

NSF
**ARCTIC
Data
Center**

Overview and NSF Standards and Policies

Matthew B. Jones

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



Operation Metrics



4,700
DATA SETS



631K
DATA FILES



56K
FILE DOWNLOADS



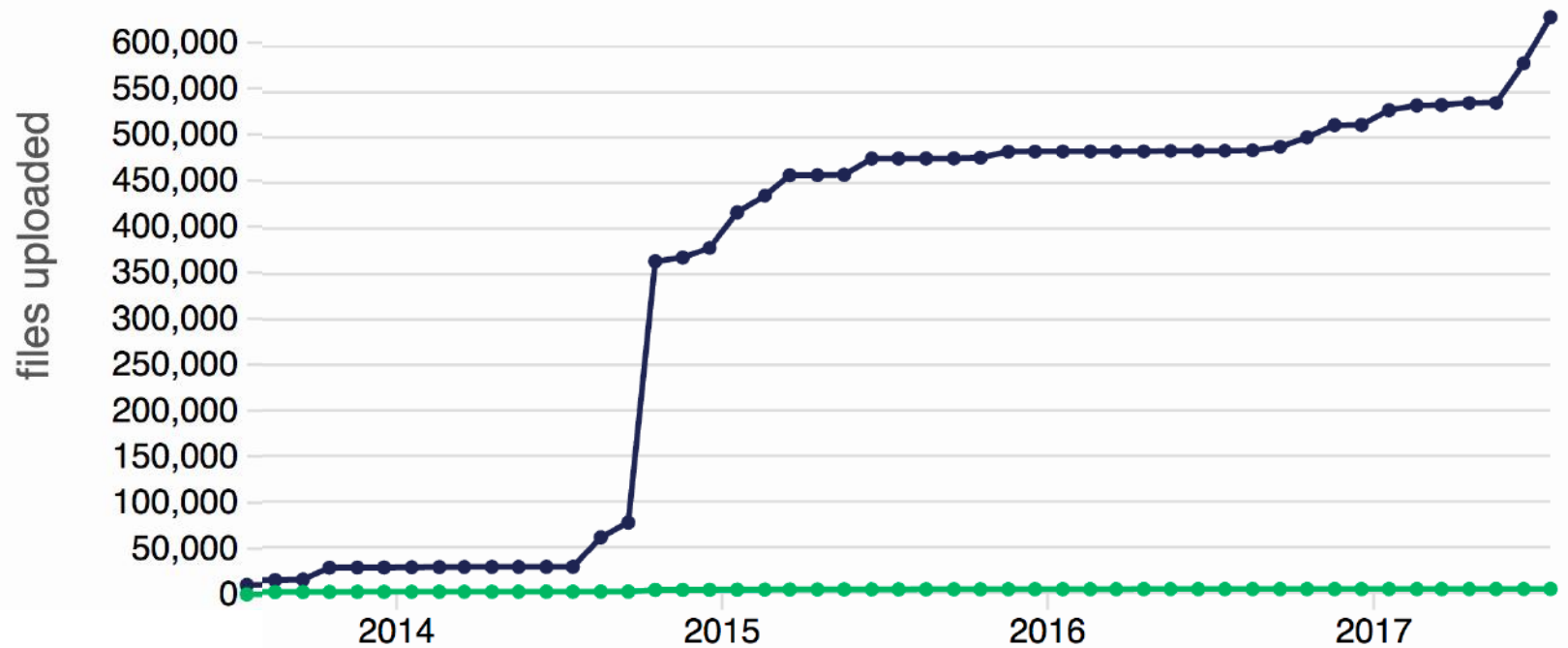
1,900
CREATORS



3,800
