NSF Arctic Data Center

https://arcticdata.io

@arcticdatactr

Best Practices: Data and Metadata Submission

Matthew B. Jones

jones@nceas.ucsb.edu
https://orcid.org/0000-0003-0077-4738
@metamattj

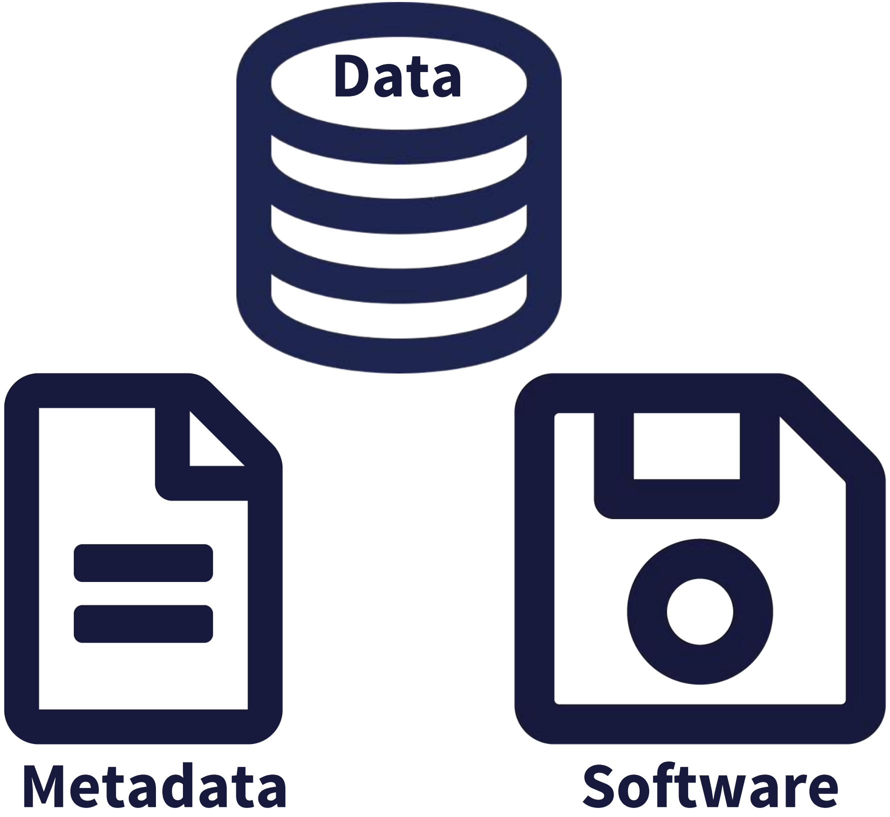
NSF Award #1546024

NCEAS

DataONE

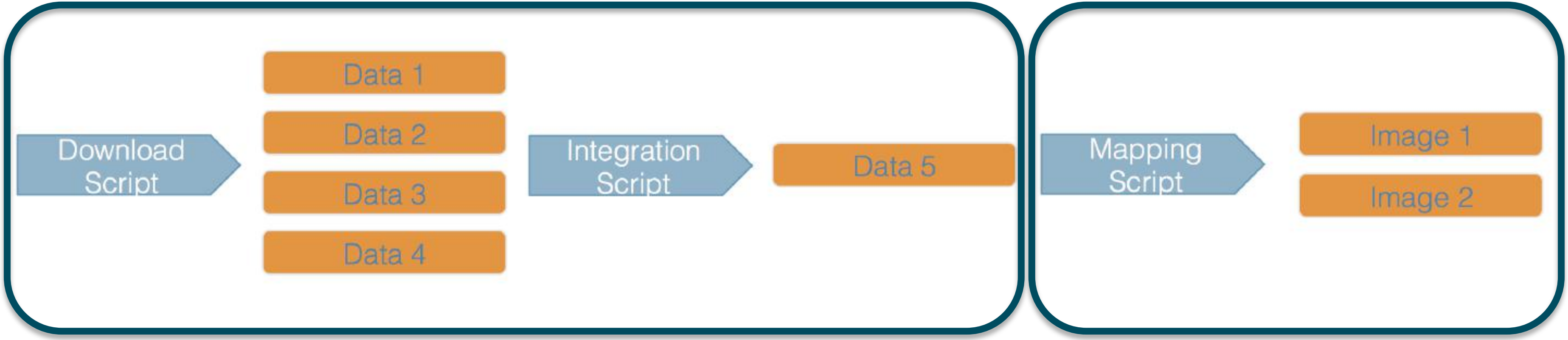
Computational Reproducibility

- Preservation enables:
  - Understanding
  - Evaluation
  - Reuse
- Future You!



Computational Workflows

Data Packages



Raw data package

Derived data package

NSF ARCTIC DATA CENTER

Data Support About Submit Data Sign in with Orcid

Home / Search / Metadata

Anna-Maria Virkkala and Miska Luoto. 2018. Arctic Chamber Metadata, 2000-2018. Arctic Data Center.
doi:10.18739/A28C6Q.

Copy Citation | Quality report

Files in this dataset Package: resource_map_doi:10.18739/A28C6Q

Name		File type	Size	Downloads	Download All
Metadata: science_metadata.xml		EML v2.1.1	33 KB	50 views	Download
Virkkala_ArcticChamber_2018.csv	More info	text/csv	191 KB	12 downloads	Download

General

Identifier: doi:10.18739/A28C6Q

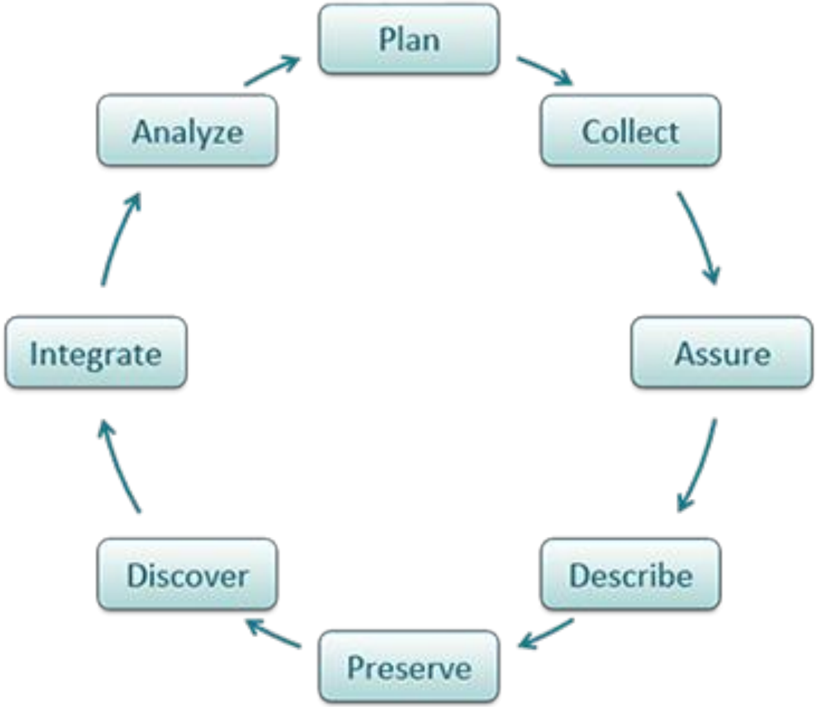
Abstract: This data summarizes the metadata of terrestrial Arctic or sub-Arctic CO2 flux chamber studies published in the 21st century. It provides descriptive information regarding the studies in general (title, keywords, authors), sites (coordinates, region), measurements (chamber size, measurement device, measurement period, fluxes), and measured plots (species, vegetation type). We aim to update the table every few years to keep track of the current state and distribution of chamber studies.

