

ADSA Data Science Community Newsletter

Data Science Community Newsletter features journalism, research papers and tools/software for March 24, 2022.

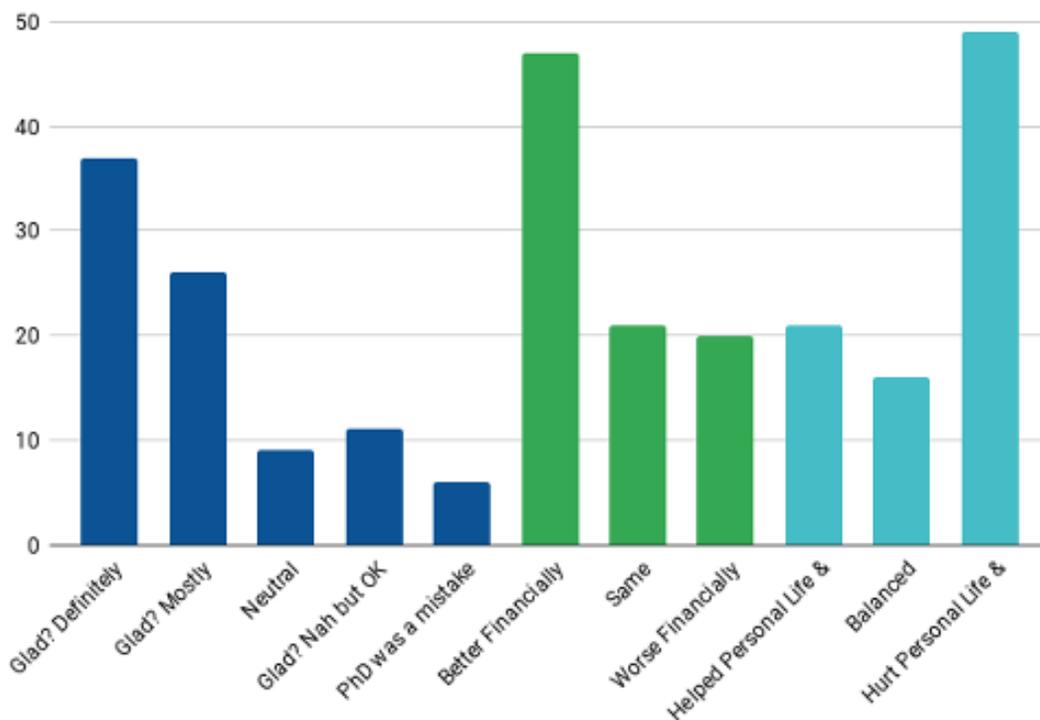
Please let us ([Micaela Parker](#), [Catherine Cramer](#), [Brad Stenger](#), [Laura Norén](#)) know if you have something to add to next week's newsletter. We are grateful for the generous financial support from the [Academic Data Science Alliance](#).

GRAD SCHOOL: WORTH IT.

Is grad school "worth it"? is our most popular poll yet In summary, most people are glad that they got an advanced degree, feel that it has benefitted them financially, and feel they made sacrifices in their personal lives and relationships as a result of getting their degree.

