

A future where data attribution Counts

Heather Piwowar [@researchremix](#)
DataONE postdoc with NESCent and Dryad



some photos NC, SA

#idcc11



























© Eric Prida 2010

[http://www.flickr.com/photos/commissariat/4829261601/
in/faves-30112411@N02/](http://www.flickr.com/photos/commissariat/4829261601/in/faves-30112411@N02/)



© 2006 Sameer Gharat





Gleditsch et al. 2003. Posting Your Data: Will You Be Scooped or Will You Be Famous?, International Studies Perspectives 4(1): 89–97.

Piwowar et al. 2007. Sharing Detailed research data is associated with increased citation Rate. PLoS ONE.

Ioannidis et al. Repeatability of published microarray gene expression analyses. Nature Genetics 41, 149 - 155

Pienta et al. 2010. NSR Social Science Secondary Use. Michigan IR.

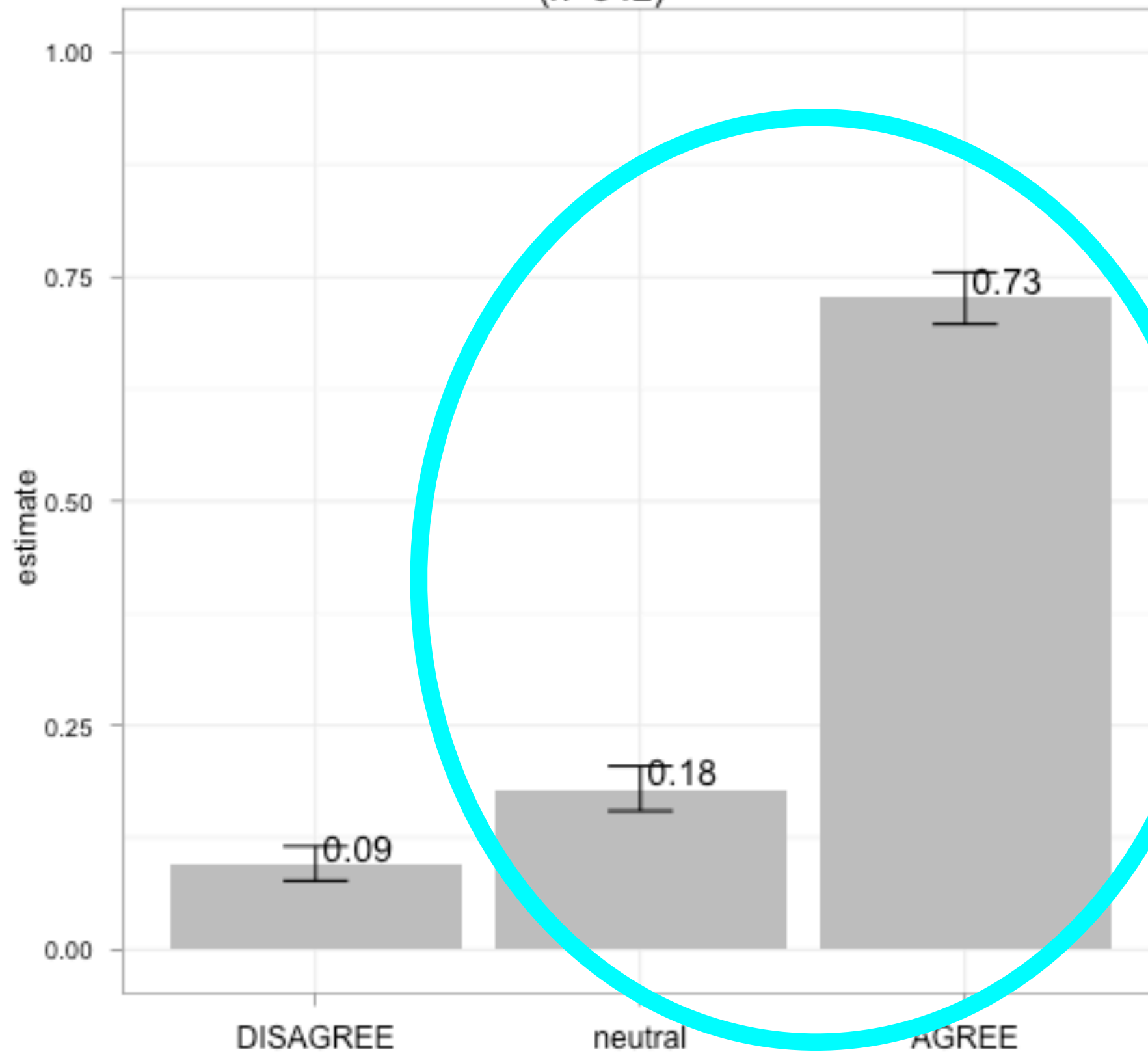
Henneken et al. 2011. Linking to Data – Effect on Citation Rates in Astronomy. ESO.

Sears 2011. Data Sharing Effect on Article Citation rate in Paleoceanography. AGU.