Practical Reproducibility

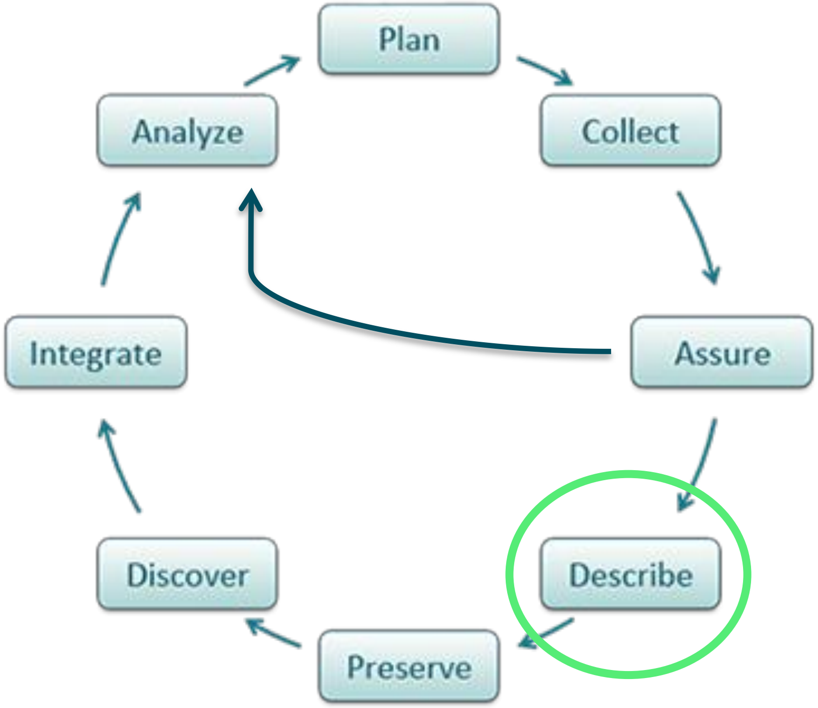
- Preserve the data
- Preserve the software workflow
- Document what you did
- Describe how to interpret it all

Data and Metadata Guidelines

A Data Life Cycle



A Data Life Cycle

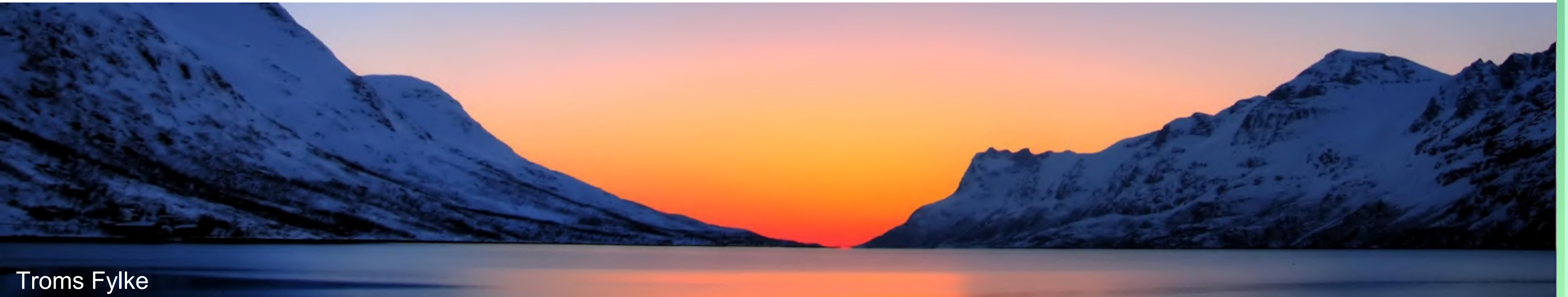


https://www.dataone.org/best-practices

Guidelines

https://arcticdata.io/submit/

- Organizing Data
- File Formats
- Large Data Packages
- Metadata
- Data Identifiers
- Provenance



Organizing Data

- Understand basics of “tidy” data models
- Design and create effective data tables

- **Benefits of tidy data systems**
- Powerful search and filtering
- Handle large, complex data sets
- Enforce data integrity
- Decrease errors from redundant updates

Not Tidy: Multiple Tables

AtlasGroveCOMPLETE.xls

Table 1

species	tree	main trunks kg	reiterated trunks kg	limbs kg	branches kg	leaves kg
SESE	Atlas	255144.9	46020.6	5477.7	13433.2	1101.2
SESE	Ballantine	221966.4	7651.6	5922.9	11210.0	1084.8
SESE	Bell	253246.4	5454.3	5792.6	48500.7	1043.4
SESE	Broken Top	130928.9	4805.2	1608.1	5137.4	729.9
SESE	Buena Vista	128833.0	3486.5	0.0	8552.1	518.4
SESE	Demeter	155896.0	11085.6	3204.3	10054.1	768.7
SESE	Epimetheus	226987.0	12915.7	1797.2	13585.2	1029.4
SESE	Iluvatar	349586.6	65003.9	12315.6	13987.0	1461.8
SESE	Kronos	134154.1	12204.4	7232.7	5036.1	597.3
SESE	Pleiades I	182385.2	3735.0	1935.2	10846.6	762.2
SESE	Pleiades II	235838.8	11183.4	4306.0	11306.5	877.7
SESE	Prometheus	239414.0	25228.9	1612.6	12458.2	1086.0
SESE	Rhea	[illegible]	487.8	730.1	5524.2	691.2
SESE	Zeus	[illegible]	[illegible]	1620.4	19104.7	954.3
SESE	3	[illegible]	0.0	0.0	87.6	41.4
SESE	4	6312.0	356.0	73.5	214.1	43.8
SESE	5	206.0	0.0	0.0	8.7	2.5
SESE	6E	18697.4	0.0	0.0	1055.2	66.3
SESE	6W	14651.5	7.7	0.0	626.3	49.6
SESE	11	614.4	0.0	0.0	28.1	17.0
SESE	12	232.1	0.0	0.0	11.2	10.3
SESE	18	15632.0	0.0	0.0	946.3	106.8
SESE	19	11805.5	0.0	0.0	770.1	80.3
SESE	20	309.5	0.0	0.0	12.5	5.9
SESE	22	25618.3	0.0	0.0	1504.0	120.2
SESE	23	463.7	0.0	0.0	18.9	4.5
SESE	25	87.7	0.0	0.0	4.1	1.3
SESE	30	512.1	1.8	0.0	18.7	8.7

Table 2

type	species	main trunk (dry masses (kg))	reiteration	limb	branch	leaf	TOTAL	% total
tree	SESE	3569312	213247	53714	230945	17192	4084409	95.3491
tree	PSME	135815	0	0	8338	961	145114	3.3876
tree	THSE	31799	0	0	6343	864	39006	0.9105
tree	ACMA	4444	0	0	925	264	5634	0.1315
tree	UMCA	2921	0	0	937	273	4131	0.0964
shrub	RUSP	0	0	0	1974	686	2660	0.0620
fern	POMU	0	0	0	[illegible]	1271	1271	0.0296
shrub	VAOV	0	0	0	526	26	552	0.0129
shrub	COCO	0	0	0	284	6	289	0.0067
fern	POSC	0	0	0	107	89	196	0.0045
tree	RHPU	100	0	0	44	18	162	0.0037
herb	OXOR	0	0	0	0	112	112	0.0026
shrub	VAPA	0	0	0	94	4	99	0.0023
tree	PISI	0	0	0	1	0	1	0.0000
tree	CHLA	0	0	0	1	0	1	0.0000
shrub	GASH	0	0	0	0	0	0	0.0000
shrub	SACA	0	0	0	0	0	0	0.0000
		3744390	213247	53714	250519	21767	4283636	

