Mar 30th, 10:15 AM - 11:00 AM

Convergence of Data and Scholarship: Open Access and Reproducibility

Samantha Teplitzky
UC Berkeley, steplitz@berkeley.edu

Anna Sackmann
UC Berkeley, asackmann@berkeley.edu

Convergence of Data and Scholarship:

Open Access and Reproducibility

Sam Teplitzky and Anna Sackmann

Motivating question: how do workflows change?

2010

I see that Geophysical Research Letters and Journal of Geophysical Research A (Space Physics) are on the proposed list of 'electronic only' journals, starting 2011. I strongly oppose this move. We have >100 scientists and students here at [redacted] who publish and read primarily these two journals (we have had a recent JGR senior editor here too). They are the lifeblood of our field and [redacted] is a leader in this field - it seems like a very poor choice to discontinue the print version of these journals.

Motivating question: how do workflows change?

2010

I see that Geophysical Research Letters and Journal of Geophysical Research A (Space Physics) are on the proposed list of 'electronic only' journals, starting 2011. I strongly oppose this move. We have >100 scientists and students here at [redacted] who publish and read primarily these two journals (we have had a recent JGR senior editor here too). They are the lifeblood of our field and [redacted] is a leader in this field - it seems like a very poor choice to discontinue the print version of these journals.

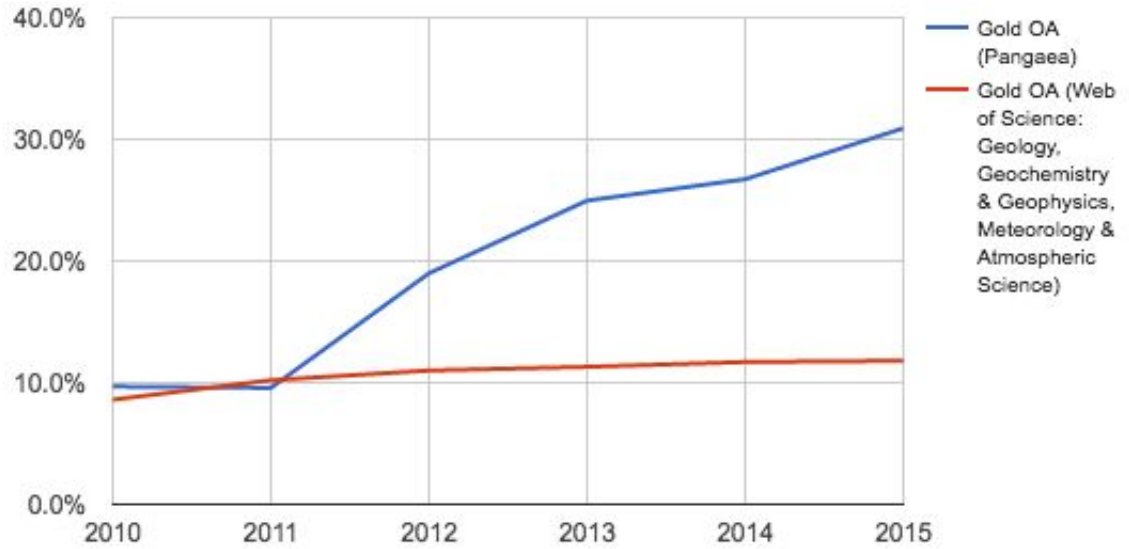
2018

Thanks for soliciting feedback on this. It's a bummer that the library has to cut back on its journals subscriptions but, after looking at the list, I think the impact on my research will be minimal.

