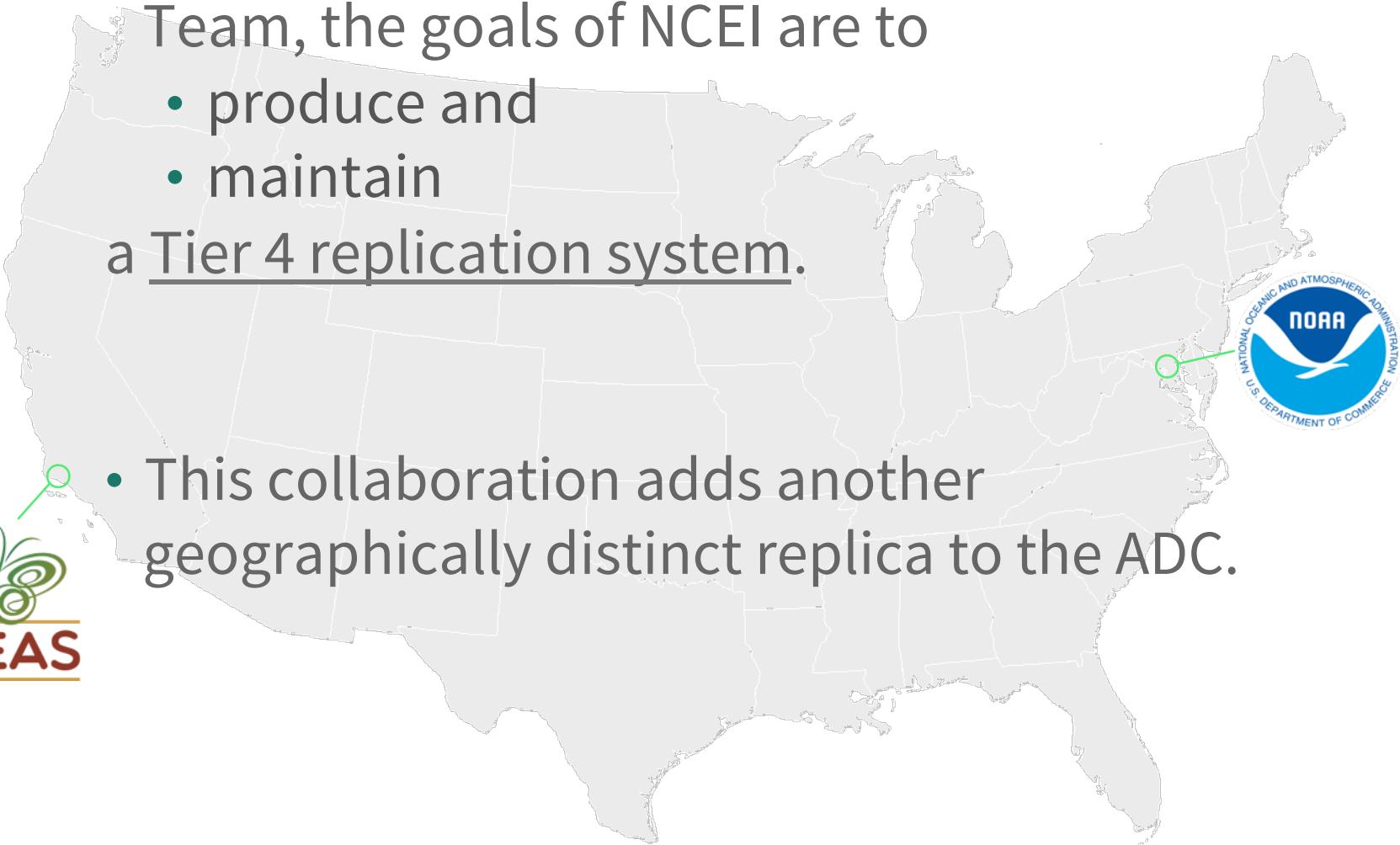




# Replication Goals

- In collaboration with the Arctic Data Center Team, the goals of NCEI are to
    - produce and
    - maintain
- a Tier 4 replication system.

- This collaboration adds another geographically distinct replica to the ADC.