Table 3

	main trunk	reiteration	limb	branch	leaf	total	proportion geophytic
SESE geo	3569312	213247	53714	230945	17192	4084409	1.00
SESE epi	0	0	0	0	0	0	
PSME geo	135815	0	0	8338	961	145114	1.00
PSME epi	0	0	0	0	0	0	
TSHE geo	31740	0	[illegible]	6332	[illegible]	38932	0.99
TSHE epi	59	0	0	12	4	74	
ACMA geo	4444	0	0	925	264	5634	1.00
ACMA epi	0	0	0	0	0	0	

Not Tidy: Inconsistent observations

AtlasGroveCOMPLETE.xls

A	B	C	D	E	F	G
		main trunks	reiterated trunks	limbs	branches	leaves
species	tree	kg	kg	kg	kg	kg
SESE	Atlas	255144.9	46020.6	5477.7	13433.2	1101.2
SESE	Ballantine	221966.4	7651.6	5922.9	11210.0	1084.8
SESE	Bell	253246.4	5454.3	5792.6	48500.7	1043.4
SESE	Broken Top	130928.9	4805.2	1608.1	5137.4	729.9
SESE	Buena Vista	128833.0	3486.5	0.0	8552.1	518.4
SESE	Demeter	155896.0	[illegible]	3204.3	10054.1	768.7
SESE	Epimetheus	226987.0	12915.7	1797.2	13585.2	[illegible]
SESE	Iluvatar	349586.6	65003.9	[illegible]	13987.0	[illegible]
SESE	Kronos	134154.1	12204.4	[illegible]	5036.1	[illegible]
SESE	Pleiades I	182385.2	3735.0	1935.2	10846.6	[illegible]
SESE	Pleiades II	235838.8	11183.4	4306.0	[illegible]	[illegible]
SESE	Prometheus	239414.0	25228.9	1612.6	[illegible]	[illegible]
SESE	Rhea	143710.4	487.8	730.1	5524.2	[illegible]
SESE	Zeus	243365.7	2885.5	1620.4	19104.7	[illegible]
SESE	3	1761.3	0.0	0.0	87.6	[illegible]
SESE	4	6312.0	356.0	73.5	214.1	[illegible]
SESE	5	206.0	0.0	0.0	8.7	[illegible]
SESE	6E	18697.4	0.0	0.0	1055.2	[illegible]
SESE	6W	14651.5	7.7	0.0	626.3	[illegible]
SESE	11	614.4	0.0	0.0	28.1	17.0
SESE	12	232.1	0.0	0.0	11.2	10.3
SESE	18	15632.0	0.0	0.0	946.3	106.8
SESE	19	11805.5	0.0	0.0	770.1	80.3
SESE	20	309.5	0.0	0.0	12.5	5.9
SESE	22	25618.3	0.0	0.0	1504.0	120.2
SESE	23	463.7	0.0	0.0	18.9	4.5
SESE	25	87.7	0.0	0.0	4.1	1.3
SESE	30	512.1	1.8	0.0	18.7	8.7

I	J	K	L	M	N	O	P	Q
				dry masses (kg)				
type	species	main trunk	reiteration	limb	branch	leaf	TOTAL	% total
tree	SESE	3569312	213247	53714	230945	17192	4084409	95.3491
tree	PSME	135815	0	0	8338	961	145114	3.3876
tree	THSE	31799	0	0	6343	864	39006	0.9105
tree	ACMA	4444	0	0	925	264	5634	0.1315
tree	UMCA	2921	0	0	937	273	4131	0.0964
shrub	RUSP	0	0	0	1974	686	2660	0.0620
[illegible]	[illegible]	[illegible]	0	0	0	1271	1271	0.0296
[illegible]	[illegible]	[illegible]	0	0	526	26	552	0.0129
[illegible]	[illegible]	[illegible]	0	0	284	6	289	0.0067
[illegible]	[illegible]	[illegible]	0	0	107	89	196	0.0045
[illegible]	[illegible]	[illegible]	0	0	44	18	162	0.0037
[illegible]	[illegible]	[illegible]	0	0	0	112	112	0.0026
[illegible]	[illegible]	[illegible]	0	0	94	4	99	0.0023
[illegible]	[illegible]	[illegible]	0	0	1	0	1	0.0000
[illegible]	[illegible]	[illegible]	0	0	1	0	1	0.0000
[illegible]	[illegible]	[illegible]	0	0	0	0	0	0.0000
[illegible]	[illegible]	[illegible]	0	0	0	0	0	0.0000
		[illegible]	[illegible]	53714	250519	21767	4283636	
								proportion
		main trunk	reiteration	limb	branch	leaf	total	geophytic
	SESE geo	3569312	213247	53714	230945	17192	4084409	1.00
	SESE epi	0	0	0	0	0	0	
	PSME geo	135815	0	0	8338	961	145114	1.00
	PSME epi	0	0	0	0	0	0	
	TSHE geo	31740	0	0	6332	860	38932	0.99
	TSHE epi	59	0	0	12	4	74	
	ACMA geo	4444	0	0	925	264	5634	1.00
	ACMA epi	0	0	0	0	0	0	

All the same observation?

No.