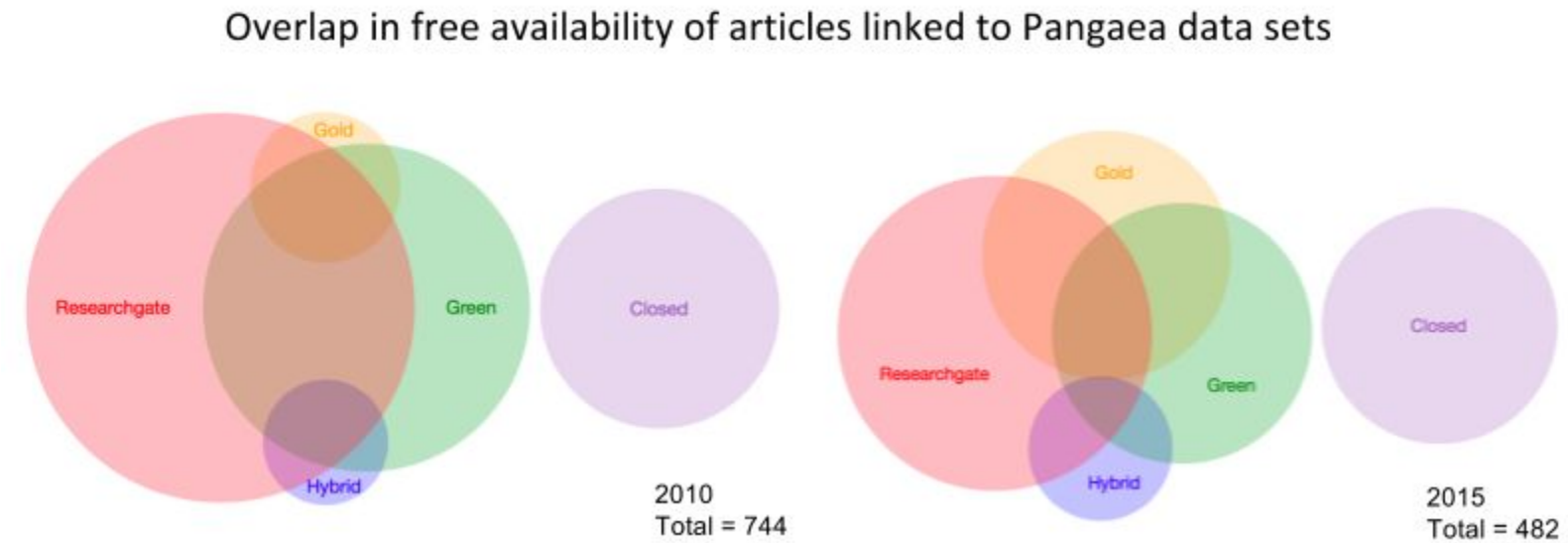
Given that this is solely a print cancellation [redacted] and that we will retain digital access to GSA Special Papers/Field Guides as well as the Lyell Collection [redacted], I have no objections to anything on the list.

Background

Gold Open Access Articles in the Earth Sciences



Background


Overlap in free availability of articles linked to Pangaea data sets
Gold
Researchgate
Green
Closed
Hybrid
2010
Total = 744
Gold
Researchgate
Green
Closed
Hybrid
2015
Total = 482

Trends in open research, open science

- Many types of open access exist concurrently

Trends in open research, open science

- Many types of open access exist concurrently
- Discipline specific preprint servers are on the rise

Trends in open research, open science

- Many types of open access exist concurrently
- Discipline specific preprint servers are on the rise
- Multidisciplinary workflows challenge traditional boundaries

Trends in open research, open science

- Many types of open access exist concurrently
- Discipline specific preprint servers are on the rise
- Multidisciplinary workflows challenge traditional boundaries
- There is a push to publish data sets, with accompanying metadata, and marked up code

Trends in open research, open science

- Many types of open access exist concurrently
- Discipline specific preprint servers are on the rise
- Multidisciplinary workflows challenge traditional boundaries
- There is a push to publish data sets, with accompanying metadata, and marked up code
- The rise of the data article elevates data to the focus rather than a supplement of research

