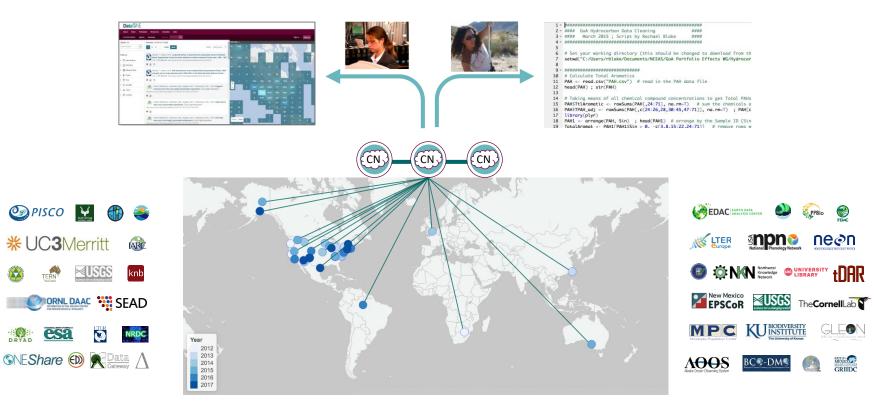


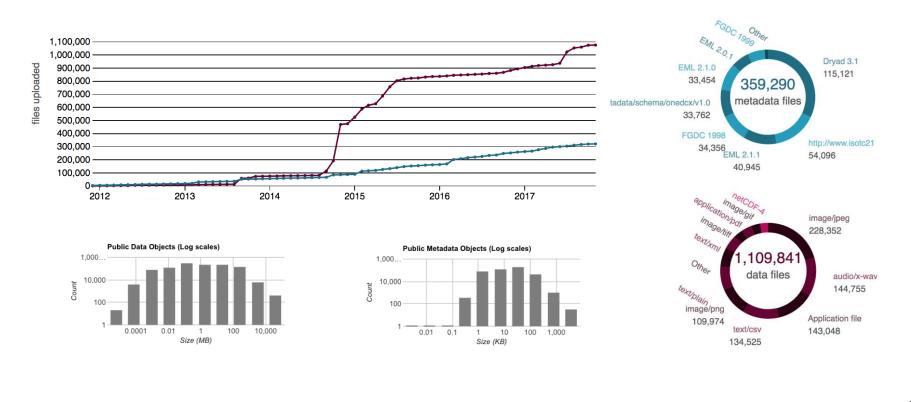
BCoN Data Integration and Attribution February 2018 Dave Vieglais University of Kansas



### Federation of Earth Science Repositories



### Production in 2012, Diverse Content



### Data Package

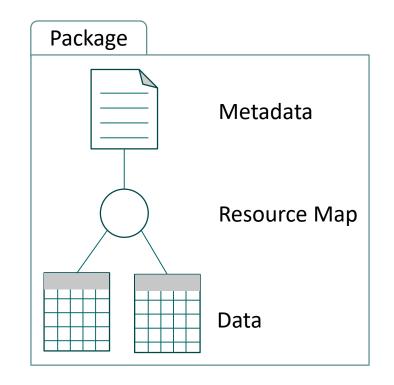
Flexible Structure for Data

Data Package components:

- Metadata
- Data
- Resource Map binding components

All components:

- Immutable
- Uniquely identified
- Resolvable
- Retrievable



### Data Package

Metadata

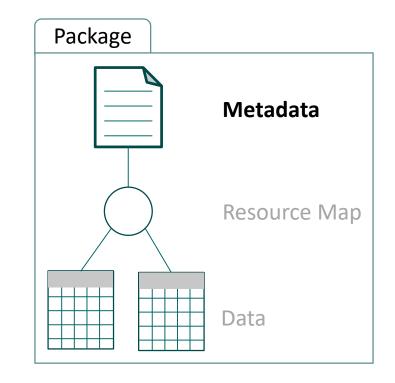
### Many Standards

- Dublin Core
- FGDC

. . .