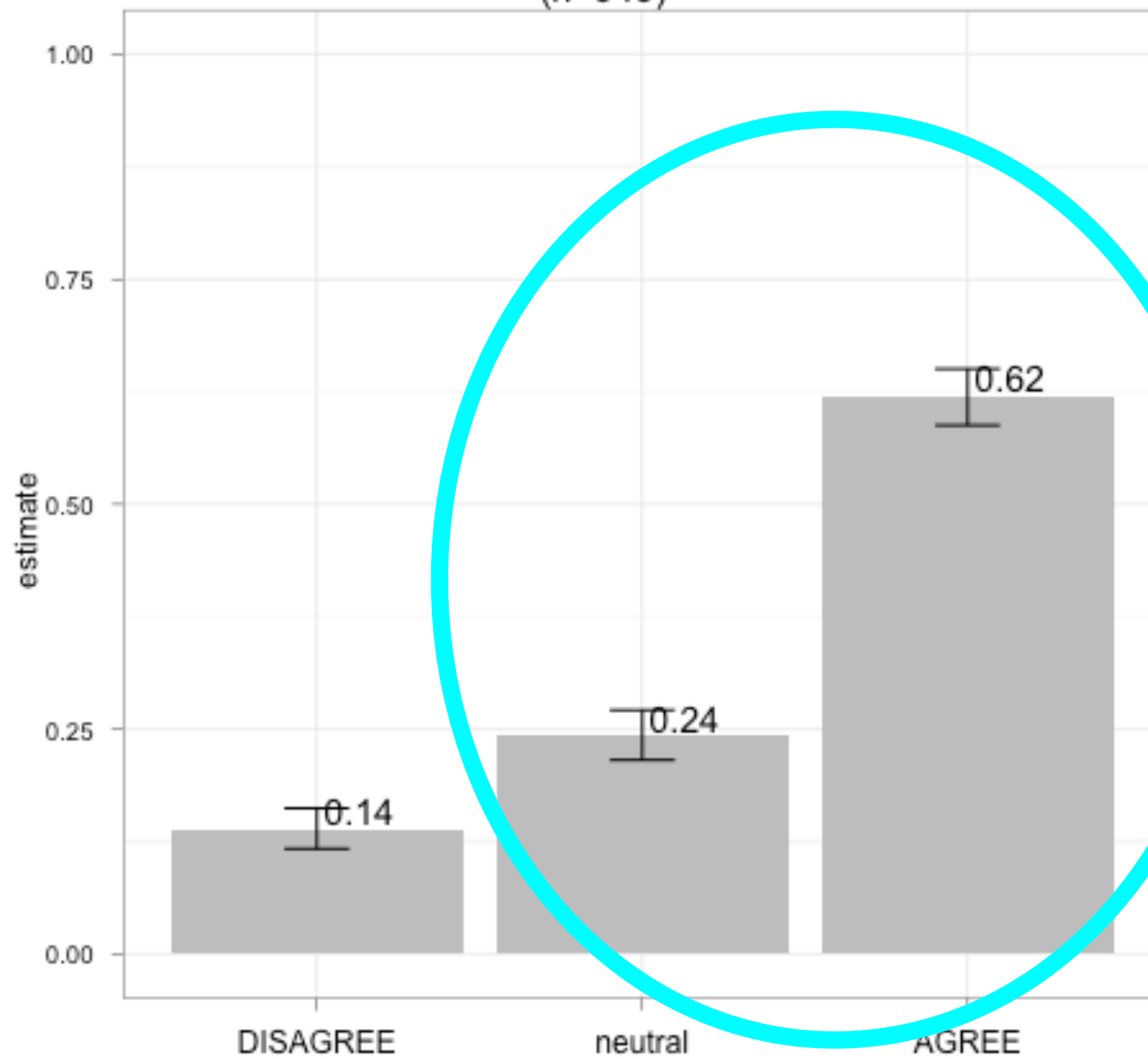


http://farm4.staticflickr.com/3319/4606049592_f780ea0fa4.jpg

I think it is the right thing to do
(n=942)