Trends in open research, open science

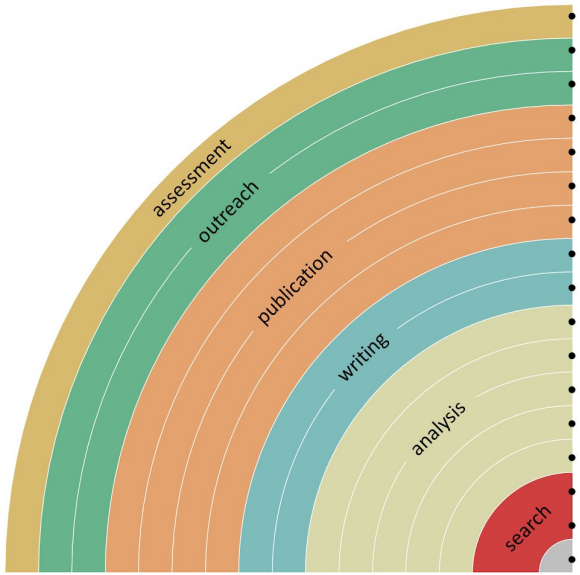
- Many types of open access exist concurrently
- Discipline specific preprint servers are on the rise
- Multidisciplinary workflows challenge traditional boundaries
- There is a push to publish data sets, with accompanying metadata, and marked up code
- The rise of the data article elevates data to the focus rather than a supplement of research
- Researchers are incorporating reproducibility practices in early phases of research to prepare open materials

Research Life Cycle

Source: http://acrl.libguides.com/scholcomm/toolkit/

Rainbow of open science

You can make your workflow more open by …



- adding alternative evaluation, e.g. with altmetrics
- communicating through social media, e.g. Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. CC0 or CC-BY
- publishing open access, 'green' or 'gold'
- using open peer review, e.g. at journals or PubPeer
- sharing preprints, e.g. at OSF, arXiv or bioRxiv
- using actionable formats, e.g. with Jupyter or CoCalc
- open XML-drafting, e.g. at Overleaf or Authorea
- sharing protocols & workfl., e.g. at Protocols.io
- sharing notebooks, e.g. at OpenNotebookScience
- sharing code, e.g. at GitHub with GNU/MIT license
- sharing data, e.g. at Dryad, Zenodo or Dataverse
- pre-registering, e.g. at OSF or AsPredicted
- commenting openly, e.g. with Hypothes.is
- using shared reference libraries, e.g. with Zotero
- sharing (grant) proposals, e.g. at RIO

 Bianca Kramer & Jeroen Bosman https://101innovations.wordpress.com

DOI: 10.5281/zenodo.1147025

Case Studies

Definitions

Open Access: free, immediate, online availability of research articles coupled with the rights to use these articles fully in the digital environment. (https://sparcopen.org/open-access/)

Open Data: Structured data that are accessible, machine-readable, usable, intelligible, and freely shared. (CASRAI)

Data Reuse: The practice of reusing previously generated data. (CASRAI)

Supplementary Materials: Additional data files that contain information directly supportive of the document, for example, an audio clip, movie, database, spreadsheet, applet, or other external file.
(https://dtd.nlm.nih.gov/publishing/tag-library/1.0/n-r2x0.html)

Open licensing: a description of usage terms that grants permission to access, re-use and redistribute a work with few or no restrictions. (http://opendefinition.org/guide/)

Sharing code, protocols or notebooks: Opening the underlying methods necessary to reproduce published research.

