

Making Data Count

DataCite/EPIC PID Meeting

Make Data Count Overview

Partners

California Digital Library, PLOS, DataONE

NSF Grant Record

Grant No. 1448821 proposal (eScholarship repository)

Project page mdc.lagotto.io

Prototype dlm.labs.datacite.org

Software github.com/articlemetrics/lagotto



Making Data Count Project

Field Research

Usage activity

Data metrics aggregation

Reporting & visualizations

Bibliometric analysis

Aims & Goals of MDC

- prototype metrics as a first step to actually collecting data, learn from the data
- connect with existing efforts: RDA, CASRAI, NISO
- leverage existing software (DataONE, Lagotto)

How do scholars access, cite, share, mention, recommend data?

Usage

PLOS Journals
(HTML, PDF, XML)

PubMed Central
(HTML, PDF)

figshare

Citations

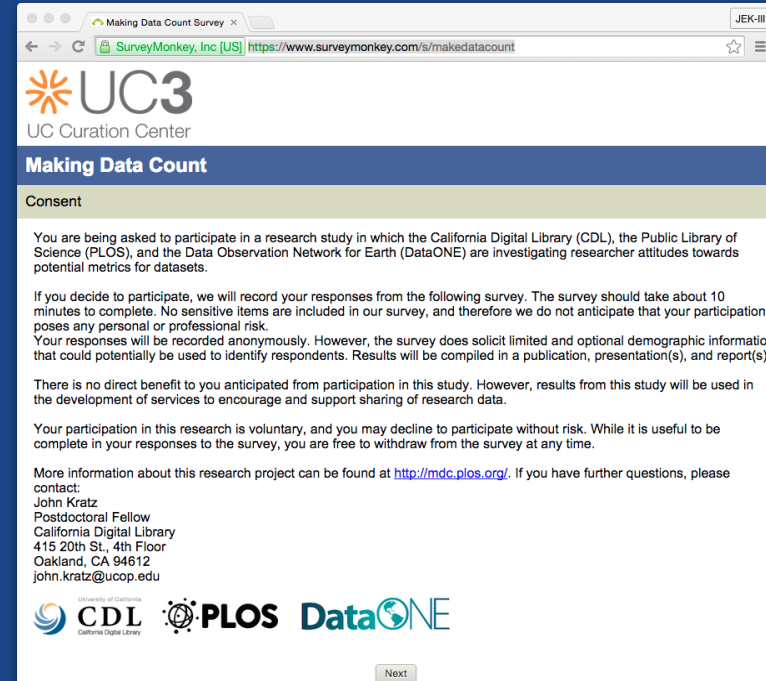
CrossRef
Scopus
Web of Science
PubMed Central
PMC Europe
PMC Europe
database Citations
DataCite

Altmetrics

PLOS Comments
F1000Prime
Mendeley
CiteULike
ResearchBlogging
ScienceSeeker
Facebook
Twitter
Wikipedia
Reddit

Field research

- Literature Review
- Focus Groups
 - DataONE All Hands Meeting
 - Society for Neuroscience
 - American Geophysical Union
- Online Surveys
 - Researchers
 - Data managers



The screenshot shows a web browser window with the URL <https://www.surveymonkey.com/s/makedatacount>. The page header includes the UC3 logo and 'UC Curation Center'. The main heading is 'Making Data Count' and the section is 'Consent'. The text explains that participants are being asked to take a 10-minute survey for a research project involving CDL, PLOS, and DataONE. It states that participation is voluntary, responses are anonymous, and there is no direct benefit. Contact information for John Kratz is provided. Logos for CDL, PLOS, and DataONE are at the bottom.

UC3
UC Curation Center

Making Data Count

Consent




You are being asked to participate in a research study in which the California Digital Library (CDL), the Public Library of Science (PLOS), and the Data Observation Network for Earth (DataONE) are investigating researcher attitudes towards potential metrics for datasets.

If you decide to participate, we will record your responses from the following survey. The survey should take about 10 minutes to complete. No sensitive items are included in our survey, and therefore we do not anticipate that your participation poses any personal or professional risk. Your responses will be recorded anonymously. However, the survey does solicit limited and optional demographic information that could potentially be used to identify respondents. Results will be compiled in a publication, presentation(s), and report(s).

There is no direct benefit to you anticipated from participation in this study. However, results from this study will be used in the development of services to encourage and support sharing of research data.