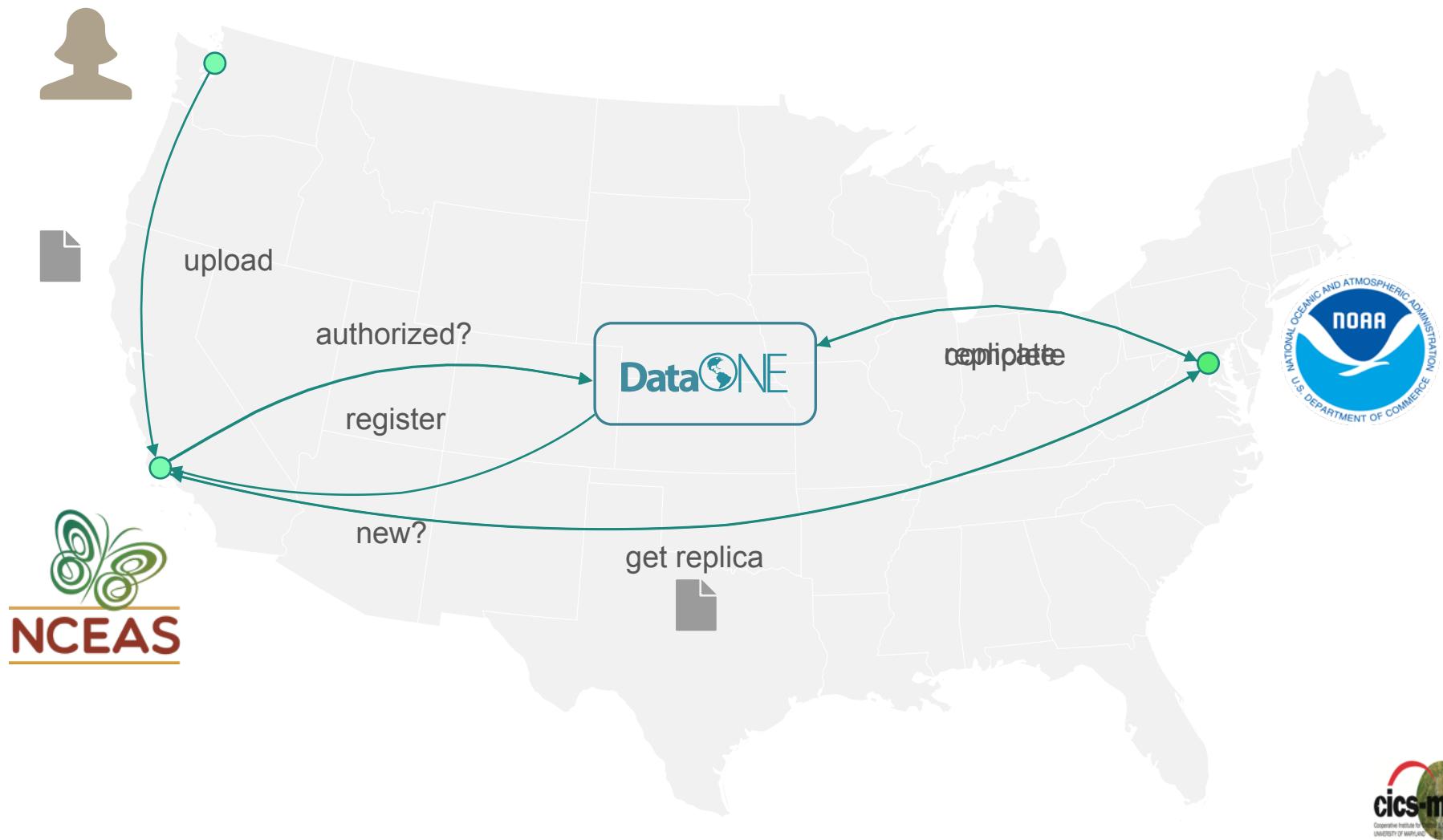
# Replication Year-2 Milestones

- In Year 2, NCEI will
  - complete the prototype replication system
  - move the prototype into production