USERS

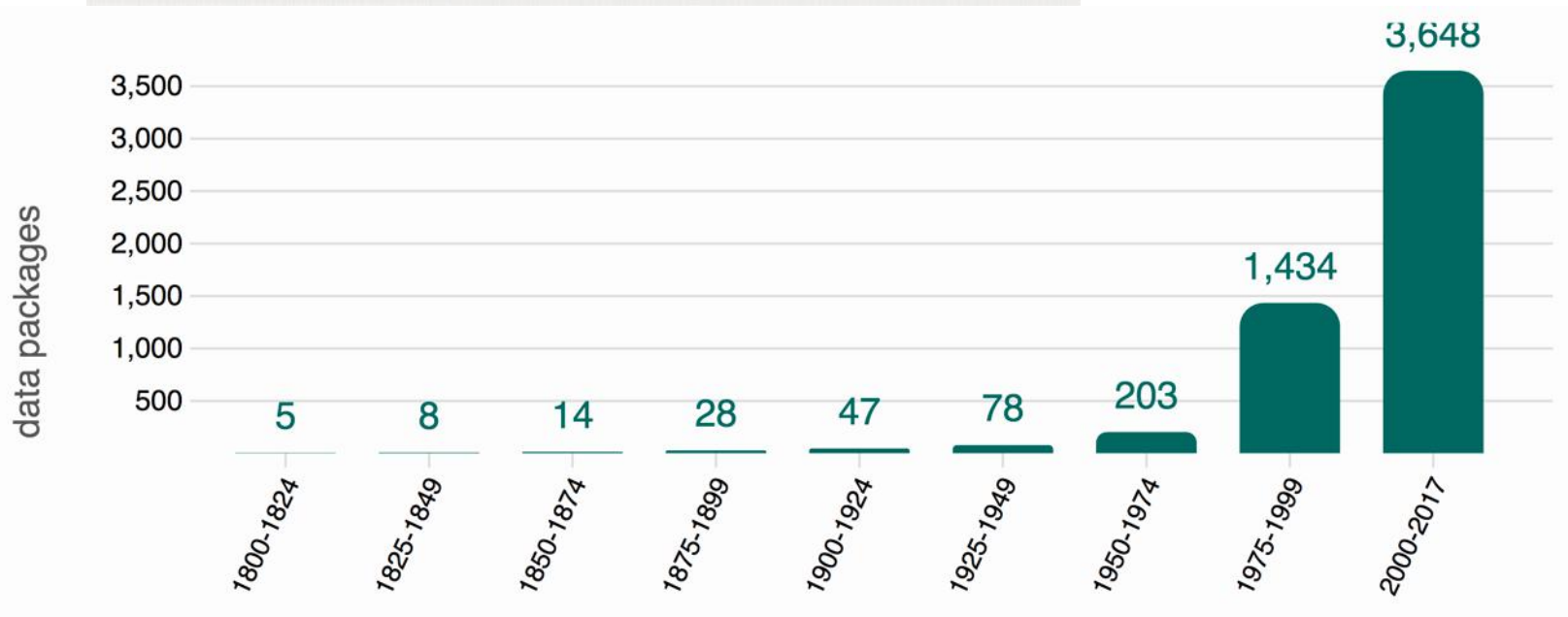
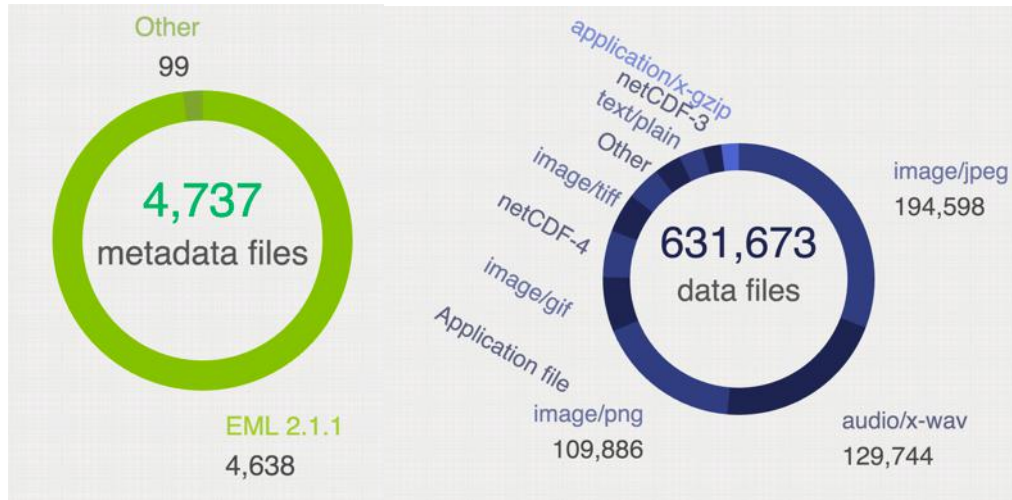


An Arctic Research Data Archive





Data Holdings





Data Discovery Portal



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 Bodenborn. 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.