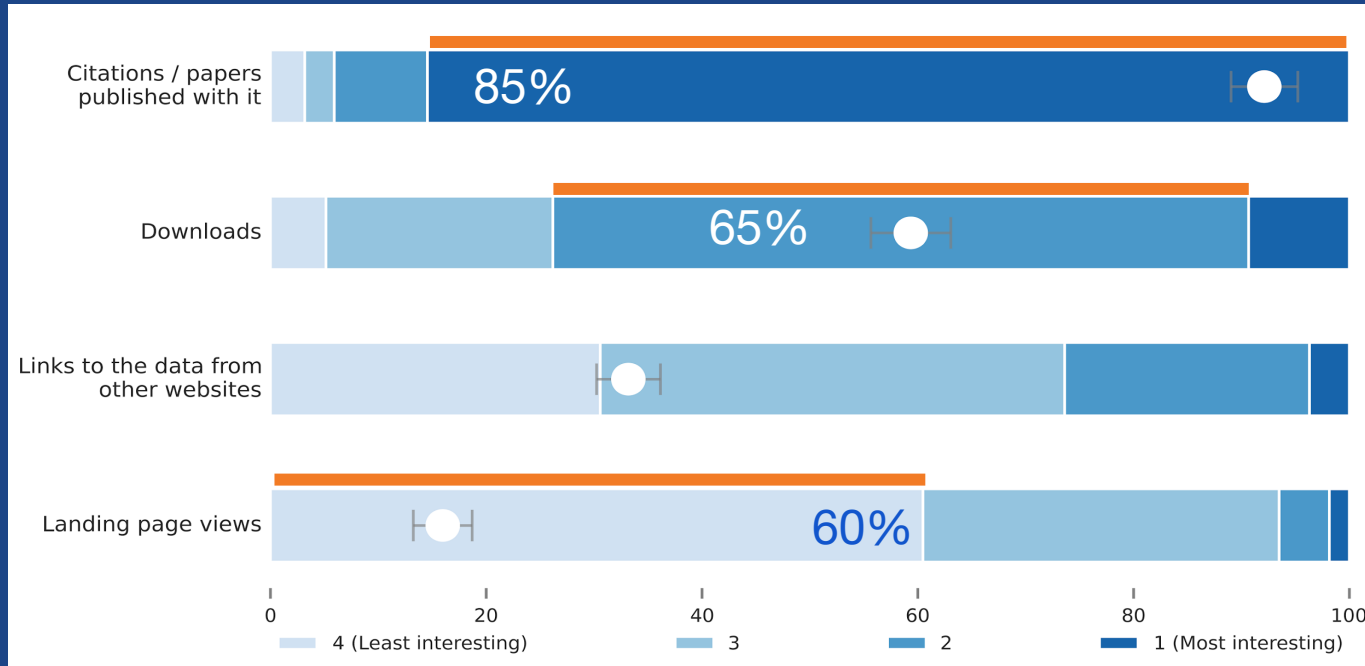
Your participation in this research is voluntary, and you may decline to participate without risk. While it is useful to be complete in your responses to the survey, you are free to withdraw from the survey at any time.

More information about this research project can be found at <http://mdc.plos.org/>. If you have further questions, please contact:
John Kratz
Postdoctoral Fellow
California Digital Library
415 20th St., 4th Floor
Oakland, CA 94612
john.kratz@ucop.edu