Preprint publication: Preliminary version of an article that has not undergone review but that may be shared for comment. Preprints may be considered as grey literature. (CASRAI)

Introduction to case studies

Examples of Open Science	Related roles	Key Issues
Open Access	Researcher	*Ex. Funder Mandates*
Open Data	Funder	
Data Reuse	Collaborators	
Supplemental Materials	Study participant	
Open licensing	Government Agency	
Sharing code, protocols or notebooks	Librarian/ Research data manager	
Preprint publication	Reporter	
Other	Other	

URLs for case studies

Case Study #1: http://ucblib.link/iF

Case Study #2: http://ucblib.link/iC

Case Study #3: http://ucblib.link/iG

Case Study #4: http://ucblib.link/iA

Case Study #5: http://ucblib.link/iz

Definitions

Open Access: free, immediate, online availability of research articles coupled with the rights to use these articles fully in the digital environment. (https://sparcopen.org/open-access/)

Open Data: Structured data that are accessible, machine-readable, usable, intelligible, and freely shared. (CASRAI)

Data Reuse: The practice of reusing previously generated data. (CASRAI)

Supplementary Materials: Additional data files that contain information directly supportive of the document, for example, an audio clip, movie, database, spreadsheet, applet, or other external file.
(https://dtd.nlm.nih.gov/publishing/tag-library/1.0/n-r2x0.html)

Open licensing: a description of usage terms that grants permission to access, re-use and redistribute a work with few or no restrictions. (http://opendefinition.org/guide/)

Sharing code, protocols or notebooks: Opening the underlying methods necessary to reproduce published research.

Preprint publication: Preliminary version of an article that has not undergone review but that may be shared for comment. Preprints may be considered as grey literature. (CASRAI)

Case Study #1:

Assessing Social Contagion in Body Mass Index...

Original Investigation

March 2018

Assessing Social Contagion Index, Overweight, and Obe Natural Experiment

Ashlesha Datar, PhD[1]; Nancy Nicosia, PhD[2]

» Author Affiliations | Article Information

JAMA Pediatr. 2018;172(3):239-246. doi:10.1001/jamapediatrics.201

The New York Times

Your Neighbors, Your Waistline

By Nicholas Bakalar Jan. 24, 2018

Is obesity contagious?

Researchers studied 1,519 military families who were assigned to 38 military bases across the country and living in counties with higher or lower rates of obesity.

The scientists collected data on height and weight of children and parents using questionnaires and in-person examinations. The study, in JAMA Pediatrics, used data on county obesity prevalence from the Centers for Disease Control and Prevention.

ARTICLE INFORMATION

Accepted for Publication: October 25, 2017.

Published Online: January 22, 2018. doi:10.1001/jamapediatrics.2017.4882

Author Contributions: Dr Datar had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Datar.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Datar.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Datar.
Obtained funding: All authors.
Administrative, technical, or material support: Datar.
Study supervision: All authors.

Conflict of Interest Disclosures: None reported.

Funding/Support: This research was funded by National Institute of Child Health and Human Development grant R01HD067536.

Role of the Funder/Sponsor: The funding source and the US Army played no role in the design and conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: We thank Bonnie Ghosh-Dastidar, PhD, and Ann Haas, MS, MPH, RAND Corporation, for constructing corrected BMI measures; and Victoria Shier, MPhil, Pardee-RAND Corporation, and Sarah White, MPP, and Elizabeth Wong, MPH, University of Southern California, for their excellent research assistance. All contributors were supported by the National Institute of Child Health and Human Development grant R01HD067536. We also gratefully acknowledge the US Department of the Army for facilitating data collection for this study.

Case Study #2:

China CO2 emission accounts

Altmetric: 30 Citations: 1 More detail »

Data Descriptor | OPEN

China CO_2 emission accounts 1997–2015

Yuli Shan, Dabo Guan, Heran Zheng, Jiamin Ou, Yuan Li, Jing Meng, Zhifu Mi, Zhu Liu & Qiang

tific Data 5, Article number: 170201
0.1038/sdata.2017.201
Received: 30 May 2017
Accepted: 27 November 2017
Published online: 16 January 2018

figshare search on figshare Browse

China CO2 emission accounts 1997-2015

Published on 21 Nov 2017 - 12:00 by Yuli Shan

This data-set provides the energy and CO2 emissions of China and its 3
1997 to 2015.