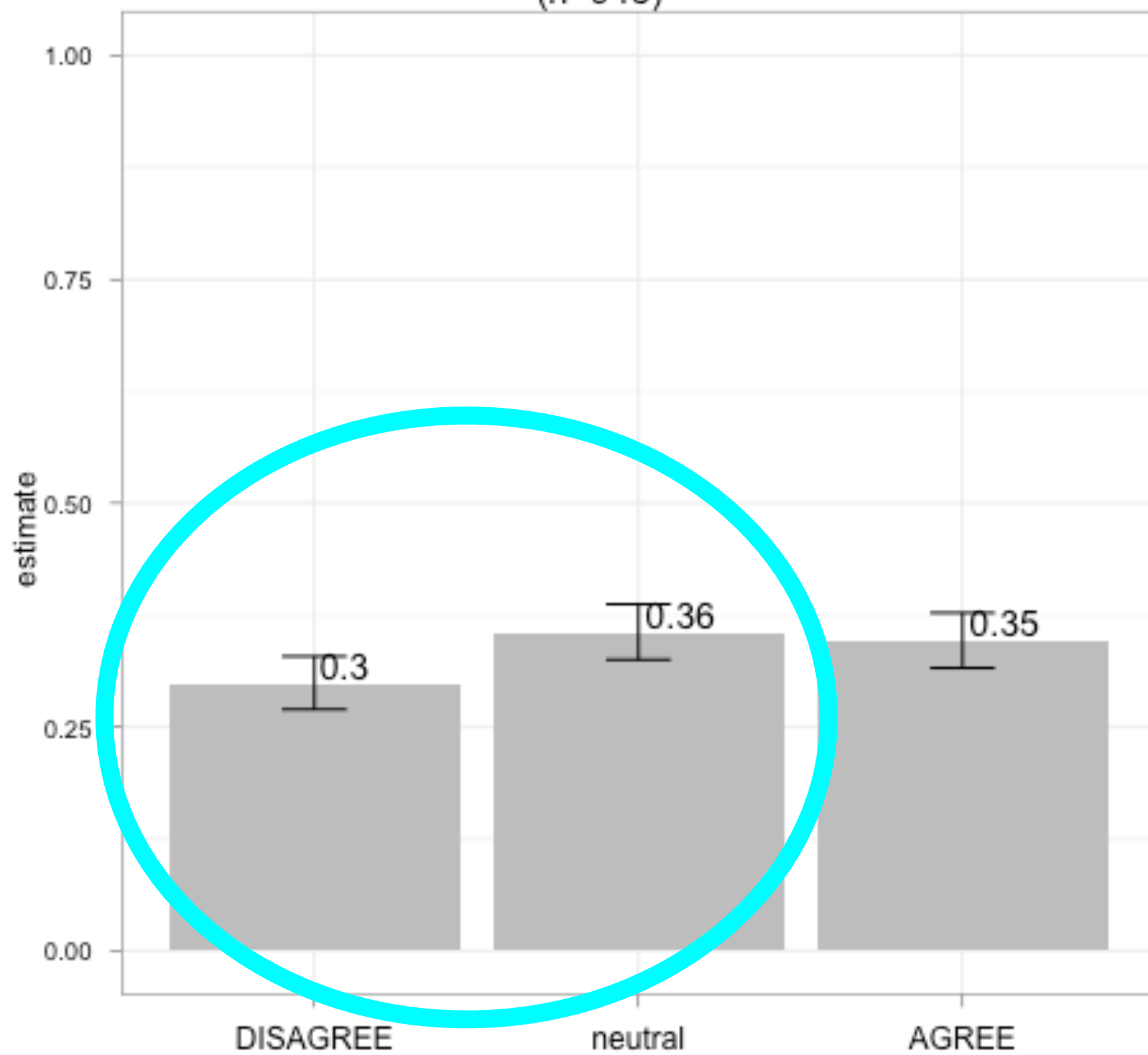


I will get more citations
(n=949)

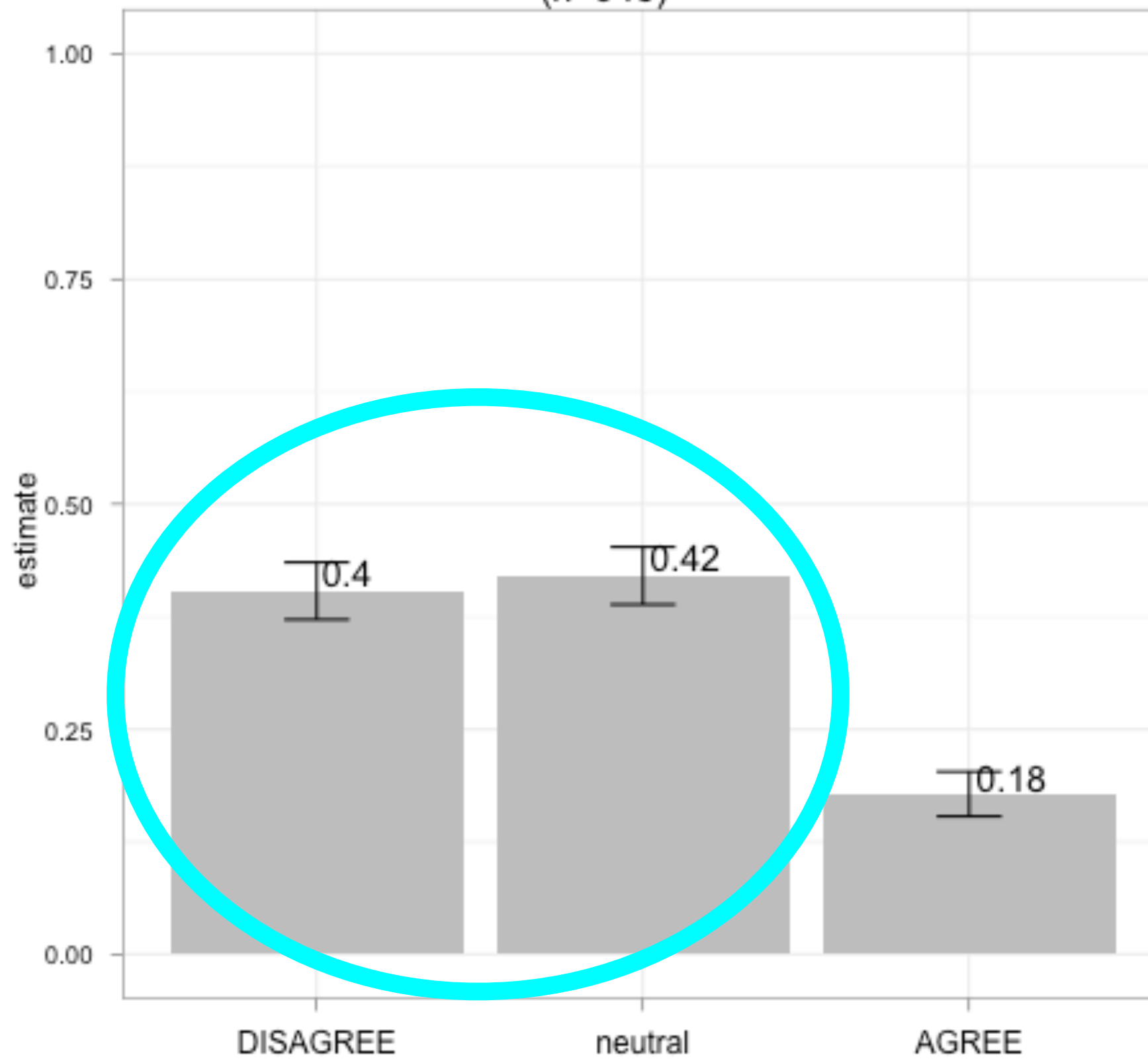


but....

the contribution will be valued by my funders
(n=945)



the contribution will be valued by my promotion or tenure committee
(n=945)





http://farm4.staticflickr.com/3473/4592014327_a6f4d4618d.jpg

So.

