

CBIIT TechScouts

'Inspiring innovation through forums for idea-sharing'

CBIIT TechScouts is a forum for promoting **continuous improvement across CBIIT** through the cross-fertilization of ideas, experiences and recommendations. This forum is designed to foster new collaborations, learn about opportunities to **better serve our customers**, engage CBIIT more broadly, and raise awareness of new techniques and technologies that **promote innovation**. Central to the CBIIT Tech Scouts is the focus on gathering insight from the community with ideas and experiences on how emerging information can be used to **improve scientific productivity and accelerate cancer research**.

Table of Contents

- [Table of Contents](#)
- [Topics Archive](#)
- [Show and Tell Archive](#)
- [Attachments](#)

Topics Archive

Date	Topic (CLICK FOR MORE INFO)	Author	Topic ID	Summary
10 Oct 2018	Arlington VA AWS Meetup list: "Call for presentors"	Sean Davis	TS-0057	There might be room for CBIIT to participate? I think that other government agencies would be very interested in hearing how NCI is approaching AWS. Natasha Clark (Co-Organizer) sent a message to the Arlington VA AWS Meetup mailing list - call for presenters Hi Everyone! I wanted to reach out to see if any of you would be interested in presenting at our October session, scheduled for Thursday, October 26th at Excella. Depending on interest, we can either make this a lightning talk session or keynote depending on length. As always, we are always interested in all things AWS and below are a few topics we have had interest for in the past. If any of these sound like they might be right up your alley, or if you have another topic in mind, please reach out to me via the messaging feature on the meetu p.com site and I can help coordinate. Looking forward to seeing you all later this month! Natasha • Intro into lambda/ serverless • Cost effective AWS practices • All things migration • Container management
04 Oct 2018	The big hack	Richard Finney	TS-0056	Bloomberg is headlining a hardware hack : The Big Hack: How China Used a Tiny Chip to Infiltrate U.S. Companies https://www.bloomberg.com/news/features/2018-10-04/the-big-hack-how-china-used-a-tiny-chip-to-infiltrate-america-s-top-companies?smd=premium. <i>Nested on the servers' motherboards, the testers found a tiny microchip, not much bigger than a grain of rice, that wasn't part of the boards' original design. Amazon reported the discovery to U.S. authorities, sending a shudder through the intelligence community.</i> <i>The companies' denials are countered by six current and former senior national security officials, who- [...]—detailed the discovery of the chips and the government's investigation.</i> BUT ... Amazon and Apple are denying they've been compromised ... https://www.thestreet.com/amp/markets/amazon-and-apple-deny-bloomberg-report-on-china-hardware-hack-14733776 <i>Amazon.com Inc. (AMZN) and Apple Inc. (AAPL) have denied claims that a secret microchip was found embedded in servers linked to Elemental Technologies, a video compressing service purchased by Amazon in 2015, amid concerns that government hackers in China were able to infiltrate U.S. corporate data.</i> <i>Bloomberg reported Thursday that the chip, found on a server made by San Jose, Calif.-based Super Micro Computer Inc (SMCI) via subcontractors in China through a contract with Elemental, could be used to infiltrate a host of computer networks linked to both major U.S. companies as well as portions of the U.S. government's national security system.</i>
04 Oct 2018	NISTIR 8202, Blockchain Technology Overview CSRC	Carl McCabe	TS-0055	https://csrc.nist.gov/publications/detail/nistir/8202/final If you're interested in learning more about blockchain technology and its potential applicability in federal work, be aware that NIST just released NISTIR 8202. This "Blockchain Technology Overview" is a technical publication that examines the history, scope, and other characteristics of blockchain technology. NISTIR 8202 discusses various blockchain implementation approaches, existing limitations and misconceptions surrounding blockchain, and several areas of consideration for federal agencies and organizations seeking to understand and manage blockchain technology. It is also an introductory document meant to provide the foundation for a planned series of publications on more specific aspects of blockchain.