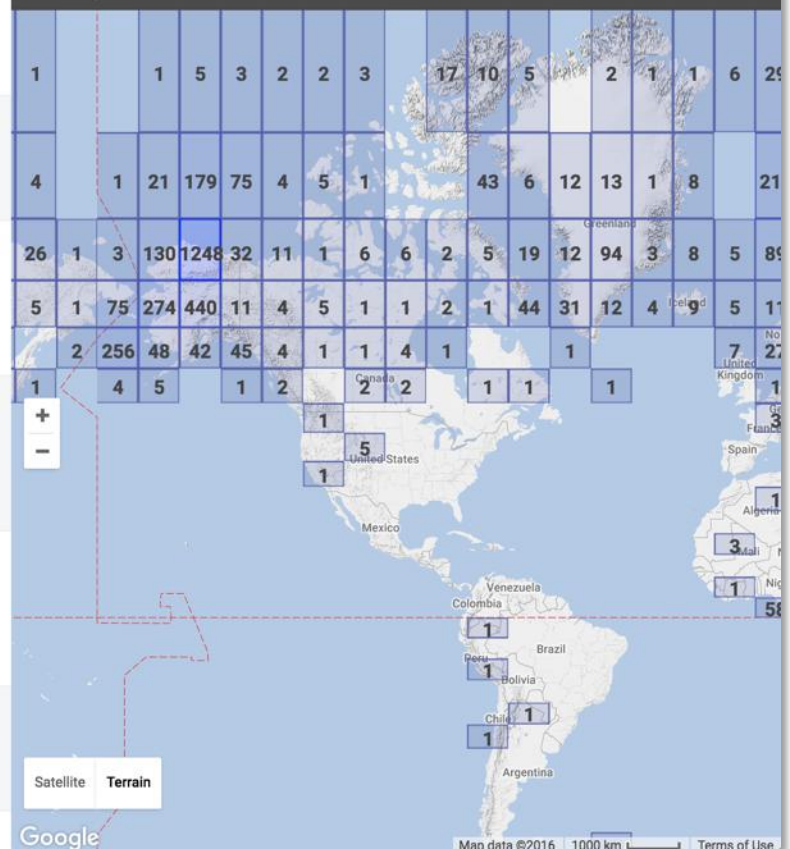
5

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

5

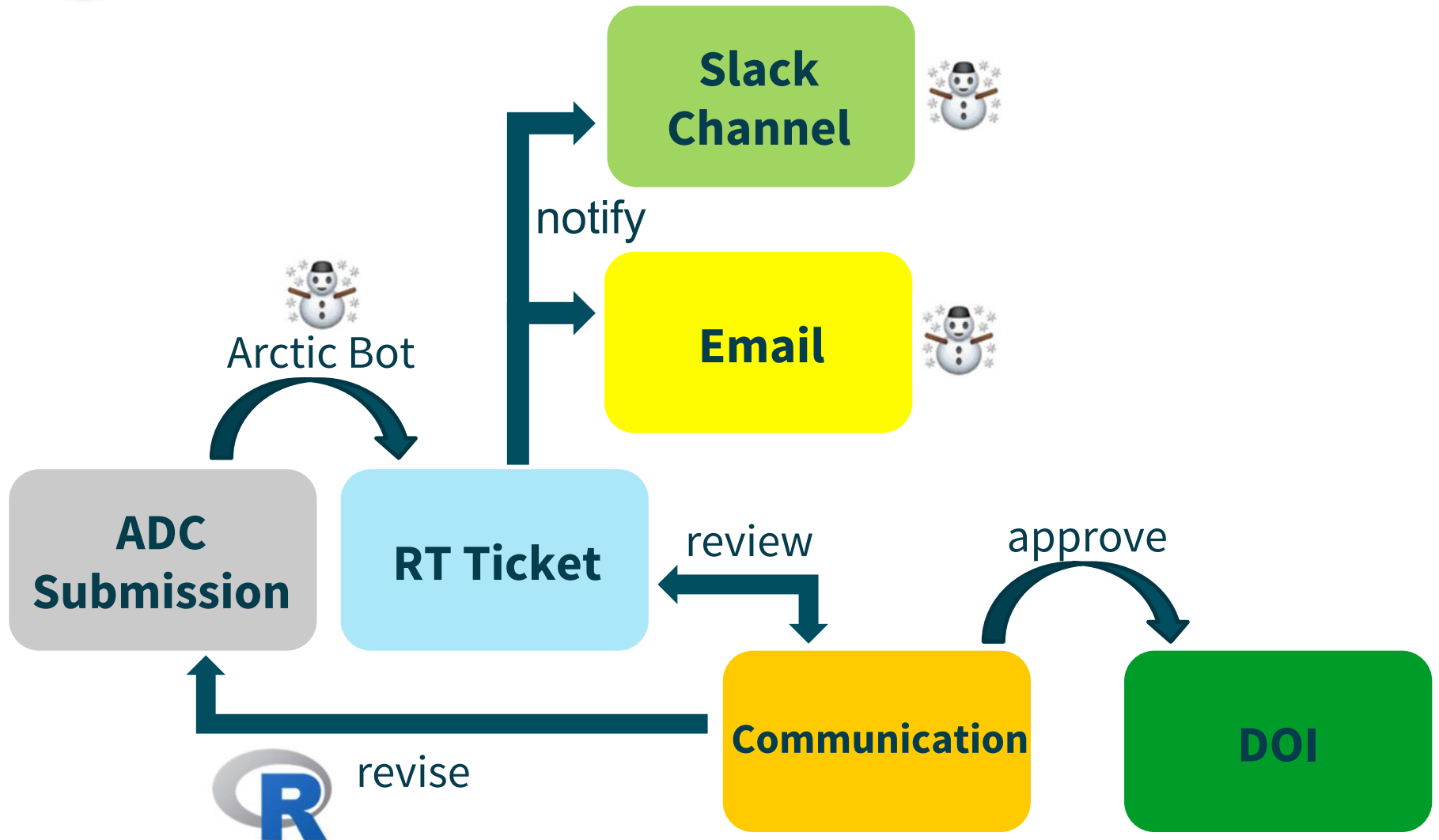
Yasushi Fukamachi, Kay I. Ohshima, Daiske Simizu, T. Takatsuka, K. Iwamoto, A.

Hide Map »





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-a584-9d1aafacca63.

- 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		Downloads	
Name	File type	3 views	Download All 
 Metadata: iso19139.xml	http://www.isotc211.org/2005/gmd	852 downloads	Download 
 dispatches_imnavait_apr2012.pdf	PDF	274 downloads	Download 
 depth_happyvalleylines_apr2012.xlsx	Microsoft Excel OpenXML	209 downloads	Download 
 depth_imnav_apr2012_1by1grid.xlsx	Microsoft Excel OpenXML		Download 
► Show 4 more items in this data set			