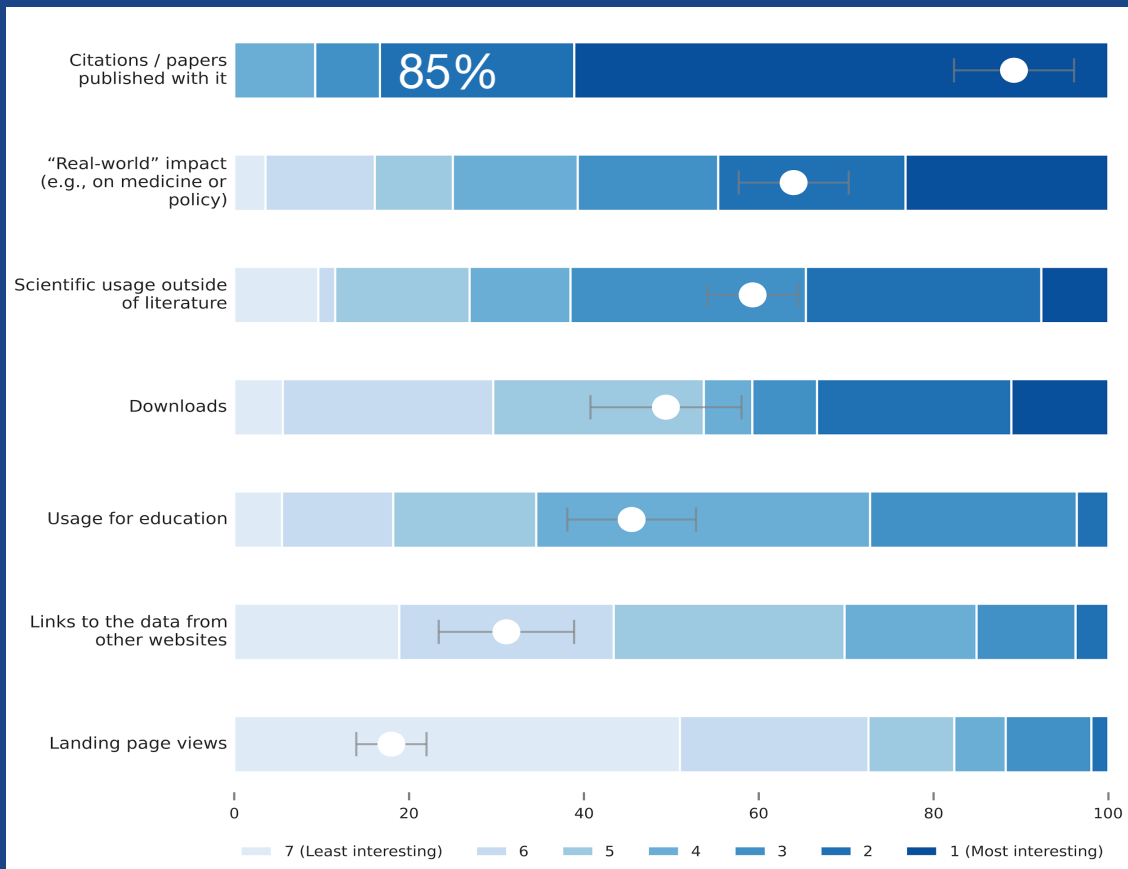
  

Next

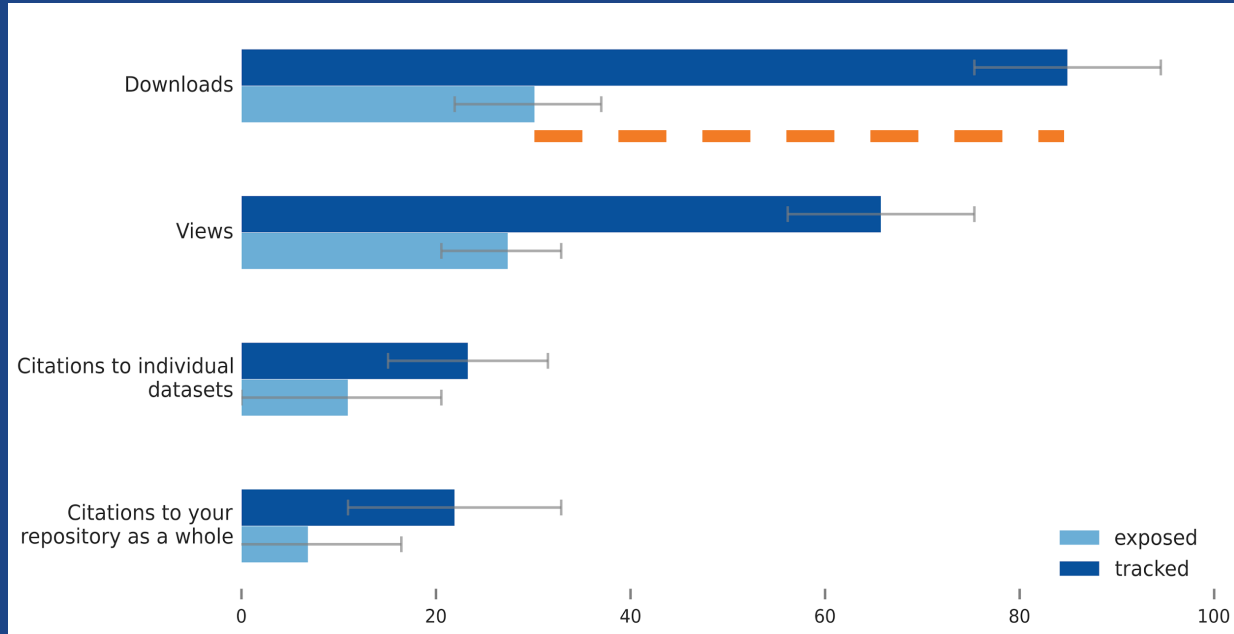
Impact: How interested you would be to know each of the following about the impact of your data?



How interested you would be to know each of the following about the impact of the data you hold?



What metrics / statistics do you currently track / expose? (e.g., on landing pages or via API)



Research findings

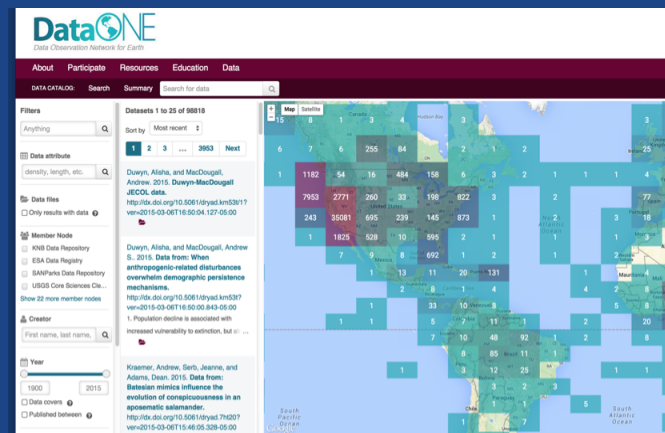
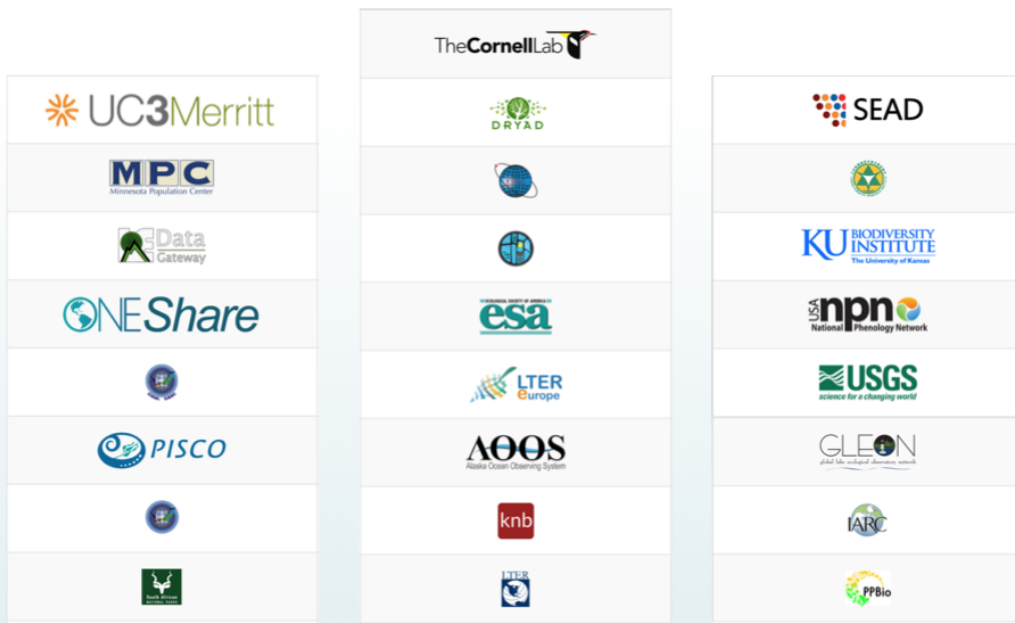
Results in Scientific Data: DOI [10.1038/sdata.2015.39](https://doi.org/10.1038/sdata.2015.39)

- Page views are considered to be highly valued at the moment.
- There is little social media activity to capture.
- Citations are considered to be the most valuable metric.
- Citations suffer from technical and cultural obstacles.
- Downloads are most tractable and valued.

