

## ADSA Data Science Community Newsletter

**Data Science Community Newsletter** features journalism, research papers and tools/software for April 29, 2021.

Please let us ([Micaela Parker](#), [Steve Van Tuyl](#), [Catherine Cramer](#), [Brad Stenger](#)) 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](#).

**At last check** there were 569 responses to our audience survey. If you responded - thank you! If you have not weighed in, please search your inbox for an email from [laura.noren@nyu.edu](mailto:laura.noren@nyu.edu) and help us figure out who you are, what you like/dislike, and what you're planning to do next year in terms of conferences and data science training.

### Academic Data Science News

One of the hubs evolving most quickly in the university research data ecosystem is the library. Responding to the pandemic is one factor, resulting in increased use of e-books and streaming services that provide remote access. Another is a recent push to make libraries [more responsive](#) to BIPOC students and broader communities in general, as well as highlighting the rise in diversity of academic library staff. **Brown University's** new [Center for Library Exploration and Research](#) will provide enhanced training in the use of research data, but will also extend beyond the university walls and work with local middle and high schools and other community partners to improve data literacy and make the university's digital assets accessible. And the library at **University of North Carolina at Chapel Hill** has [launched Story Archive](#), providing an opportunity for members of underrepresented groups to record their personal stories and have the recordings permanently archived at the university library. However, according to [a recent survey](#) conducted by **Ithaka S+R**, while these changes are highly valued, change comes to libraries slowly. According to **Trevor A. Dawes**, vice provost for libraries and museums at the **University of Delaware**, the number of Black library leaders remains where it was 30 years ago.

An area in which university libraries are successfully seeing change is in the development of new programs that specifically target both their research and data science goals. The move toward e-books is bolstered by a new partnership allowing [libraries to buy university press collections in e-format](#), creating a sustainable revenue stream for presses. **Brandeis Library** announces a [new repository for scholarly research](#). And a [student consultation program](#) at **Virginia Tech** helps researchers find new data tools and resources and it gives students experiential learning opportunities in uses for research data. **VTech University Libraries** also house the VTechWorks Open Repository, which has just been [granted open access policy](#). Open access has been shown to increase societal impact as well as citation rates. Data curation needs are also evolving quickly, often out-pacing the knowledge and skills of even the most seasoned digital librarian. To bridge the gap, the data curation community has been developing a new set of tools – [data curation primers](#).

The **University of Chicago Department of Computer Science** [wrote a long profile](#) of **Victor Veitch**, the school's

first hire in its data science initiative. Veitch will split time between UC and **Google Cambridge** where he works in the causality group.

**Woods Hole Oceanographic Institution** and **Analog Devices, Inc.**, both based in Massachusetts, [launched the Ocean and Climate Innovation Accelerator](#) (OCIA). The consortium will focus on "the networked ocean" and plans to provide grants and awards for seed-funding and bridge-funding selected projects. The **Carnegie Institution of Science** which, like WHOI, is a private research non-profit, is undergoing a major reorganization. The Institution [sold its landmark Washington DC headquarters building](#) to the **government of Qatar**, despite Qatar's dismal human rights record.

**Penn State University** has reorganized the way it approaches Artificial Intelligence with a new initiative called [AI Hub](#). "**Penn State's Institute for Computational and Data Sciences** (ICDS) and the **Social Science Research Institute** (SSRI) are collaborating to grow AI across all Penn State colleges and campuses. The AI Hub will serve as an entry point for Penn State faculty interested in finding collaborators working on AI," said **Lora Weiss**, senior vice president for research. Within the Hub and housed in the **School of Electrical Engineering and Computer Science**, Penn State [will also establish](#) the **Center for Artificial Intelligence Foundations and Engineered Systems** (CAFE), directed by **Vijaykrishnan Narayanan**. Penn State also [recently joined](#) the **Public Interest Technology University Network** (PIT-UN), an umbrella organization managed by **New America Foundation** with 43 U.S. universities all working on public interest technology.