Not Tidy: Inconsistent variables

AtlasGroveCOMPLETE.xls

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
		main trunks	reiterated trunks	limbs	branches	leaves						dry masses (kg)				
species	tree	kg	kg	kg	kg	kg		type	species	main trunk	reiteration	limb	branch	leaf	TOTAL	% total
SESE	Atlas	255144.9	46020.6	5477.7	13433.2	1101.2		tree	SESE	3569312	213247	53714	230945	17192	4084409	95.3491
SESE	Ballantine	221966.4	7651.6	5922.9	11210.0	1084.8		tree	PSME	135815	0	0	8338	961	145114	3.3876
SESE	Bell	253246.4	5454.3	5792.6	48500.7	1043.4		tree	THSE	31799	0	0	6343	864	39006	0.9105
SESE	Broken Top	130928.9	4805.2	1608.1	5137.4	729.9		tree	ACMA	4444	0	0	925	264	5634	0.1315
SESE	Buena Vista	128833.0	3486.5	0.0	8552.1	518.4		tree	UMCA	2921	0	0	937	273	4131	0.0964
SESE	Demeter	155896.0	11085.6	3204.3	10054.1	768.7		shrub	RUSP	0	0	0	1974	686	2660	0.0620
SESE	Epimetheus	226987.0	12915.7	1797.2	13585.2	1029.4		fern	POMU	0	0	0	0	1271	1271	0.0296
SESE	Iluvatar	349586.6	65003.9	12315.6	13987.0	1461.8		shrub	VAOV	0	0	0	526	26	552	0.0129
SESE	Kronos	134154.1	12204.4	7232.7	5036						0	0	284	6	289	0.0067
SESE	Pleiades I	182385.2	3735.0	1935.2	10846						0	0	107	89	196	0.0045
SESE	Pleiades II	235838.8	11183.4	4306.0	11306						0	0	44	18	162	0.0037
SESE	Prometheus	239414.0	25228.9	1612.6	12458						0	0	0	112	112	0.0026
SESE	Rhea	143710.4	487.8	730.1	5524						0	0	94	4	99	0.0023
SESE	Zeus	243365.7	2885.5	1620.4	19104						0	0	1	0	1	0.0000
SESE	3	1761.3	0.0	0.0	87						0	0	1	0	1	0.0000
SESE	4	6312.0	356.0	73.5	214						0	0	0	0	0	0.0000
SESE	5	206.0	0.0	0.0	8						0	0	0	0	0	0.0000
SESE	6E	18697.4	0.0	0.0	1055						213247	53714	250519	21767	4283636	
SESE	6W	14651.5	7.7	0.0	626											proportion
SESE	11	614.4	0.0	0.0	28						eration	limb	branch	leaf	total	geophytic
SESE	12	232.1	0.0	0.0	11.2	10.3			SESE geo	3569312	213247	53714	230945	17192	4084409	1.00
SESE	18	15632.0	0.0	0.0	946.3	106.8			SESE epi	0	0	0	0	0	0	
SESE	19	11805.5	0.0	0.0	770.1	80.3			PSME geo	135815	0	0	8338	961	145114	1.00
SESE	20	309.5	0.0	0.0	12.5	5.9			PSME epi	0	0	0	0	0	0	
SESE	22	25618.3	0.0	0.0	1504.0	120.2			TSHE geo	31740	0	0	6332	860	38932	0.99
SESE	23	463.7	0.0	0.0	18.9	4.5			TSHE epi	59	0	0	12	4	74	
SESE	25	87.7	0.0	0.0	4.1	1.3			ACMA geo	4444	0	0	925	264	5634	1.00
SESE	30	512.1	1.8	0.0	18.7	8.7			ACMA epi	0	0	0	0	0	0	

All the same
variable?
No.

Not Tidy: Marginal info

AtlasGroveCOMPLETE.xls

		main trunks	reiterated trunks	limbs	branches	leaves
species	tree	kg	kg	kg	kg	kg
SESE	Atlas	255144.9	46020.6	5477.7	13433.2	1101.2
SESE	Ballantine	221966.4	7651.6	5922.9	11210.0	1084.8
SESE	Bell	253246.4	5454.3	5792.6	48500.7	1043.4
SESE	Broken Top	130928.9	4805.2	1608.1	5137.4	729.9
SESE	Buena Vista	128833.0	3486.5	0.0	8552.1	518.4
SESE	Demeter	155896.0	11085.6	3204.3	10054.1	768.7
SESE	Epimetheus	226987.0	12915.7	1797.2	13585.2	1029.4
SESE	Iluvatar	349586.6	65003.9	12315.6	13987.0	1461.8
SESE	Kronos	134154.1	12204.4	7232.7	5036.1	597.3
SESE	Pleiades I	182385.2	3735.0	1935.2	10846.6	762.2
SESE	Pleiades II	235838.8	11183.4	4306.0	11306.5	877.7
SESE	Prometheus	239414.0	25228.9	1612.6	12458.2	1086.0
SESE	Rhea	143710.4	487.8	730.1	5524.2	691.2
SESE	Zeus	243365.7	2885.5	1620.4	19104.7	954.3
SESE	3	1761.3	0.0	0.0	87.6	41.4
SESE	4	6312.0	356.0	73.5	214.1	43.8
SESE	5	206.0	0.0	0.0	8.7	2.5
SESE	6E	18697.4	0.0	0.0	1055.2	66.3
SESE	6W	14651.5	7.7	0.0	626.3	49.6
SESE	11	614.4	0.0	0.0	28.1	17.0
SESE	12	232.1	0.0	0.0	11.2	10.3
SESE	18	15632.0				
SESE	19	11805.5				
SESE	20	309.5				
SESE	22	25618.3				
SESE	23	463.7				
SESE	25	87.7				
SESE	30	512.1				

		dry masses (kg)						
type	species	main trunk	reiteration	limb	branch	leaf	TOTAL	% total
tree	SESE	3569312	213247	53714	230945	17192	4084409	95.3491
tree	PSME	135815	0	0	8338	961	145114	3.3876
tree	THSE	31799	0	0	6343	864	39006	0.9105
tree	ACMA	4444	0	0	925	264	5634	0.1315
tree	UMCA	2921	0	0	937	273	4131	0.0964
shrub	RUSP	0	0	0	1974	686	2660	0.0620
fern	POMU	0	0	0	0	1271	1271	0.0296
shrub	VAOV	0	0	0	526	26	552	0.0129
shrub	COCO	0	0	0	284	6	289	0.0067
fern	POSC	0	0	0	107	89	196	0.0045
tree	RHPU	100	0	0	44	18	162	0.0037
herb	OXOR	0	0	0	0	112	112	0.0026
shrub	VAPA	0	0	0	94	4	99	0.0023
tree	PISI	0	0	0	1	0	1	0.0000
tree	CHLA	0	0	0	1	0	1	0.0000
shrub	GASH	0	0	0	0	0	0	0.0000
shrub	SACA	0	0	0	0	0	0	0.0000
		3744390	213247	53714	250519	21767	4283636	

	main trunk	reiteration	limb	branch	leaf	total	proportion geophytic
SESE	3569312	213247	53714	230945	17192	4084409	1.00
SE epi	0	0	0	0	0	0	
ME geo	135815	0	0	8338	961	145114	1.00
ME epi	0	0	0	0	0	0	
HE geo	31740	0	0	6332	860	38932	0.99
HE epi	59	0	0	12	4	74	
MA geo	4444	0	0	925	264	5634	1.00
MA epi	0	0	0	0	0	0	

Marginal sums and totals

Data Modeling 101

id	date	site	elev	sp1code	sp1height	sp2code	sp2height
1	2017-10-10	1	3.7	DAPU	4.6	DAMA	4.5
2	2017-09-05	2	3.2	DAMA	3.5	DAPU	3.9

- Denormalized data (aka, not Tidy)
- Observations about different entities combined

Tidy Data (observe one entity per table)

- Species observations

id	date	site	spcode	height
1	2017-10-10	1	DAPU	4.6
2	2017-09-05	2	DAMA	3.5
3	2017-10-10	1	DAMA	4.5
4	2017-09-05	2	DAPU	3.9

- Site observations

site	name	elev	temp
1	Taku	3.7	21.2
2	Lituya	3.2	23.1

Tidy Data (Relational)

Join Key

- Species observations

id	date	site	spcode	height
1	2017-10-10	1	DAPU	4.6
2	2017-09-05	2	DAMA	3.5
3	2017-10-10	1	DAMA	4.5
4	2017-09-05	2	DAPU	3.9

- Site observations

site	name	elev	temp
1	Taku	3.7	21.2
2	Lituya	3.2	23.1

Organizing Data: Best Practices

- **Some Simple Guidelines for Effective Data Management.**
  - Borer et al. 2009. Bulletin of the Ecological Society of America. https://doi.org/10.1890/0012-9623-90.2.205

- **Nine simple ways to make it easier to (re)use your data.**
  - White et al. 2013. Ideas in Ecology and Evolution 6. https://doi.org/10.4033/iee.2013.6b.6.f

Organizing Data: Best Practices

- **Scripts** for all data manipulation
  - Uncorrected raw data file
  - Document processing in scripts
- **Design to add rows, not columns**
  - Each column one variable
  - Each row one observation
- **Nonproprietary file formats**
  - Descriptive names, no spaces
  - Header line

File Formats

https://arcticdata.io/submit/#file-format-guidelines

- **Open Formats**
  - **Text** - support long term access and preservation
  - **Open binary** formats (NetCDF, HDF5)

Always bet on text!