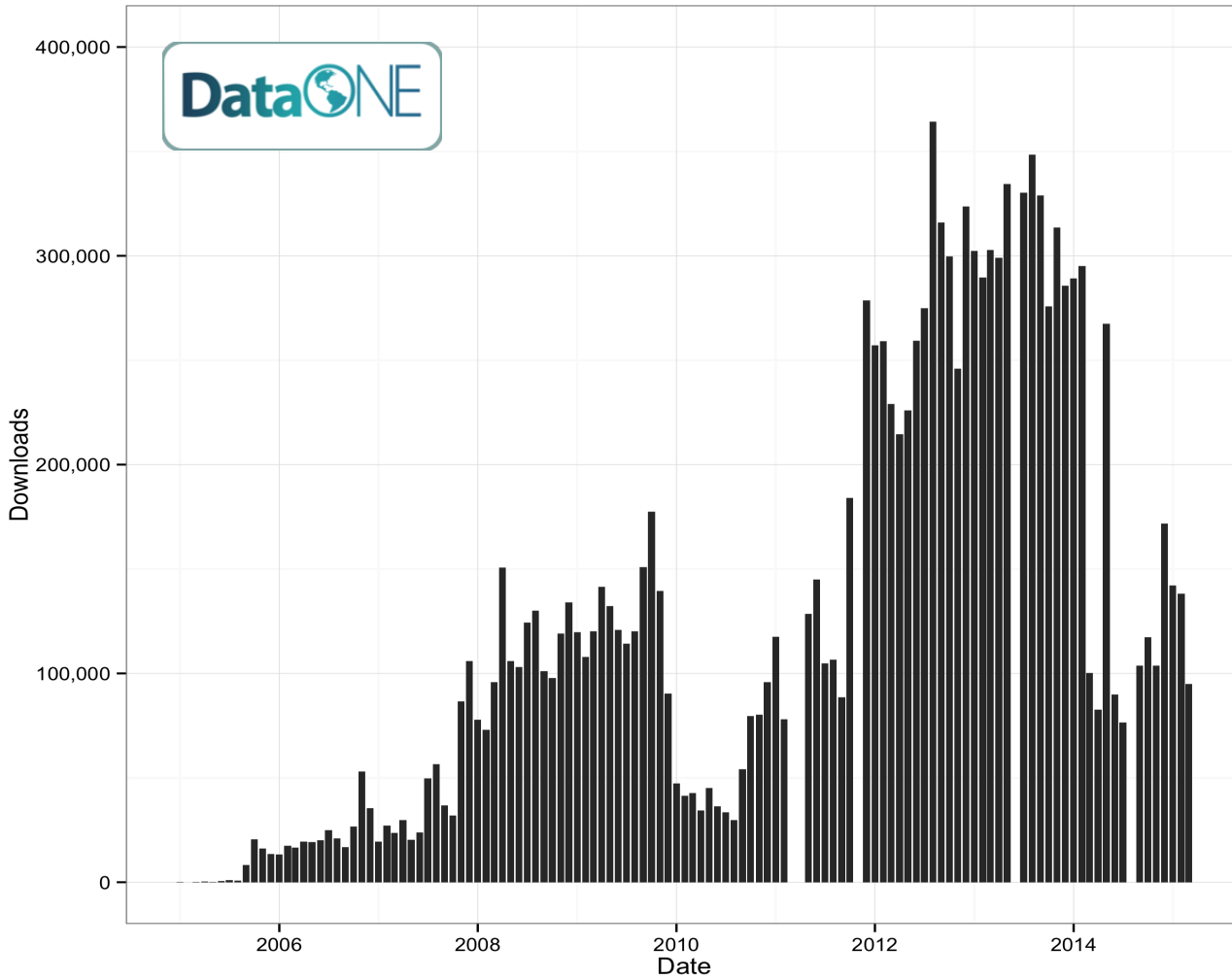
Usage activity

25 DataONE Repository Members

North America (21), Europe (1), Africa (1), Asia (1), South America (1)