26 Sep 2018	Software disenchantment - "Everything is unbearably slow"	Richard Finney	TS-0054	A lament on the current state of software: http://tonsky.me/blog/disenchantment/ <i>Modern text editors have higher latency than 42-year-old Emacs. Text editors! What can be simpler? On each keystroke, all you have to do is update tiny rectangular region and modern text editors can't do that in 16ms. It's a lot of time. A LOT. A 3D game can fill the whole screen with hundreds of thousands (!!!) of polygons in the same 16ms and also process input, recalculate the world and dynamically load/unload resources. How come?</i>
20 Sep 2018	John Hancock adds fitness tracking to all policies - BBC News	Carl McCabe	TS-0053	https://www.bbc.com/news/technology-45590293 With the release of the gen 4 Apple Watch, which includes (or will include) an ECG and AFib monitoring features, we will probably be seeing a lot more stories like this.
18 Sep 2018	Local news: Montgomery County Hearing on ZTA 18-11	Robert Wynne	TS-0052	(The following public information may be of interest to anyone living and/or working in Montgomery County.) <u>Wireless Facilities Hearing on ZTA 18-11 and Map</u> A public hearing will be held regarding ZTA 18-11 on Sept. 25 at 7:30 p.m. in the third-floor hearing room of the Council Office Building at 100 Maryland Avenue in Rockville. https://www.montgomerycountymd.gov/towers The Wireless Facilities Map describing new towers and 5G mini-towers, as proposed to the County Council, is publicly available. There are multiple locations planned for 5G mini-towers and in many communities across Montgomery County less than 30' from homes, as well as new mobile utility poles. https://gis3.montgomerycountymd.gov/WirelessAntennasAndTowers/ (long load time) NIH resources https://www.nih.gov/news-events/news-releases/high-exposure-radiofrequency-radiation-linked-tumor-activity-male-rats https://ntp.niehs.nih.gov/results/areas/cellphones/ ACS https://www.cancer.org/cancer/cancer-causes/radiation-exposure/cellular-phones.html Scientific American https://www.scientificamerican.com/article/new-studies-link-cell-phone-radiation-with-cancer/ Not linked: The Ramazzini study
12 Sep 2018	Announcing Globus Support for Protected Data	Sean Davis	TS-0051	We're excited to announce availability of new Globus features for managing protected data, including HIPAA-regulated data and personally identifiable information. With higher assurance levels for protected data, subscribers can easily manage this data and share it securely and appropriately with collaborators. These new features especially benefit organizations and projects where protected data is shared by multiple researchers, such as institutions with secure data enclaves, multi-institutional studies using clinical data, and facilities distributing sensitive data to investigators and their collaborators. Read the announcement to get more details, or register for a live Q&A webinar on October 24.
04 Sep 2018	Lecture Announcement: Containerization for Reproducible Bioinformatics Research	Sean Davis	TS-0050	Containerization for Reproducible Bioinformatics Research Date: Tuesday, September 4, 2018 Time: 11:00 am – 12:00 pm Location: Room E1/E2, Natcher Conference Center (NIH Building 45) Registration: No pre-registration is required. Seating is first come first serve. As computational work becomes increasingly embedded in biomedical research practices, computational reproducibility has become an issue of increasing importance. Computational reproducibility requires that other researchers are able to deploy and use software and analysis workflows in their own computing environments. Platforms like Docker and Singularity allow the creation and configuration of software containers, which can be distributed and deployed across a range of systems. This lecture, presented by Steve Tsang, will give an introductory overview of containerization and how containers can facilitate reproducible bioinformatics research, providing examples from the NCI Cloud Resources and various hackathons. Seating is limited, but the presentation will be available through Webex (calendar invite attached).
