

Make Data Count

DataONE Webinar

J Lin*, M Fenner*, M Jones** & J Kratz ***

*PLOS

**NCEAS, DataONE

***California Digital Library

Agenda

- Make Data Count Overview - Jennifer Lin, PLOS
- Unit 1 research - John Kratz, CDL
- Unit 2 DataONE Usage - Matt Jones, DataONE
- Unit 3 DLM Application - Martin Fenner, PLOS

Overview

Make Data Count NSF Project Introduction

Jennifer Lin

Senior Product Manager, PLOS

Make Data Count Overview

Partners

California Digital Library, PLOS, and DataONE/NCEAS

NSF Grant Record: Grant No. 1448821

Proposal PDF in eScholarship repository

Project page

<http://articlemetrics.github.io/MDC/>

Prototype <http://dlm.plos.org>

Lagotto software is Open Source <https://github.com/articlemetrics/lagotto>

Make Data Count Project Plan

1. DLM Field Research	2. Data Usage Tracking	3. Data Activity Aggregation	4. DLM Integration & Presentation
<p>What: Surveys, interviews, focus groups to determine requirements for DLM</p> <p>Output: metrics design & requirements</p>	<p>What: Extend DataONE usage tracking capacity</p> <p>Output: COUNTER-based usage API</p>	<p>What: Formulate a set of test metrics, extend technology</p> <p>Output: DLM application</p>	<p>What: Develop tools for the community to use metrics</p> <p>Output: DLM Reports application & widgets, Bibliometric Analysis</p> <p>What: Analyze, write up results from project</p> <p>Output: final report &recommendations</p>

Scholars access, cite, share, mention, recommend papers

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

But, data...

- Also 1st class scholarly object
- Broader role in research process
- Has its own use & reuse profile
- Infrastructure services to collect metrics are lagging
- Common best practices are not enshrined in research communities

Aims & Goals

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

Unit 1 - Research

Gathering DLM requirements

John Kratz

Data Curation Postdoctoral Fellow, California Digital
Libraries

Make Data Count Project Plan

1. DLM Field Research

What: Surveys, interviews, focus groups to determine requirements for DLM

Output: metrics design & requirements

Literature Review

Focus Groups

- DataONE All Hands Meeting
- Society for Neuroscience
- American Geophysical Union

Online Surveys

- Researchers
- Data managers

Making Data Count Survey JEK-III

SurveyMonkey, Inc [US] https://www.surveymonkey.com/s/makedatacount

 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

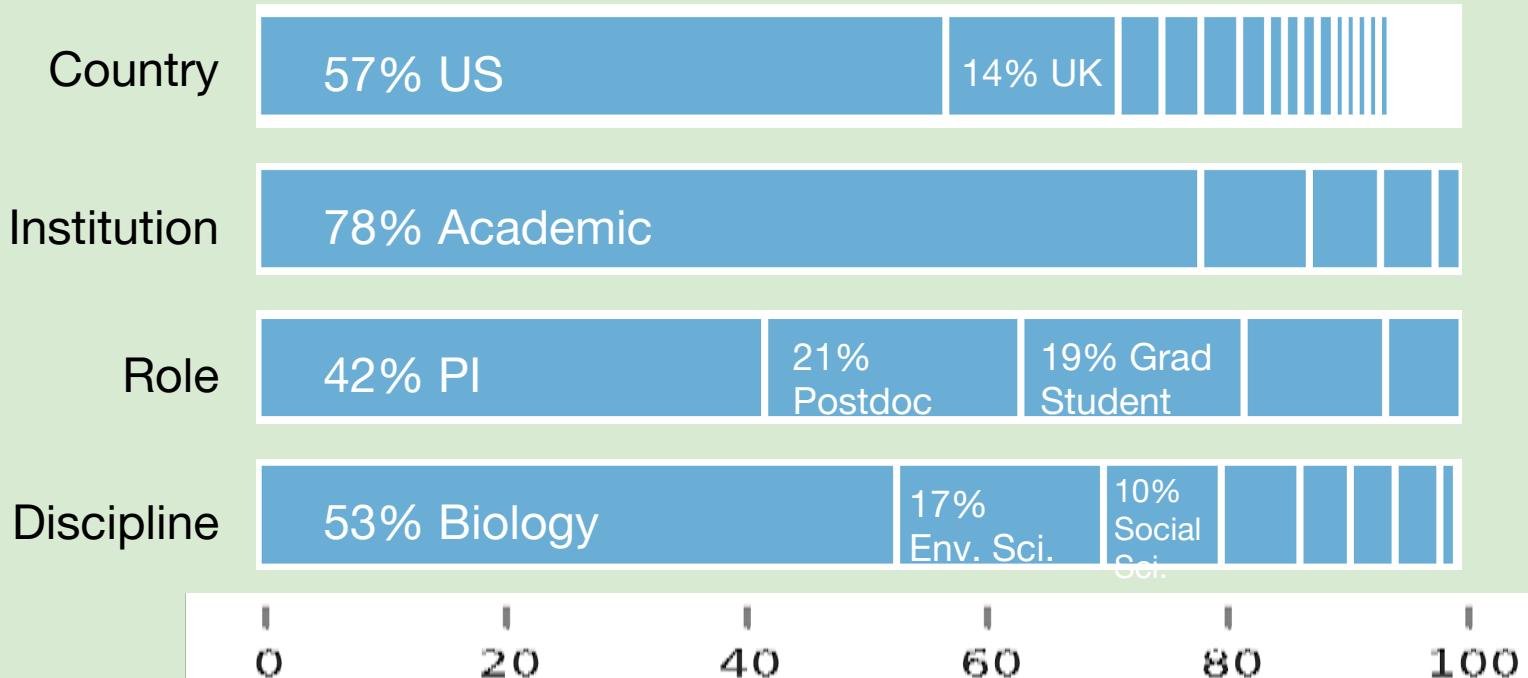
 University of California
CDL
California Digital Library

