

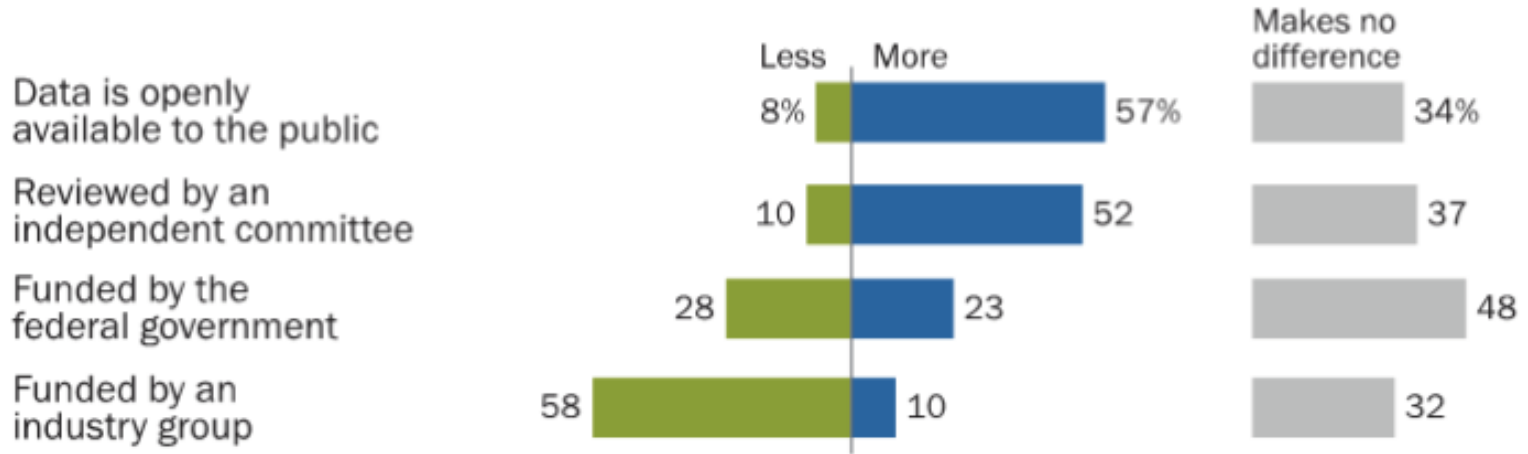
DataSeer

Dr. Tim Vines PhD

tim@dataseer.ai



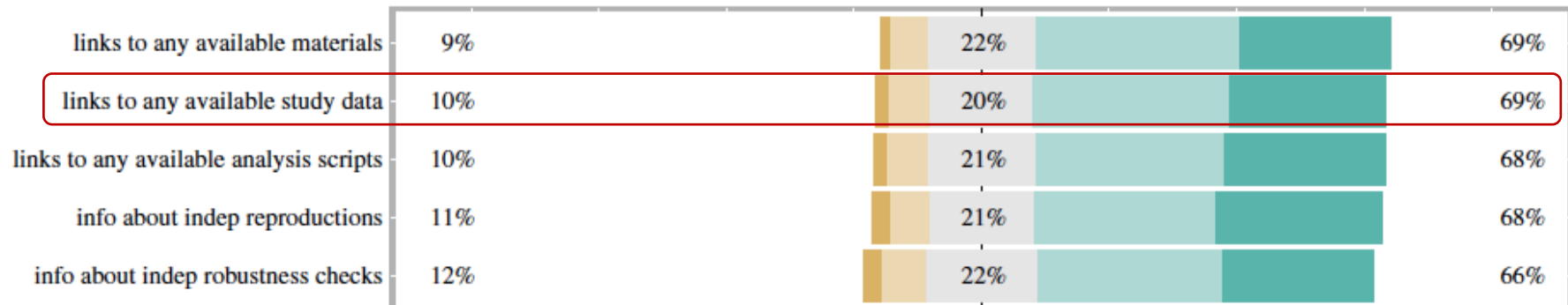
% of U.S. adults who say when they hear each of the following, they trust scientific research findings ...



from 'Trust and Mistrust in Americans' Views of Scientific' by the Pew Research Center (2019)



When assessing the credibility of a preprint
how important would it be to have each of the following pieces of information?



Soderberg CK, Errington TM, Nosek BA. 2020 Credibility of preprints: an interdisciplinary survey of researchers. R. Soc. Open Sci. 7 : 201520.



Promoting open research data

- Essential for trust in science
- But all too often unavailable
 - Only ~10% of published articles share their data
 - Rises to only ~40% when authors *must* share data
- Better open data = quick win for science
 - enables verification
 - enables re-use & synthesis
 - builds trust



Promoting open research data

- To better promote open data we need 2 bits of info:
 - Which datasets are associated with this article?
 - Which of those datasets have been shared?
- Natural Language Processing is ideal for this



DataSeer as a solution

- This is a sentence describing data collection:

Anthropometric measurements

With complete physical examinations, anthropometric measurements and blood samples for participants in the TCHS-E were collected in 2009. Weight and height were measured with an autoanthropometer (super-view, HW-666), with participants shoeless and wearing light clothing.



DataSeer as a solution

- This is a sentence describing data sharing

4). The different responsive pathways indicated that WSSV and the synthetic viral analogue poly (I:C) could induce some different host immune reactions.

