



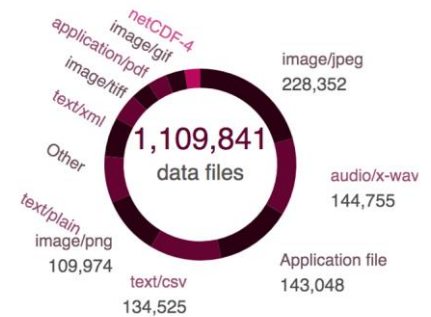
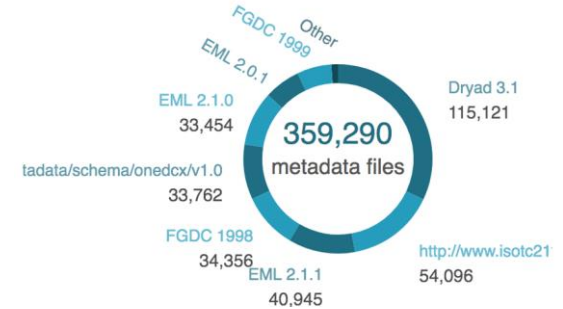
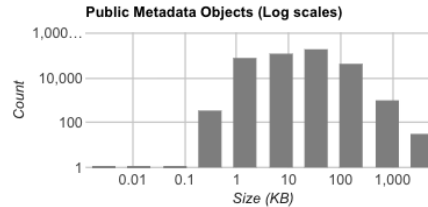
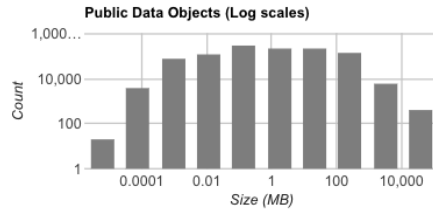
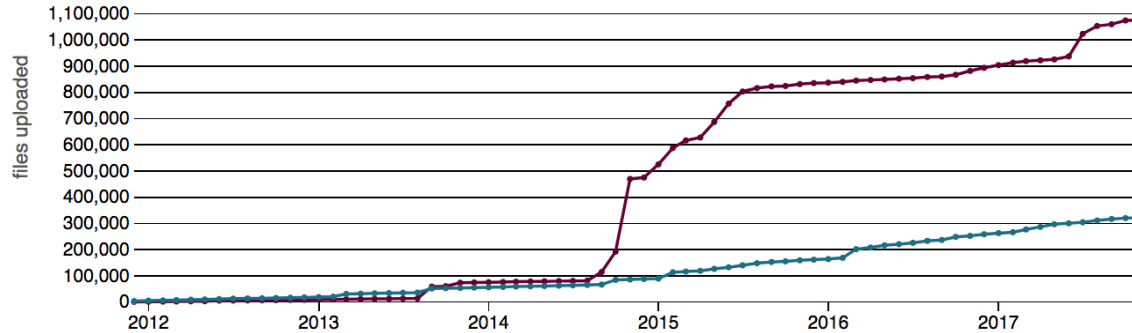
BCoN Data Integration and Attribution
February 2018

Dave Viegla
University of Kansas



Grant Numbers
0830944 and
1430508

Production in 2012, Diverse Content



Data Package

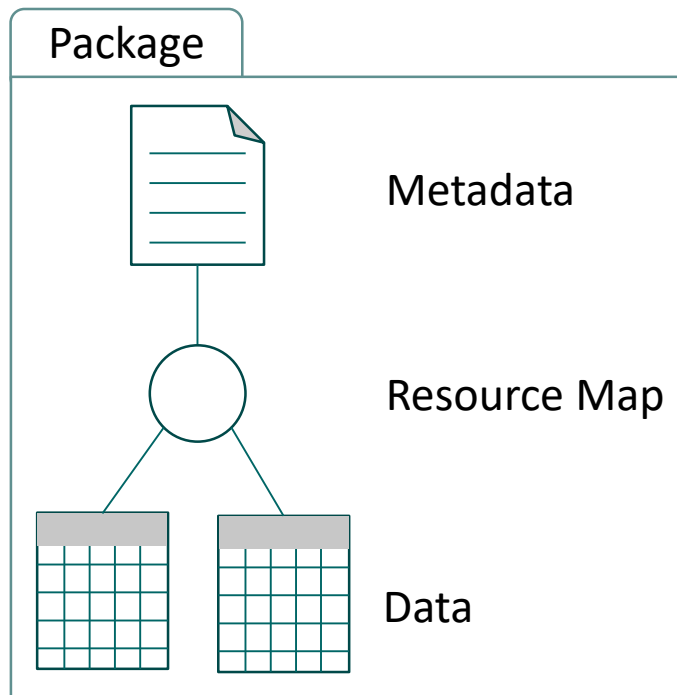
Flexible Structure for Data

Data Package components:

- Metadata
- Data
- Resource Map binding components

All components:

- Immutable
- Uniquely identified
- Resolvable
- Retrievable

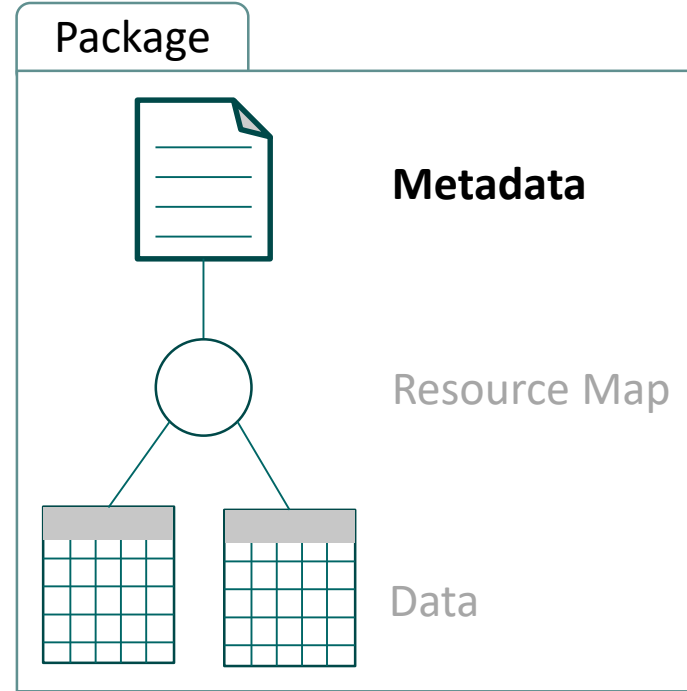
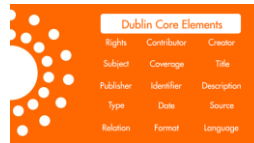


Data Package

Metadata

Many Standards

- Dublin Core
- FGDC
- Dryad Metadata Profile
- Ecological Metadata Language
- ISO 19000 family
- ...



6

Federated Search Across Repositories

DataONE

About News Participate Resources Education Data

DATONE SEARCH: Search Summary Jump to: DOI or ID Go

Search ?

Search phrase

Filter by:

- Data attribute
 - density, length, etc.
- Data files
- Member Node
- Creator
 - Name
- Year
- Identifier
- Taxon
- Location

My Search

Creator: Peterson ✕

Filter by:

- Data attribute
 - density, length, etc.
- Data files
- Member Node
- Creator
 - Name
- Year
- Identifier
- Taxon
- Location

Search Results:

- DRYAD** Fabes, Jez, Seligman, William, Barrett, Carolyn, McKee. Intensive care discharge risk score and nurse-led units in the United Kingdom. Dryad Digital Repository. <https://doi.org/10.5061/dryad.pr1h027>
- DRYAD** Park, Hyeree, Cho, Sooyoung, Woo, Hyeontaek, Park, In. The Korean Multi-center Cancer Cohort. Dryad Digital Repository. <https://doi.org/10.5061/dryad.pr1h027>
- LTER** Jason Schatz, Carly Ziter, and Christopher Kucharik. 2016. Madison, Wisconsin from 2012-2016. LTER Network. <https://lter.psu.edu/package/metadata/eml/krib-ter-cap4>
- LTER** Heather Bateman, Dan Childers, Madhusudan Katti, Eye bird abundance and diversity in central Arizona-Phoenix. <https://pasta.lternet.edu/package/metadata/eml/krib-ter-cap4>
- knb** Peter Bergholz. 2017. Escherichia coli isolated from a... <https://knb.berkeley.edu/package/metadata/eml/krib-ter-cap4>
- PPBio** Márcia Cristina Mendes Marques, Anamaria Ceguelin, B... naturaldo estrato arbóreo em um gradiente edáfico-Biodiversidade (PPBio). PPBio/MA.31.3
- Antarctic Data Center** Melinda Neevonger, Kristal Verhulst, Murat Aydin, and f... Antarctic ice cores. Arctic Data Center. <https://arcticdatacenter.org/package/metadata/eml/krib-ter-cap4>
- DRYAD** Carlo, Michael. 2017. Coriaria EcoLett. Isabandfield, ver=2017-11-21T17:02:43.51005.00.
- DRYAD** Carlo, Michael A., Riddell, Eric A., Levy, Ofir, and Sears... inhibits juvenile growth, and alters species distribut... <https://doi.org/10.5061/dryad.pr1h027>