We need to facilitate
deep recognition of the
labour of dataset creation.



http://farm6.staticflickr.com/5106/5692787622_e6f9367f3b.jpg





Researchers

Heather Piwowar

hpiwowar@gmail.com

Research Passion *Accelerating research progress by facilitating more effective reuse of research data*

Current Position **Postdoctoral Research Associate** **Vancouver BC, Canada**
Duke University, funded by NSF through DataONE, NESCent, and Dryad
June 2010 – present. PI: Dr Todd Vision

- Conducting research studies in science data sharing and reuse

Education **University of Pittsburgh** **Pittsburgh, PA**
PhD in 2010, Masters of Science in 2006, concentration in Bioinformatics
Department of Biomedical Informatics, School of Medicine.
2005-2010 (maternity leave: April-June 2006). Advisor: Dr Wendy Chapman.
Dissertation: Foundational studies for measuring the impact, prevalence, and patterns of publicly sharing biomedical research data

Massachusetts Institute of Technology **Cambridge, MA**
Masters of Engineering in 1996, Bachelors of Science in 1995.
Majored in Electrical Engineering and Computer Science, GPA 4.9/5.0
Concentration in Digital Signal Processing (DSP)
Humanities and mathematics each comprised 20% of curriculum

Academic Experience & Accomplishments Piwowar (2011). Who shares? Who doesn't? Bibliometric factors associated with open data archiving. PLoS ONE 6(7): e18657.
Garnett, Whiteley, Piwowar, Rasmussen, Illes (2011) Neuroethics and fMRI: Mapping a Fledgling Relationship. PLoS ONE 6(4): e18537.

Peer-reviewed Journal Articles Piwowar, Chapman (2010). Recall and bias of retrieving gene expression microarray datasets through PubMed identifiers. J Biomed Discov Collab. Mar 28;5:7-20
Piwowar, Chapman (2010). Public sharing of research datasets: A pilot study of associations. Journal of Informetrics 4: 2. 148-156 April (also presented at Symposium on Informetrics & Scientometrics ASIS&T and ISSI Pre-conference 2009)

Piwowar, Chapman (2008) Envisioning a Biomedical Data Reuse Registry AMIA 2008
Piwowar, Chapman (2008) Prevalence and Patterns of Microarray Data Sharing PSB 2008
Piwowar, Fridsma (2007) Examining the uses of shared data ISMB 2007

Datasets

Piwowar, Vision, Whitlock (2011) Data from: Data archiving is a good investment. Dryad Digital Repository. [doi:10.5061/dryad.j1fd7](https://doi.org/10.5061/dryad.j1fd7)
Piwowar (2011) Data from: Who shares? Who doesn't? Factors associated with openly archiving raw research data. Dryad Digital Repository. [doi:10.5061/dryad.mf1sd](https://doi.org/10.5061/dryad.mf1sd)
Piwowar, Chapman (2009) Data from: Public sharing of research datasets: a pilot study of associations. Dryad Digital Repository. [doi:10.5061/dryad.3td2f](https://doi.org/10.5061/dryad.3td2f)
Piwowar, Day, Fridsma (2007) Data from: Sharing detailed research data is associated with increased citation rate. Dryad Digital Repository. [doi:10.5061/dryad.j2c4g](https://doi.org/10.5061/dryad.j2c4g)

Invited Presentations

(2011) International Digital Curation Conference (upcoming), UK.
 (2011) Tracking Data Reuse: Motivations, Methods, and Obstacles. IASSIST Vancouver.
 (2011) Citation Effects of Open Access Publishing Research Data. UK.

Heather Piwowar

run update

download data

54 artifacts; updated 5 Dec, 2011

article

Permalink: <http://total-impact.org/collection/sURpEg> copy
 Tweet 2

05f22f70-6d0b-11df-afb8-0026b95d30b2

Piwowar, Day, Fridsma (2007) **Sharing Detailed Research Data Is Associated with Increased Citation Rate** *PLoS ONE*.

Piwowar HA, Day RS, Fridsma DB **Data from: Sharing detailed research data is associated with increased citation rate.** *Dryad Data Repository*.

10.1038/npre.2008.1701.1

Piwowar, Chapman (2008) **Prevalence and Patterns of Microarray Data Sharing** *Nature Precedings*.