30 Aug 2018	The State of Agile Software in 2018 - Martin Fowler	Carl McCabe	TS-0049	The State of Agile Software in 2018 - Martin Fowler https://martinfowler.com/articles/agile-aus-2018.html
09 Aug 2018	Doctor Data: How Computers Are Invading the Clinic / AI for Biomedical Research	Carl McCabe	TS-0048	This article was included in the most recent issue of NIH's IRP Weekly: Doctor Data: How Computers Are Invading the Clinic / AI for Biomedical Research https://irp.nih.gov/blog/post/2018/08/doctor-data-how-computers-are-invading-the-clinic
31 Jul 2018	Cloud computing approaches to Genomic Data Science	Sean Davis	TS-0047	FWIW, an introductory talk that I gave at the American Statistical Association Joint Stats Meeting on the topic: https://seandavi.github.io/talk/2018/07/31/cloud-computing-approaches-to-genomic-data-science/
30 Jul 2018	Generating high-quality workshop materials using open-source tooling	Sean Davis	TS-0046	We (Bioconductor) have created an online and published set of workshop resources that we used for our annual conference. We did so using the open source Bookdown package (https://bookdown.org/yihui/bookdown/) in a collaborative editing effort that resulted in 388 pages from 19 contributors in just over 8 weeks. There is an associated Amazon Machine Image that was used to test build the materials and then each conference participant received his/her own instance for the duration of the conference. Materials are here: https://bioconductor.github.io/BiocWorkshops/ (html) https://bioconductor.github.io/BiocWorkshops/BioC2018.pdf (pdf) Feel free to contact me to discuss the process in more detail.

26 Jul 2018	Spectre Hits the Net	Richard Finney	TS-0045	Three things in computers are hard: cache invalidation and off by one errors. The original spectre attack from earlier this year took advantage of privileged data being available to ordinary user's code because the cache wasn't cleared. Some Austrian researchers have identified a new angle on this ... <i>"That impact is now a little larger. Researchers from Graz University of Technology, including one of the original Meltdown discoverers, Daniel Gruss, have described NetSpectre: a fully remote attack based on Spectre. With NetSpectre, an attacker can remotely read the memory of a victim system without running any code on that system". (- from ars technica)</i> News: https://duckduckgo.com/?q=netspectre+&t=ffsb&iar=news&ia=news original reporting paper: https://misc0110.net/web/files/netspectre.pdf ouch.
26 Jul 2018	Why Multiple Database Types	Brent Coffey	TS-0044	A one size fits all database doesn't fit anyone https://www.allthingsdistributed.com/2018/06/purpose-built-databases-in-aws.html
24 Jul 2018	Artificial Intelligence for Government Services	Sean Davis	TS-0043	Now, registration is open for the next edition of DigitalGov University Emerging Technology Leadership Series, a new pilot to enhance the modern federal workforce with training, education, and awareness of emerging technologies including Artificial Intelligence, Robotic Process Automation, Blockchain, Social Technologies, and Virtual/Augmented Reality. https://digital.gov/event/2018/07/30/emerging-technology-leadership-series-mina-hanna-ai-for-government-services/
19 Jul 2018	Packer for reproducible research	Sean Davis	TS-0042	Packer (https://packer.io) is a toolkit that implements "infrastructure-as-code" (https://en.wikipedia.org/wiki/Infrastructure_as_Code) for building Amazon Machine Images (AMIs). At the annual Bioconductor conference, 50% of the conference is devoted to hands-on coding and we supply every participant with a custom AMI. To ensure reproducibility and reusability of the AMI, we used packer to automate the creation of the AMI directly from a json file, eliminating any hand-editing or configuration of the machine. I describe this process very briefly in a blog post here: http://bit.ly/2JDztmZ I thought it might be useful for a few folks.