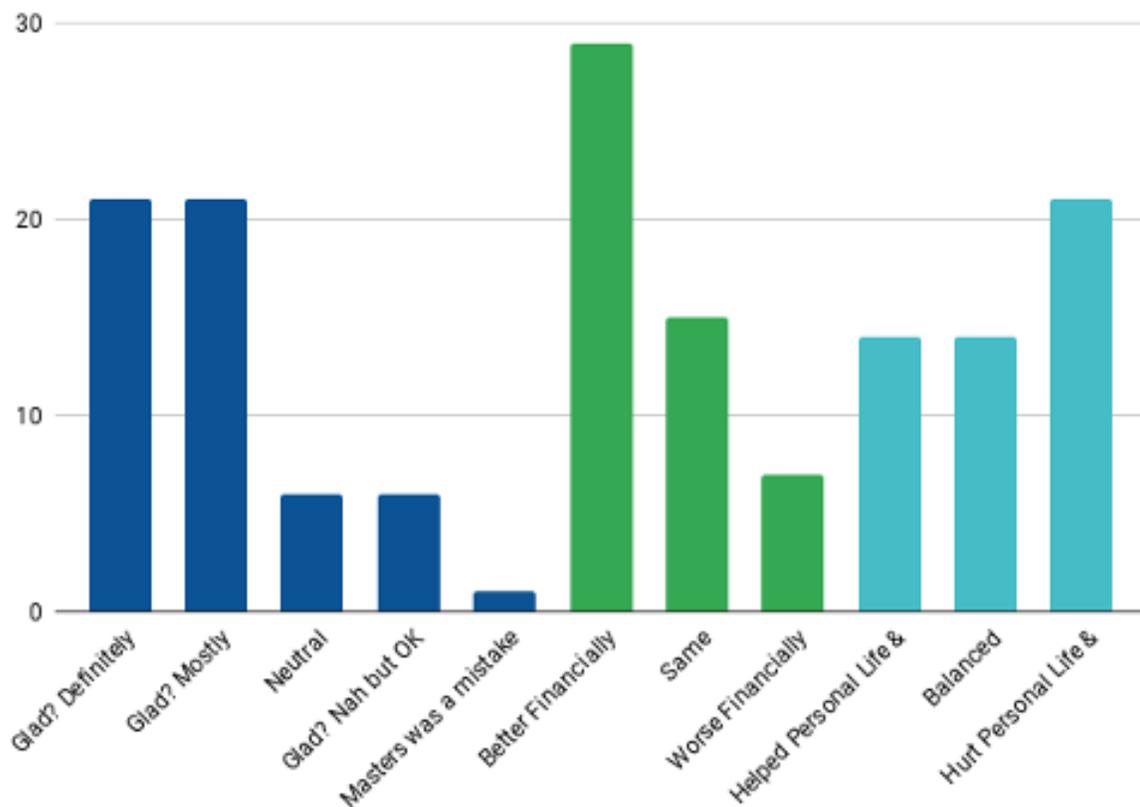
Getting a PhD

PhD Acquisition - Overall Glad?, Impact on Finances, on Personal Life & Relationships (N=89)



Getting a Masters

Masters Acquisition, impact on financial future, and on personal life & relationships (N=55)



Worth noting: 11% of respondents reported experiencing sexual harassment and/or sexual assault while in grad school. These are mostly people who don't identify as men (80%), but it is important to note that it happens to men, too.

From conversations with those contemplating going back for an advanced degree, the biggest stumbling blocks appear to be: the requirement to uproot oneself for 2-6 years and whatever collateral damage that will have on personal lives and relationships; being cash-strapped at a time when family formation is underway — getting married, adopting pets, having children, finding housing that works for those new families. For universities that are eager to keep their graduate cohorts large and happy, it may be time to contemplate learning from the "executive masters" degrees that require regular, periodic in-person interactions on campus without requiring students to fully relocate.

If you didn't take the poll, please take it now so we have enough data to report on cross-tabs such as: Are scholars in certain fields more likely to be happy/unhappy than scholars in other fields? More likely to report financial upsides? More likely to report sexual harassment/assault? Are men/women/non-binary people more likely to report making sacrifices in their personal lives for grad school?

[Respond to Poll](#) ... if you haven't already.

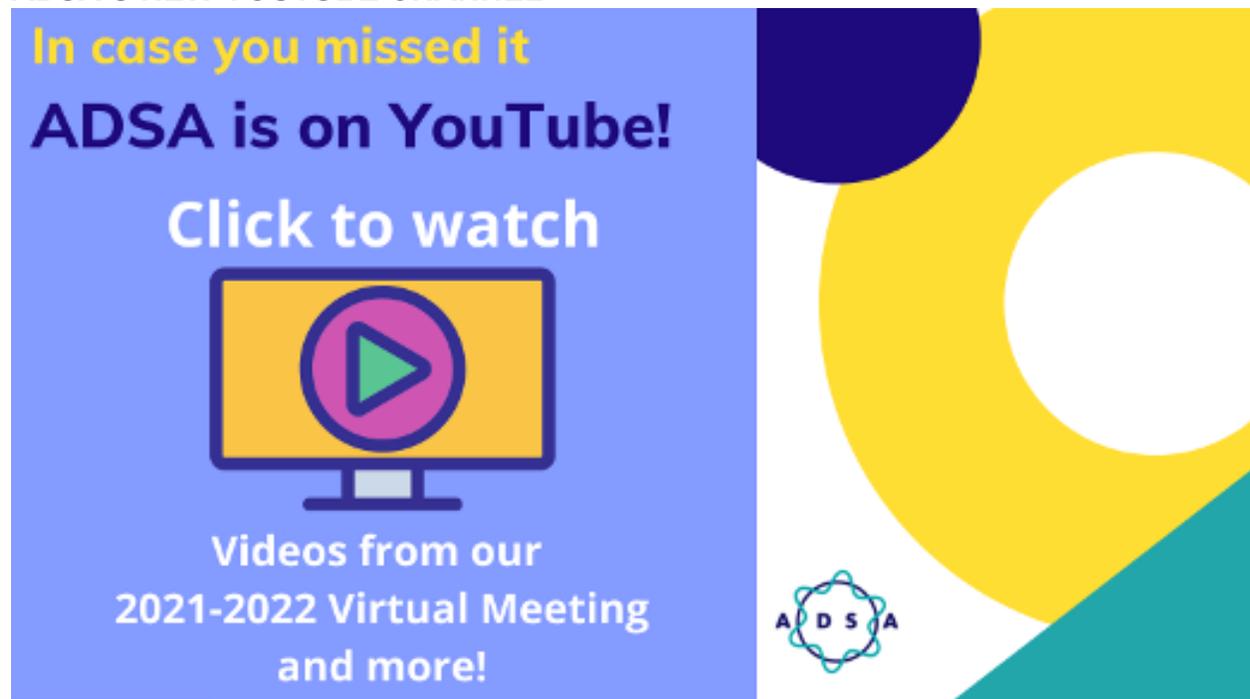
ACADEMIC CONFERENCES: ALSO WORTH IT.