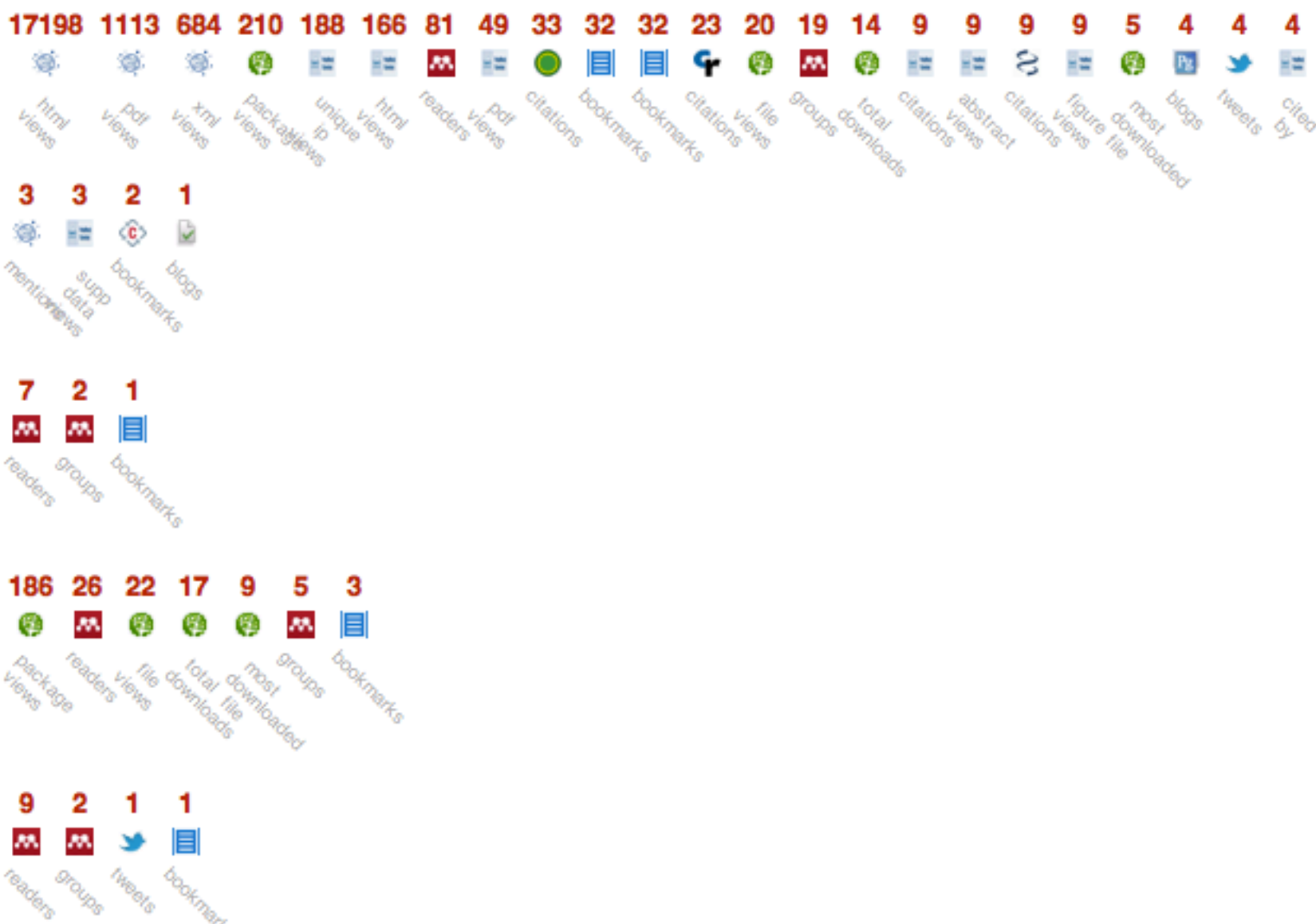
10.1016/j.joi.2009.11.010

Piwowar, Chapman (2010) **Public sharing of research datasets: A pilot study of associations** *Journal of Informetrics*.

Piwowar HA, Chapman WW **Data from: Public sharing of research datasets: a pilot study of associations.** *Dryad Data Repository*.

10.1525/bio.2011.61.8.8

McDade, Maddison, Guralnick, Piwowar, Jameson, Helgen, Herendeen, Hill, Vis (2011) **Biology Needs a Modern Assessment System for Professional Productivity** *BioScience*.



dataset

10.5061/dryad.j1fd7

Piwowar, Vision, Whitlock (0) Data from: Data archiving is a good use of research funds .

Piwowar HA, Vision TJ, Whitlock MC Data from: Data archiving is a good investment. *Dryad Data Repository.*



10.5061/dryad.j2c4g

Piwowar HA, Day RS, Fridsma DB Data from: Sharing detailed research data is associated with increased citation rate. *Dryad Data Repository.*



10.5061/dryad.mf1sd

Piwowar HA Data from: Who shares? Who doesn't? Factors associated with openly archiving raw research data. *Dryad Data Repository.*



10.5061/dryad.3td2f

Piwowar HA, Chapman WW Data from: Public sharing of research datasets: a pilot study of associations. *Dryad Data Repository.*



dataset

10.5061/dryad.j1fd7

Piwowar, Vision, Whitlock (0) Data from: Data archiving is a good use of research funds .

Piwowar HA, Vision TJ, Whitlock MC Data from: Data archiving is a good investment. *Dryad Data Repository.*

807 96 36 27 1 1

package views total file most blogs readers

10.5061/dryad.j2c4g

Piwowar HA, Day RS, Fridsma DB Data from: Sharing detailed research data is associated with increased citation rate. *Dryad Data Repository.*

210 20 14 5

package views total file most

10.5061/dryad.mf1sd

Piwowar HA Data from: Who shares? Who doesn't? Factors associated with openly archiving raw research data. *Dryad Data Repository.*

596 162 112 107 1 1

package total file most blogs mentions

10.5061/dryad.3td2f

Piwowar HA, Chapman WW Data from: Public sharing of research datasets: a pilot study of associations. *Dryad Data Repository.*

186 22 17 9

package views total file most

Repositories




Statistical Analyzer

71 views 33 downloads [View File Details](#)

Mathematica file that contains all of the growth rate data obtained from the Bioscreen as well as a sporulation data. This file performs all of the major statistical analyses on these datasets reported and Otto (2007).