CITE THIS COLLECTION

Shan, Yuli; Guan, Dabo; Zheng, Heran; Meng, Jing; Mi, Zhifu; Liu, Zhu (2017): Chi
accounts 1997-2015. figshare.

https://doi.org/10.6084/m9.figshare.c.3936484.v1

MATLAB code

We take Anhui 2007 as an example to show the calculation of emission inventory with MATLAB R2014a using the energy data.

```
% Read emission factors and energy data from the excel files
NCV=xlsread('Emission factors', 'NCV', 'A2:Q2'); %NCV_i refers to Table 1
NCV=repmat(NCV,68,1);
CC=xlsread('Emission factors', 'CC', 'A2:Q2'); %CC_i refers to Table 1
CC=repmat(CC,47,1);
O = xlsread('Emission factors', 'Oxygenation Efficiency', 'B2:R48'); %O_ij refers to Table 3
Energy=xlsread('Province Energy inventory 2007','Anhui','B3:R70');
```

Case Study #3:

Major trauma in winter sports

Download PDF | Add To Library | References 37 | Cited By 5 | RCR: 3.80 | SharedIt

Online access to this article has been provided by Springer Nature SharedIt.
Back to Springer Nature | What's this?

Eur J Trauma Emerg Surg
DOI 10.1007/s00068-015-0596-7

ORIGINAL ARTICLE

Major trauma in winter sports: an international trauma database analysis

C. D. Weber[1,2] · K. Horst[1] · R. Lefering[3] · M. Hofman[1,2] · T. Dienstknecht[1] · H.-C. Pape[1] · TraumaRegister DGU[3]

POPULAR SCIENCE
WANT MORE?

SCIENCE

The Olympic sport most likely to kill you

It's probably not skeleton.

By Erin Blakemore February 23, 2018

Download Unavailable

You are viewing a complimentary shared article.

Complimentary shares do not allow saving. If you have another access method, please visit the regular publisher article page in order to download.

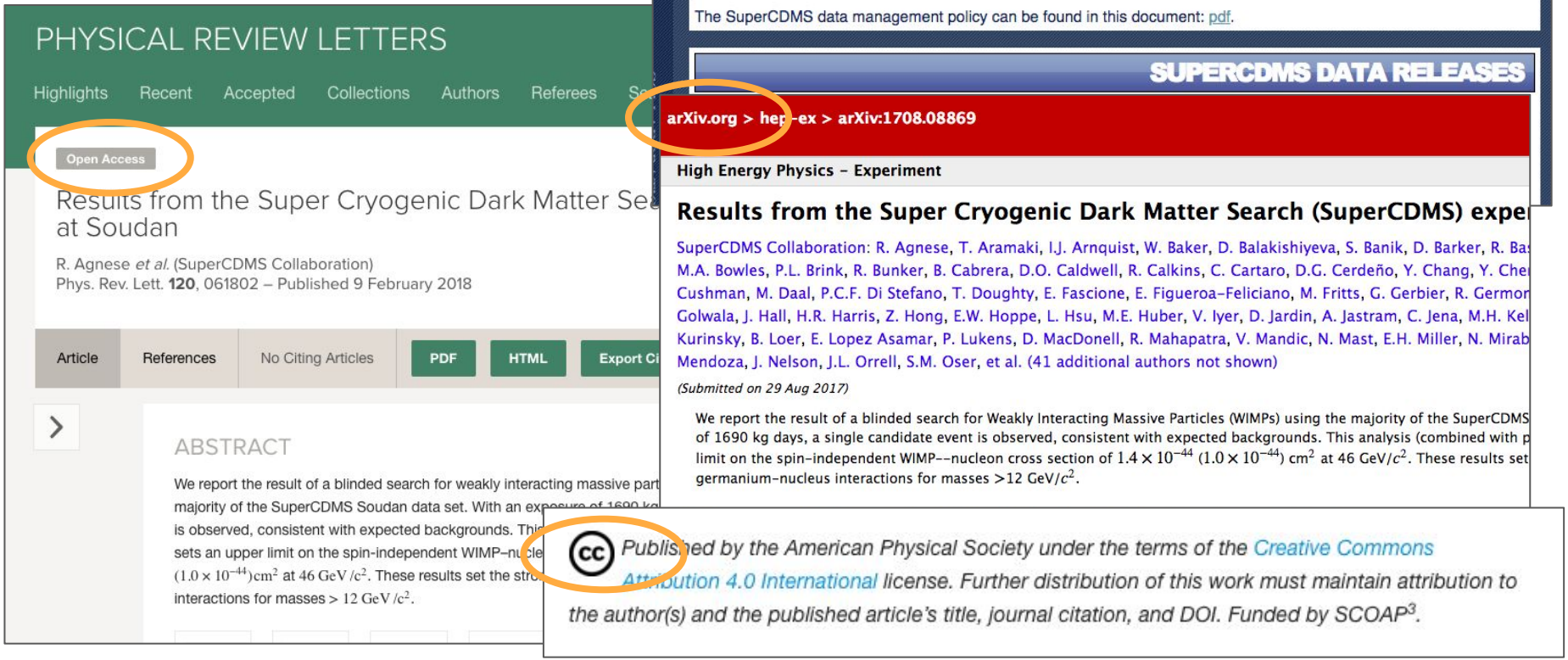
GO TO PUBLISHER PAGE | GOT IT!