- Any (meta)data is better than none
  - Microsoft Excel: common but proprietary
  - Export GIS data to ESRI shapefiles
  - Export MATLAB, IDL, etc. to NetCDF

Large Data Packages (> Terabytes)

- Talk to the data center early
- Tile data structures by subset
  - Spatial regions
  - Temporal windows
  - Measured variables
- Use efficient tools (NetCDF, HDF)
  - Compact data format
  - Parallel read/write libraries

Metadata Guidelines

Metadata: the Goal

- Target a typical researcher (maybe you!)
- 30+ years from now

- Goal
  - Understand
  - Interpret
  - Re-use

Metadata

Metadata: the Goal

- **What** was measured?
- **Who** did it?
- **When** and **where**?
- **How**? (data structure & methods)
- **Why**? (science context)
- **Attribution** & **Licensing**

Metadata

Metadata: Bibliographic Details

- **Global Identifier (e.g., DOI)**
- **Descriptive title**
  - topic, geographic location, dates, and, if applicable, the scale of the data
- **Descriptive abstract**
  - brief overview of the specific contents and purpose of the data package.
- **Funding** information (award number and sponsor).
- **People and organizations**
  - **Creators** – who should be cited for the data set
  - Contacts
  - Contributors
  - Sponsors, and more

Metadata

Metadata: Discovery Details

- **Geospatial coverage**
  - Field and laboratory sampling locations
  - including place names and precise coordinates
- **Temporal Coverage**
  - When measurements were made
  - To what time period do measurements apply
  - Might be calendar times, or geologic times
- **Taxonomic Coverage**
  - What species were measured
  - Taxonomy standards and procedures
- Other contextual information

Metadata

Metadata: Interpretation Details

- Field and laboratory data **collection methods**
- Full **experimental and project design**, and relationship to data
- Full field and laboratory sample **processing methods**
- **Sampling quality control** procedures

- Analysis and modeling methods
  - **Provenance** information
  - **Hardware** and **software** used
    - including make, model, and version
  - **Computing quality control** procedures
    - testing, code review, etc.

Metadata

Metadata: Data Structure and Contents

- **Data model description**

- **Data object descriptions (granules)**
  - Tables
  - Images
  - Matrices
  - Spatial layers, etc.

- **Variable information** (attributes/parameters)
  - Definitions / link to methods
  - Standardized measurement types
  - Units
  - Coded values
  - Missing value codes

Metadata

Metadata: Rights and Attribution

- **Scientific rights and expectations**
  - **Citation format**
  - **Attribution expectations**
  - **Reuse rights**
    - Who may reuse data, and for what purposes
  - **Redistribution rights**
    - Who may copy and redistribute data and metadata

- **Legal terms and conditions**
  - **Licensing** terms

Metadata Standards

- Ecological Metadata Language (EML)
- Geospatial Metadata Standards
  - (ISO 19115*, ISO 19139)
- Biological Data Profile (BDP)
- Dublin Core
- Darwin Core
- PREMIS and METS
- … and the list goes on

Metadata

Data | Submit Data | Matthew Jones

Research and Analysis Section. 2017. Resident vs Nonresident Workers Wages in the Alaskan Seafood and Fishing Processing Industry. KNB Test Node. urn:uuid:d52fa737-fdc1-4192-9c60-b2ad145aa7f9.

Files	Size	Type	Status	
Resident vs Nonresident Workers Wages in the Alaskan Seafood and Fishing Processing Industry	26 KB			Add Files
AISFPOver.pdf	6 KB	Data		Describe
processingWorkersWages4.csv	6 KB	Data		Describe
ANSFPOver.pdf	6 KB	Data		Describe

- Overview *
- People
- Dates *
- Locations *
- Taxa
- Methods

Overview

Title *

A title for this dataset. Include the topic, geographic location, dates, and if applicable, the scale of the data. Write out all abbreviations.

Resident vs Nonresident Workers Wages in the Alaskan Seafood and Fishing Processing Industry

Abstract *

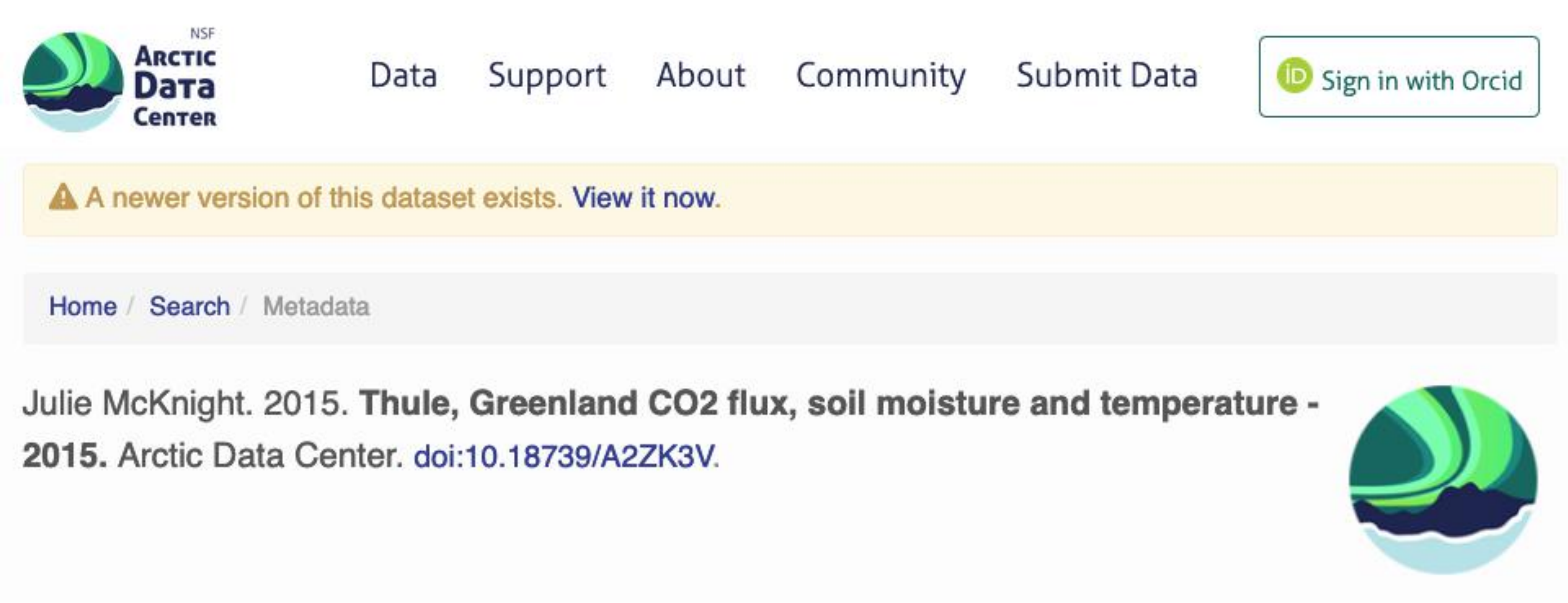
Provide a brief overview that summarizes the specific contents and purpose of this dataset.