 PLOS

 DataONE

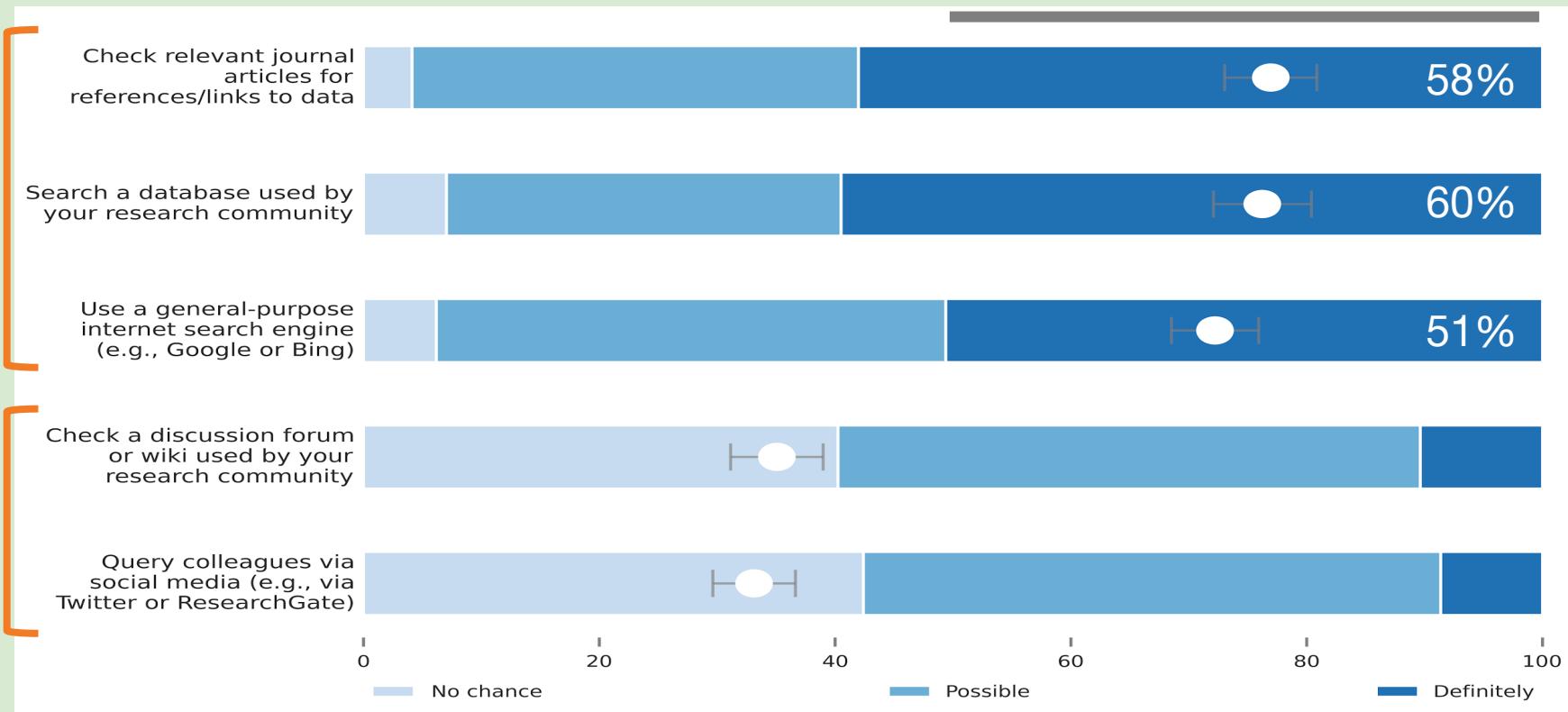
Next

247 researchers responded

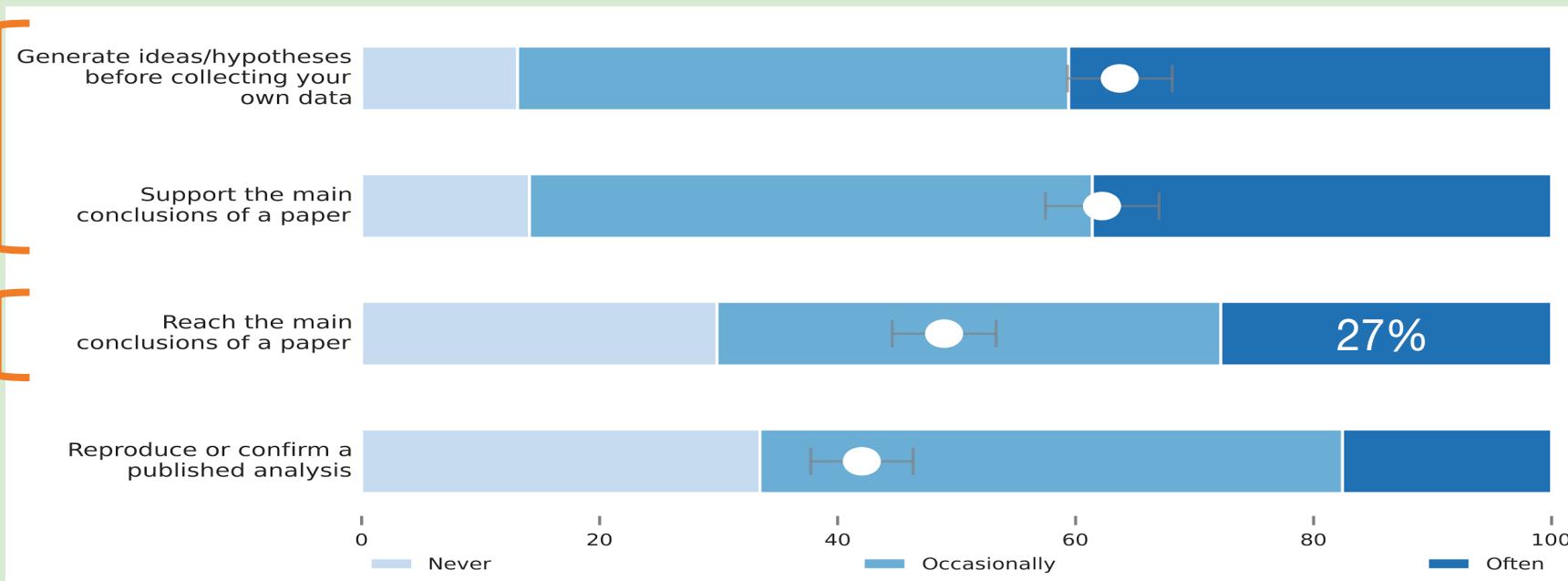


Data Workflow

When looking for public data to use, how likely are you to search in each of the following ways?