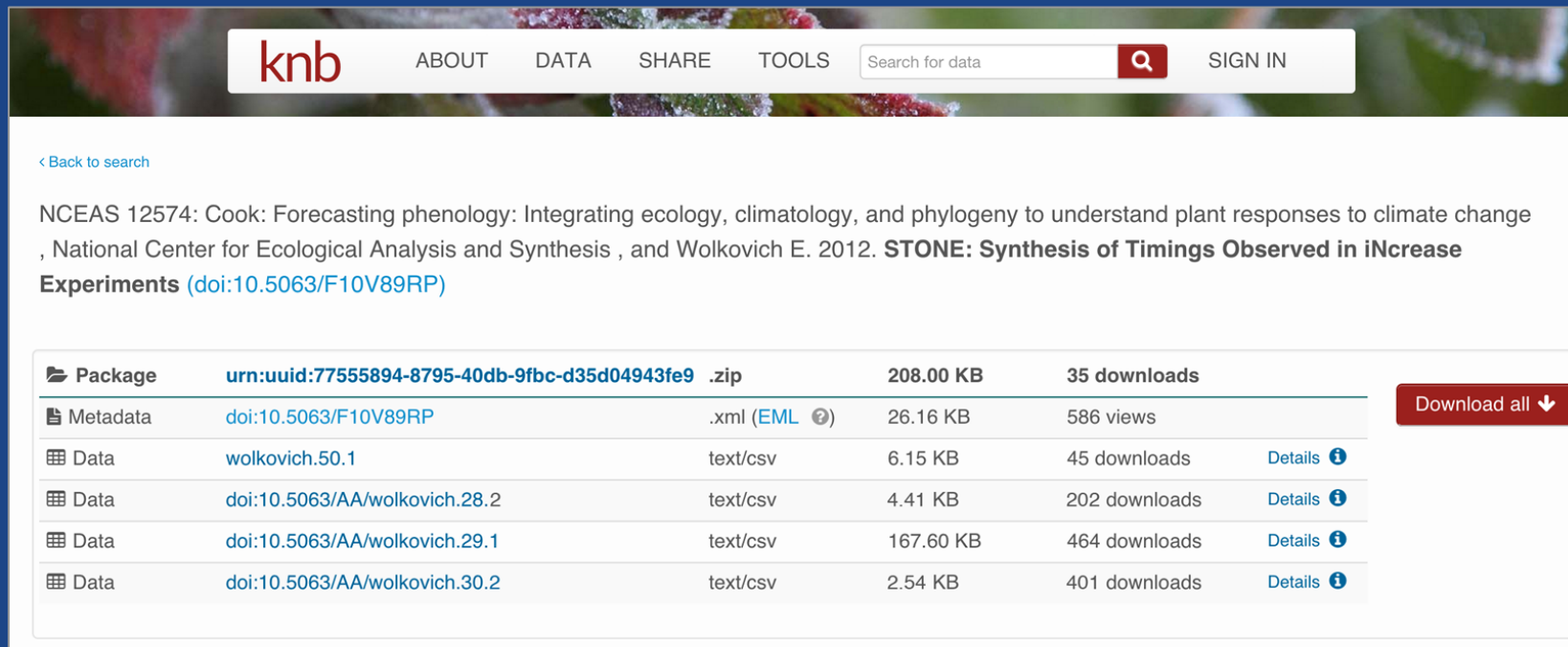
All DataONE Downloads



Data usage tracking: downloads






Independent downloads? Entire package?


- Sum downloads: 1733
- Average downloads: 346
- Maximum downloads: 586
- Whole package: 35



< Back to search

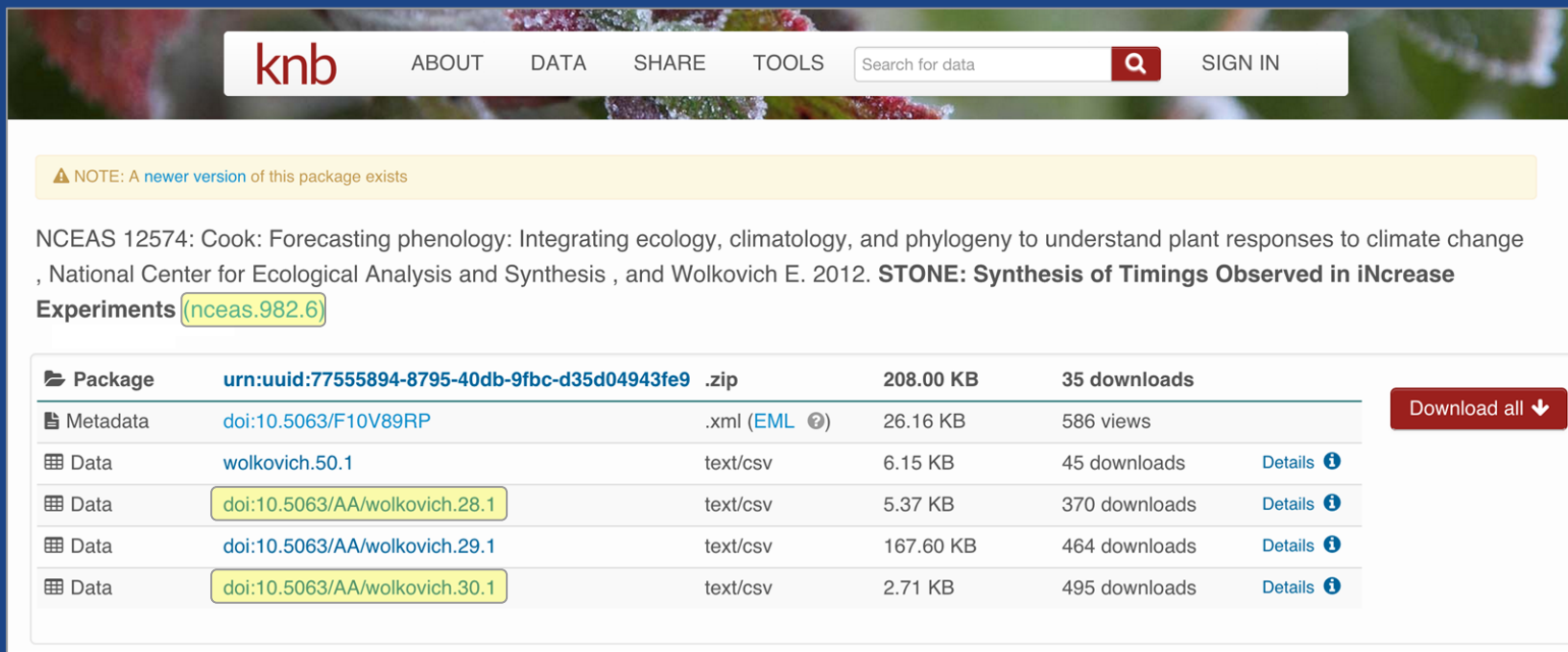
NCEAS 12574: Cook: Forecasting phenology: Integrating ecology, climatology, and phylogeny to understand plant responses to climate change , National Center for Ecological Analysis and Synthesis , and Wolkovich E. 2012. **STONE: Synthesis of Timings Observed in iNcrease Experiments** (doi:10.5063/F10V89RP)