Conflict of interest Christian Weber, Klemens Horst, Rolf Lefering, Martijn Hofman, Thomas Dienstknecht, and Hans-Christoph Pape declare that they have no conflict of interest. This study was performed without any financial or other support.

Funding statement The TraumaRegister DGU® was previously partly funded by the Deutsche Forschungsgemeinschaft (Ne 385/5) and by a grant from Novo Nordisk A/S, Bagsvaerd, Denmark. It is now supported by fees from the participating hospitals (a list of hospitals is available at http://www.traumaregister.de) and hosted by the AUC—Academy for Trauma Surgery (AUC).

Case Study #4:

Super Cryogenic Dark Matter Search



SUPER CRYOGENIC DARK MATTER SEARCH

Home | In the News | The Experiment | The Collaboration | Publications | Theses | Public Data | Limit Plotter | Photos

SUPERCDMS DATA MANAGEMENT POLICY

The SuperCDMS data management policy can be found in this document: pdf.

SUPERCDMS DATA RELEASES

PHYSICAL REVIEW LETTERS

Highlights Recent Accepted Collections Authors Referees

Open Access

Results from the Super Cryogenic Dark Matter Se
at Soudan

R. Agnese *et al.* (SuperCDMS Collaboration)
Phys. Rev. Lett. **120**, 061802 – Published 9 February 2018

Article References No Citing Articles PDF HTML Export Ci

ABSTRACT

We report the result of a blinded search for weakly interacting massive part
majority of the SuperCDMS Soudan data set. With an exposure of 1690 ke
is observed, consistent with expected backgrounds. Thi
sets an upper limit on the spin-independent WIMP–nucle
$(1.0 \times 10^{-44})\text{cm}^2$ at 46 GeV/c^2. These results set the stro
interactions for masses > 12 GeV/c^2.