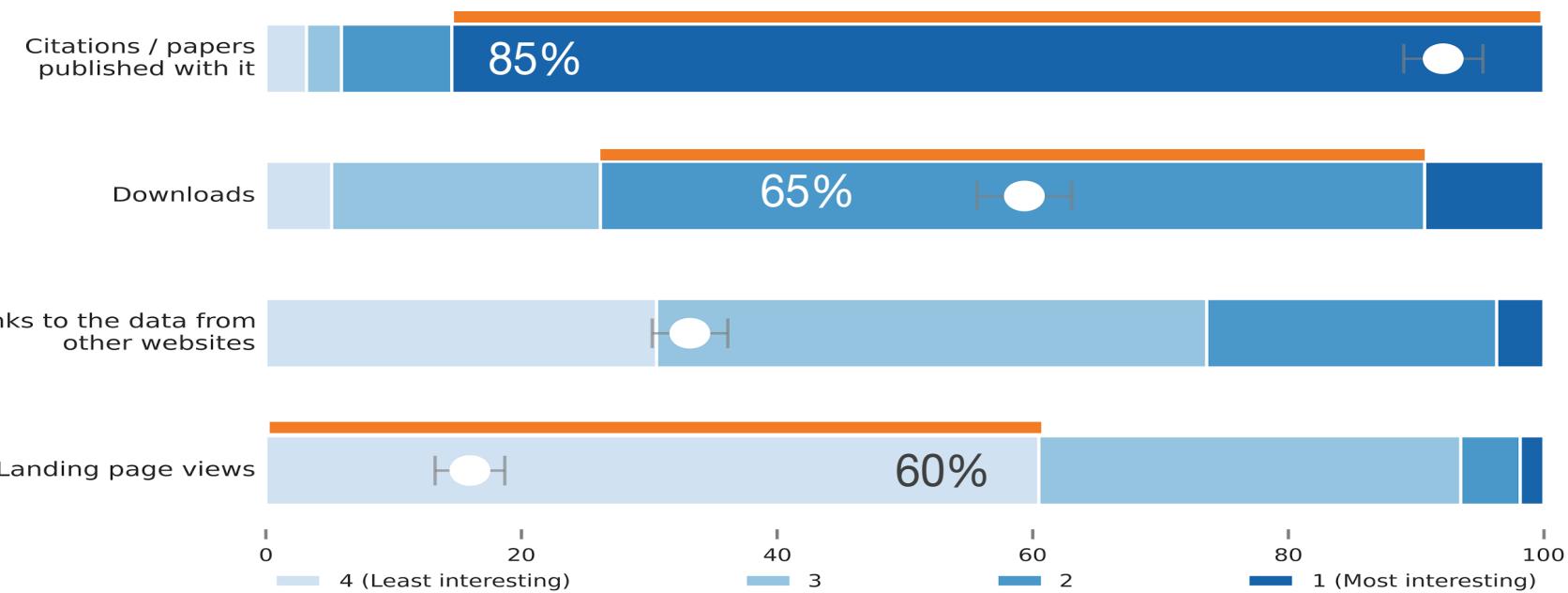


How frequently do you use data from public sources to accomplish each of the following?