11 Jul 2018	Systems Operations on AWS Course	Sean Davis	TS-0041	This might be of interest to a few folks. It is (relatively) local. We'd like to invite you to a private delivery of our new Systems Operations on AWS course, which will be publicly released later this year. This instructor-led preview is scheduled for 7 – 9 August in Herndon, VA. We're offering seats at 50% off our standard price and asking for detailed feedback on class content and delivery. In <i>Systems Operations on AWS</i>, we teach individuals how to create automatable and repeatable deployments of networks and systems on the AWS platform. We will explore the AWS features and tools related to configuration and deployment and best practices for configuring and deploying systems. You will also learn how to: • Use standard AWS infrastructure features such as Amazon Virtual Private Cloud (VPC), Amazon Elastic Compute Cloud (EC2), Elastic Load Balancing (ELB), and AWS Auto Scaling from the command line • Use AWS CloudFormation and other automation technologies to produce stacks of AWS resources • Build virtual private networks with Amazon VPC Seats for this invitation-only preview are \$900 (a 50% savings on the full price). Space is limited, so we encourage you to register today.
06 Jun 2018	NIH Strategic Plan for Data Science	Sean Davis	TS-0040	FYI.... http://bit.ly/2sLYc1Q
03 Jun 2018	Microsoft is Said to Have Agreed to Acquire Coding Site GitHub	Carl McCabe	TS-0039	Microsoft Is Said to Have Agreed to Acquire Coding Site GitHub https://www.bloomberg.com/news/articles/2018-06-03/microsoft-is-said-to-have-agreed-to-acquire-coding-site-github
22 May 2018	Rare Cancer Hackathon	Sean Davis	TS-0038	This might be of interest to a few people on this list. A couple of us from NIH helped organize a hackathon in San Francisco over the past weekend focused on understanding the cancer genome of a patient with a rare kidney cancer, papillary renal cell carcinoma. The patient attended the hackathon. http://bit.ly/2IGut1l We had about 150 participants, about 60% of whom were developers and data scientists, most from Silicon Valley. After self-assigning to "teams", the participants identified areas of interest and developed code for data visualization, data engineering, machine learning, and precision medicine. Code resources from the hackathon are available here: https://github.com/svai
07 May 2018	A small data lake resource	Sean Davis	TS-0037	"Data infuses intelligence in to every business." http://www.ibmbigdatahub.com/blog/get-out-data-swamp-governed-data-lake Perhaps a group or two will be interested in implementing. Note that CBIIT the NCI Cleversafe object storage system could be a nice technological base for a data lake.
07 May 2018	Biowulf transitioning from RHEL6 /CentOS6 to RHEL7 /CentOS7 in June	Sean Davis	TS-0036	If you are working with Biowulf users (and NCI has MANY), this may be of interest to you. If there are CBIIT groups who would like to learn more about Biowulf, let me know and we can probably organize either a tour of Biowulf facilities or a Biowulf staff visit. For general information about Biowulf, the NIH enterprise HPC system, see: https://hpc.nih.gov/
05 May 2018	Blockchain explained in 7 python functions	Sean Davis	TS-0035	Ever wonder about what blockchain actually IS? https://towardsdatascience.com/blockchain-explained-in-7-python-functions-c49c84f34ba5

03 May 2018	Python pip support ending for TLS versions less than 1.2	Rohit Paul	TS-0034	I was just bit by this on my Mac, so passing along in case anyone experiences (or has experienced) being unable to install/upgrade packages using 'pip' due to a TLS-related error: http://pyfound.blogspot.ca/2017/01/time-to-upgrade-your-python-tls-v12.html This affected my El Capitan 10.11 system; the default Python on macOS 10.12+ is hopefully unaffected. Installing the latest Python 2.x using Homebrew worked fine for me, after upgrading Ruby as well. Fun stuff.