Map

Hide Map >

3 12 34 253 143 6 6 7 2 8 66 7 13 16 3 43

1 24 342 4891 114 24 11 6 15 16 13 21 26 201 2 34

4 175 685 2738 267 6 19 2 10 5 12 63 139 63 32 213

14 379 345 959 605 102 7 14 12 4 6 13 164 30 42 40

54 57 185 205 302 277 17 270 93 12 18 40 164 129 89 48

14 51 12 69 114 2081 1140 437 5249 1588 90 73 219 192 83 65

1 31 37 62 63 9773 5027 3862 1009 3894 11912 401 202 280 71 56

17 6 52 57 124 537 40838 12633 2951 3753 8644 508 166 194 228 201

4 35 35 53 149 247 4987 17360 1378 6787 1053 464 89 112 149 173

10 79 265 122 240 26 66 89 590 8039 1172 95 92 213 107 138

2 44 1080 719 33 13 40 117 313 352 394 1370 40 39 51 156

4 25 34 17 24 14 34 84 112 150 98 145 59 50 86 78

9 20 87 22 42 32 34 40 44 316 96 194 830 60 83 114 58

13 13 37 44 151 68 68 69 69 85 66 125 88 76 95 112

11 2 17 16 150 42 16 12 21 83 38 386 135 40 46 51

6 11 7 3 23 10 3 23 21 67 80 194 49 33 19 23

95 12 10 9 8 26 24 85 20 83 44 5 22

2 7 184 13 7 13 1 8 8 32 19 15 24 22 12

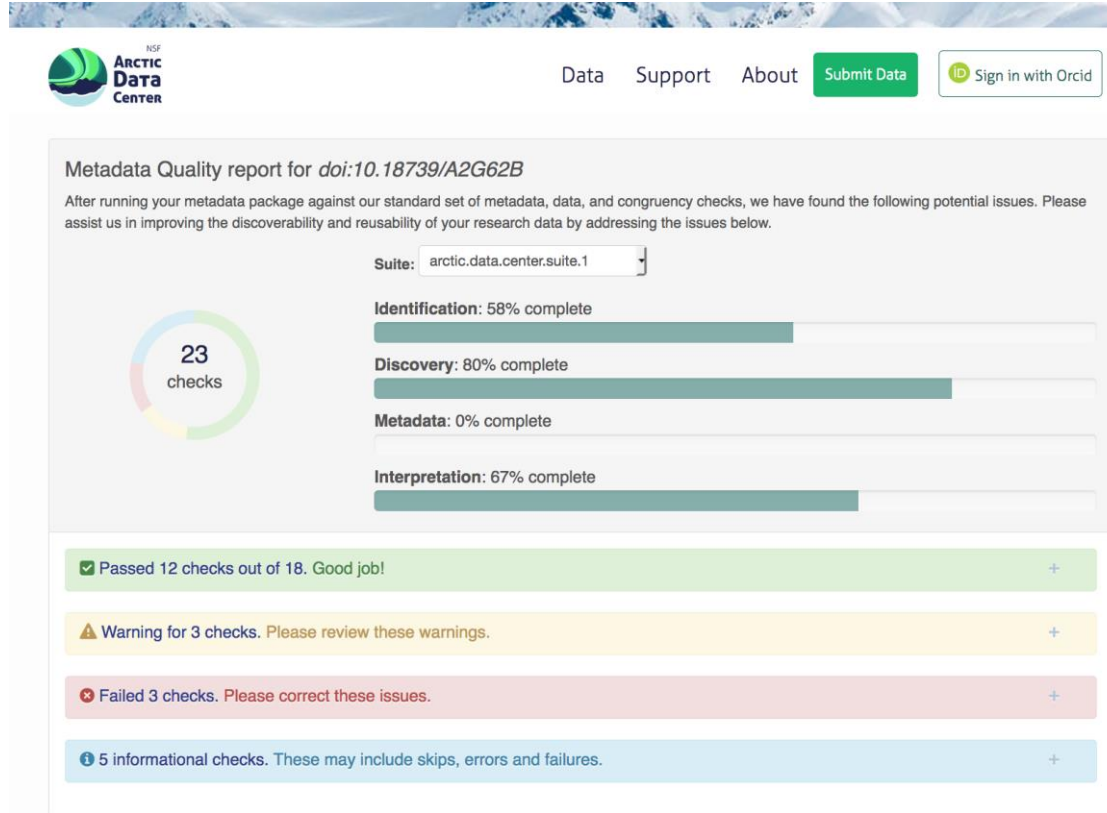
4 12 11 2 2 4 3 6 10 98 14 24 8

7 6 1

Map data ©2017 Google, INEGI 1000 km Terms of Use

Metadata Quality Improvement

Improve Metadata Quality at Level of Collection and Data Set



Data Package

OAI-ORE Binding Graph



Open Archives Initiative,
Object Reuse and Embedding

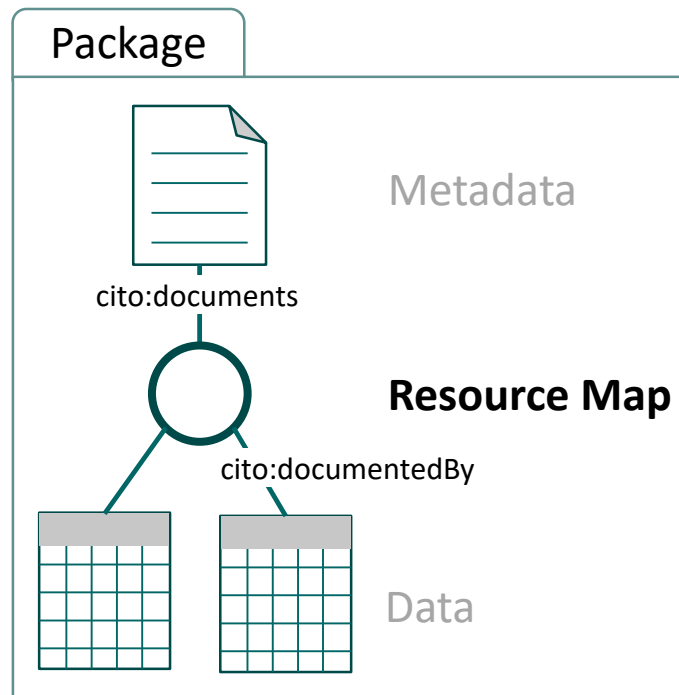
- Aggregation of components

Citation Ontology

- Component Relationships

Extensible for

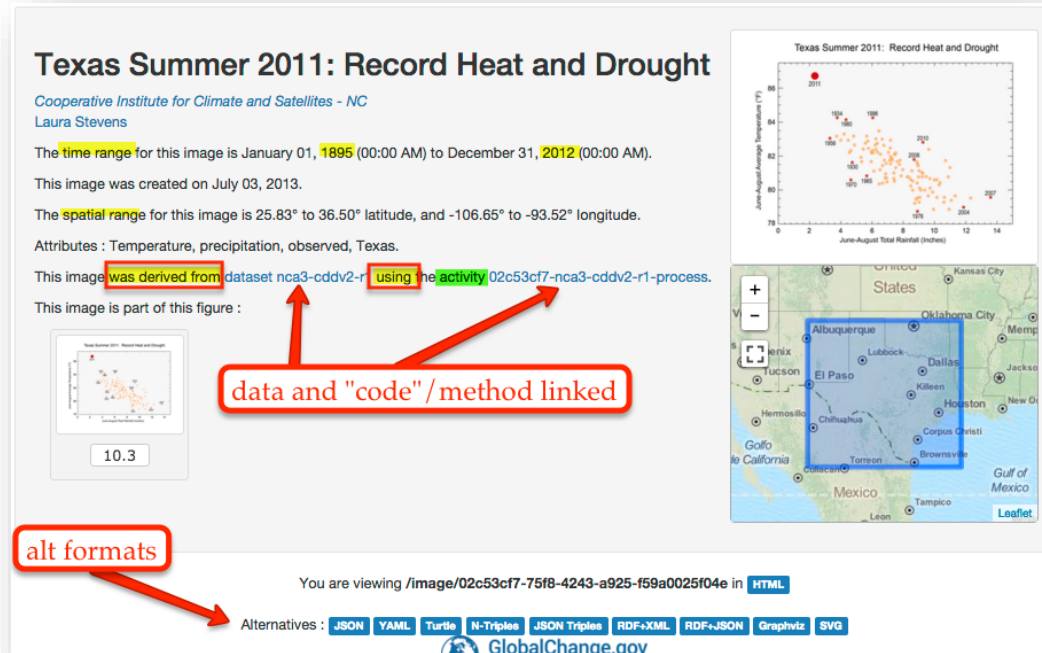
- Provenance metadata
- Semantic annotations