A new **Northwestern University** [study](#) by first author **Emma Zajdela** finds that attending academic conferences is worth it. Researchers who attend sessions have more and better collaborations. But the kicker is that it didn't seem to matter if the conferences were in-person or virtual. As we continue to live with waves of new variants and climate change, it's excellent to hear that virtual conferences are functioning well as networking opportunities.

RIT OFFERS 50% OFF MASTERS DEGREE TUITION FOR RECENT ALUMS

Rochester Institute of Technology alumni who graduated between December 2019 and December 2022 are encouraged to enroll in RIT masters programs by [getting 50% off tuition](#). The program was originally designed to support "students who were graduating during the early pandemic and into a difficult job market," according to **Diane Ellison**, RIT senior associate vice president of Graduate Enrollment Services. The program has already enrolled 300 students.

ADSA'S NEW YOUTUBE CHANNEL

A graphic with a blue background on the left and a yellow and teal background on the right. The blue section contains the text: "In case you missed it" in yellow, "ADSA is on YouTube!" in white, "Click to watch" in white, an icon of a computer monitor with a play button, and "Videos from our 2021-2022 Virtual Meeting and more!" in white. The yellow section features a large white play button icon and the ADSA logo at the bottom.

Please subscribe to [ADSA's YouTube channel](#) so you can get access to talks and panels as they become available.

NEW NIST FRAMEWORK ON AI FAIRNESS

NIST has released a new [draft framework](#) for addressing AI Fairness. There's a [comment](#) period open through April 29, 2022 and a two-part workshop on March 29-31 ([registration link](#)). It's good to see socio-technical considerations hold equal weight with technical considerations, both of which are underpinned by fairness, accountability, and transparency principles. Would have liked to see harm avoidance/reduction in that list of principles, too, but that's why there is a comment period!

MODEL ELITISM IN AI

Nithya Sambasivan, most recently of **Google Research**, has a [new ACM paper](#) arguing that "the overt emphasis on model performance" is bad for data science and bad for the living beings it ostensibly aims to serve. Many who read this newsletter already avoid some of the problematic

attributes of our field in which “the enormous focus on model development” tends to relegate critically important aspects to less-respected “operations” work. In particular, Sambasivan warns that:

- Ignoring the importance of collecting data from the ‘real world’ conditions could render a model utterly useless in the ‘real world’ even if it achieves paper-winning performance using a spotless dataset. “Eye-disease-detection models, trained on noise-free datasets to improve model performance, can result in total failure in predicting retinal disease when there are even small specks of dust on a camera lens.”
- Treating data collection, cleaning, curation, and interpretation as scut work and the people who do it as non-experts is a dangerous type of AI-elitism. “Domain experts...are necessary for AI projects in areas with poor infrastructure where there are limited datasets available. These underpaid and overworked domain experts are often drafted to perform AI data collection for free”.

She also takes a big swing at algorithmic fairness, explaining why algorithmic fairness approaches developed in the US cannot simply be exported to other contexts. She argues that “the emphasis on models in algorithmic fairness offers a veneer of credibility to system builders, but when examined closely, these frameworks can be dangerously symbolic, non-generalizable to non-Western contexts.”