These data were taken from Alaska's Department of Labor and Workforce Development website (http://live.laborstats.alaska.gov/seafood/), Research and Analysis Section. The csv data file is extracted from the pdfs included in the data package. The data file contains the average wages of resident and nonresident workers in the Alaskan seafood and fishing processing industry from 2001-2015. The data are organized into 8 regions, and 1 'Statewide' region encompassing all 8 regions. For the Northern region data, the large jump in workers in 2013 was due to an employer previously in a different industry being recoded into the seafood processing industry.

Data Identifiers

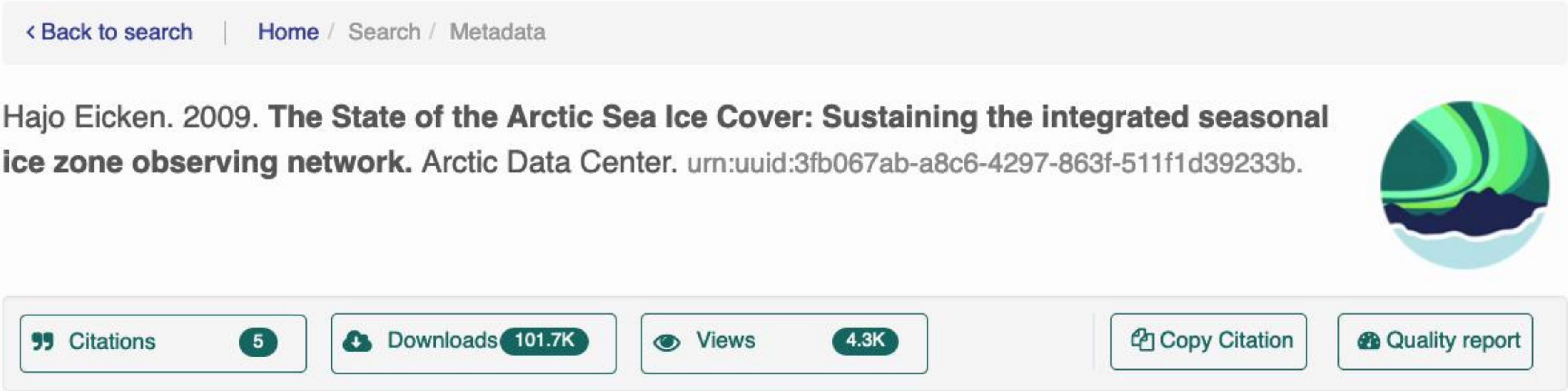
Nina J. Karnovsky and Ann M. A. Harding. 2016. At-sea density of foraging little auks (Alle alle) near Hornsund Fjord. Arctic Data Center. doi:10.5065/D6MK6B17.

- DOI == Digital Object Identifier
- We assign a DOI to each published data set
- Researchers should cite data they use



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

Data Usage Metrics



MAKE DATA COUNT

5 Citations

I.J. Smith, H. Eicken, A.R. Mahoney, R. Van Hale, A.J. Gough, et al. 2016. Surface water mass composition changes captured by cores of Arctic land-fast sea ice. Continental Shelf Research. Vol. 118. pp. 154-164. https://doi.org/10.1016/j.csr.2016.02.008.

Daisuke Hirano, Yasushi Fukamachi, Eiji Watanabe, Kay I. Ohshima, Katsushi Iwamoto, et al. 2016. A wind-driven, hybrid latent and sensible heat coastal polynya off Barrow, Alaska. Journal of Geophysical Research: Oceans. Vol. 121. pp. 980-997. https://doi.org/10.1002/2015JC011318.

Megan O'Sadnick, Malcolm Ingham, Hajo Eicken, and Erin Pettit. 2016. In situ field measurements of the temporal evolution of low-frequency sea-ice dielectric properties in relation to temperature, salinity, and microstructure. The Cryosphere. Vol. 10. pp. 2923-2940. https://doi.org/10.5194/tc-10-2923-2016.

Megan O'Sadnick, Malcolm Ingham, Hajo Eicken, and Erin Pettit. 2016. In situ field measurements of the temporal evolution of low-frequency sea-ice dielectric properties in relation to temperature, salinity, and microstructure. The Cryosphere. Vol. 10. pp. 2923-2940. https://doi.org/10.5194/tc-10-2923-2016.

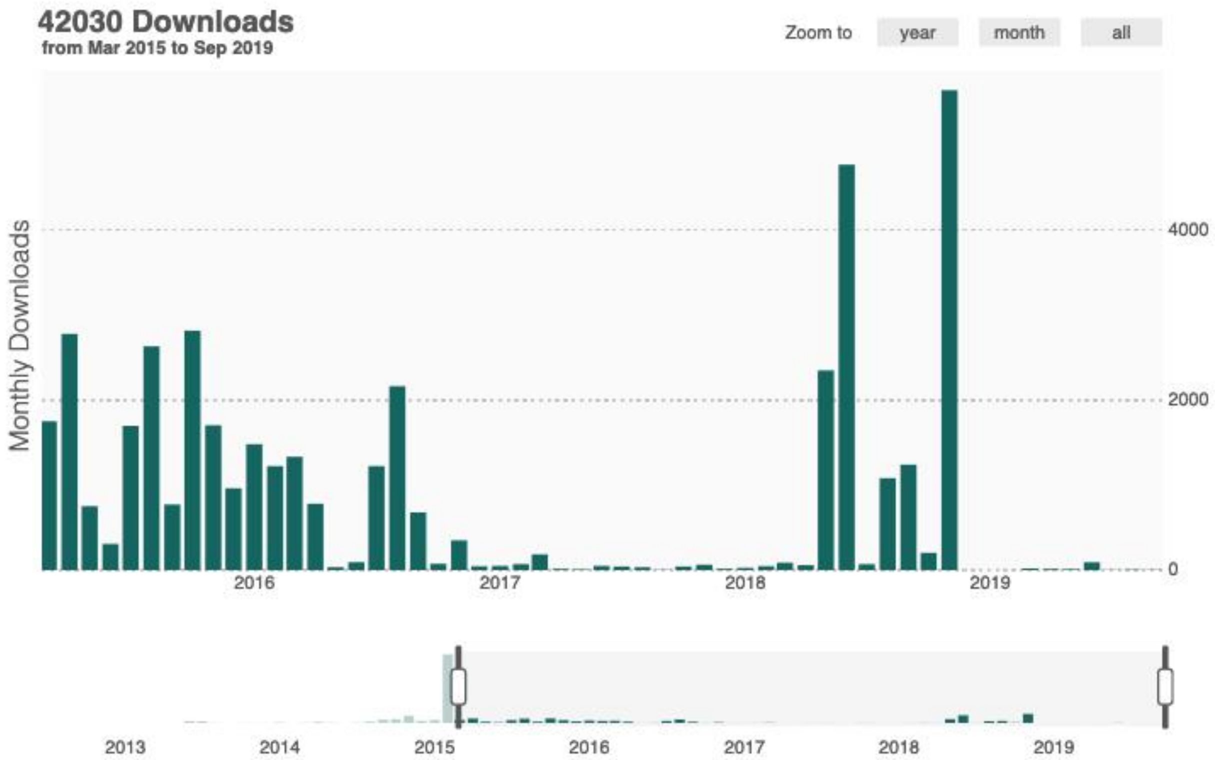
P. J. Griewank and D. Notz. 2015. A 1-D modelling study of Arctic sea-ice salinity. The Cryosphere. Vol. 9. pp. 305-329. https://doi.org/10.5194/tc-9-305-2015.