11 Apr 2018	NCI Containers and Workflow Seminar	Carl McCabe	TS-0033	Here's a talk to the NIH Data Science SIG that several people across CBIIT may be interested in. NCI Containers and Workflows Interest Group / NIH Data Science Lectures Joint Seminar Bio-Docklets: virtualization containers for single-step execution of NGS pipelines Presenter: Konstantinos Krampis - Associate Professor, Hunter College, City University of New York; Faculty, Weill Cornell Medical College WebEx URL: https://cbit.webex.com/cbit/j.php?MTID=ma3f0d94985bbd365d114c5a469359aca Meeting number (access code): 731 927 985
07 Mar 2018	Debian GNU/Linux for WSL	Carl McCabe	TS-0032	Debian GNU/Linux for WSL now available in the Windows Store https://blogs.msdn.microsoft.com/commandline/2018/03/06/debian-gnulinux-for-wsl-now-available-in-the-windows-store/ This is an update of a previous announcement about Linux on Linux on WSL: New distros coming to Bash/WSL via Windows Store https://blogs.msdn.microsoft.com/commandline/2017/05/11/new-distros-coming-to-bashwsl-via-windows-store/

26 Feb 2018	Accelerating Agency IT Modernization	Jeff Shilling	TS-0031	The White House's American Technology Council and Office of American Innovation on modernizing federal technology published specific recommendations to jumpstart a new wave of modernization efforts by accelerating cloud adoption, consolidating networks and prioritizing key applications for needed upgrades. Now the daunting task of implementing these recommendations sits within agencies, and it is not a one size fits all proposition. Where do they begin to successfully move away from expensive legacy infrastructure? How do they transition to a more secure, agile, and cost-effective technology ecosystem, much of which will be supported by shared services? Join i360Gov and senior level technology leaders from government and industry as we provide an overview of current initiatives and solutions to the many IT modernization challenges agencies face, such as: • Maintenance that often requires immediate attention and runs the risk of breaking integrations and upgrades • Legacy solutions that are unable to properly communicate between on-premises, mobile, and the cloud • Citizen facing services not designed for today's technology environment You will also learn about Identity, the hidden accelerator to IT modernization, and how by creating a single solution for identity, agencies can speed up digital and cloud programs that will enable you to: • Reduce costs and architecture complexity • Securely connect any employee, vendor, partner or citizen to any resource, on premise or in the cloud • Make administrators self-sufficient and decrease reliance on customization • Scale seamlessly as you move services into the cloud <u>Webinar Presenters</u> • Dr. Ronald Ross, Computer Scientist, NIST Fellow • David Hogue, Technical Director, Cybersecurity Threat Operations Center, NSA • Joe Diamond, Director, Cybersecurity Strategy, Okta Register now for this complimentary, educational webinar. As long as you register now, you will receive the link. If you are unable to attend the live webinar, you can use the link to watch the recording at your convenience.
23 Feb 2018	ATARC Federal DevOps Summit	Tim Harvey	TS-0030	Interested in Agile? Keen on DevOps? Please join the Advanced Technology Academic Research Center next Thursday, March 1 for the premier government Agile and DevOps event of the year, the ATARC Federal DevOps Summit at the Marriott Metro Center in Washington, D.C. This educational symposium for Federal IT practitioners is free to government and eligible for 7.5 CPE credits. To view the agenda and register, please visit: www.fedsummits.com/devops/ A Visionary Panel on DevOps and Government Transformation will include Cris Brown, Master Data Management Program Manager, NRC; Peter Burkholder, Innovation Specialist, GSA 18F; Jennifer Hoover, Digital Services Expert, DHS; and Evan Lee, Chief Technology Officer, HHS OIG. Tom Temin of Federal News Radio will serve as moderator. An all-star list of Federal IT thought leaders include: David Larrimore, Chief Technology Officer, DHS ICE; Simmons Lough, Software Architect, USPTO; Navin Vembar, Chief Technology Officer, GSA; and Hasan Yasar, Technical Manager & Adjunct Faculty Member, Software Engineering Institute, Carnegie Mellon University. The afternoon will feature the MITRE-ATARC DevOps Collaboration Symposium with government, academic and private industry SMEs who will brainstorm and whiteboard during five concurrent sessions on: SecDevOps; DevOps Implementation with Cloud; DevOps Culture; DevOps Testing; and DevOps in Health IT. Session topics: 1. SecDevOps 2. DevOps Implementation with Cloud 3. DevOps Culture 4. DevOps Testing 5. DevOps in Health IT

20 Feb 2018	NIH Supercomputers Have Come a Long Way	Carl McCabe	TS-0029	In the February 9th issue of the NIH Record: NIH Supercomputers Have Come a Long Way BY DANA TALESIK https://nihrecord.nih.gov/newsletters/2018/02_09_2018/story1.htm
02 Feb 2018	Free Pro Git 2 book (Kindle edition)	Carl McCabe	TS-0028	I don't know how long this deal will last, but Amazon currently (Friday 5pm) has Pro Git 2, by Scott Chacon and Ben Staub, on sale at a very affordable price of \$0.00. It is the Kindle edition, not a paper copy. If you're interested, here's the link: https://www.amazon.com/gp/product/B01ISNIKES Melina Scotto: 'A reminder that the hash used in Git is SHA1. Deprecated for government use since 2013. Great to play with, though.' Jeff Shilling: 'Since Git is not using SHA1 for cryptographic functions but as a hash of the files to determine if changes to the files have taken place, it doesn't pose a security concern.'