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

[About](#) [News](#) [Participate](#) [Resources](#) [Education](#) [Data](#)

DATAONE SEARCH:

[Search](#) [Summary](#)

Jump to: [DOI or ID](#) [Go](#)











[Sign in](#) or [Sign up](#)

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

Mark Carls. 2017. Analysis of hydrocarbons following the Exxon Valdez oil spill, Gulf of Alaska, 1989 - 2014. Gulf of Alaska Data Portal. urn:uuid:3249ada0-afe3-4dd6-875e-0f7928a4c171.

[Copy Citation](#)

Files in this dataset Package: urn:uuid:1d23e155-3ef5-47c6-9612-027c80855e8d

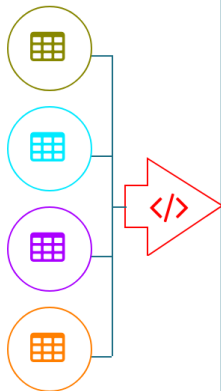
 Name	File type	Size	Download all 
 Metadata: metadata.xml	EML v2.1.1	140 KB 112 views	Download 
 Total_Aromatic_Alkanes_PWS.csv	More info text/csv	3 MB 3 downloads	Download 
 CollectionMethods.csv	More info text/csv	793 B 2 downloads	Download 
 Non-EVOS_SINs.csv	More info text/csv	3 KB	Download 

[► Show 8 more items in this data set](#)



Data Table, Image, and Other Data Details

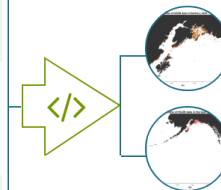
4 sources



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	<table><tr><td>Number of Header Lines</td><td>1</td></tr><tr><td>Record Delimiter</td><td>#x0A</td></tr><tr><td>Attribute Orientation</td><td>column</td></tr><tr><td colspan="2">Simple Text</td></tr><tr><td>Field Delimiter</td><td>,</td></tr></table>	Number of Header Lines	1	Record Delimiter	#x0A	Attribute Orientation	column	Simple Text		Field Delimiter	,
Number of Header Lines	1										
Record Delimiter	#x0A										
Attribute Orientation	column										
Simple Text											
Field Delimiter	,										
Number Of Records	12142										

2 derivations



Data Table, Image, and Other Data Details

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).

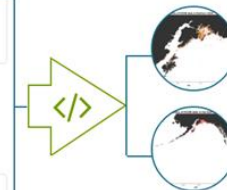
Alkanes_PWS.csv

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

Alkanes_PWS.csv

<https://datacatalog.doe.noaa.gov/org/cn/v2/resolve/urn:uuid:44108e76-405d-4d58-b1b3-fb4b55e3fff9>

2 derivations



Text Format

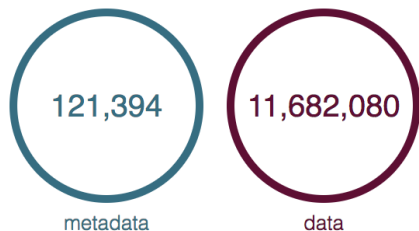
Number of Header Lines	1
Record Delimiter	#x0A
Attribute Orientation	column
Simple Text	
Field Delimiter	,