< Views Downloads >

101.7K Downloads

For all versions of this data set, the number of times that all or part of this data set was downloaded over time.

These download counts are COUNTER compliant, meaning that downloads from some Internet robots and repeat downloads within a certain time window are excluded.



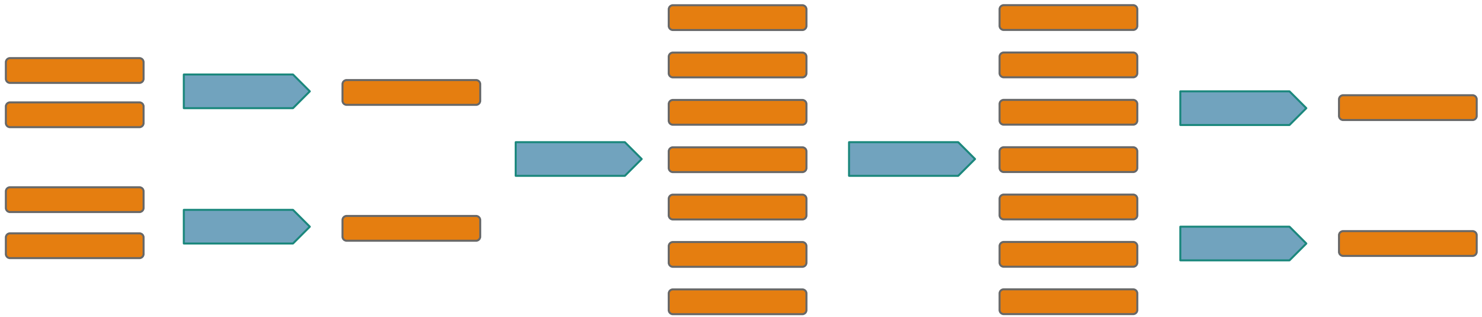
Drag the slider to visualize a specific time window for the download events.

Citations

Views

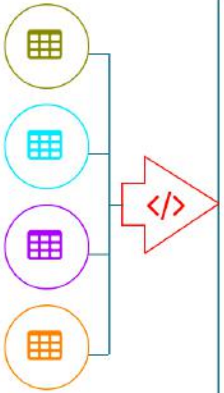
Provenance Metadata

- Simplified view of complex workflows



Data Table, Image, and Other Data Details

4 sources



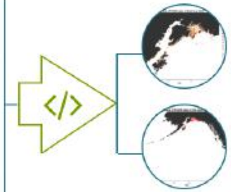
Data Table

Entity Name	**Total_Aromatic_Alkanes_PWS.csv**
	Download
Description	Combined dataset from PAH, Alkane and Sample tables documenting samples collected after the Exxon Valdez oil spill in Prince William Sound, AK
Object Name	Total_Aromatic_Alkanes_PWS.csv
Online Distribution Info	https://cn.dataone.org/cn/v2/resolve/urn:uuid:44108e76-405d-4d58-b1b3-fb4b55e3fff9
Size	2801033 byte
Text Format	
Number Of Records	12142

Text Format

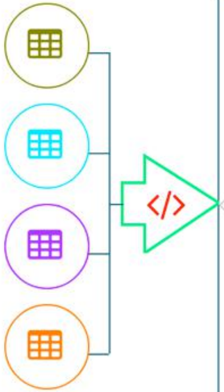
Number of Header Lines	1
Record Delimiter	#x0A
Attribute Orientation	column
Simple Text	
Field Delimeter	,

2 derivations

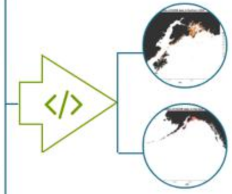


Data Table, Image, and Other Data Details

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 ⊞)**.

kanes_PWS.csv

rom PAH, Alkane and Sample tables documenting samples collected after the ll in Prince William Sound, AK

anes_PWS.csv

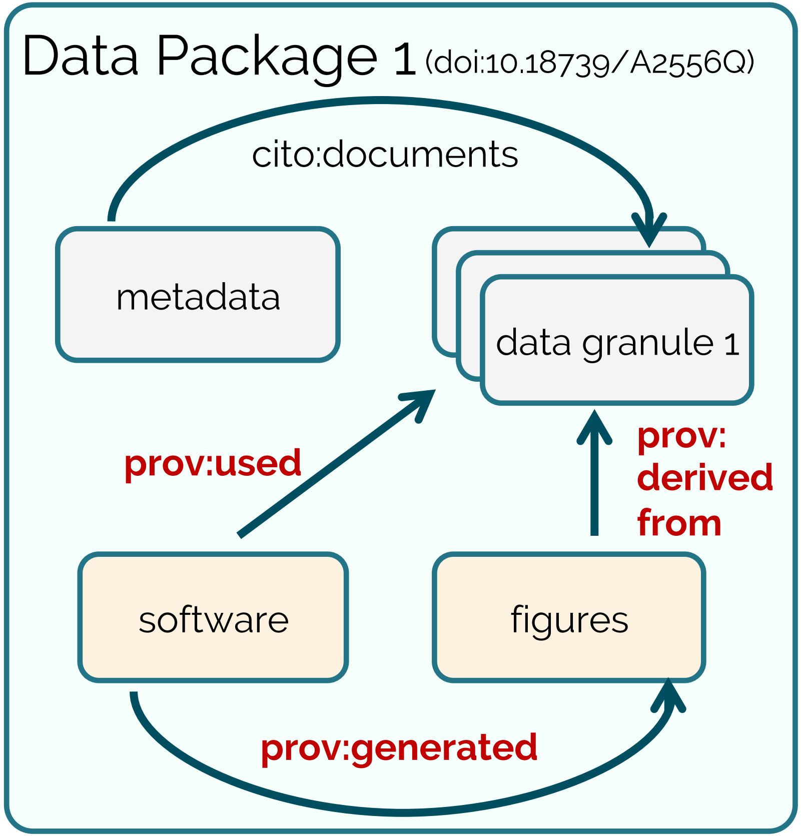
org/cn/v2/resolve/urn:uuid:44108e76-405d-4d58-b1b3-fb4b55e3fff9

Text Format

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

Number Of Records: 12142

Data package with Provenance



Rmarkdown as Provenance

01-brood-table-integration.Rmd

```
## Datasets

As part of the SASAP project, brood tables for 48 Sockeye salmon stocks were collected.
Table 2.1 shows a list of these stocks, along with other regional and location
information.

```{r, echo = FALSE}
stocks <- read.csv('data/original/StockInfo.csv', stringsAsFactors = F)
```

```{r, echo = FALSE}
datatable(stocks[, c('Stock.ID','Stock' ,'Region', 'Sub.Region')], rownames = FALSE,
caption = "Stock information")
```

These stocks range geographically from Washington to Alaska. Although temporal coverage
varies by stock, many of the brood tables were updated in 2016, and some have
reconstructions dating back to 1922.

Figure 2.1 indicates the approximate location of the salmon stocks in Table 2.1.