- Dryad Metadata Profile
- Ecological Metadata Language
- ISO 19000 family





### Federated Search Across Repositories

#### **Data**SNE

DATAONE SEARCH: Search	Summary Jump to: DOI or ID Go														Sign	in or Sig	n un
	Cummay Sumpto, Second Co	Hide Map	526W												Orgin		ιup
Search @	0 🖻	нюе мар	*							100	DAVING!	ASTIC:					
Search phrase Q	👘 Fabes, Jez, Seligman, William, Barrett, Carolyn, McKechnie, Stuart, and Griffiths, John. 2017. Data from: Does the implementation of a novel								- 784 - 78	1 2 2							
ilter by:	• #YTA: Intensive care discharge risk score and nurse-ted inpatient review tool improve outcome? A prospective cohort study in two intensive care units in the United Kingdom. Dryad Digital Repository. https://doi.org/10.5061/dryad.7/f7q/verve017-11-22T0911159.034-05.00.	3	12	34	253	143	6	6	7	2	and the second	66	INT 7	13	16	3	43
ensity, length, etc. Q	-: 10 Park, Hyeree, Cho, Sooyoung, Woo, Hyeongtaek, Park, Sue K., Shin, Hai-Rim, et al. 2017. Data from: Fasting glucose and risk of colorectal cancer		1					Res."		-	Al and					1 AN	
	in the Korean Multi-center Cancer Cohort. Dryad Digital Repository. https://doi.org/10.5061/dryad.f1858?ver=2017-11-22T09.03.35.154-05.00.							E	11.	Q.	Em	CARLE .		and the		and the second	
Data files	0	1	24	342	4891	114	24	11	6	15	16	13	21	26	201	2	34
Member Node	Jason Schatz, Carly Ziter, and Christopher Kucharik. 2016. WSC - Temperature and relative humidity data from 150 locations in and around			(pris)	and the		-	19 2		1.1	15	13	der Star	2	-	A A A A A A A A A A A A A A A A A A A	
Creator Q	assurt schatz, vany ziter, and ornsopher roorians. core: woo'- temperature and relative number of too costions in and around Madison, Wisconsin from 2012 -2016. LTER Network Member Node. https://pasta.temet.edu/package/metadata/emi/knb-iter-nt/324/12.	4	175	685	2738	267	6	<b>19</b>	2	10	5	12	63	139	and 63	32	211
Year	Heather Bateman, Dan Childers, Madhusudan Katti, Eyal Shochat, and Paige Warren. 2017. Point-count bird censusing: long-term monitoring of bird abundance and diversity in central Arizona-Phoenix, ongoing since 2000. LTER Network Member Node.	14	379	345	959	605	102	7 00	<sup>ada</sup> 14	12	4	6	13	164	30	42	40
Identifier	https://pasta.iternet.edu/package/metadata/emi//rob-iter-cap/48/15.	54	57	185	205	302	277	17	270	93	12	- 18	40	164	129	89	48