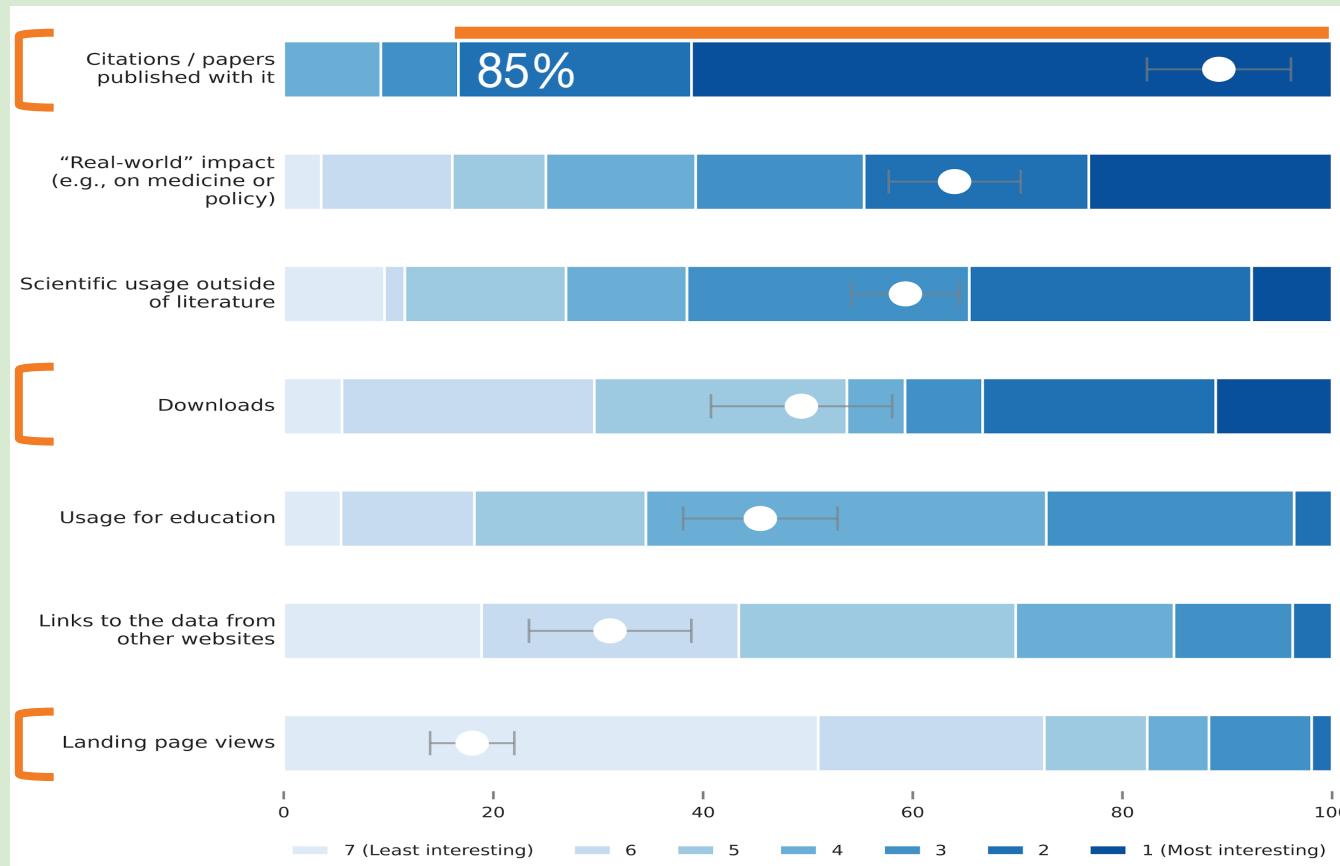


Impact

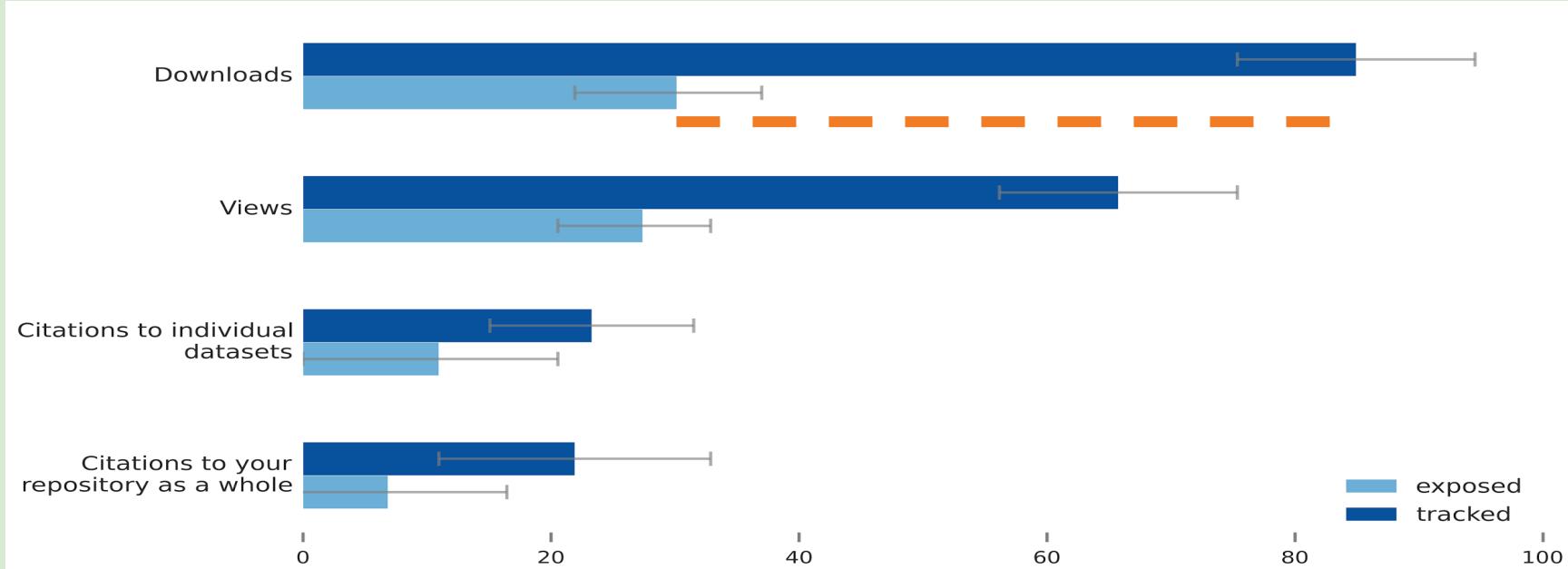
How interested you would be to know each of the following about the impact of your data?



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)



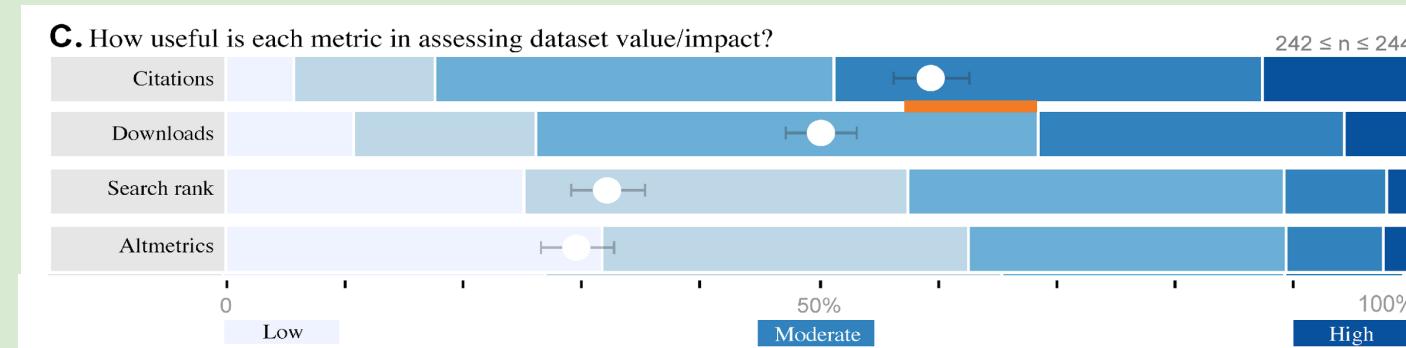
Page views are not valued by anyone.

There is little social media activity to capture.

Citations are the most valuable metric...

...but they suffer from technical and cultural obstacles.

Downloads are both tractable and valued.