```{r, echo = FALSE}
salmon = makeIcon('images/salmon_tiny.png',
                  'images/salmon_big.png',
                  26, 14)

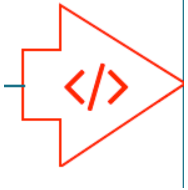
m <- leaflet(stocks) %>%
  setView(-median(stocks$Lon), median(stocks$Lat), zoom = 4) %>%
  addTiles() %>%
  addMarkers(~-Lon, ~Lat, icon = salmon)

m

```

Figure 2.1: Location of stocks used in this data integration. Salmonid icon by Servien
(vectorized by T. Michael Keesey)
[CC-BY-SA](https://creativecommons.org/licenses/by-sa/3.0/), available at
[PhylopicI](http://phylopic.org/)
```

37:72 Chunk 2 R Markdown



2.2 Datasets

As part of the SASAP project, brood tables for 48 Sockeye salmon stocks were collected. Table 2.1 shows a list of these stocks, along with other regional and location information.

Show 10 entries Search:

Stock information

Stock.ID	Stock	Region	Sub.Region
101	Washington	WA	WA
102	E.Stuart	Fraser River	Fraser Early Stuart
103	Bowron	Fraser River	Fraser Early Summer
104	Fennell	Fraser River	Fraser Early Summer
105	Gates	Fraser River	Fraser Early Summer
106	Nadina	Fraser River	Fraser Early Summer
107	Pitt	Fraser River	Fraser Early Summer
108	Raft	Fraser River	Fraser Early Summer
109	Scotch	Fraser River	Fraser Early Summer
110	Seymour	Fraser River	Fraser Early Summer

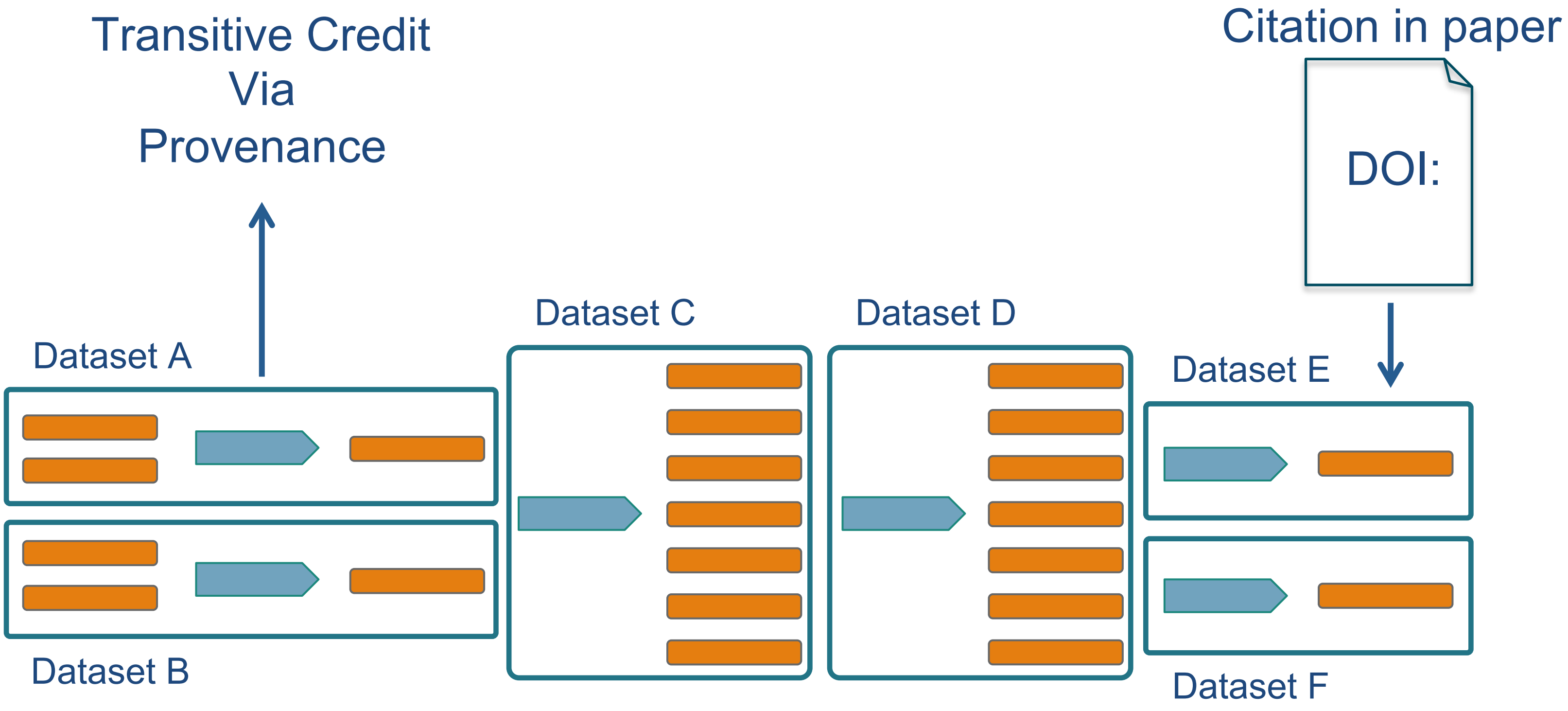
Showing 1 to 10 of 54 entries Previous 1 2 3 4 5 6 Next

These stocks range geographically from Washington to Alaska. Although temporal coverage varies by stock, many of the brood tables were updated in 2016, and some have reconstructions dating back to 1922.

Figure 2.1 indicates the approximate location of the salmon stocks in Table 2.1.

Figure 2.1: Location of stocks used in this data integration. Salmonid icon by Servien (vectorized by T.

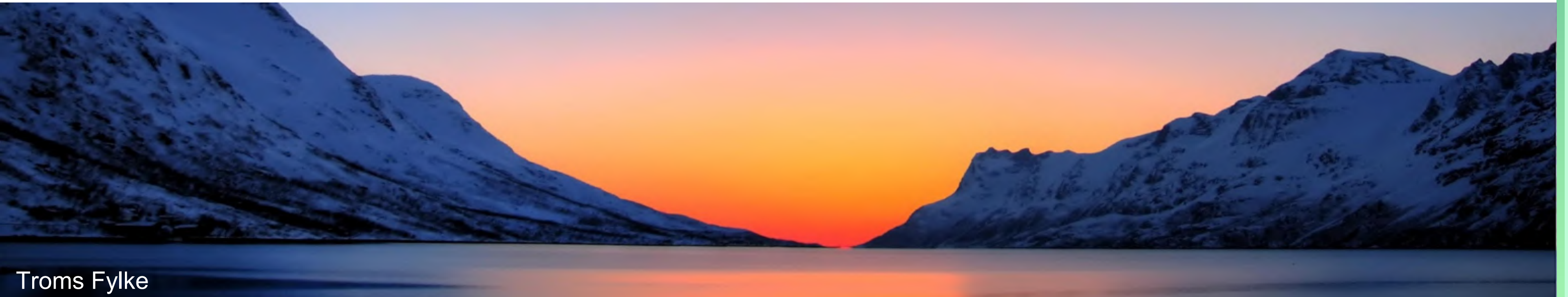
Citing multi-generational workflows



Guidelines

https://arcticdata.io/submit/

- Organizing Data
- File Formats
- Large Data Packages
- Metadata
- Data Identifiers
- Provenance



Troms Fylke

Arctic Data Center Support Team

support@arcticdata.io

Clark

Mullen

Chong

?

?

Sun

Student Interns

NSF ARCTIC DATA CENTER

https://arcticdata.io