Contion	Peter Bergholz. 2017. Escherichia coli isolated from surface soli in an agricultural landscape. Knowledge Network for Biocomplexity. knb.92105.1.	14	51	12	69	114	2081	1140	437	5249	1588	90	73	219	192	83	65
	Ø ≧ ♀ 1 ●	- 1	31	37	62	63	9773	5027	3862	states Kill MD	3894	11912		202	280	71	56
	Márcia Cristina Mendes Marques, Anamaria Cequinel, Elivane Salete Capellesso, and Fernanda C. G. Cardoso. 2017. valiação da regeneração	17	6	52	57	124	537	40838	12633	2951 AR	3753	** 8644	508	166	194	228	20
	🖏 👘 - naturaldo estrato arbóreo em um gradiente edáfico-sucessional em Floresta Atlântica, Antonina, Paraná, Brasil. Programa de Pesquisa em Biodiversidade (PPBio). PPBoMA.31.3. O	4	35	35	53	149	247	4987	17360	*1378	6787	1053	464	89	112	149	17:
		10	79	265	122	240	26	66	89 Mexi	590	8039	1172	95	92	213	107	13
		2	44	1080	719	33	13	40	117	313	352	394	1370	40	39	51	15
	Melinda Nicewonger, Kristal Verhulst, Murat Aydin, and Eric Steven Saltzman. 2017. Ethane measurements in air extracted from Greenland and Antarctic ice cores. Arctic Data Center. umauid:/?46cd06a-1563-4981-a1f7-0257be1130eb.	4	25	34	17	24	14	34	84	112	150 Niceraque	98	145	59	50	86	78
	0 🗎 9 12 15 👁	9	20	87	22	42	32	34	40	44	316	96 Ner	Garrent	60	83	114	58
		13	13	37	44	151 150	68	68	69	69	85 83 <sup>Lev</sup>	Co 66 <sup>bia</sup>	386	135	76 40	95 46	11
	Carlo, Michael. 2017. Carloetal_EcolLett_labandfield_rearing_data2014and2015. Dryad Digital Repository. https://doi.org/10.5061/dryad.pr1h0/2?	6	11	7	3	23	42 10	16 3	12 23	21 21	67	80	194 Be	135 mail 49	40 Pi 33	40	23
	0		95	12	10		10	9	8	26	24	85	20	83	44	5	22
	Code Maked A. Diddell Frick Laws Of and Party Michael W. 2017 Bala Arm. Baumant autority and an and an and an and	2	7		184	13	7	13	1	8	8	32	19	15	24	22	12
	Carlo, Michael A., Riddell, Eric A., Levy, Ofir, and Sears, Michael W 2017. Data from: Recurrent sublethal warming reduces embryonic survival, DRYAD inhibits juvenile growth, and alters species distribution projections under climate change. Dryad Digital Repository.	4	12		11		2	2	4		3	6	Paragua 10	98	14	24	8
	https://doi.org/10.5061/dryad.pr1h0?ver=2017-11-21T17:02:39.643-05:00.	Satellite	Terrain	7				6		2	12	20	2	** 49	18	25	23
	0	Google										N.	Unip	party for the second se			