Download: [sliding window 06 NEW.nb](#) (7.235Mb)

To the extent possible under law, the authors have waived all copyright and related or neighboring rights in this data.  [OPEN DATA](#)

Sporulation Rate Data

67 views 51 downloads [View File Details](#)

Download: [sporedata.xls](#) (174.0Kb)

To the extent possible under law, the authors have waived all copyright and related or neighboring rights in this data.  [OPEN DATA](#)

Find & Analyze Data

Find ICPSR Data

Bibliography of Data-
Related Literature

Variables Database

Analyze Data Online

Thematic Collections

Restricted Data

Publication-Related
Archive

Utilization for American National Election Study, 2000: Pre

- [General Information](#)
- [Unique Sessions](#)
- [Unique Users](#)
- [Downloads by Member Institution](#)

By default, the reports below show utilization for the last 12 months

from: to:

Please note that this study was first released on 2001-07-03, and t

General Information

- Initial release date: 2001-07-03
- Last updated: 2008-04-11
- [Related citations: 328](#)

Utilization for American National Election Study, 2000: Pre- and Post-Election Sur

- [General Information](#)
- [Unique Sessions](#)
- [Unique Users](#)
- [Downloads by Member Institution](#)

By default, the reports below show utilization for the last 12 months. To view a different ran

from: to:

Please note that this study was first released on 2001-07-03, and that our current reporting

General Information

- Initial release date: 2001-07-03
- Last updated: 2008-04-11
- [Related Publications](#): 328

Unique Sessions

	Total Data/Documentation Sessions		Data & Documentation	Documentation Only
Any Datasets/Files	389	=	101	288

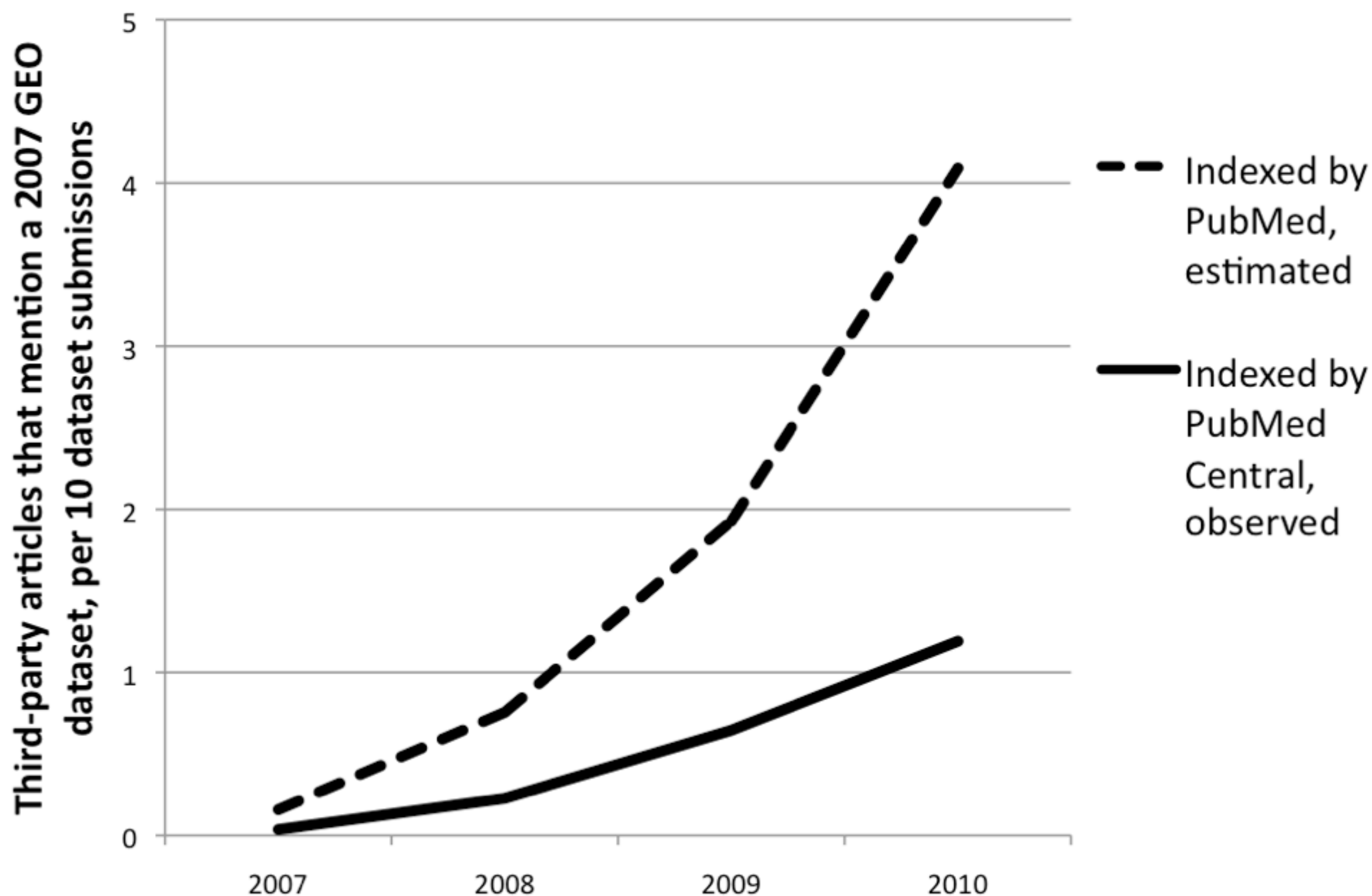
This study was downloaded by users at 57 institutions during the time period specified.