# Replication Steps

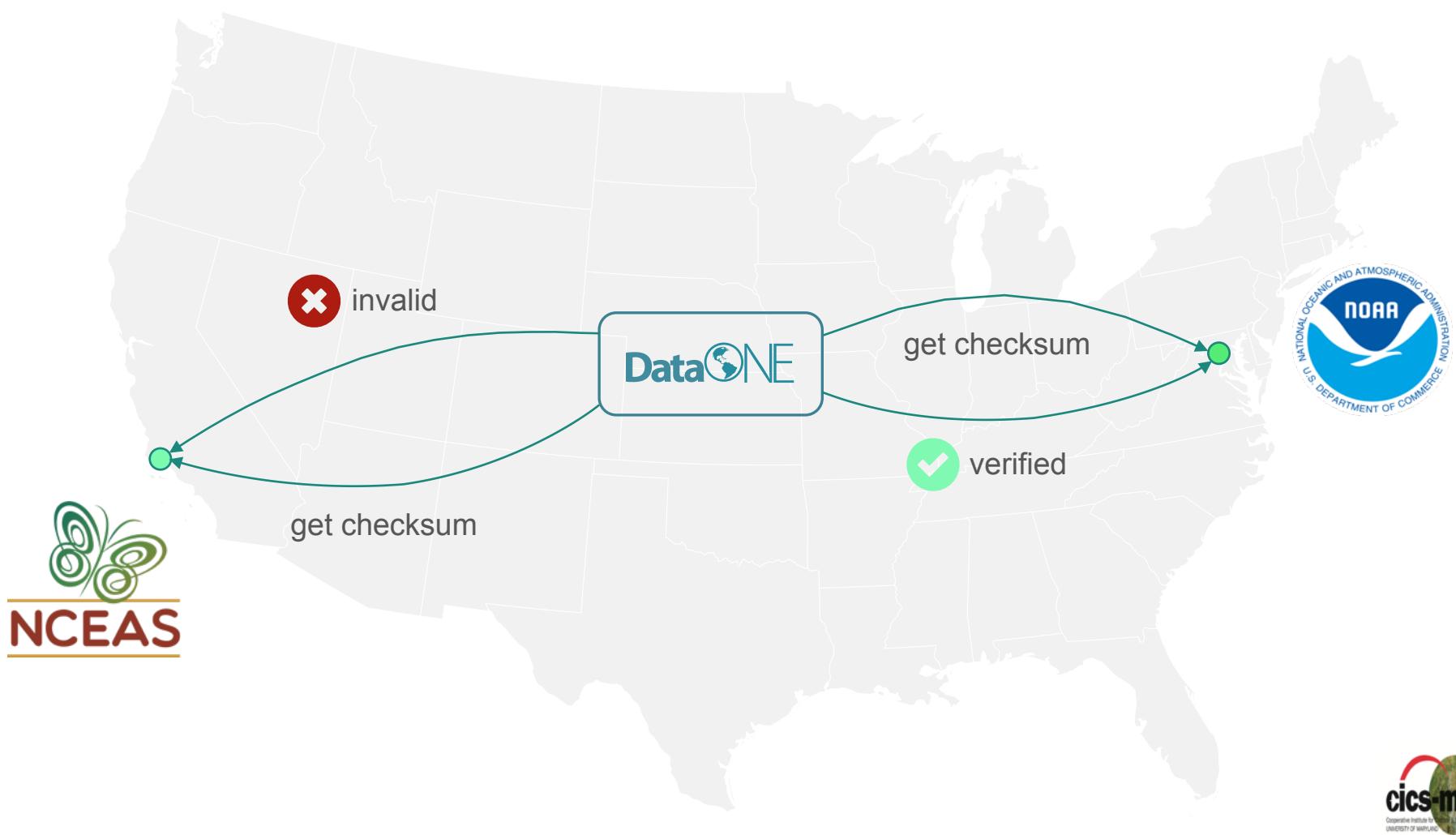


NCEAS



# Replica Audit Steps

- All files are audited every 3 months





# Replication with NCEI Member Node

- Automated replication via DataONE

- Curation step into NCEI catalog



DataONE



package	verified
arcticdata.1.1	✓
arcticdata.2.1	✓
arcticdata.3.1	—



# The New Web Submission System

NSF Arctic Data Center    x    NSF Arctic Data Center    x    LPD

Secure <https://test.arcticdata.io/#share/urn:uuid:c2d6e8d2-8b68-47b0-8211-17e16429c594>

NSF Arctic Data Center    Data    Support    About    Submit Data    Lauren Walker

Alaska Department of Natural Resources - Information Resource Management. 2017. Well Log Tracking System (WELTS). NSF Arctic Data Center Test Repository. urn:uuid:c2d6e8d2-8b68-47b0-8211-17e16429c594.