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



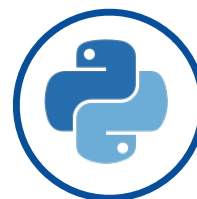
MATLAB



R



WEB PORTAL



PYTHON



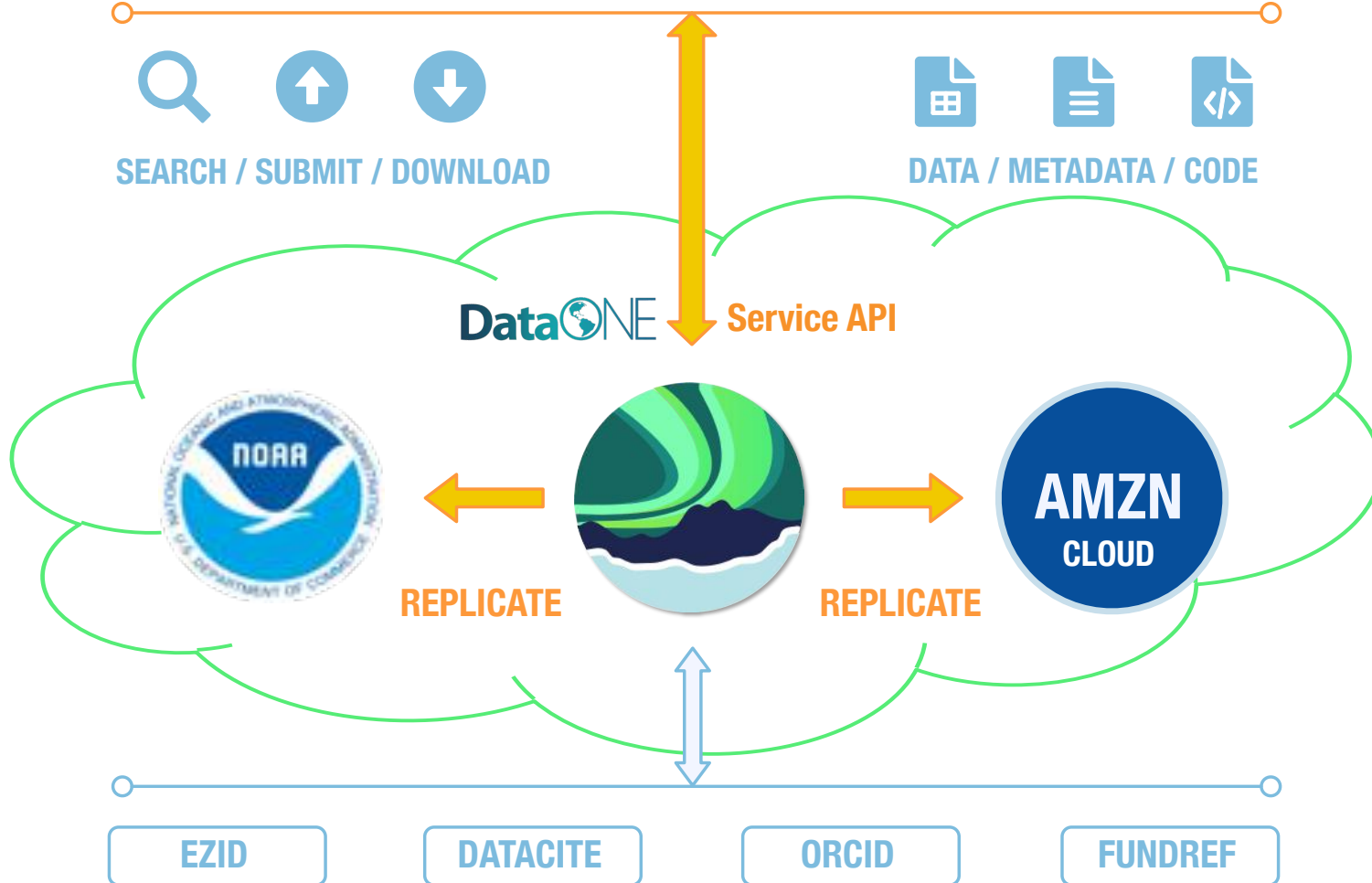
METACATUI



SEARCH / SUBMIT / DOWNLOAD



DATA / METADATA / CODE



Complementary Services



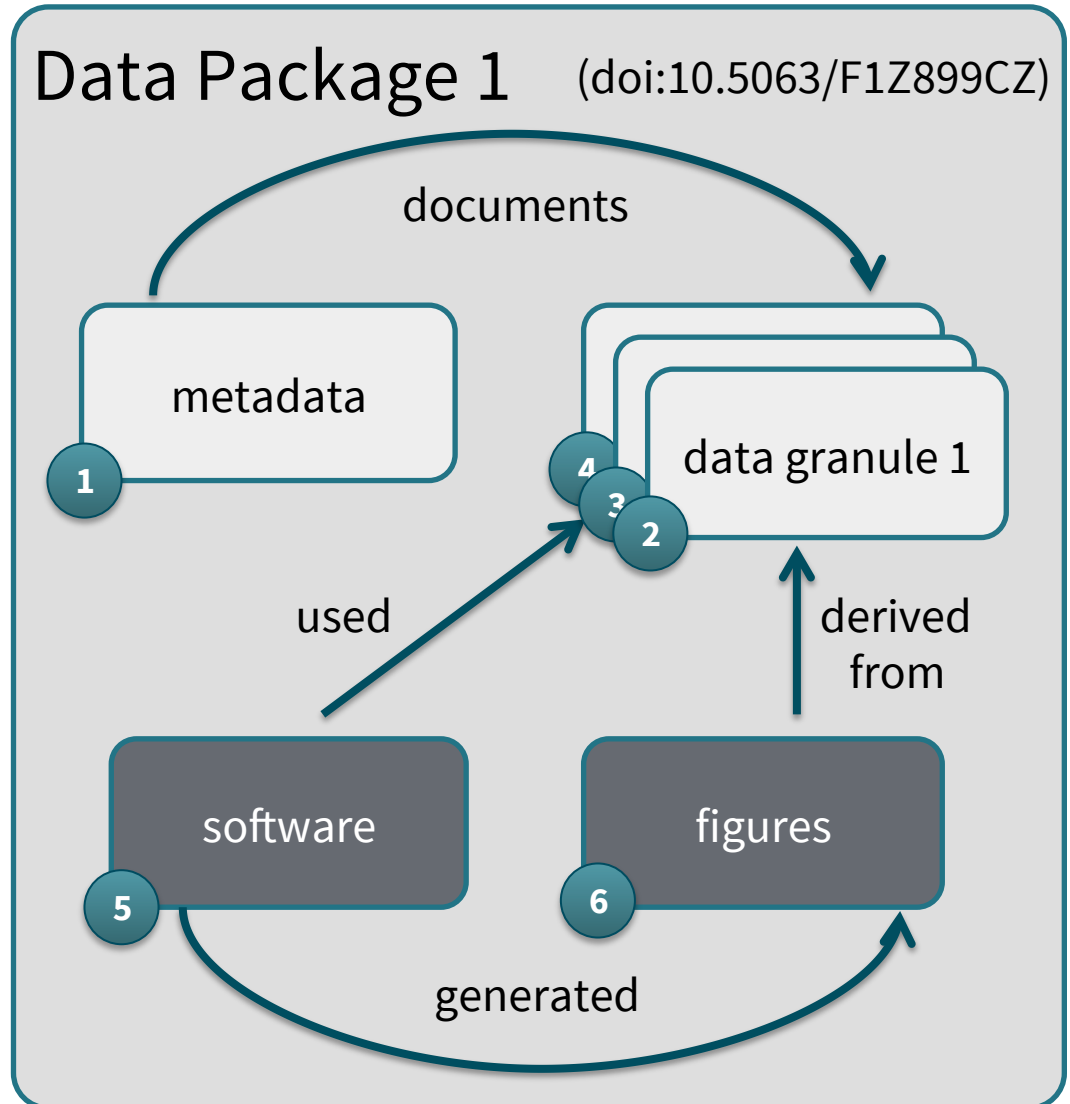
Key Features

- 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





 Sign in with Orcid

Map data ©2016 Google 5 km Terms of Use Report a map error



Metadata Record



[Data](#) [Support](#) [About](#)

[Submit Data](#)

[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.

[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](#)



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

Creator (Author/Owner/Originat ⌵)

First Name

Last Name



Data Submission

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