### Federated Search Across Repositories

DataSNE		Search 😧																	
	pate Resources Education Data	Search phrase Q																	
DATAONE SEARCH: Search	h Summary Jump to: DOI or ID Go																		Sign up
earch @	0			Hide Map															
Search phrase Q	- Fabes, Jez, Seligman, William, Barrett, Carolyn, McKech	My Search	tion of a novel																
liter by:	units in the United Kingdom. Dryad Digital Repository. https://doi	-	two intensive care																
III Data attribute	•	Creator: Peterson	×	3	12	34	253	143	6	6	7	2	8	66	7	13	16	3	4
tensity, length, etc. Q	Park, Hyeree, Cho, Sooyoung, Woo, Hyeongtaek, Park,		k of colorectal cancer																
Data files	DRYAD in the Korean Multi-center Cancer Cohort. Dryad Digi	Filter by:	54-05:00.		24	342	4001	114	24	1		15	16	13	21	26	201	2	
Member Node		-		13.00	24	342	4091		2	1	å	13	10	13	21	1.20	201	J.H.S.	
a Creator	Jason Schatz, Carly Ziter, and Christopher Kucharik. 20 Madison, Wisconsin from 2012 -2016. LTER Network	▼	s in and around 12.																
ime Q	• 🖻 V		_	4	175	685	2738	267	6	19 NT	2	10	5	12	63	139	63	32	21
M Year	Heather Bateman, Dan Childers, Madhusudan Katti, Eya	density, length, etc.	Q -term monitoring of	14	379	345	959	605	102	7 0	ada 14	12	4	6	13	164	30	42	4
Identifier	https://pasta.lternet.edu/package/metadata/eml/knb-lter-cap/46			54	57	185	205	302	277	17	270	93	12	- 18	40	164	129	89	
🚠 Taxon	Image: A state of the state			04	57		200	302	211	No.	270	123	ON	00 0	40	104	127	07	
@ Location	knb Peter Bergholz. 2017. Escherichia coli isolated from s	<ul> <li>Data files</li> <li>Member Node</li> </ul>	mplexity. knb.92105.1.	+ 14	51	12	69	114	2081	1140	437	5249	1588	90	73	219	192	83	6
	V 1 ·			- 1	31	37	62	63	9773	5027	3862	st 1009	3894	11912	401	202	280	71	50
	Márcia Cristina Mendes Marques, Anamaria Cequinel, E		da regeneração	17	6	52	57	124	537	40838	12633	2951 M	3753	8644	508	166	194	228	20
	PPBio naturaldo estrato arbóreo em um gradiente edáfico-s Biodiversidade (PPBio). PPBioMA.31.3.	Creator	na de Pesquisa em	4	35	35	53	149	247	4987	17360	<sup>22</sup> 1378,	6787	1053	464	89	112	149	17
	Biodiversidade (PPBio), PPBioMA.31.3.			10	79	265	122	240	26	66	89 Mex	590	8039	1172	95	92	213	107	13
		Name	Q om Greenland and	2	44	1080	719	33	13	40	117	313	352		= 1370	40	39	51	15
	Melinda Nibewonger, Kristal Verhulst, Murat Aydin, and I Antarctic Ice cores. Arctic Data Center. umauid;746od0	Name	Greenland and	4	25	34	17	24	14	34	84	112	150 Miceregue 316	98 96.Net	145	59 60	50 83	86 114	78
	0 🖻 9 ½ 15 ®			13	13	37	44	151	68	68	69	69	85	Col66bia	125	88	76	95	11
	Carlo, Michael. 2017. Carloetal_EcolLett_labandfield_	Year	61/dryad.pr1h0/2?	11	2	17	16	150	42	16	12	21	83	38	386	135	40	46	51
	DRYAD ver=2017-11-21T17:02:43.511 05:00.		-	6	11	7	3	23	10	3	23	21	67	80	194 в	nzil 49	* 33	19	23
	•	s ldentifier			95	12	10			9	8	26	24	85	20	83	44	5	22
	Carlo, Michael A., Riddell, Eric A., Lvy, Ofir, and Sears, Inhibits juvenile growth, and alters species distribut		embryonic survival,		7		184		7	13	1	8	8	32	19 Parao	15	24	22	12
	https://doi.org/10.5061/dryad.pr1h0?ver=2017-11-21T1772:39.643-05:00	Taxon		4 Satellite	12 Terrain		11		2	2	4		3	6 Chile	10	98	14	24	8
	•			Google		7				6		2	12	20	2	49	18	25	23
		Subscription	ce Foundation (NSF) under a Coop rdings, and conclusions or recomm	rathen Accessmen	6 t. Acknowledge	A ment: This met	erial is based a	pon work supp	arted by the				1	1 Map	data ©2017 Go	ogle, INEGI	1000 km 📖		erms of U

7

### Metadata Quality Improvement

### Improve Metadata Quality at Level of Collection and Data Set

NSF	
	Data Support About Submit Data Data Sign in wit
Metadata Quality report	for <i>doi:10.18739/A2G62B</i>
	ge against our standard set of metadata, data, and congruency checks, we have found the following potential issues. I ability and reusability of your research data by addressing the issues below.
	Suite: arctic.data.center.suite.1
	Identification: 58% complete
23	Discovery: 80% complete
checks	Metadata: 0% complete
	Interpretation: 67% complete
Passed 12 checks out of 1	18. Good job!
A Warning for 3 checks. Ple	base review these warnings.
8 Failed 3 checks. Please co	orrect these issues.