The **University of Maryland** will have a [new undergraduate major in social data science](#), based at the school's **Joint Program in Survey Methodology** and the **College of Information Studies**.

**University of Florida** now offers undergraduates a [new AI Fundamentals and Applications Certificate](#). Or it will offer the certificate once the required courses – one in AI Fundamentals, another in AI Ethics and a third course in the student's home college – are available to students this Fall.

The **American University of Paris** has launched a [new Masters degree in Human Rights and Data Science](#). The program combines a rigorous foundation in data science with the legal and philosophical considerations necessary to implement appropriate protective policies and regulations.

**Mellon College of Science at Carnegie Mellon University** launched a [new Master of Science in Data Analytics for Science](#) (MS-DAS). "It is a program created by scientists to equip the next generation of scientists with modern data analytics tools, machine learning skills and relevant mathematics and statistics backgrounds," according to **Manfred Paulini**, who spearheaded the development of the degree program.

**Del Mar College**, a community college in Corpus Christi, Texas, is set to offer its five-course, 14-hour [Artificial Intelligence Certificate program](#) to students starting this Fall.

Prospective students eager to earn their PhD in Quantum Science and Engineering can apply to **Harvard University's** [brand new program](#). Let's hope that the scholars/pioneers can bolster the dormant research program at **Haggar**, where 20 years ago [physicists invented Quantum Slacks](#), wrinkle-free pants that "paradoxically behave like both formal and casual wear," as reported by satirists at **The Onion**.

**Capital One** has [gifted \\$2 million](#) to the **University of Virginia School of Data Science**. The new 60,000-square-foot data science building going up at the corner of Emmet Street and Ivy Road in Charlottesville will henceforth be called the Capital One Hub at the School of Data Science.

**Rutgers University Center for Discrete Mathematics and Theoretical Computer Science** [received its first Simons Foundation](#) grant to establish the Center's Simons Postdoctoral Leadership Initiative. There is substantial funding available for four postdocs, plus a "Hot Topics Fund" to support future large events.

The home improvement retailer **Lowe's** [donated \\$1.5 million](#) to the **University of North Carolina-Charlotte**, earmarking the gift for the school's **College of Computing and Informatics**. The donation will establish an endowed faculty position whose research will focus on machine learning.

The **University of Toronto** is launching a [new global consortium](#) "dedicated to using artificial intelligence to accelerate the design and discovery of advanced materials." The new **Acceleration Consortium** will be led by **Alan Aspuru-Guzik**. No information about funding has been made public.

**Icahn School of Medicine at Mount Sinai** will offer [a new PhD concentration](#) in Artificial Intelligence and Emerging Technologies in Medicine (AIET). **Hayit Greenspan** and **Alan C. Seifert** are the new AIET Co-Directors. Enrollment will begin in the Fall 2022.

**Arizona State University School of Arts, Media and Engineering** received a [three-year, \\$100,000 grant](#) from **National Endowment for the Humanities** to create an undergraduate certificate program to study artificial intelligence in culture.

Arizona State is also partnering with **Starbucks** on [a joint research facility](#). It's hard to tell exactly where the research lab starts and the coffeehouse stops, but maybe that's the idea.



Efforts are increasing to do away with the GRE as an evaluation and entrance requirement for graduate programs in the US. Join the Academic Data Science Alliance for a discussion with Dr. Sandra Petersen, Professor in the Department of Veterinary and Animal Sciences at UMass Amherst and coordinator with [Beyond the GRE](#), hosted jointly by ADSA's Education and DEI [Special Interest Groups](#). Learn more and [REGISTER HERE](#)

## Editor's Picks

The **European Commission** [announced](#) new regulations for artificial intelligence applications. Facial recognition and other real-time biometric identifiers are subject to constraints. Companies that fail to comply face fines up to 6% of revenue. These new rules are subject to approval by the **European Parliament**. **NYU's Meredith Broussard** put her notes on the [proposal](#) to regulate Artificial Intelligence on **Twitter**. One Broussard highlight: "Less legal wrangling about what tech firms can and can't do with AI/data. This is a relief; we're two decades into the internet revolution, it's about time we stopped treating tech like an upstart." **Kate Crawford** (currently the subject of a **Wired** [profile](#)) also [weighs in](#), "There's a big role for datasheets here. For high-risk AI, the regs require datasheets – e.g provenance of datasets, what's in them, how the data was obtained, selected, labelled, and cleaned."