Package	urn:uuid:77555894-8795-40db-9fbc-d35d04943fe9	.zip	208.00 KB	35 downloads	
Metadata	doi:10.5063/F10V89RP	.xml (EML )	26.16 KB	586 views	
Data	wolkovich.50.1	text/csv	6.15 KB	45 downloads	Details 
Data	doi:10.5063/AA/wolkovich.28.2	text/csv	4.41 KB	202 downloads	Details 
Data	doi:10.5063/AA/wolkovich.29.1	text/csv	167.60 KB	464 downloads	Details 
Data	doi:10.5063/AA/wolkovich.30.2	text/csv	2.54 KB	401 downloads	Details 

Download all 

How do we count?

- Sum downloads across all versions?
- Average downloads across versions?
- Only some objects change in a new package



NOTE: A newer version of this package exists

NCEAS 12574: Cook: Forecasting phenology: Integrating ecology, climatology, and phylogeny to understand plant responses to climate change , National Center for Ecological Analysis and Synthesis , and Wolkovich E. 2012. **STONE: Synthesis of Timings Observed in iNcrease Experiments** ([nceas.982.6](#))

Package	urn:uuid:77555894-8795-40db-9fbc-d35d04943fe9	.zip	208.00 KB	35 downloads	
Metadata	doi:10.5063/F10V89RP	.xml (EML ?)	26.16 KB	586 views	
Data	wolkovich.50.1	text/csv	6.15 KB	45 downloads	Details ⓘ
Data	doi:10.5063/AA/wolkovich.28.1	text/csv	5.37 KB	370 downloads	Details ⓘ
Data	doi:10.5063/AA/wolkovich.29.1	text/csv	167.60 KB	464 downloads	Details ⓘ
Data	doi:10.5063/AA/wolkovich.30.1	text/csv	2.71 KB	495 downloads	Details ⓘ

[Download all](#) ↓

Standardize collection

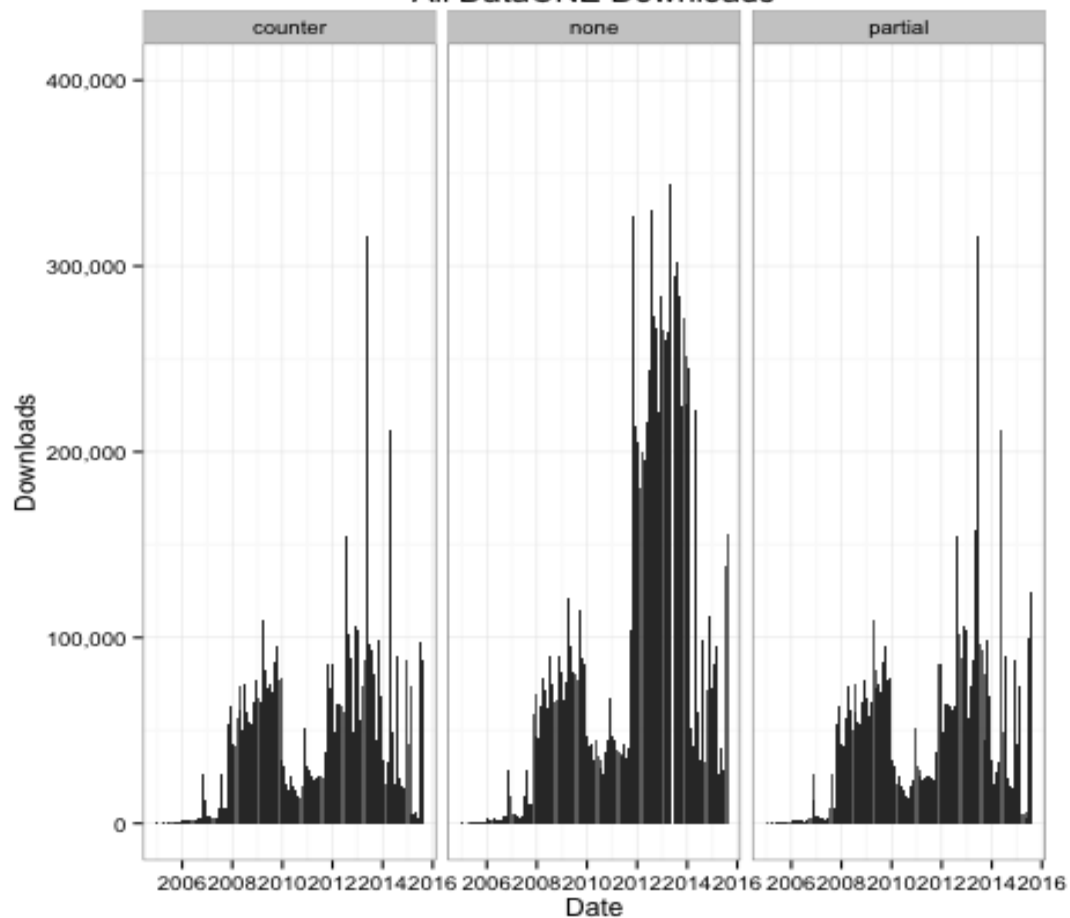


<http://www.projectcounter.org/>

- Standard means of reporting usage of articles
- COUNTER reports remove:
 - Web robots from search engines
 - Repeat visits in short time window (double clicks)
 - All accesses from Python, Java, curl, wget, etc.
 - Scientists frequently use these to access data
- COUNTER issues with composite objects
- COUNTER issues with versioning

We propose changes to COUNTER for data and data package downloads.