17 Jan 2018	Golem open source, decentralized supercomputer	Carl McCabe	TS-0027	This is not something we can use right now (and maybe not ever in the Federal govt), but it is an interesting peek into the future. Golem https://golem.network/ "Golem is a global, open source, decentralized supercomputer that anyone can access. It is made up of the combined power of users' machines, from PCs to entire data centers. Golem is capable of computing a wide variety of tasks, from CGI rendering, through machine learning to scientific computing. Golem's limitations are only defined by software developers' creativity. Golem creates a decentralized sharing economy of computing power and supplies software developers with a flexible, reliable and cheap source of computing power." Golem does have competitors, e.g. https://iex.ec/
29 Dec 2017	R and an NCI Genomic Data Comms Use Case	Sean Davis	TS-0026	NCI has established the Genomic Data Commons (GDC, https://gdc.cancer.gov) as a home for NCI cancer genomics datasets. One feature of the GDC is that it uses UUIDs for "everything". However, most cancer researchers think in terms of legacy "barcodes" when working with these datasets. Given the increasing importance of web-based APIs and their use for NCI data, I wrote a quick blog post using an R client for these data and, specifically, for translating from UUID back to "barcodes": https://seandavi.github.io/post/2017/12/genomicdatacommons-example-uuid-to-tcga-and-target-barcode-translation/
15 Dec 2017	PUTTY Begone! Microsoft Will Ship an OpenSSH Client	Carl McCabe	TS-0025	https://techcrunch.com/2017/12/13/putty-begone-microsoft-will-ship-an-openssh-client/
05 Dec 2017	NIH Hour of Code	Sean Davis	TS-0024	The international hour of code is an effort to bring computer programming to the masses. We at the NIH DataScience Special Interest Group have put together four sessions: R, python, natural language processing, and Shiny. Webinar slots are still available. Details are here: https://nihlibrary.nih.gov/about-us/news/hour-code-classes-nih-library Feel free to pass along. My session on truly introductory R is from 10-12 today. Materials (we won't cover all of them) are here: https://seandavi.github.io/ITR/
04 Dec 2017	Bioconductor, Software for Genomic Data Science	Sean Davis	TS-0023	I am giving a talk at 11am about Bioconductor, a large software project partially funded by NCI and the CBIIT ITCR program. The talk will be the TE406 at 11am and is only 20 minutes long--part of a morning-long conference open to anyone sponsored by the Center of Excellence in Integrative Cancer Biology and Genomics. Slides are posted here: http://bit.ly/2jbKgKG Here is the abstract: Progress in biotechnology is continually leading to new types of data, resulting in data sets that are rapidly increasing in volume, resolution and diversity. The promise of unprecedented advances in our understanding of biological systems and in medicine is challenged by complexity and volume of data also challenge scientists' ability to analyze them. Meeting this challenge requires continuous improvements in analytical methods and capable, usable software tools implementing them. Bioconductor is a well-established open-source, open-development software project for the analysis and comprehension of high-throughput data in genomics and molecular biology. The project aims to enable interdisciplinary research, collaboration and rapid development of scientific software. Based on the statistical programming language R, Bioconductor comprises 1473 interoperable packages contributed by a large, diverse community of scientists. These packages undergo formal initial review and continuous automated testing. Each package includes documentation and working example use cases. Bioconductor supports many types of high-throughput sequencing data (including DNA, RNA, chromatin immunoprecipitation, Hi-C, methylomes and ribosome profiling) and associated annotation resources; contains mature facilities for microarray analysis; and covers proteomic, metabolomic, flow cytometry, quantitative imaging, cheminformatic and other high-throughput data. Bioconductor package interoperability enables the rapid creation of workflows combining and integrating multiple data types and tools for statistical inference, regression, network analysis, machine learning and visualization at all stages of a project from data generation to publication. A large and growing community of researchers and users contribute to ongoing development, online support, and education. The influence of the project is evidenced by more than 250,000 downloads per year and tens of thousands of citations in the literature. I will present an overview of the project for prospective users and contributors.