There is bi-partisan legislation making its way through the **U.S. Congress** to [expand](#) the **National Science Foundation**, establishing a new directorate with forward-looking responsibility, similar to how **DARPA** is the **Defense Department's** primary research and development group. [Another new bill](#) seeks to stop government agencies from purchasing American citizens' data from third-party brokers without a warrant. According to reporter **Karin Fischer**, lobbyists for higher education [have sent a letter](#) to the **Senate Foreign Relations committee** "expressing concern about

a bill that would greatly expand government oversight over universities' reporting of foreign gifts and contracts."

Personalization can be indistinguishable from discrimination. Not always but often, too often if one believes [the recent warning](#) from **Elisa Jillson**, a privacy attorney with the **U.S. Federal Trade Commission**, who recently blogged at [ftc.gov](#). Her message: The FTC is going to report, evaluate, advise and police the organizations developing AI that creates bias or unfairness, whether inadvertently or on purpose. (**Ryan Calo** thinks that this is [just the beginning](#) for FTC enforcement.) Healthcare delivery, because of its ample data and its distinctly positive and negative outcomes, has clearly shown that disparities arise when biased systems work for and against different populations. One hospital system, **Rush University Medical Center**, has [prioritized correcting health inequities](#), recognizing the 14-year life expectancy gap between the Chicago Loop area (median household income, \$107,246) and the East Garfield Park neighborhood (median household income, \$22,818), both "just two stops on the **Chicago Transit Authority** Blue Line from Rush." **Stanford University** researcher **James Zhou** has retested government-approved AI medical devices, and found that past **U.S. Food and Drug Administration** testing lacked thoroughness and relied on retrospective analysis. It's a problem now, and [looms as an ongoing issue](#). "These are learning algorithms, and they keep learning," Zhou says. "They're also prone to biases. If we don't rigorously monitor these devices, the biases could get worse."

**Disneyland** and other California theme parks have fought customer bottlenecks with personal apps for years, and more technology is [the plan](#) for scaling up post-Covid. The concerns here match issues that some have with the vaccine passports, which are [popping up](#) in major metros that are eager to restore sports and entertainment. (**Vox Recode** has [an excellent backgrounder](#) on vaccine passports.) The apprehension isn't about the effectiveness of the solution; it comes out of the lack of trust that the customer data won't be re-purposed in the future in surprising, unpleasant ways. Apps-related signage and QR codes eliminate queues, offer faster gratification and "cut down on aimless intermixing" at amusement parks. Are vaccine passports going to make cityscapes into funlands? Maybe.

Hyperpersonalization is [a realistic scenario](#) for music finding and listening in the future, according to **Danny Wright** of **Vice magazine**. Biometrics will play a crucial role, and Wright points to ongoing tech investments at **Spotify**, **Apple** and **Amazon**. But don't get too excited; new research shows that [music recommendation algorithms are failing](#) fans of "beyond-mainstream music."