8



The HDF Group

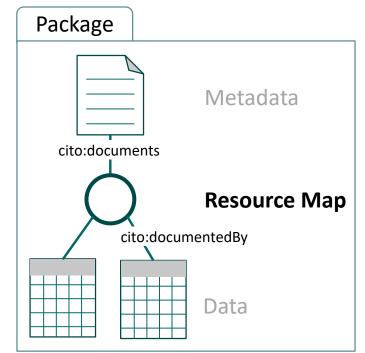
# Data Package

OAI-ORE Binding Graph



Open Archives Initiative, Object Reuse and Embedding

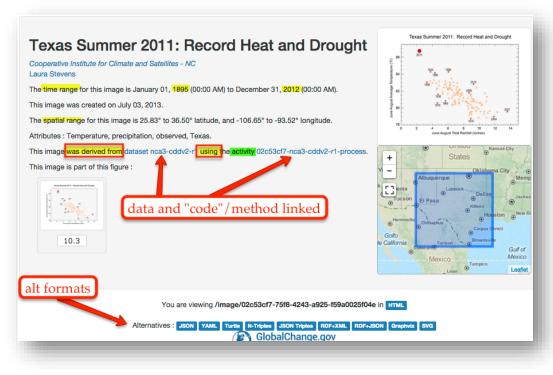
- Aggregation of components
- **Citation Ontology** 
  - Component Relationships
- Extensible for
  - Provenance metadata
  - Semantic annotations



# **Computational Provenance**

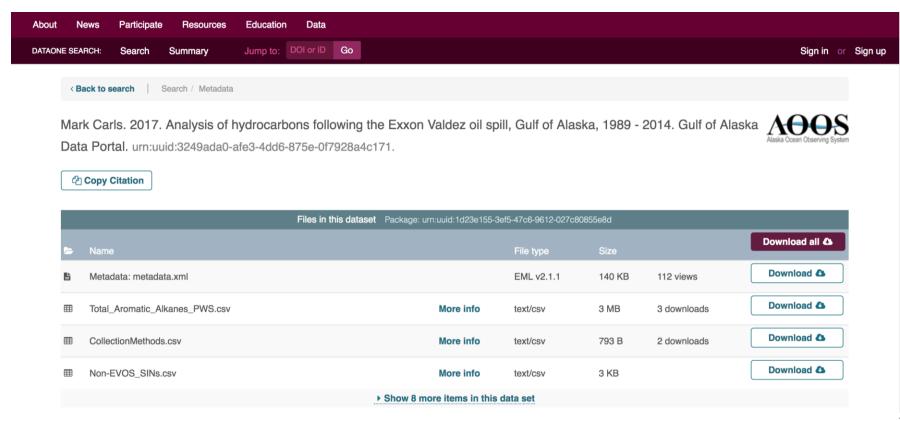
Origin, processing history of artifacts

- data products, figures
- workflow/script evolution
- understand methods, dataflow, and dependencies



# Provenance Display in DataONE

### Search User Interface



#### Data Table, Image, and Other Data Details

4 sources	Data Table			2 derivations					
	Entity Name	Total_Aromatic_Alkanes_PWS.csv							
	Download 🔹								
	Description								
	Object Name	Object Name Total_Aromatic_Alkanes_PWS.csv							
	Online Distribution Info	on Info https://cn.dataone.org/cn/v2/resolve/urn:uuid:44108e76-405d-4d58-b1b3-fb4b55e3fff9							
	Size	2801033 byte							
	Text Format	Number of Header Lines	1						
		Record Delimiter	#x0A						
		Attribute Orientation	column						
		Simple Text							
		Field Delimeter	3						
	Number Of Records	12142							

#### Data Table, Image, and Other Data Details

4 sources	Source Pro	gram			2 derivations
	Total_PAH_and_Alk drocarbons_Clean.F	1	kanes_PWS.csv		
	This program generated currently viewing, III Total_Aromatic_Alkane This program used III PA Sample.csv, III Non-E and (and 1 more I).	es_PWS.csv.	om PAH, Alkane and Sample tables documenting sa Il in Prince William Sound, AK anes_PWS.csv org/cn/v2/resolve/urn:uuid:44108e76-405d-4d58-b		
	Text Format	Number of Head		1 #x0A	
		Attribute Oriental		column	
		Simple Text			
		Field Delimeter		,	
	Number Of Records	12142			