- Bowling Green State University
- Brandeis University
- Brown University
- California State University at Fullerton
- Clemson University
- Creighton University
- George Mason University
- Georgia State University
- Grinnell College
- Hunter College
- Louisiana State University
- Medgar Evers College
- Michigan State University
- New York University
- Niagara University
- North Carolina State University
- Northwestern University
- Ohio State University
- Oregon State University
- Pennsylvania State University
- Princeton University
- Richard Stockton College of New Jersey

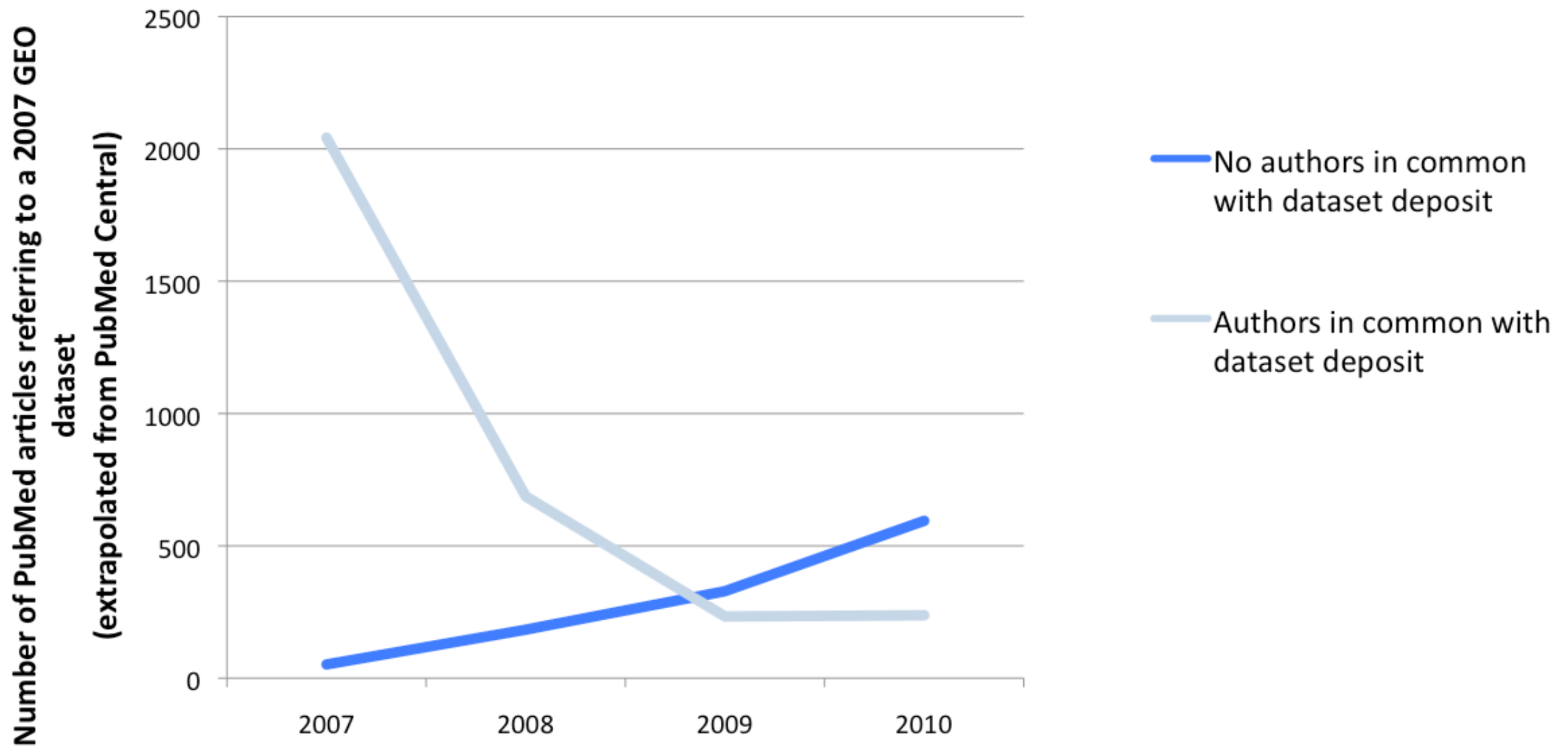
Unique Users

	Total Unique Users		Faculty	Grad. Students	Undergrads	Univ. Staff	Other	Anon. Downloads
Any Datasets/Files	180	=	15	35	45	4	4	77

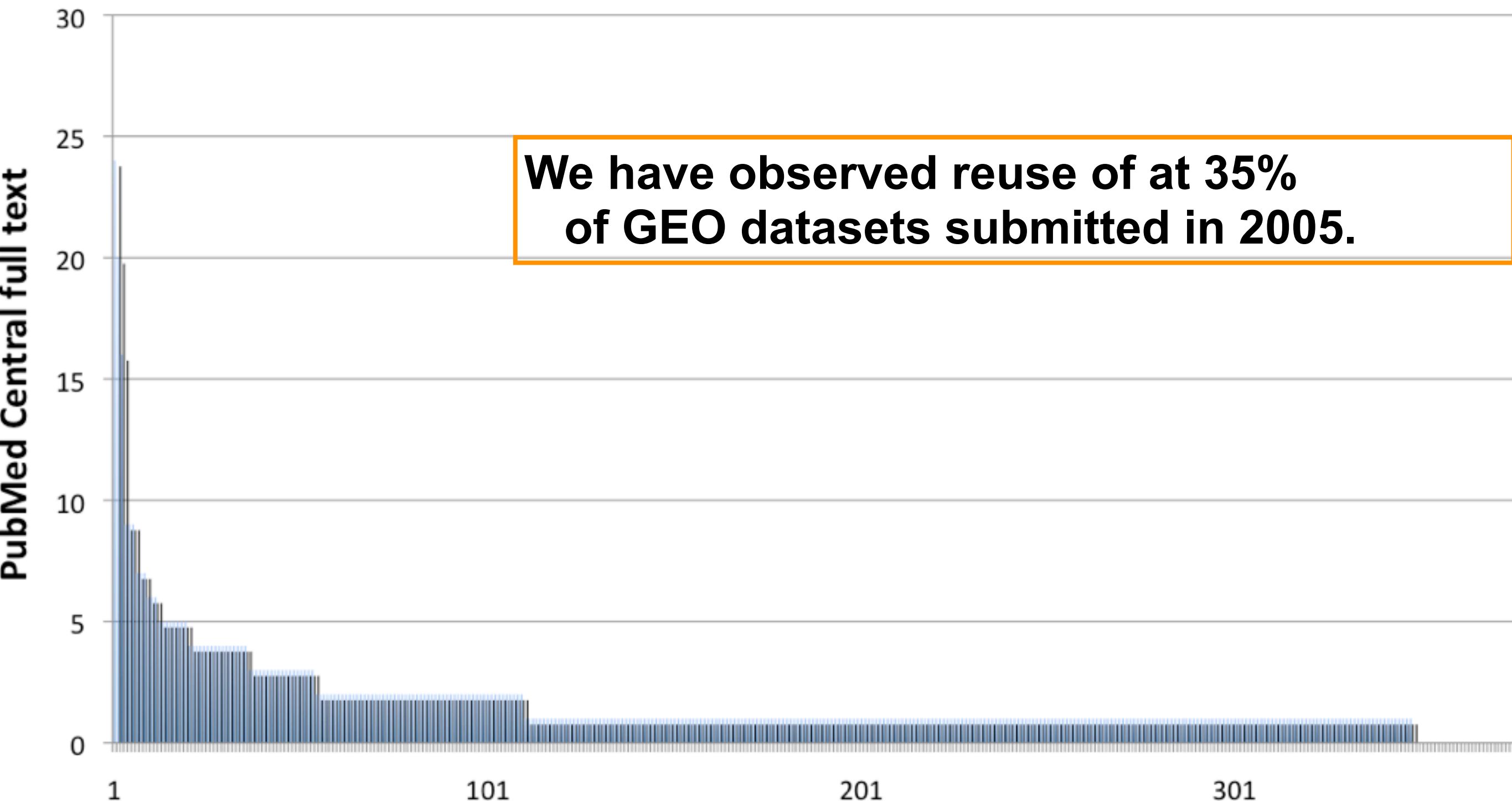
Cumulative third-party use of datasets submitted to the GEO data repository in 2007