**Kate Crawford** (here writing for **Nature**) is [advocating regulation](#) of software that interprets human emotions as it monitors employees, customers or school children. Monitoring emotions is unproven and – when combined with intrusive biometrics – doubles down on what Crawford calls the phrenological impulse: "drawing faulty assumptions about internal states and capabilities from external appearances, with the aim of extracting more about a person than they choose to reveal." (Also, **The Atlantic** has an adapted excerpt from Crawford's recent book, *Atlas of AI*, discussing [how AI misreads human emotions](#). She [recently joined](#) the faculty at **University of Southern California Annenberg School of Communications and Journalism**.) **Reuters** reports that large commercial banks are [rolling out elaborate surveillance technologies](#). Gender classification, once thought to be a core element of facial recognition, [has become deeply problematic](#) as its effectiveness becomes increasingly dubious. Researcher **Os Keyes** told **James Vincent** from **The Verge**, "What these technologies are doing is making those decisions a lot more efficient, a lot more automatic, and a lot more difficult to challenge." It hurts the marginalized and it forces "everyone into narrower forms of self-expression."

## Research News

A research team from **San Jose State University** has found that fuel moisture content (FMC) in California has fallen to new record lows, indicating [higher fire potential and fears of wildfires](#), stoked by data from catastrophe risk modeler **RMS**. Atlantic hurricane activity is once again also [expected to be higher in 2021](#) – making it 6 years in a row, say researchers from the **University of Arizona**. It's unclear whether there will be more earthquakes, but at least the [ShakeAlert early warning system](#) developed by **U.S. Geologic Survey** and the **Pacific Northwest Seismic**

**Network** might help you get out of the way. Don't confuse ShakeAlert with [DeepShake](#), another, complimentary earthquake warning system developed at **Stanford** that uses deep learning trained on 36,000+ past earthquakes. And it's particularly heartening to know that potential damage from volcanoes is being lessened by the successful [collaboration of a network of volcanologists](#). "More networks like these are needed to allow for further, active exchange of knowledge and experience within the global volcanology community," say the authors of this **Nature** editorial.

Since **NASA** launched the first Landsat satellite in 1972 the ability to monitor changes happening on the surface of the Earth, its inhabitants and the atmosphere in which they live has produced a [continual stream of data](#) that has been recorded and archived, and is immensely useful to research, says **Sachidananda Babu**, in charge of NASA's Sustainable Land Imaging Technology (SLI-T) program. Researchers at **Duke University Pratt School of Engineering** are using machine learning, satellite imagery and weather data to [autonomously find air pollution hotspots](#). The **State of California** is going to launch its own satellites to monitor greenhouse pollution with the help of big money philanthropists like **Michael Bloomberg**. The \$100 million [Carbon Mapper project](#) endeavors to be "like a weather service for methane and carbon dioxide." Epidemiologists from **Boston University** and **UCLA** have joined with NASA to form the agency's [first-ever mission devoted to health](#), using satellite data to parse which types and sources of air pollution particulate matter are most harmful.

Don't overlook the importance of terrestrial mapping projects. **Mario Moura** and **Walter Jetz** from **Yale** have [mapped the probabilities for finding undiscovered species](#) across the entire globe. **Nordregio**, a planning and development research center in Stockholm, has [combined permafrost maps with Census data](#) to identify communities located near thawing permafrost, which in turn may help them to prepare for change. **LandCover6k** has **University of Pennsylvania** archeologists working with climate modelers to create a single, accessible database on historical human land use, a tool to help predict the Earth's response to climate change. The relationship between aquatic biodiversity and human nutrition is being studied by researchers from **Yale University** and the **University of British Columbia**, highlighting the complicated need to find a balance between the risks and benefits of seafood biodiversity and consumption. "Ecological concepts of biodiversity can deepen our understanding of nature's benefits to people and unite sustainability goals for biodiversity and human well-being," said **Joey Bernhardt**, lead author.

The reproducibility crisis keeps **University of Michigan** computational social scientist **Scott Page** up at night. He [recently tweeted](#) a recent study where multiple research teams were given the same data to test the same hypothesis. Results varied widely among the teams. "Evidence that empiricists rely on different models to organize data." Page wrote.