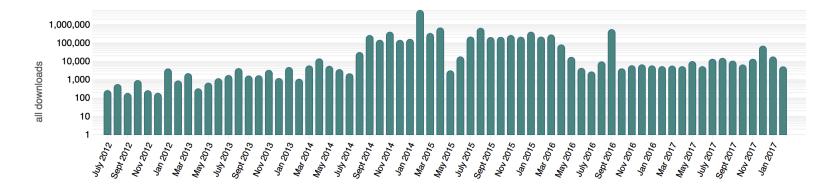
# **Event Tracking**

### Object events aggregated by Coordinating Nodes



#### **Downloads**

The number of individual metadata and data files downloaded over time. These download counts are partially **COUNTER** compliant, meaning that downloads from some Internet robots and repeat downloads within a certain time window are excluded.



# Making Data Count

2015

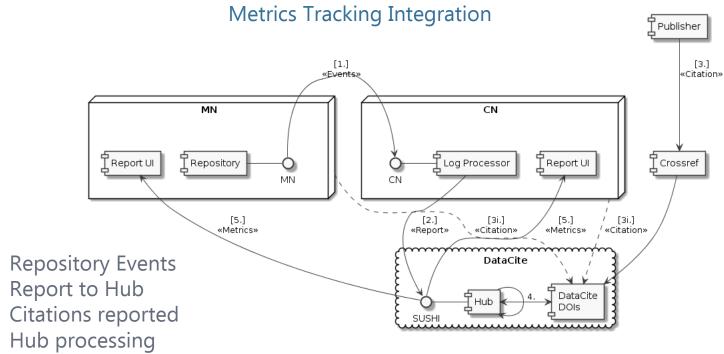
Majorities of both researchers and data managers ranked **landing page views** as the **least interesting metric**. Researchers consistently ranked **downloads as the second most interesting**, while data managers (who were given a longer list to rank) put them in the middle. Resoundingly, both researchers and data managers measure scholarly prestige in citations; 85% of researchers and 61% of data managers **ranked citations as the most interesting metric**. This ordering-viewing, downloading, citing-is consistent with the results our previous researcher survey.

Kratz, J. & Strasser, C. doi:10.1038/sdata.2015.39





### Make Data Count



5. UI display

1.

2.

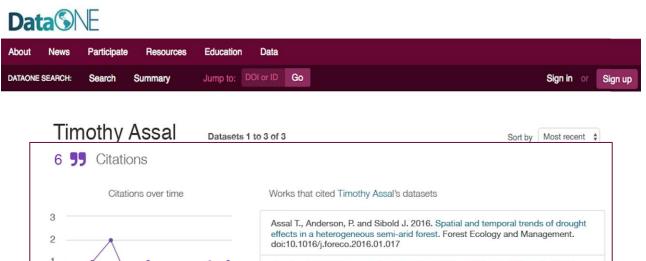
3.

4.



### Make Data Count

### **Data Level Metrics**



0

Tim Assal. 2016. Documenting Spatial and Temporal Trends of Drought. NASA.gov

Smith, Joe. 2016. Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque audantium. Ecology. doi:1234/123JK

Douglas et al. 2016. At vero eos et accusamus et iusto odio dignissimos ducimus qui blanditiis praesentium voluptatum deleniti atque corrupti. Nature. doi:1J78/397HF9

See 2 more



Make Data Count

**Reference Material** 

doi.org/10.7287/peerj.preprints.26505v1 Code of Practice: http://bit.ly/DU-COP-v1

rd-alliance.org/groups/data-usage-metrics-wg RDA Group: Project: makedatacount.org

@MakeDataCount









# DataONE

In Summary...

Member Nodes provide content Uniquely identified, immutable Indexed by Coordinating Nodes Common model for discovery W3C-PROV documents products Semantic annotations for reuse