02 Dec 2017	Protect Against Secrets in Git Repositories	Sean Davis	TS-0022	I wrote a blog post about an embarrassing but educational experience including the solution I implemented for myself to keep it from happening again. Perhaps someone else will find my experience useful. http://bit.ly/2i9co06
17 Nov 2017	GitHub - Security Alerts and Dependency Graphs	Carl McCabe	TS-0021	Github recently rolled out a new feature allowing you to see a graph of your project's dependencies. And yesterday they took this a step further by providing you with alerts when vulnerabilities are detected in any of those dependencies. Dependency Graphs are enabled automatically for public repos, but must be enabled optionally for private repos. Currently, this works for Ruby and Javascript, but Python is coming sometime in 2018. Dependency Graphs https://help.github.com/articles/listing-the-packages-that-a-repository-depends-on/ Security Alerts https://help.github.com/articles/about-security-alerts-for-vulnerable-dependencies/
16 Nov 2017	Agricultural Data Ecosystem Talk	Sean Davis	TS-0020	I recently visited the USDA to give a talk with thoughts on an Agricultural Data Ecosystem, inspired in part by the work done by pioneering work done by CBIIT and NIH on data commons and cloud pilots. The slides are available on my website and might interest a few folks: https://seandavi.github.io/talk/2017/11/15/thoughts-on-an-agricultural-data-ecosystem/ Thanks to Ishwar Chandramouliswaran, Steve Tsang, and Durga Addepalli for several of the slides.

20 Sep 2017	Bored With Your Fitbit? These Cancer Researchers Aren't	Carl McCabe	TS-0019	Here is an overview of activity tracking devices (specifically Fitbits) use in clinical research.\ BORED WITH YOUR FITBIT? THESE CANCER RESEARCHERS AREN'T (Wired.com) https://www.wired.com/story/bored-with-your-fitbit-these-cancer-researchers-arent/
11 Aug 2017	Malware Encoded Into DNA can Hack the Computer that Reads It	Joel Yobouet	TS-0018	What if someone stores a malicious program into DNA, just like an infected USB storage, to hijack the computer that reads it? A team of researchers from the University of Washington in Seattle have demonstrated the first successful DNA-based exploit of a computer system that executes the malicious code written into the synthesized DNA strands while reading it. Here is the white paper : http://dnasec.cs.washington.edu/dnasec.pdf
02 Aug 2017	DigiCert to Acquire Symantec's Website Security Business	Jesse Bocinski	TS-0017	Digicert is acquiring Symantec's web security business, a consolidation of the public commercial certificate market. http://bit.ly/2AUyc2w
06 Jul 2017	Semiconductor Feature Size Roadmap	Warren Kibbe	TS-0016	Pretty amazing feats of science and engineering to get us to 2020. https://en.wikipedia.org/wiki/5_nanometer For comparison, a single benzene ring is about 500 picometers, or 0.5 nanometers
29 Jun 2017	Monarch Made Twitter	Amy Gentzel	TS-0015	I wanted to share a session that was presented at the AWS Public Sector Summit by NIAID that discussed Next-Generation Medical Analysis leveraging the cloud. Our CSRA team at NIAID assisted by developing a PaaS solution they've named Monarch, discussed in the presentation, that assists with this work leveraging a full DevOps pipeline. For background, Nephele, discussed in the presentation, is a project that a team in BCBP has been working on for several years now. Due to the infrastructure group (OEB) not having a defined service offering for public cloud at NIAID years ago when Nephele started, the project team went out on their own and architected a solution using AWS. Within the last year, CSRA consulted with them and provided some guidance for best practices and improvements now that we are launching a custom platform as a service using open source technologies and AWS services. In the future, we'll be moving Nephele components to the custom PaaS we created called Monarch and hosting it under our architecture. https://www.youtube.com/watch?v=rLkBWVv0Hdc&list=PLhr1KZpdzukePsKIUofhgp50b63-5yr1V&index=66