REAL-WORLD APPLICATIONS OF AI & TRIUMPH OF VALIDATION

Doctors in two different Pennsylvania hospitals using two different stroke-predictor models [both saw incremental improvements](#) in treating stroke patients. “Doctors tracked the time from when patients got their CT scans to when they actually received incisions for a procedure to unblock the artery, both before the introduction of the artificial intelligence and after. The time dropped significantly, from an average of 93 minutes before use of artificial intelligence to 68 minutes after.”

This stroke diagnostic success might be an exception relative to other AI-based diagnostics especially if those apps tout “explainable AI,” [writes Jeremy Kahn](#), who pens *Fortune's Eye on AI* newsletter. Kahn cites **Carnegie Mellon** professor **Zachary Lipton** in saying that explainable AI is useless at best to medical imaging clinicians and their patients. Bias isn't accounted for with these interfaces. Mostly it's human bias where users believe that the AI problem-solves like they do, only better, and so misinterpret the AI *and* the explanation. What matters, Kahn concludes, is validation, not explanation. Validation, like we just saw with the strokes app.

Mismatch problems between clinicians and their AIs also occur with training data in clinical apps. **The Mayo Clinic's John Halamka**, **Johns Hopkins' Suchi Saria**, and **Stanford's Nigam Shah** [warn](#) of the dire consequences data shifts — “which happen when an algorithm must process data that differ from those used to create and train it” — can have in health care settings.

STANFORD HAI'S 2022 AI INDEX [NEW]

Stanford's Human-Centered Artificial Intelligence Institute [published](#) its [2022 AI Index](#) which finds that ethics and fairness are now mainstream concerns based on the number of industry papers published on the topic. They also flagged the big leaps made in natural language processing, noting that one of the final big challenges is to rid them of the toxicity and social bias present in naturally spoken languages. Asking a model to do as we think we should say, not what we actually say, is quite difficult.

ANDREJ KARPATHY REPRODUCED YANN LECUN'S DEEP NEURAL NET ZIP CODE RECOGNITION PAPER

The new AI crew — **Andrej Karpathy** — reproduces work by the old guard — **Yann LeCun**. Karpathy tweeted, "we reproduce what I think may be the earliest real-world application of a neural net trained end-to-end with backprop (LeCun et al. 1989), try to improve it with time travel, and reflect." LeCun [quote-tweets: "fun blog post by Andrej."](#)

CLIMATE CHANGE NEWS

In an astounding co-occurrence both the Arctic and the Antarctic [experienced](#) heat waves of 50 and 70 degrees F above normal since our last issue. That's hard to pull off considering it's spring in one hemisphere and fall in the other (summer in neither).

In better [news](#), suburban trees were found to suck up more carbon dioxide than those deep in the forest while urban soils release less than soil in more ecologically active areas farther out. It's great when there are small things individuals can do that make a difference.

NIH ALL OF US RELEASES 100,000 HUMAN GENOMES

Following the [world's largest release](#) of 200,000 whole human genomes by the **UK Biobank's** last November, the **National Institutes of Health** in the US just [released](#) a batch of 100,000 whole human genomes from its All of Us project. The All of Us genomes are only available to researchers based at US institutions. The UK Biobank data is available globally.

BUT NEW DEFENSE APPROPRIATIONS GUT NIH LINE ITEM

The war in Ukraine is having a direct impact on the future of science in the US. The Biden Administration just [signed](#) a large defense appropriations bill, hollowing out a proposal that would have sent \$6.5 billion to stand up **ARPA-H**, but will now include only \$1 billion. Overall, the "shift in priorities shrunk Biden's proposed 21% hike, much of it for a new Advanced Research Projects Agency for Health (ARPA-H) to develop cutting-edge medical treatments, to 5%."

AVONET — AN OPEN SOURCE BIRD ATLAS

When doctoral candidate **Catherine Sheard** decided to [create a catalog](#) of the anatomical measurements of all 6000 passerine (perching) birds, it ended up catalyzing a [large international project](#) to "publish anatomical measurements of all 11,009 living bird species". The AVONET data will enable many research projects, including those investigating the impact of climate change on bird behaviors and populations.

Magpies may not mind being measured, but they dislike wearing trackers. There's evidence from studies in Australia that [they work together](#) to remove tiny backpack-like scientific tracking devices from each other.

Another project involving tracking animals by attaching sensors to them has been going much better. Spearheaded by researchers at **Yale University**, but [launched globally](#), the goal is to track 10,000 birds, bats, other mammals, insects, and reptiles across 15 species.