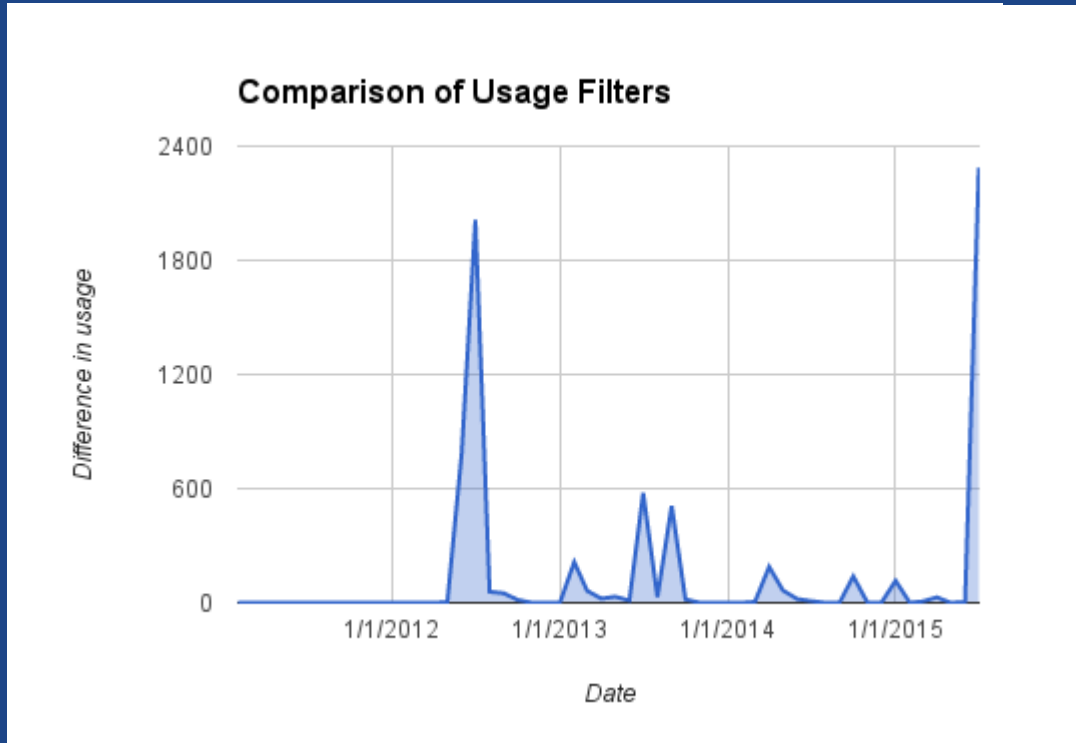
All DataONE Downloads



Usage activity - comparing filters

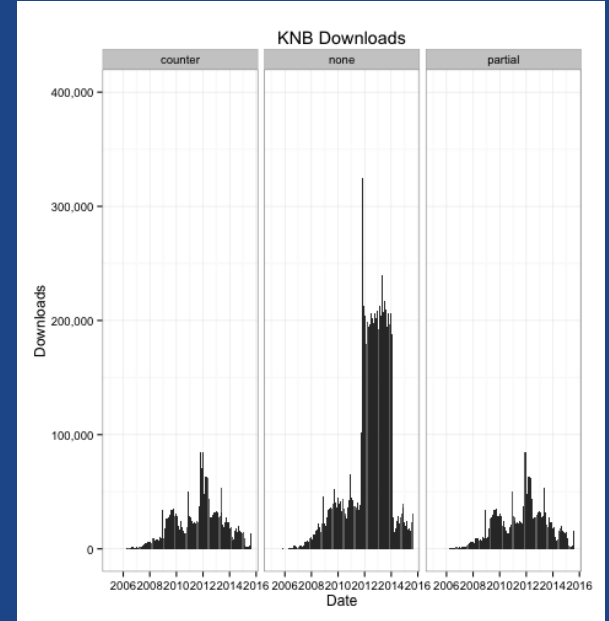
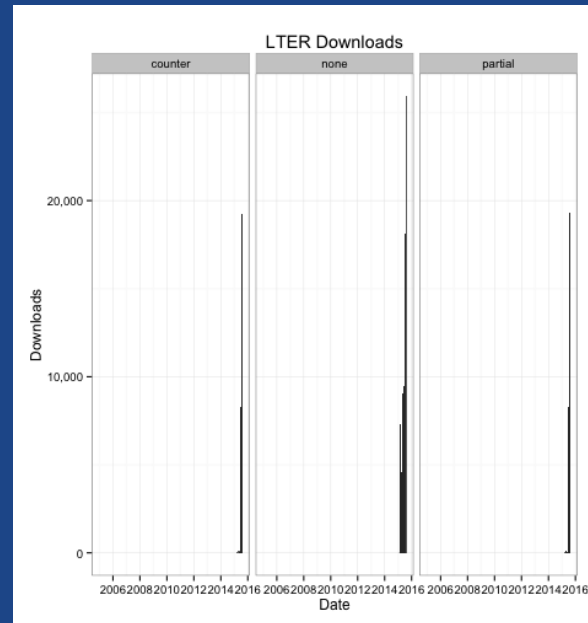
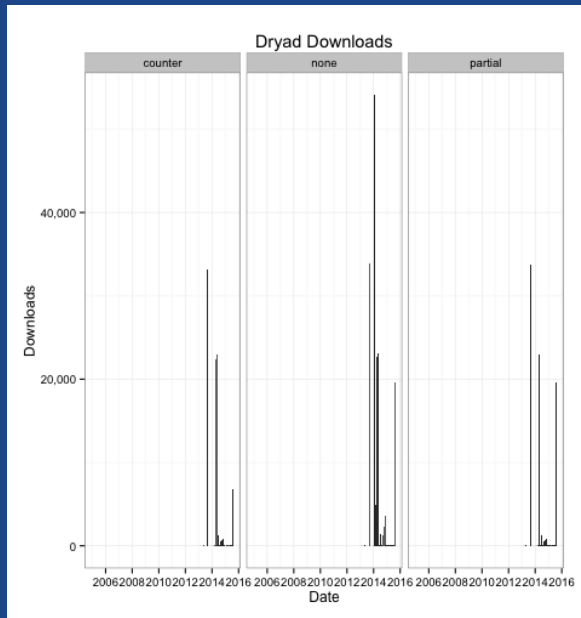
	Average % (not filtered)	
since 2005	COUNTER	63.57%
	Partial	63.59%
this past year	COUNTER	44.88%
	Partial	47.05%