Temporal Coverage

Start Date

*Year (yyyy)

Month

MM



Day

DD



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

Year (yyyy)

Month

MM



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. Maori. NSF Arctic Data Center. doi:10.5065/D6X63K0J.

[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!

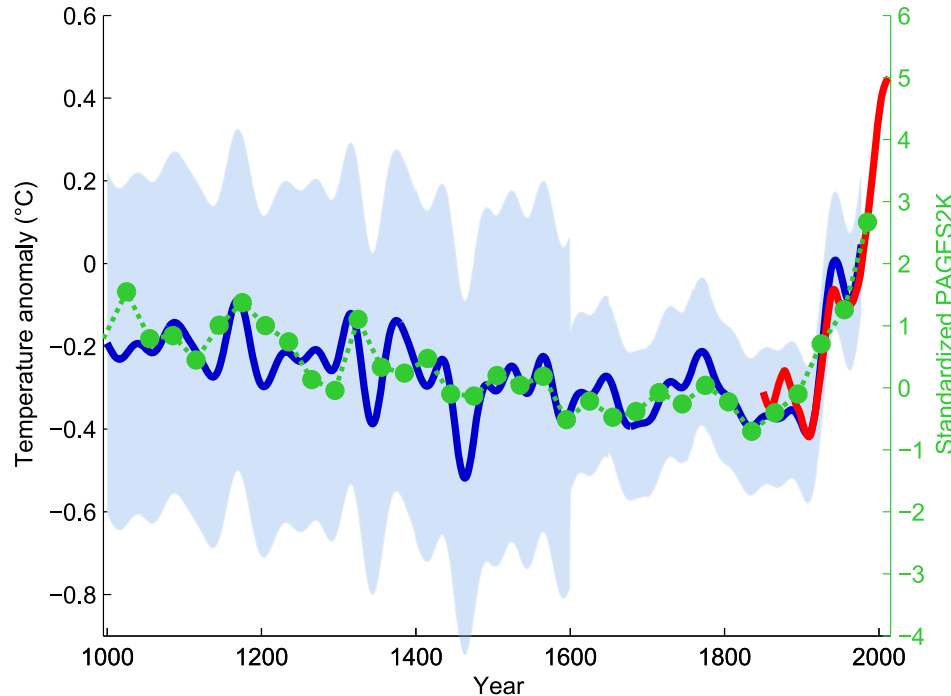
▶ Warning for 2 checks. Please review these warnings.

▼ 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?

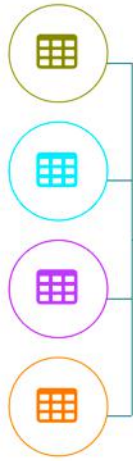
- **Provenance** (*lineage*): track **origin** and **processing history** of data → trust, data quality ~ audit trail for attribution, credit



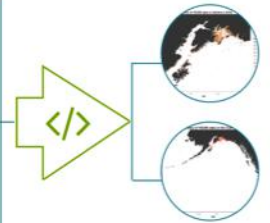
Provenance Web UI

Data Table, Image, and Other Data Types

4 sources



2 derivations



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**.