Number Of Records

12142

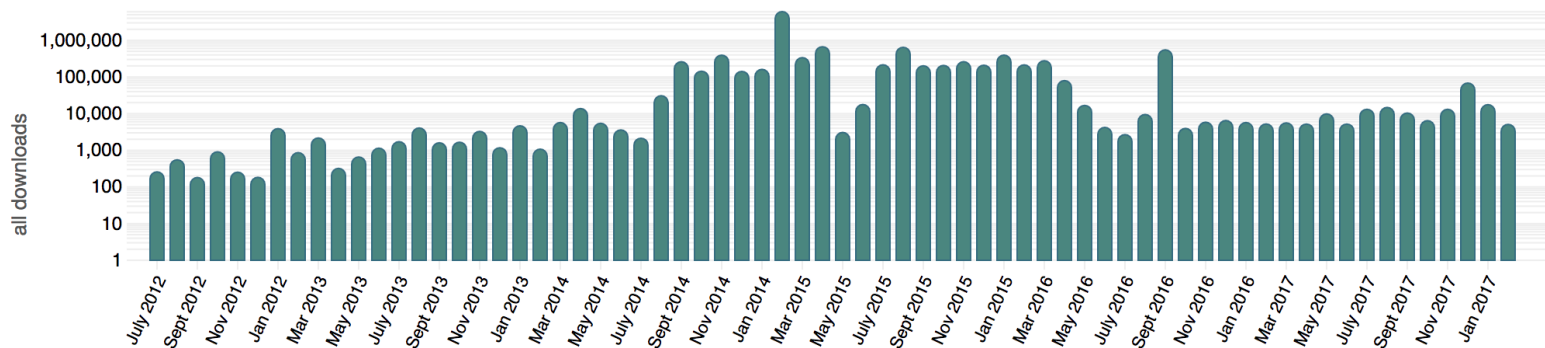
Event Tracking

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

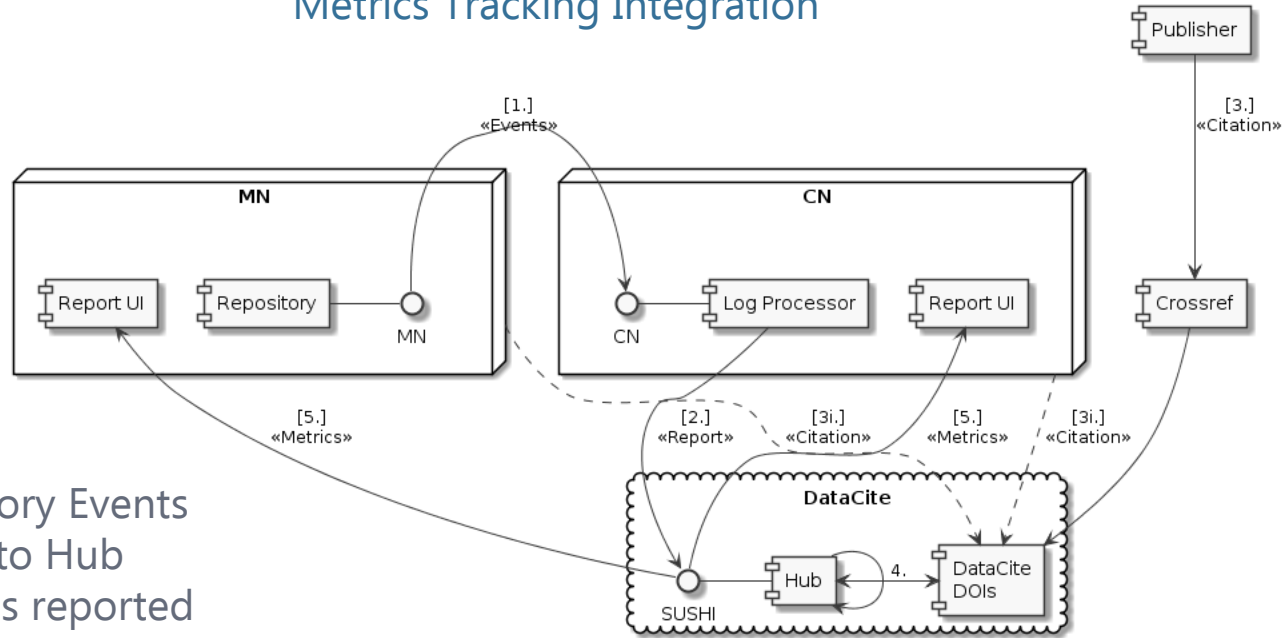
2015

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

Metrics Tracking Integration



1. Repository Events
2. Report to Hub
3. Citations reported
4. Hub processing
5. UI display

Make Data Count

Data Level Metrics



About News Participate Resources Education Data

DATAONE SEARCH:

Search

Summary

Jump to:

DOI or ID

Go

Sign in

or

Sign up

Timothy Assal

Datasets 1 to 3 of 3

Sort by Most recent

6 Citations

Citations over time



Works that cited Timothy Assal's datasets

Assal T., Anderson, P. and Sibold J. 2016. [Spatial and temporal trends of drought effects in a heterogeneous semi-arid forest](#). Forest Ecology and Management. doi:10.1016/j.foreco.2016.01.017

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

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

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

Project: makedatacount.org
[@MakeDataCount](#)



ALFRED P. SLOAN
FOUNDATION



University of California
CDL
California Digital Library



DataCite
FIND, ACCESS, AND REUSE DATA

DataONE

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