Unit 2 - DataONE Usage

Providing COUNTER-based usage stats for DLM

Matt Jones

Director of Informatics Research & Development,
NCEAS

Filters

Datasets 1 to 25 of 98818

Sort by [Most recent](#)
[1](#)
[2](#)
[3](#)
[...](#)
[3953](#)
[Next](#)

Duwyn, Alisha, and MacDougall, Andrew. 2015. **Duwyn-MacDougall JECOL data.**

<http://dx.doi.org/10.5061/dryad.km53t/1>
ver=2015-03-06T16:50:04.127-05:00



Duwyn, Alisha, and MacDougall, Andrew S.. 2015. **Data from: When anthropogenic-related disturbances overwhelm demographic persistence mechanisms.**

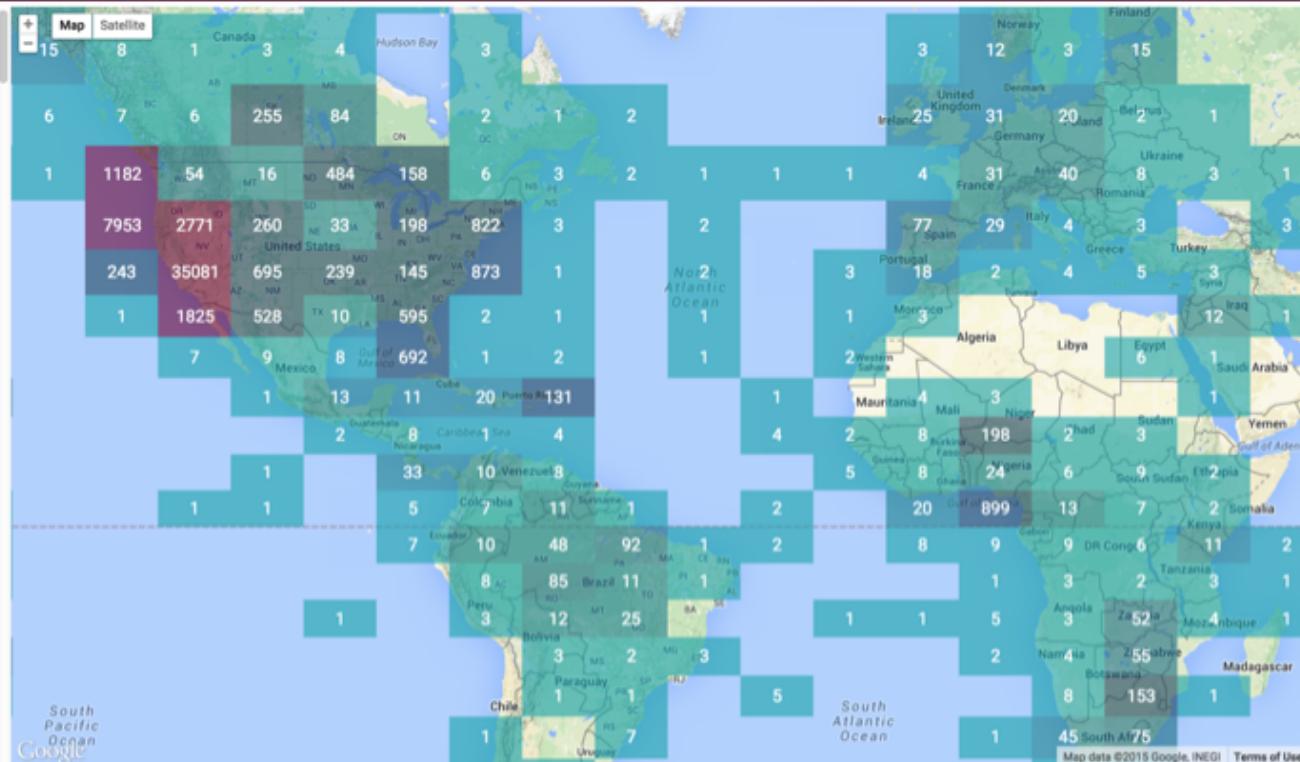
<http://dx.doi.org/10.5061/dryad.km53t?ver=2015-03-06T16:50:00.843-05:00>

1. Population decline is associated with increased vulnerability to extinction, but all



Kraemer, Andrew, Serb, Jeanne, and Adams, Dean. 2015. **Data from: Batesian mimics influence the evolution of conspicuousness in an aposematic salamander.**

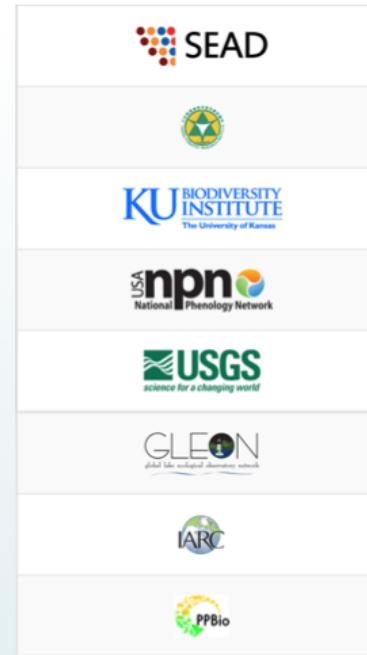
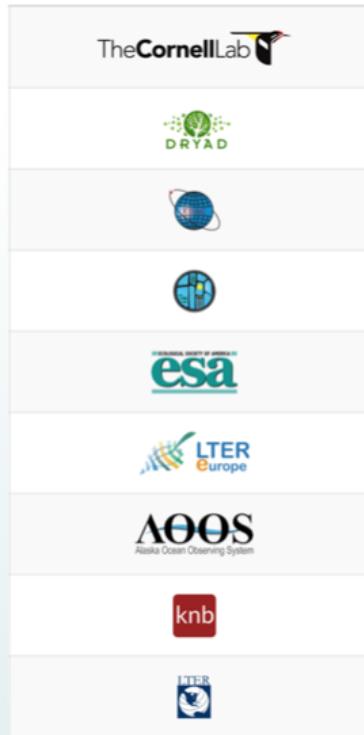
<http://dx.doi.org/10.5061/dryad.7ht20?ver=2015-03-06T15:46:05.328-05:00>



25 DataONE Repository Members



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



Data usage tracking: downloads



Independent downloads? Entire package?