Files	Size	Type	
Well Log Tracking System (WELTS)	68 KB	Metadata	+ Add Files
WELTS.zip	940 KB	Data	edit X
WELTS_flatfile.csv	10.65 MB	Data	edit X

**Overview**

**People**

**Dates**

**Locations**

**Abstract**

The Well Log Tracking System (WELTS) contains lithologic information submitted to the Division of Mining, Land and Water, Alaska Hydrologic Survey by water well contractors as required per Alaska State Statute 41.08.020(b4) authority delegated to the Alaska Hydrologic Survey per Department Order 115, require of water well contractors, the filing with it of basic water and aquifer data normally obtained, including but not limited to well location, estimated elevation, well driller's logs, pumping tests and flow measurements, and water quality

Submit



# The New Web Submission System

The screenshot shows a web browser window for the NSF Arctic Data Center. The URL is <https://test.arcticdata.io/#share/urn:uuid:c2d6e8d2-8b68-47b0-8211-17e16429c594>. The page displays a file listing for the Well Log Tracking System (WELTS) test repository. The files listed are:

Files	Size	Type	Actions
Well Log Tracking System (WELTS)	68 KB	Metadata	
WELTS.zip	940 KB	Data	
WELTS_flatfile.csv	10.65 MB	Data	

A large teal arrow points from the "Add Files" button in the top right corner of the file list to a callout box at the bottom right of the screen. The callout box contains the text "Upload many files at once".

**Overview**

**People**

**Dates**

**Locations**

**Abstract**

The Well Log Tracking System (WELTS) contains lithologic information submitted to the Division of Mining, Land and Water, Alaska Hydrologic Survey by water well contractors as required per Alaska State Statute 41.08.020(b4) authority delegated to the Alaska Hydrologic Survey per Department Order 115, require of water well contractors the filing with it of basic water and aquifer data normally obtained, including but not limited to



# The New Web Submission System

client  
deposit  
Files  
Well  
WELT  
WELT

**Describe WELTS.zip**

Overview Attributes

**Log\_ID**

**Attribute Name \***  
The attribute name is usually the name of the variable that is found in the header of a data file.

Log\_ID

**Attribute Label \***  
A descriptive label that is used to display the name of an attribute.

A short, descriptive label

**Definition**  
Explain the contents of the attribute fully so that a data user could interpret the attribute accurately.

Well log record identifier

PDESC

Keywords

Done

Describe each data file  
AND each measurement



# The New Web Submission System

More than one location? No problem!

The screenshot shows a web browser window for the NSF Arctic Data Center. The left sidebar has links for Overview, People, Dates, Locations (which is selected), Taxa, and Methods. The main content area is titled 'Locations' and contains three entries:

- Description:** Alaska  
**Bounding Box Coordinates:** Northwest coordinates: 73.875, Southeast coordinates: 47.375, -125.5
- Description:** Cascade Lake, Brooks Range, Alaska.  
**Bounding Box Coordinates:** Northwest coordinates: 68.383°, -154.57, Southeast coordinates: Lat, Long
- Description:** (empty)  
**Bounding Box Coordinates:** Northwest coordinates: Lat, Long, Southeast coordinates: Lat, Long

A teal callout box with the text "More than one location? No problem!" points to the first location entry. A large teal arrow points from the text to the first entry's description field.

Secure <https://test.arcticdata.io/#share/urn:uuid:c2d6e8d2-8b68-47b0-8211-17e16429c594>

Submit



# Comparison

	Current	New	In future versions
We are able to quickly make improvements based on user feedback	✓ (sometimes)	✓	✓
Upload multiple files at once		✓	✓
Max. number of files that can be submitted	10	~100	1000+
Describe the characteristics of data files (file type, description of contents)		✓	✓
Describe the measurements inside each data file		✓	✓
Number of geographic regions that can be described	1	unlimited	unlimited
Provide ORCIDs for each person		✓	✓
Add nested folders			✓
Save a draft of your dataset and edit later			✓
Even more file and attribute-level metadata (geographic regions per data file, temporal coverage per data file, measurement precision, etc.)			✓



# Agenda

<b>Agenda</b>	<b>Day 1.</b>	<b>Day 2.</b>
8:30-9:00	Welcome and introductions	Writing Good Data Management Plans
9:00-9:45	Arctic Data Center and NSF Standards and Policies	Writing Good Data Management Plans
9:45-10:00	Break	Break
10:00-12:00	Effective data modeling and management	Data packaging and file hierarchies
Noon-1:15	Lunch	Lunch
1:15-2:15	Authoring Quality metadata	Authoring large data sets
2:15-2:30	Break	Break
2:30-4:30	Authoring Quality metadata	Large data and Tracking data provenance
4:30-5:00	Question and Answer	Discussion