Use and reuse of 2007 GEO data attributed by accession reference, extrapolated to all of PubMed



Number of times a 2007 GEO accession was mentioned in PMC, 2007-2010, excluding mentions by original authors



One bar per 2007 GEO dataset
(349 of 2711 GEO datasets had at least one mention in PMC)

Piwowar, Vision, Whitlock (2011)
Data archiving is a good investment.
Nature letter to the editor: 473, p285.

<http://researchremix.wordpress.com/2011/05/19/nature-letter/>

Indispensible

Funders



<http://www.flickr.com/photos/n2artsapes/3527520456/>

innovation and
experimentation



open access to citation data

open access to full text

open access to other use

1. raise our expectations



<http://www.flickr.com/photos/quinnanya/2055471833>

2. raise our voices



[Home](#) • [The Administration](#) • [Office of Science and Technology Policy](#)

Search WhiteHouse.gov

Search



Office of Science and Technology Policy

[About OSTP](#) | [OSTP Blog](#) | [Pressroom](#) | [Divisions](#) | [R&D Budgets](#) | [Resource Library](#) | [NSTC](#) | [PCAST](#) | [Contact Us](#)

Request for Information on Public Access to Digital Data and Scientific Publications

Subscribe

Posted by [Phil Larson](#) on November 07, 2011 at 11:40 AM EST

[E-Mail](#) [Tweet](#) [Share](#) [+](#)

The America COMPETES Reauthorization Act of 2010, signed by President Obama earlier this year, calls upon OSTP to coordinate with agencies to develop policies that assure widespread public access to and long-term stewardship of the results of federally funded unclassified research. Towards that goal, OSTP last week released two Requests for Information (RFI) soliciting public input on long term preservation of and public access to the results of federally funded research, including digital data and peer-reviewed scholarly publications.

GIVE FEEDBACK
ABOUT THIS PAGE

YOUR FEDERAL
TAXPAYER RECEIPT



Launch the Receipt

**3. get excited and
make things**



**GET
EXCITED
AND
MAKE
THINGS**



<http://www.flickr.com/photos/huzzahvintage/4577075021/>

**A future where
data attribution Counts**

**The future is about
what kind of impact
a dataset makes,
not just a citation number.**

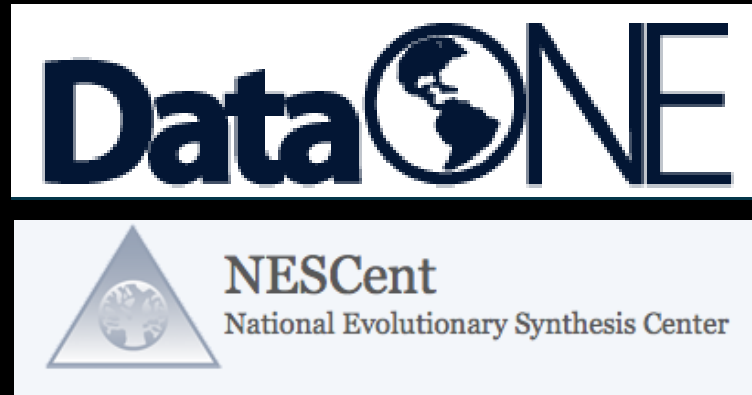
The future is



<http://www.flickr.com/photos/myklorentine/892446624/>

Open data.

Open data about our data.



thank you

Todd Vision,
Jonathan Carlson, Estephanie Sta Maria,
Jason Priem, total-Impact and Beyond Impact
Dryad and DataONE teams

The open science online community and those who
release their articles, datasets and photos openly

*blog: **ResearchRemix.wordpress.com***
@researchremix

- 1. raise our expectations**
- 2. raise our voices**
- 3. get excited and make things**