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

The screenshot shows the knb data portal interface. At the top, there is a navigation bar with the knb logo, links for ABOUT, DATA, SHARE, and TOOLS, a search bar with a magnifying glass icon, and a SIGN IN button. Below the navigation bar, a banner displays the title of a dataset: "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)". The main content area shows a table of download statistics for the dataset's components. The table has columns for Package, DOI, File Type, Size, Downloads, and Details. A red "Download all" button is located at the bottom right of the table.

Package	DOI	Type	Size	Downloads	Details
Metadata	doi:10.5063/F10V89RP	.xml (EML)	26.16 KB	586 views	Download all
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

How do we count?



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

The screenshot shows a package page on the knb platform. At the top, there is a navigation bar with links for ABOUT, DATA, SHARE, TOOLS, a search bar containing "Search for data" with a magnifying glass icon, and a SIGN IN button. Below the navigation bar, a yellow banner displays a warning: "⚠ NOTE: A newer version of this package exists". The main content area shows a package titled "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)". Below the title, there is a table listing the contents of the package:

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	Download all ↴
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 ⓘ

How do we standardize collection?



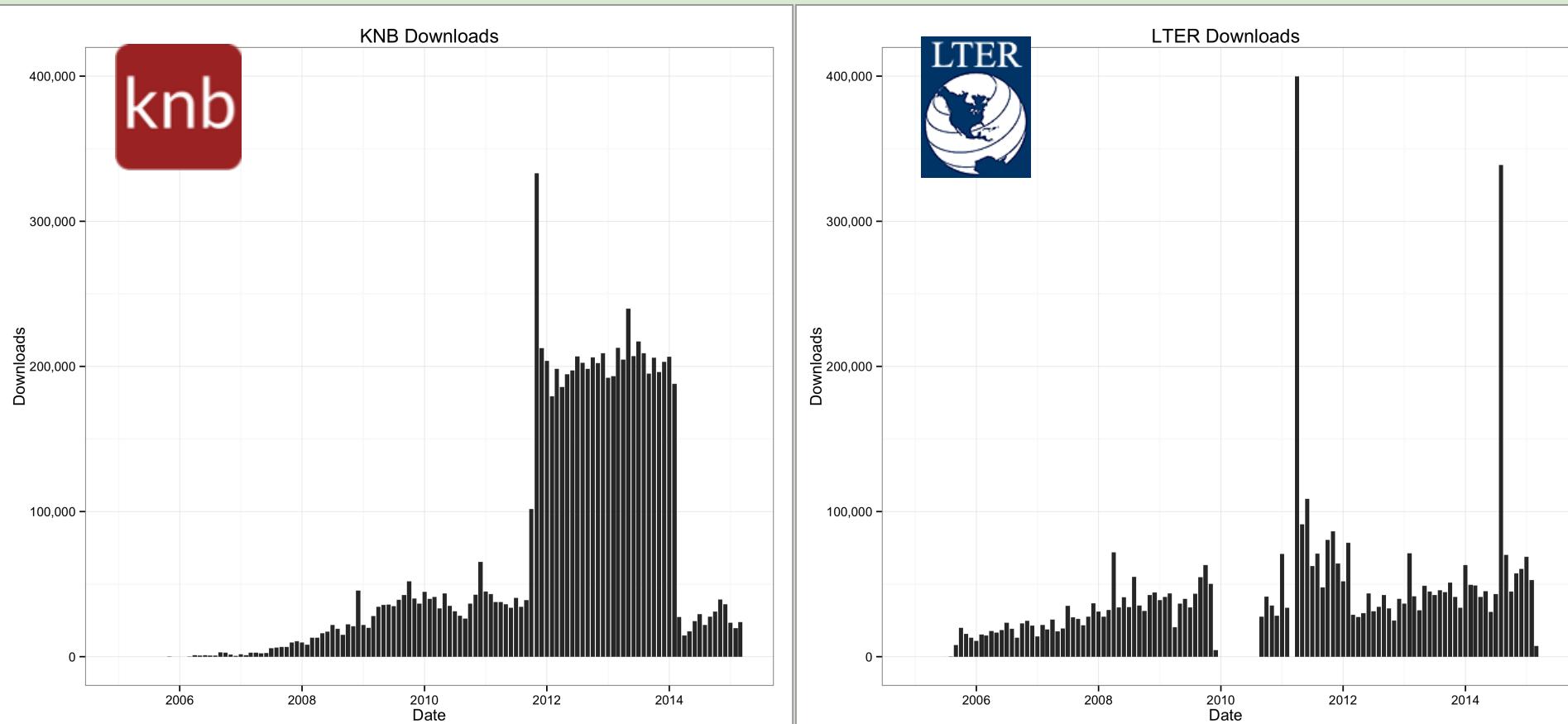
Counting Online Usage of Networked Electronic Resources