Alkanes_PWS.csv

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

Alkanes_PWS.csv

[org/cn/v2/resolve/urn:uuid:44108e76-405d-4d58-b1b3-](#)

Text Format

Number of Header Lines	1
Record Delimiter	#x0A
Attribute Orientation	column
Simple Text	
Field Delimiter	,



Replication Goals

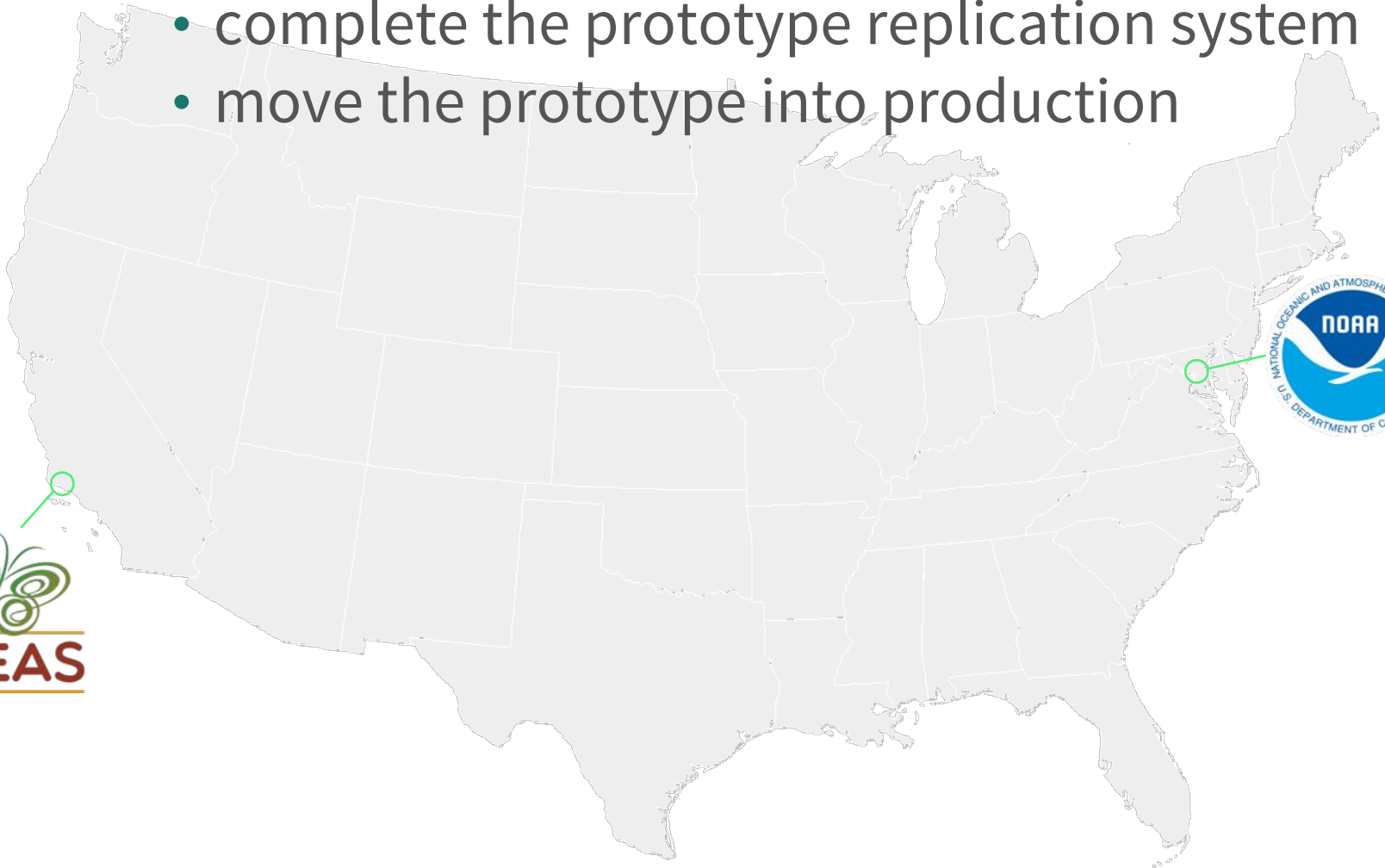
- In collaboration with the Arctic Data Center Team, the goals of NCEI are to
 - produce and
 - maintaina Tier 4 replication system.
- This collaboration adds another geographically distinct replica to the ADC.