All the raw data including the expressed gene lists and the differentially expressed genes (DEGs) lists were supplemented in the Dryad Digital Repository: <https://doi.org/10.5061/dryad.53f1j4d>.



DataSeer's Solutions

1. Open Science Audits

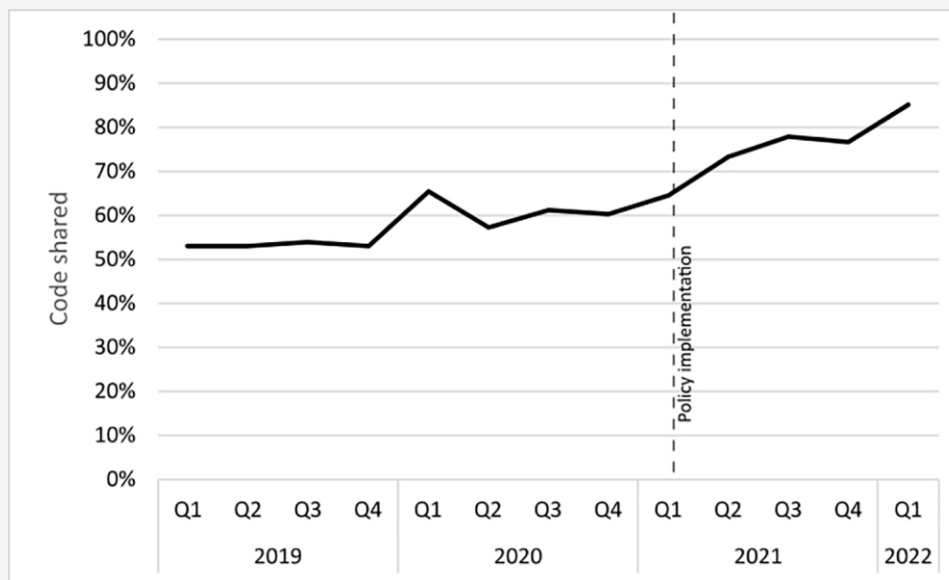
- Does (e.g.) data sharing change through time?
- Did my new policy have any effect?
- Activate ongoing monitoring



DataSeer's Solutions

1. Open Science Audits

- e.g. PLOS Computational Biology Code availability



Cadwallader *et al.* PLOS
Comp Biol
June 2022

DataSeer's Solutions

2. High Fidelity Article Checks
 - Show authors how policies **apply to their articles**
 - Surface & link all research outputs
 - Achieve FAIR open science



Dataset Discovery

Article Metadata

Article Title	Contamination effects on sexual selection in wild dung beetles
Manuscript Number	JEB-2022-00056.R1
Authors	Sebastián Villada-Bedoya, Alex Córdoba-Aguilar, Federico Escobar, Daniel González- Tokman
DataSeer Link	JEB-2022-00056.R1.PDF
Report Date	5/16/2022
Has ReadMe	

Datasets - Image

	Associated Sentence from Article Text	Repository	Notes
Pronotum photos	Pronotum width (linear distance between the outer edges of the pronotum), and cephalic horn length were measured for each beetle from digital photographs taken at 2.0x under a Leica Z16 APO stereomicroscope (with an adapted Leica DMC 2900 camera) using Leica Application Suite version 4.7 software.		Not provided with data repository
Horn Photos	Side-view photographs were used for measuring horn length for <i>C. incertus</i> and <i>E. intermedius</i> (Appendix S2a and b), and for <i>D. gazella</i> , frontal photographs were used because they present a pair of horns on both sides of the head (Appendix S2c).		Not provided with data repository

Datasets - Tabular Data

	Associated Sentence from Article Text	Repository	Notes
contaminate use by ranch	The study was conducted in August 2017 in 19 cattle ranches in Papantla, Veracruz, Mexico, that were selected based on the ease offered by the owners to conduct the samplings. For each ranch, information was obtained by interviewing ranch managers and owners about the use of chemical compounds (i.e., veterinary drugs and herbicides) (Appendix S1).	Appendix S1	Binary use data for herbicides and ivermectin in data repository. (https://datadryad.org/stash/share/FLRjcaoBh9vRp-y1MJz2UE5WVwuh8qknwOScC2oPFvg)
sampling locations	The study was conducted in August 2017 in 19 cattle ranches in Papantla, Veracruz, Mexico, that were selected based on the ease offered by the owners to conduct the samplings. Beetles were collected with pitfall traps baited with cow dung. A total of 3578 individuals were collected from the 19 ranches: 337 <i>C. incertus</i> , 1326 <i>E. intermedius</i> , and 1915 <i>D. gazella</i> .	Appendix S1	Specific location data and sampling numbers in appendix S1
Sexual trait size	Pronotum width (linear distance between the outer edges of the pronotum), and cephalic horn length were measured for each beetle from digital photographs taken at 2.0x under a Leica Z16 APO stereomicroscope (with an adapted Leica DMC 2900 camera) using Leica Application Suite version 4.7 software. Side-view photographs were used for measuring horn length for <i>C. incertus</i> and <i>E. intermedius</i> (Appendix S2a and b), and for <i>D. gazella</i> , frontal photographs were used because they present a pair of horns on both sides of the head (Appendix S2c).	https://datadryad.org/stash/share/FLRjcaoBh9vRp-y1MJz2UE5WVwuh8qknwOScC2oPFvg	Body length was also measured and included in data though this wasn't mentioned in the methods.



DataSeer

Dr. Tim Vines PhD

tim@dataseer.ai