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

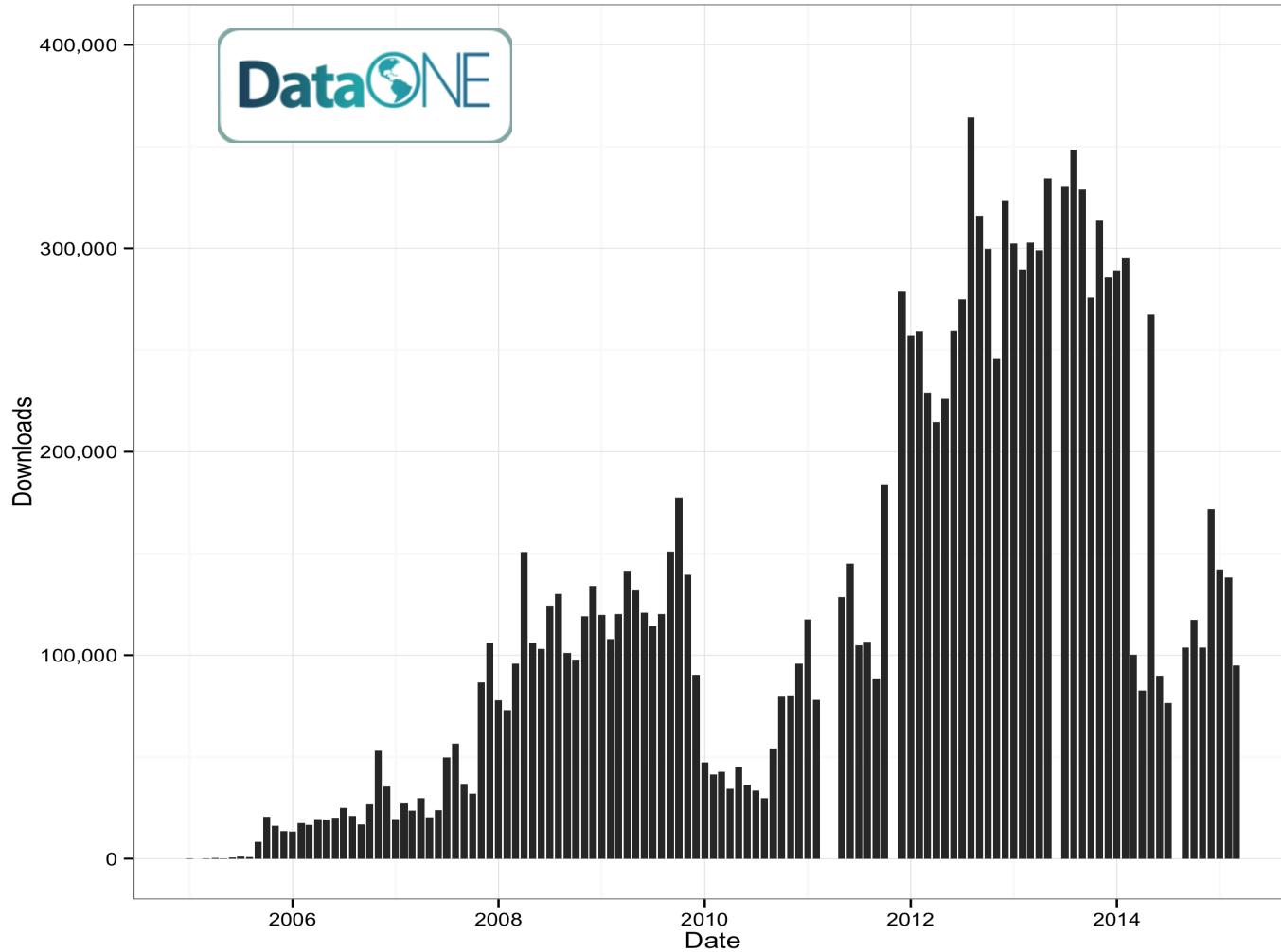
- Standard means of reporting usage of articles
- COUNTER reports remove:
 - Web robots from search engines
 - Repeat visits in short time window (double clicks)
 - All accesses from Python, Java, curl, wget, etc.
 - Scientists frequently use these to access data
- COUNTER issues with composite objects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

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

Downloads from 2006-2015



All DataONE Downloads



Downloads versus Impact

- Conclusions
 - Download data are plentiful and quantitative
 - COUNTER needs work to be applicable to DATA
- Caveats
 - Downloads ≠ Usage
 - Downloads ≠ Citation
 - Downloads ≠ Impact
- Understanding **impact** requires understanding usage
 - Citations are the currency

Unit 3 - DLM Application

Using the Lagotto open source application

Martin Fenner

Technical Lead Article-Level Metrics
PLOS

Lagotto tracks citations, usage and altmetrics for scholarly articles

articlemetrics / lagotto

Unwatch 32 Unstar 77 Fork 13

Article-Level Metrics <http://articlemetrics.github.io> — Edit

4,777 commits 5 branches 141 releases 10 contributors

branch: master lagotto / +

don't automatically import works in default .env configuration

mfenner authored a day ago latest commit 8a6c324184

.librarian/chef	file mode changed to regular file	5 months ago
app	updated digital ocean image name	2 days ago
bin	switch to teaspoon for javascript tests. #276.	20 days ago
code_example_templates	added css style guide, generated with hologram gem. Closes #240.	2 months ago
config	filter authentication_token from log. Closes #285	11 days ago
data	file mode changed to regular file	5 months ago
db	added data migrations. Closes #280.	17 days ago

Code Issues 35 Pull requests 0

Pulse Graphs Settings

SSH clone URL
git@github.com:artic... [Copy](#)

You can clone with [HTTPS](#), [SSH](#), or [Subversion](#).

DLM

Status

Works

Sources

Data from: aTRAM - automated Target Restricted Assembly Method:
loci across divergent taxa from next-generation sequencing data

Publication Date April 01, 2015

Type dataset

DOI [10.5061/dryad.kh886](https://doi.org/10.5061/dryad.kh886)

Publisher URL <http://datadryad.org/resource/doi:10.5061/dryad.kh886>

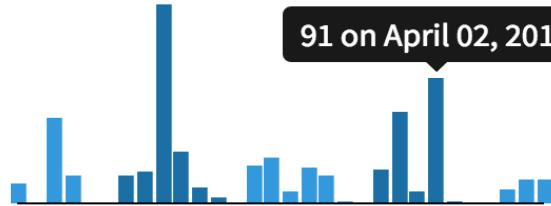
Rewrote Lagotto open source application to handle research outputs beyond journal articles
(data model, API, admin frontend)

Status 08 Apr 19:06 UTC

Works

91,977

91 on April 02, 2015



Wrote import script to regularly import new DataONE datasets. Handles persistent identifiers beyond DOIs, including URLs.