While vaccination numbers continue to go up and pandemic case numbers plateau at best, SARS-CoV-2 research carries on. A team at **Washington University in St. Louis** took another look at a biosensor they had developed years ago to gauge its ability to [detect airborne SARS-CoV-2](#). They are now on track to develop two sensors – one that continuously monitors air quality in large gathering places, and a breathalyzer that could be used to rapidly gauge the health of people entering workplaces or other semi-public areas. Wastewater monitoring to detect COVID has been used sporadically throughout the pandemic, but not on the scale being used in Louisville KY. With the support of the **CDC** as well as private foundations, researchers at the **University of Louisville** have continued to collect robust wastewater data with concurrent random sampling of coronavirus test results throughout the city, which they say will be able to pinpoint [when the city meets the critical point of herd immunity](#) – signaling the end of the pandemic, at least in the testing area. Project lead **Dr. Aruni Bhatnagar**, director of **UofL's Christina Lee Brown Envirome Institute**, hopes to set a new standard of measuring public health. Up north, the **Canadian government** announced a \$10-million grant to **University of Manitoba** to support [five multidisciplinary infectious disease modelling networks](#) to help better prepare for future public health crises.

Training neural nets involves moving data between processor & memory, requiring time and energy. Big chips, [chips the size of dinner plates](#) made by **Ceres**, keep huge amounts data adjacent to hundreds of thousands of AI processor cores. It's the size of the entire silicon wafer. Ceres calls it the "Wafer Scale Engine." WSE Version 2 should be available commercially later this year.

Processing data at the edge with lower power chips and boards is an alternative to centralized data processing. **Arm** [recently announced](#) its upcoming v9 architecture that emphasizes security and improved vector math to help with AI applications. **Arduino** [added an outdoor-rated board](#) to its online store. The Arduino Edge Control is prepped to run on solar-connected batteries for smart agriculture. French researchers at **University of Grenoble** are exploring experimental memristor (also called resistive-RAM, or ReRAM) low power processors that [leverage the devices' inherent randomness](#) by applying machine learning algorithms where the variability is a feature, not a problem.

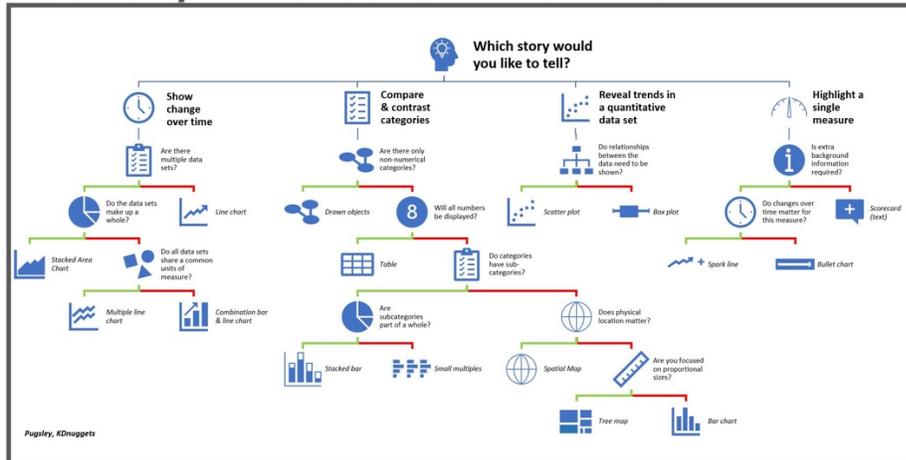
**Toyota** is [launching two consumer sedans](#) (one is a **Lexus**) in Japan with advanced-driver assistance. The vehicles require drivers to monitor the road but helps to maintain distance from other vehicles. Toyota has also [founded a new subsidiary, T-Hive](#), based in The Netherlands, that will deploy autonomous manufacturing vehicles and robots, an

opportunity created by the current e-commerce labor shortage. **Volkswagen** is establishing a [new AI Detroit research unit](#) in Michigan. The group, headed by **Daniel Weimer**, will pursue advances for vehicles, for services and for manufacturing. **BMW** keeps its AI research in Germany but the group recently [open-sourced computer vision labeling algorithms](#), part of a push for transparency and organization-wide tech literacy. **Waymo**, the **Google** self-driving vehicle spinout, is holding onto hope for fully autonomous consumer vehicles as it [reorganizes around new co-CEOs](#), **Tekedra Mawakana** and **Dmitri Dolgov**. You can find Waymo robot taxis on Phoenix' streets. Go for a ride. Or avoid at all costs.