arXiv.org > hep-ex > arXiv:1708.08869

High Energy Physics – Experiment

Results from the Super Cryogenic Dark Matter Search (SuperCDMS) expe

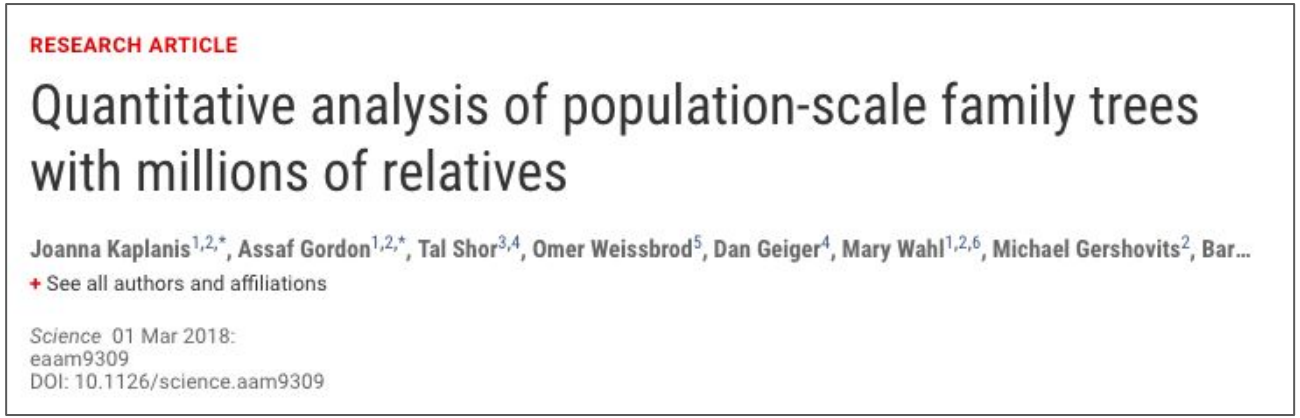
SuperCDMS Collaboration: R. Agnese, T. Aramaki, I.J. Arnquist, W. Baker, D. Balakishiyeva, S. Banik, D. Barker, R. Ba
M.A. Bowles, P.L. Brink, R. Bunker, B. Cabrera, D.O. Caldwell, R. Calkins, C. Cartaro, D.G. Cerdeño, Y. Chang, Y. Che
Cushman, M. Daal, P.C.F. Di Stefano, T. Doughty, E. Fascione, E. Figueroa-Feliciano, M. Fritts, G. Gerbier, R. Germor
Golwala, J. Hall, H.R. Harris, Z. Hong, E.W. Hoppe, L. Hsu, M.E. Huber, V. Iyer, D. Jardin, A. Jastram, C. Jena, M.H. Kel
Kurinsky, B. Loer, E. Lopez Asamar, P. Lukens, D. MacDonell, R. Mahapatra, V. Mandic, N. Mast, E.H. Miller, N. Mirab
Mendoza, J. Nelson, J.L. Orrell, S.M. Oser, et al. (41 additional authors not shown)

(Submitted on 29 Aug 2017)

We report the result of a blinded search for Weakly Interacting Massive Particles (WIMPs) using the majority of the SuperCDMS
of 1690 kg days, a single candidate event is observed, consistent with expected backgrounds. This analysis (combined with p
limit on the spin-independent WIMP--nucleon cross section of 1.4×10^{-44} (1.0×10^{-44}) cm^2 at 46 GeV/c^2. These results set
germanium-nucleus interactions for masses >12 GeV/c^2.

Case Study #5:

Analysis of population-scale family trees



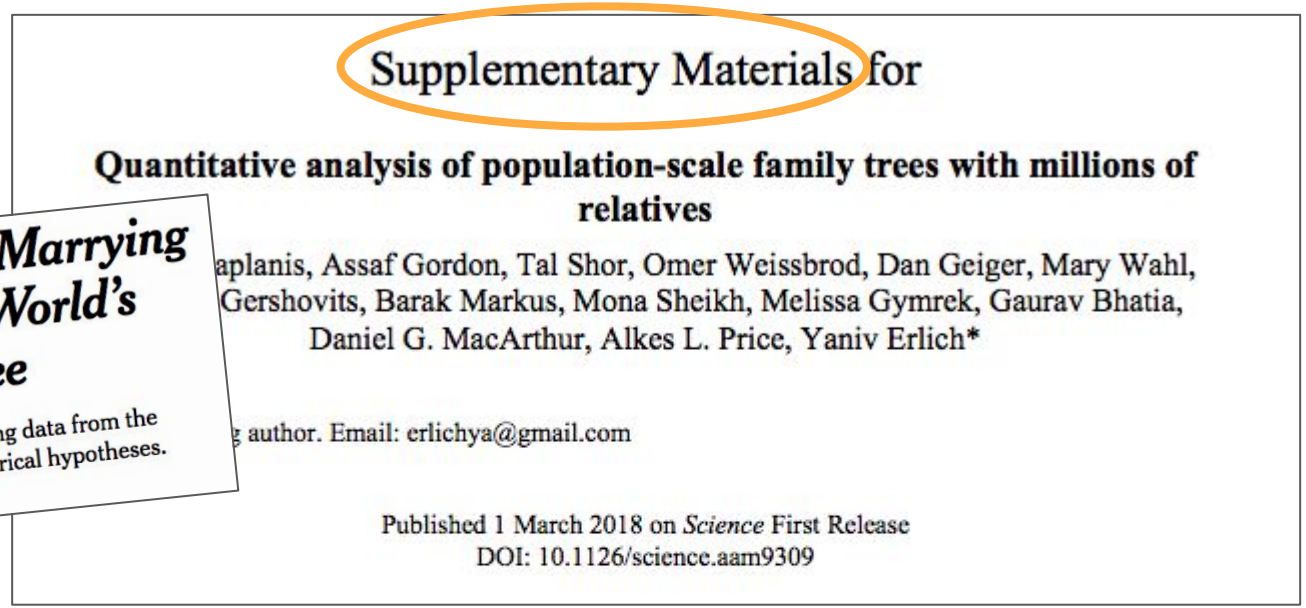
RESEARCH ARTICLE

Quantitative analysis of population-scale family trees with millions of relatives

Joanna Kaplanis[1,2,*], Assaf Gordon[1,2,*], Tal Shor[3,4], Omer Weissbrod[5], Dan Geiger[4], Mary Wahl[1,2,6], Michael Gershovits[2], Bar...

+ See all authors and affiliations

Science 01 Mar 2018:
eaam9309
DOI: 10.1126/science.aam9309



Supplementary Materials for

Quantitative analysis of population-scale family trees with millions of relatives

aplanis, Assaf Gordon, Tal Shor, Omer Weissbrod, Dan Geiger, Mary Wahl,
Gershovits, Barak Markus, Mona Sheikh, Melissa Gymrek, Gaurav Bhatia,
Daniel G. MacArthur, Alkes L. Price, Yaniv Erlich*

g author. Email: erlichya@gmail.com

Published 1 March 2018 on *Science* First Release
DOI: 10.1126/science.aam9309

When Did Americans Stop Marrying Their Cousins? Ask the World's Largest Family Tree

Researchers assembled 5 million family trees using data from the website Geni.com to test several genetic and historical hypotheses.

Common Themes

Questions?

Contact information:

Sam Teplitzky steplitz@berkeley.edu
Anna Sackmann asackmann@berkeley.edu

Slides: http://ucblib.link/iE

References

Teplitzky, S. (2017). Open data, [open] access: linking data sharing and article sharing in the Earth Sciences. Journal of Librarianship and Scholarly Communication, 5(1), eP2150. DOI: http://doi.org/10.7710/2162-3309.2150

Rainbow of Open Science: https://zenodo.org/record/1147025#.WoTHxZM-dE6

Research Life Cycle: http://acrl.libguides.com/scholcomm/toolkit/

Case Study Links:

- Case Study #1: http://ucblib.link/iF
- Case Study #2: http://ucblib.link/iC
- Case Study #3: http://ucblib.link/iG
- Case Study #4: http://ucblib.link/iA
- Case Study #5: http://ucblib.link/iz