<http://dlm.plos.org/status>

2013

Polymorphic microsatellite loci for *Virola sebifera* (Myristicaceae) derived from shotgun 454 pyrosequencing

N Wei, Cw Dick, Aj Lowe & Mg Gardner. Appl Plant Sci. Journal article. 2013. <http://dx.doi.org/10.3732/apps.1200295>

Widespread horizontal transfer of retrotransposons

Am Walsh, Rd Kortschak, Mg Gardner, et al. Proc Natl Acad Sci U S A. Journal article. 2013. <http://dx.doi.org/10.1073/pnas.1205856110>

Environmental complexity and biodiversity: the multi-layered evolutionary history of a log-dwelling velvet worm in Montane Temperate Australia

Jk Bull, Cj Sands, Rc Garrick, et al. PLoS One. Journal article. 2013. <http://dx.doi.org/10.1371/journal.pone.0084559>

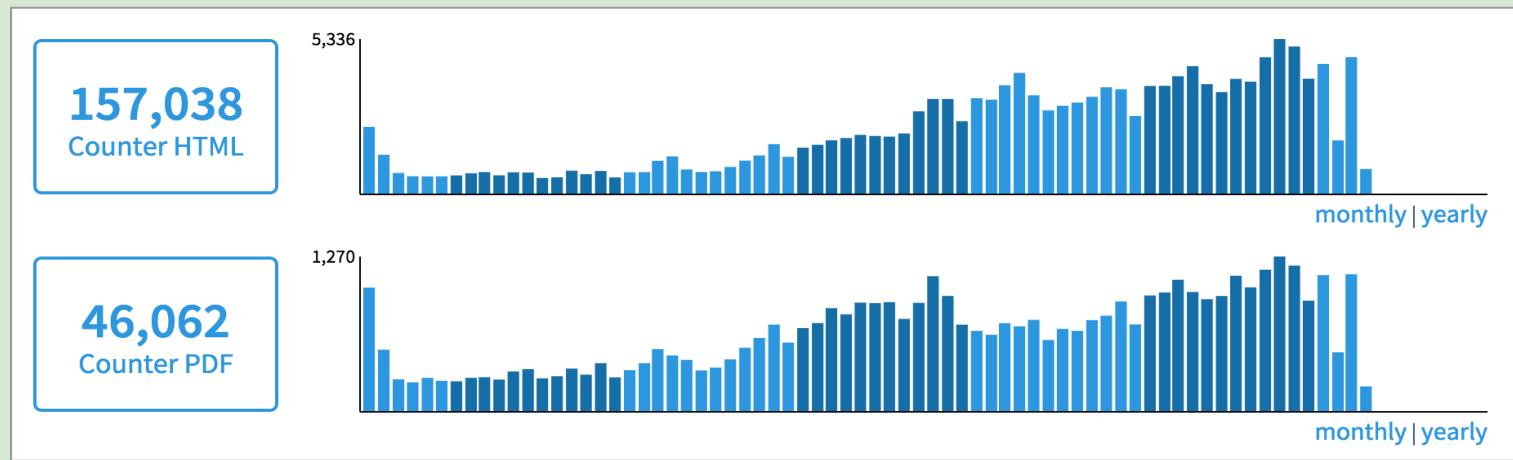
2012

Breakdown of phylogenetic signal: a survey of microsatellite densities in 454 shotgun sequences from 154 non model eukaryote species

E Meglécz, G Nèvre, E Biffin & Mg Gardner. PLoS One. Journal article. 2012. <http://dx.doi.org/10.1371/journal.pone.0040861>

Wrote sources to track citations of datasets in the scholarly literature (e.g. Europe PubMed Central Fulltext Search)

Work in Progress: Usage Stats



PLOS article

Work in Progress: Related Works

All Relations ▾

All Sources ▾

Using Hierarchical Cluster Models To Systematically Identify Groups Of Jobs With Similar Occupational Questionnaire Response Patterns To Assist Rule Based Expert Exposure Assessment In Population Based Studies



2015. <http://dx.doi.org/10.1093/annhyg/meu101>

Cites via CrossRef

ORCID profile for Anasua Sarkar

April 09, 2015. <http://orcid.org/0000-0001-7365-3924>

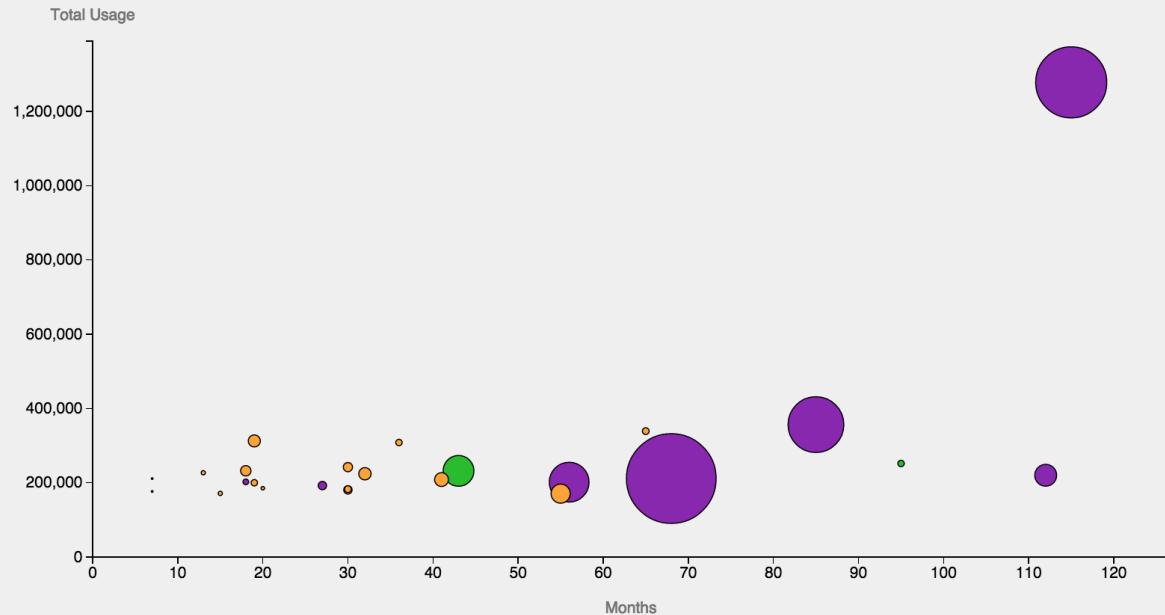
Bookmarks via ORCID

Using the DataCite Metadata Schema (IsPartOf, IsNewVersionOf, etc.)

<http://labs.lagotto.io/works/doi:10.1371/journal.pone.0046468>

Work in Progress: Reporting

1 Article Usage as a Function of Age



<http://almreports.plos.org/reports/visualizations/17115>

Total usage includes page views and downloads from PLOS and PubMed Central Usage Stats. Bubble size correlates with Scopus citations and bubble color with the journal.

Thank you

Make Data Count Team

<http://articlemetrics.github.io/MDC/>

<http://dlm.plos.org>

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



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

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