ARXIV BANS PHYSICIST FOR 6 MONTHS

A [tumultuous debate](#) among physicists about a recent preprint on room temperature super conductivity "boiled over" as "people's emotions became too affected. They got acrimonious." **University of California, San Diego** physicist **Jorge Hirsch** has been banned from posting for 6 months after "repeatedly publishing inflammatory content...making repeated submissions in response to specific papers and contacting arXiv administrators and board members to complain." Being banned is a serious consequence for Hirsch who said, "I can't work if I can't publish papers."

UNIVERSITY STAKEHOLDERS IN STUDENT SPORTS BETTING

With the legalization of sports betting, a handful of universities (e.g. **University of Colorado**), have signed deals in which [they get paid \\$30](#) for every student who downloads a betting app. Addictive behaviors like gambling are a particular concern among 18-25 year olds whose brains are not yet fully mature. Gambling is a deadly form of addiction with [high rates](#) of suicidality.

COLUMBIA PROFESSOR DISMANTLES COLUMBIA'S #2 RANKING

Columbia professor of Mathematics, **Michael Thaddeus**, has [completely, methodically, publicly obliterated](#) Columbia's #2 ranking (tied with **MIT** and **Harvard**; behind **Princeton**). If you happen to teach a data science class and want to demonstrate to students how they may reverse engineer claims by using historical records and public data, Thaddeus has provided a primer. If you happen to enjoy throwing proverbial tomatoes at popularity contests, thank us (and **Justin Wolfers'** [tweet](#)) for surfacing this one.

COASTAL CITIES MAY HAVE HIT PEAK TECH

With new work from home policies and continuing housing price disparities between the coast and the middle, it seems that we may have hit peak coastal tech. **Brookings Institution** analysts [took a look](#) at the dispersion of tech jobs and found that in "most superstar (e.g. NYC, SF) and rising star (e.g. Seattle) metro areas, tech sector employment growth in 2020 slowed compared to the previous five years." A **LinkedIn** study also found that the number of jobs advertised in the Seattle area did not match the number of people updating their locations to Seattle, suggesting that many workers are either still remote or are permanently remote.

The cities that are gaining tech workers more quickly are all over. They include "northern business cities such as Philadelphia, Minneapolis, and Cincinnati; sizable warm-weather cities such as Charlotte, N.C., San Antonio, Nashville, Tenn., Birmingham, Ala., New Orleans, Greensboro, N.C., Jackson, Miss., and Stockton, Calif.; and a number of substantial university cities such as Chapel Hill, N.C. and Madison, Wis. Also seeing accelerated 2020 tech growth were numerous lifestyle, Sun Belt, or vacation centers such as Virginia Beach, Va., Ogden, Utah, Albuquerque, N.M., Tucson, Ariz., and El Paso, Texas."

Our Data Viz of the Week maps where tech clusters are waxing and waning.