## Data Visualization of the Week

KDnuggets, Stan Pugsly from February 15, 2021

### Data Story Visualization: A Decision Tree



## Events

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

### [Healthcare's AI Future: A Conversation with Fei-Fei Li & Andrew Ng](#)

**Online** April 29, starting at 10 a.m. Pacific. Organized by **DeepLearning.ai** and **Stanford Institute for Human-Centered Artificial Intelligence**.

### [Mothers in Science Conference, Motherhood and Career Retention in STEMM](#)

**Online** May 5, starting at 10 a.m. Eastern. [\$\$]

### [WeaSul 2021 - Accepted Papers](#)

**Online** May 7 starting at 7 a.m. Pacific. "Workshop on Weakly Supervised Learning" is an ICLR 2021 Workshop. [registration required]

### [Philosophy of Data Science virtual symposium](#)

**Online** May 7 starting at 1:30 p.m. Eastern. Organized by **University of Rochester** and **Syracuse University**. Sponsored by the **Central New York Humanities Corridor** with an award from the **Andrew W. Mellon Foundation**. [rsvp required]

## Deadlines

## Conferences

### [Information+ 2021 conference](#)

"Three more weeks to submit an abstract for your paper presentation or lightning talk about #infodesign and #datavis. We're twirling in anticipation to receive your submissions!"

## Education Opportunities

### [CIFellows 2021 Application Now Open](#)

"The CIFellows 2021 application site is now open. Applicants may start their applications here. Please read the CIFellows 2021 website and FAQs in its entirety before submitting." Deadline for applications is May 10.

## RFPs

### [SAGE Concept Grants](#)

The "program provides funding for innovative software solutions that support research in the social sciences." Deadline for applications is May 23.

## Tools & Resources

### [The Python/Jupyter ecosystem: today's problem-solving environment for computational science](#)

*Lorena Barba from Apr 19, 2021*

"In the March/April 2021 issue, **CiSE** is proud to feature several articles showcasing Jupyter in computational science. The showpiece is an invited article by Jupyter co-founders **Brian Granger** and **Fernando Pérez** (Granger 2021). They shift the focus of our conversation about problem-solving environments to the human angle: the researcher interactively exploring a scientific question or analyzing data, and the community of people collaborating and advancing their field."

### [Publishing stats papers in general science journals](#)

*Jeff Leek from Apr 09, 2021*

"There are two types of work I have had success in getting into [top] journals: (1) opinion pieces and (2) scientific work. In general my experience is it is much easier to get statistics ideas into these journals as opinion pieces rather than research."

### [James Landay: Smart Interfaces for Human-Centered AI](#)

*YouTube, Stanford HAI from Apr 7, 2021*

"AI has the potential to automate people out of their jobs, and in some cases, it will. But while we should carefully consider the risk of replacing human capabilities, it's important to realize that AI has enormous potential to augment them as well: it can boost the creativity of our work, help us learn better, deliver healthcare more effectively, and make our societies more sustainable. Like any tool, however, AI and its relationship with humans has as much to do with its interface as it does with the underlying capabilities it provides. Does it amplify our actions and remain attentive to our goals—even as we revise them—or is it a black box that accomplishes tasks autonomously?" [video, 59:40]

## Careers

See the [ADSA Jobs Page](#) for more opportunities.

[McCollum Endowed Chair in Data Sciences](#), Baylor University, Computer Science and Informatics Department; Waco, TX.

[Software & Data Engineer](#), Catalyst Cooperative; Remote, North America.

[Data Wrangler & Analyst](#), Catalyst Cooperative; Remote, North America