Usage activity - comparing filters



- no difference between filters from Jan 2005 - June 2011
- difference between filters in Aug 2015 = 37363 counts

Usage activity - filters vary by repository



Data metrics aggregation

Track citations, usage and altmetrics for datasets

- Rewrote Lagotto open source application to handle research outputs beyond journal articles
- Handles persistent identifiers beyond DOIs, including URLs
- Wrote import script to import DataONE datasets
- Wrote sources to track citations of datasets in the scholarly literature (e.g. Europe PMC Fulltext Search)

Citations

- no available dataset citation index trusted by community
- approach:
 - text mine research articles for reference to dataset identifiers (open APIs)
 - full text search is important given highly variable locations of dataset reference
 - de-duplicate results across corpuses
 - remove noise from signal

Work in Progress

Reporting tool
customize open source
application (ALM Reports)

Analysis
source correlations, usage
count recommendations, open
data activity comparisons

Parascope

Your List

5

Parascope allows you to view article-level metrics for any set of articles published since 2013 as well as summarize and visualize the data results.

[Home](#) | [About](#) | [Samples](#)

1 Select Articles

2 Preview List

3 View Report

Review your list of articles

Create Report

1-5 of 5 articles

[Clear All](#)

Add more articles

[Direct Invasion of Bones by Highly Pathogenic Fungi in an in vitro Model and Its Ecological Significance.](#)

Shahana Shamin, Fukuko Kishi, Ayako Sano, Katsuhiko Kamel, Kazuko Nishimura, Makoto Miyaji
Journal-article | | published 01 Jan 2003 | Jpn J Med MycolNippon Ishinkin Gakkai Zasshi | The Japanese Society for Medical Mycology

[Search Again](#)

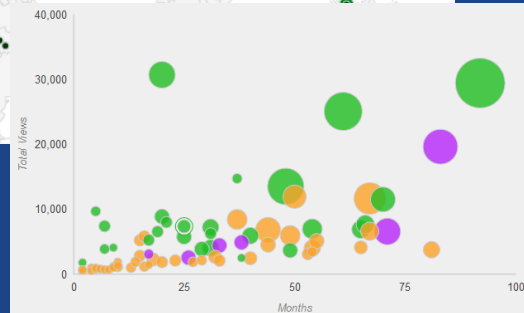
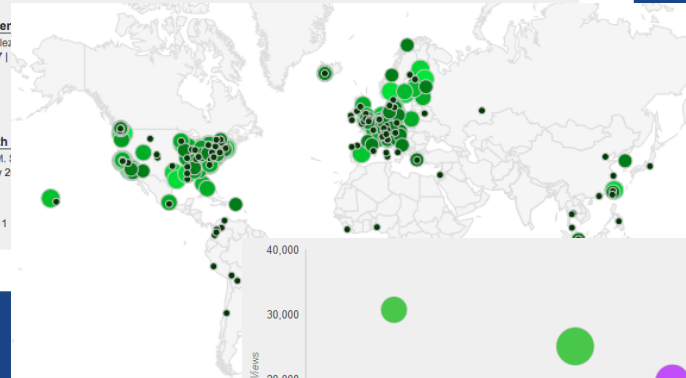
[By DOI/PMID](#)

[Fungi in the sea fan Gorgonia ver](#)
C. Toledo-Hernández, A. Bones-González
Journal-article | | published 21 Jun 2007 |

[Bare-Bones Entrepreneurship](#)
Book-chapter | Published 23 Apr 2013

[Detection of fungi colony growth](#)
F. M. Vnstorio, N. Budini, C. Mulone, M. I
Proceedings-article | | published 18 Nov 2

[Bones and teeth](#)
Reference-entry | Published 01 Jan 2011



Making Data Count Team

<http://mdc.lagotto.io>

<http://dlm.labs.datacite.org>

- Stephen Abrams
- John Chodacki
- Patricia Cruse
- Martin Fenner
- Matt Jones
- John Kratz
- Jennifer Lin
- Kristen Ratan
- Peter Slaughter
- Carly Strasser
- Dave Vieglais

This presentation is made available under a CC-BY 4.0 license.

<http://creativecommons.org/licenses/by/4.0/>