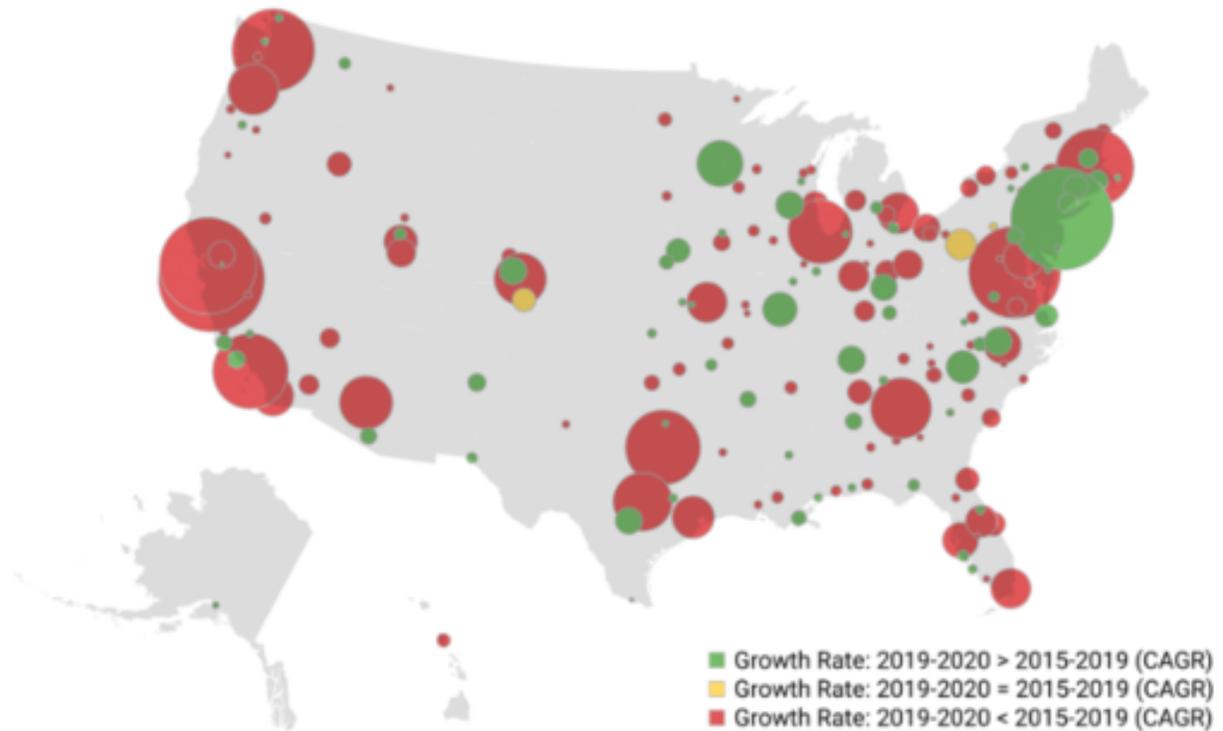
NEW PROGRAMS, FOLLOW THE MONEY

Click through to access [a structured spreadsheet](#) of New Programs and money moving around in academic data science.

DATA VISUALIZATION OF THE WEEK

The Brookings Institution blog, Mark Muro and Yang Yuo from March 8, 2021

Map 1. Change in tech sector growth rates in selected metros, 2015-2019 and 2019-2020



Source: Brookings analysis of Emsi Burning Glass data
Note: Bubble size reflects number of tech jobs in 2020. This map shows the 100 largest metropolitan areas and metros with tech workers exceeding 1,000 in 2020

B | Brookings Metro

Deadlines

Conferences

[CALL FOR ABSTRACTS: The Cascadia Symposium on Statistics in Sports \(Sept 24, 2022 in Vancouver\)](#)

"Deadline is May 15."

RFPs

[New Bloomberg Philanthropies City Data Alliance Launches in the U.S., Latin America, and Canada with \\$60 Million Investment](#)

City Data Alliance "will identify 100 cities from across North and South America that are the most sophisticated at using data to make policy decisions and propel them to an even higher standard with executive education, direct-to-city coaching, and a focused investment to improve a critical data capacity." Deadline to apply for first cohort is March 31.

Tools & Resources

[Textbook Introduces Students to Mechanistic Data Science](#)

Northwestern University, *Northwestern Engineering* from March 16, 2022

"Traditional data science methodologies require copious quantities of data to show a reliable pattern, but the amount of required data can be greatly reduced by considering the mathematical science principles. These data science and mechanistic analysis steps are presented in an intuitive manner that emphasizes practical concepts for solving problems such as tracking the spread of COVID-19 and other contagious viruses, developing autonomous vehicles, and improving computer vision."

"The whole idea is that I can teach a student on all levels," [Wing K. Liu] said. "We use MDS to integrate the concepts of STEM together."

[If you don't know Open Storage Network https://openstoragenetwork.org, you should.](https://openstoragenetwork.org)

Twitter, Ian Foster from March 14, 2022

"OSN pods across the US host many datasets, including Level 1 Terra Basic Fusion files <https://digirolamo.web.illinois.edu/projects/terra-fusion/>. Here I describe how we use @Globus to enable easy and rapid access to Terra data on OSN."

[Excellent blog post by @cHHillee walking through the basics of actually making models scale across GPUs and across nodes.](#)

Twitter, Stella Rose Biderman, Horace He from March 15, 2022

"Highly recommended reading on an important topic that doesn't get explicitly taught to a lot of ML people."

Featured Events

See the [ADSA Events Page](#) for more details and more opportunities.

About Us: The Data Science Community Newsletter was founded in 2015 in the Moore-Sloan Data Science Environment at NYU's Center for Data Science. We continue to be supported by the Gordon and Betty Moore Foundation and the Alfred P. Sloan Foundation through the [Academic Data Science Alliance](#). Our archive of newsletters is at <https://academicdatascience.org/resources/newsletter>. Our mailing address is [1037 NE 65th St #316; Seattle, WA 98115](#).