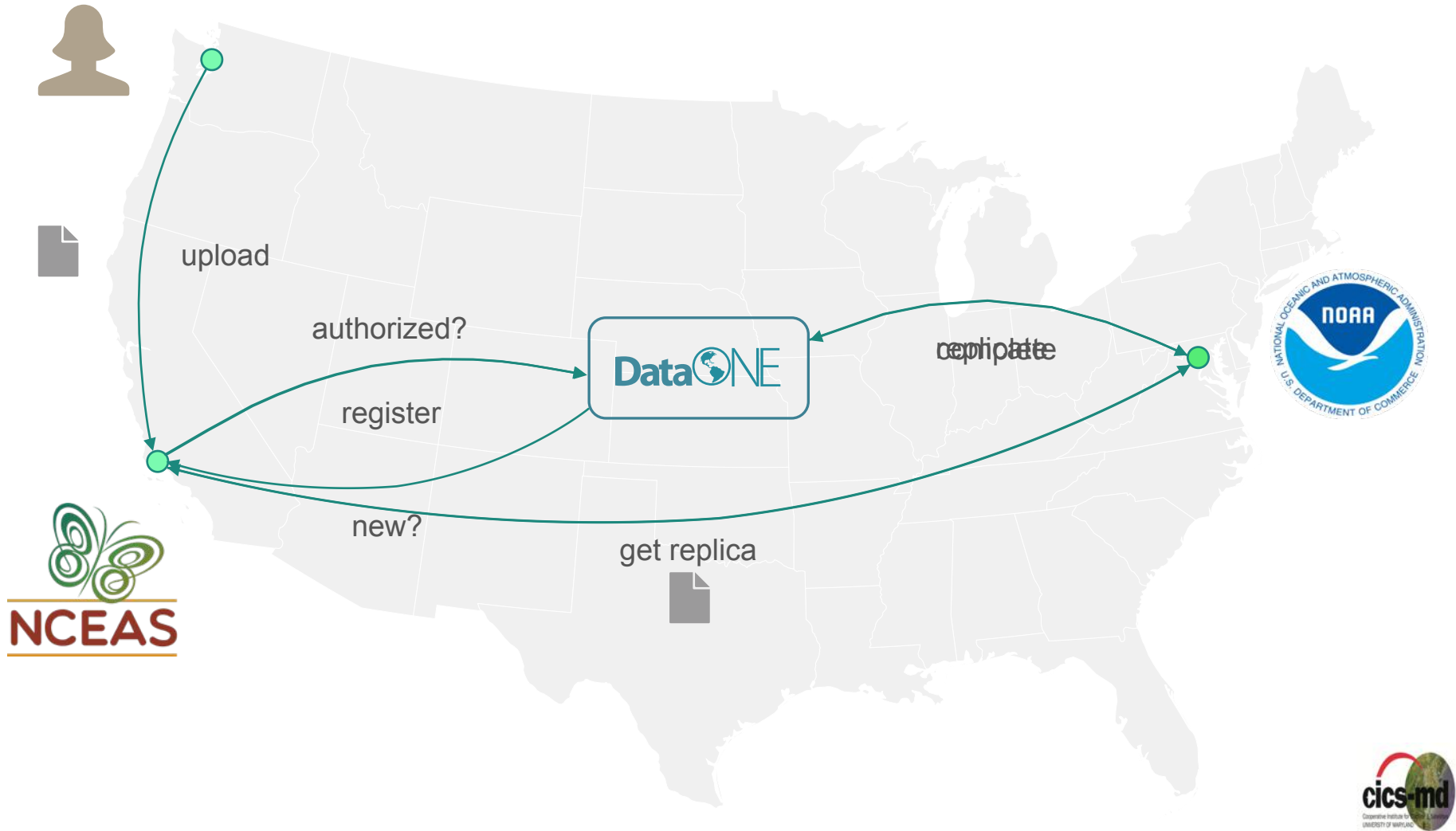
Replication Year-2 Milestones

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





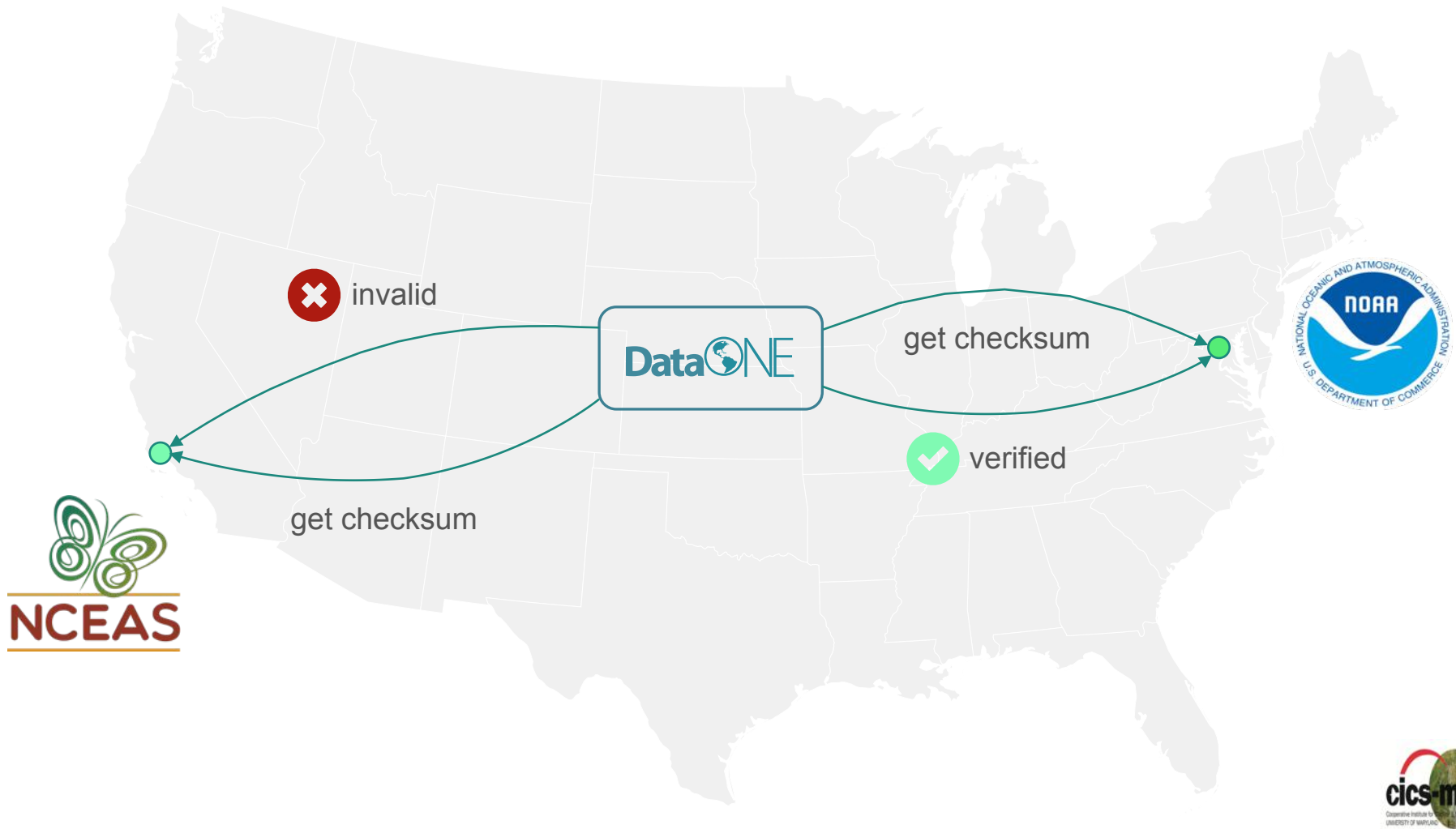
Replication Steps





Replica Audit Steps

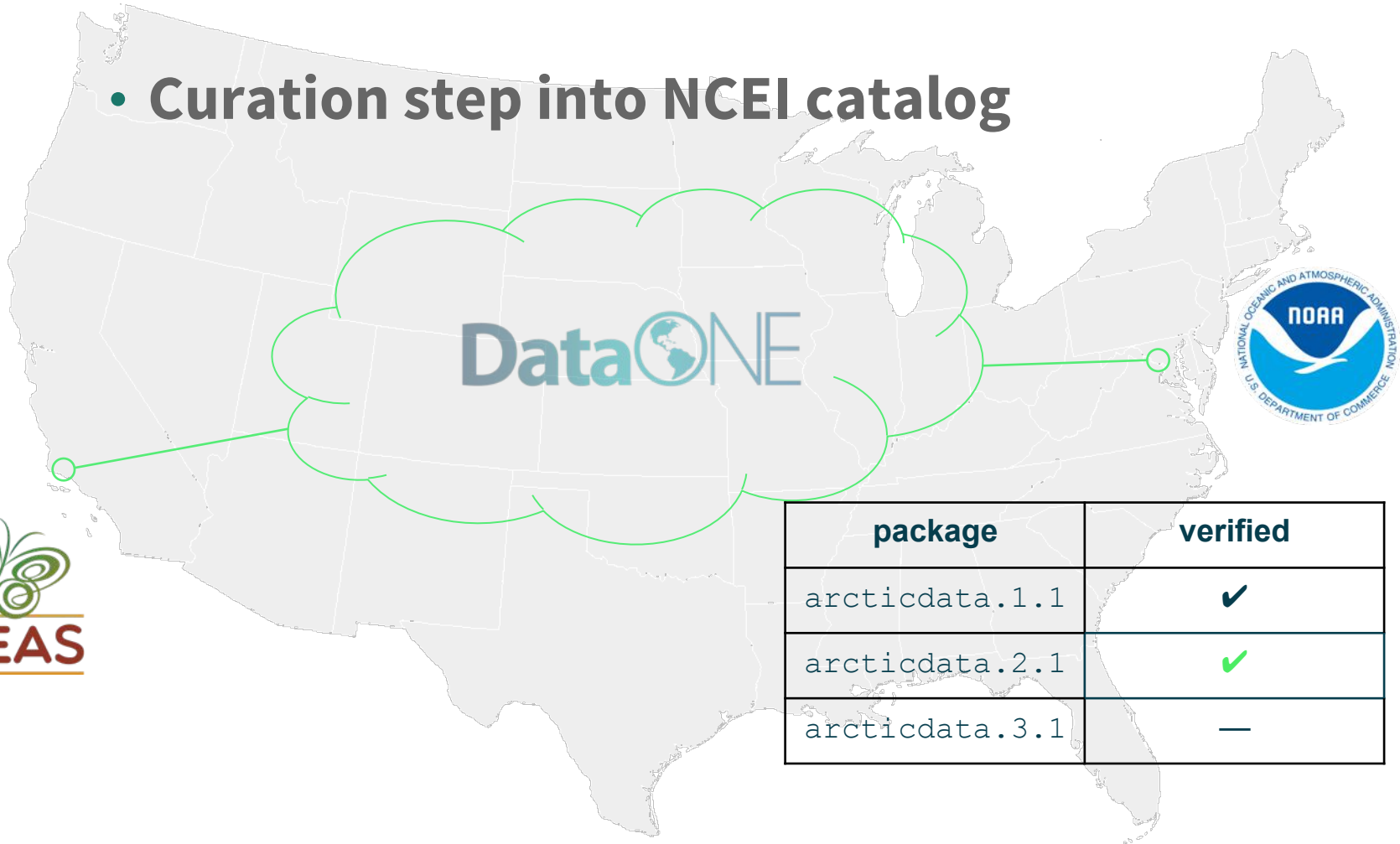
- All files are audited every 3 months





Replication with NCEI Member Node

- Automated replication via DataONE
- Curation step into NCEI catalog



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



The New Web Submission System

NSF Arctic Data Center

NSF Arctic Data Center

LPD

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

DataSupportAboutSubmit DataLauren 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
WELTS.zip	940 KB	Data
WELTS_flatfile.csv	10.65 MB	Data

Overview

People

Dates

Locations

Overview

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

NSF Arctic Data Center

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
WELTS.zip	940 KB	Data
WELTS_flatfile.csv	10.65 MB	Data

[+ Add Files](#)

Overview

People

Dates

Locations

Overview

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

Upload many files at once



The New Web Submission System

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

Done

Describe each data file
AND each measurement



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

Overview

People

Dates

Locations

Taxa

Methods

Locations

Description

Bounding Box Coordinates
Enter only one lat,lng pair for a single point

Short geographic description

Northwest coordinates

Southeast coordinates

Alaska

73.875170.375

47.375-125.5

Short geographic description

Northwest coordinates

Southeast coordinates

Cascade Lake, Brooks Range. Alaska.

68.383-154.57

LatLong

Short geographic description

Northwest coordinates

Southeast coordinates

LatLong

LatLong

Submit

More than one location? No problem!





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

Agenda	Day 1.	Day 2.
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