29 Jun 2017	DCEG Linkage: 3D Printing in Radiation Research	Carl McCabe	TS-0014	Here's a unique example of 3D printing right on site in the NCI Shady Grove building. This is from Choonsik Lee's group in DCEG's Radiation Epidemiology Branch. https://dceg.cancer.gov/news-events/research-news-highlights/2017/3d-printed-phantoms?cid=eb_govdel
16 Jun 2017	SciAm: China Shatters 'Spooky Action at a Distance' Record, Preps for Quantum Internet	Carl McCabe	TS-0013	Here's more on the long-term future (ignore the immediate or mid-term geopolitical implications). https://www.scientificamerican.com/article/china-shatters-ldquo-spooky-action-at-a-distance-rdquo-record-preps-for-quantum-internet/
16 May 2017	MaruOS - Your Phone is Your Computer	Carl McCabe	TS-0012	https://maruos.com With Maru, your phone is your PC -- you connect it to a keyboard and monitor whenever you need the desktop environment. Obviously this isn't ready for government use yet, but it illustrates the path toward convergence into single device personal computing.
10 May 2017	Why Windows Must Die. For the Third Time	Mark Cunningham	TS-0011	Interesting article about the history—and potential future—of Windows: http://www.zdnet.com/article/why-windows-must-die-for-the-third-time/ Great article as Microsoft tries to stay relevant in an age of minimal OS's, immutable servers and relentless competition from, some would argue, more mature operating systems such as those found on MAC's. It brought a tear to my eye as I took a walk down memory lane
09 May 2017	Biowulf HPC Expansion Details	Sean Davis	TS-0010	Biowulf now has >180,000 CPUs, including 100 GPU nodes, 24 very large memory nodes (up to 3TB per machine), and a very high-performance, low latency network (FDR infiniband) with an 80 Gb/sec connection to the outside world. Additional dedicated resources (with priority usage) are available to CCR researchers (10,000 cpus) and, soon, to DCEG (4000 cpus). <i>This expansion will likely place Biowulf in the top 100 most performant HPC systems in the world.</i> If you have questions, feel free to contact me or the Biowulf staff.
11 Apr 2017	Globus Data Management Solution	Miles Kimbrough	TS-0001	- Globus, a cloud-authenticated data management and transfer platform, will be hosting a user-focused webinar on Tuesday, May 16th, to benefit those interested in exchanging datasets across a variety of sources. The webinar will provide a high-level overview of Globus, steps to start using the service, and common use cases along the following topics: 1. When, where, and why to use Globus? 2. NIH account specs - distinction from Globus Plus 3. What do system administrators need to set up managed endpoints? 4. Which endpoints are already set up? 5. How to set up Globus on your own desktop 6. How to transfer and share data 7. If sharing with collaborator, what info does collaborator need? What do you need to give to collaborator? 8. New Globus command line interface, allowing users to script their transfers 9. Encryption, verification, and expected data transfer speeds as compared to other resources (e.g. FTP)
05 Apr 2017	The Algorithm Will See You Now	Melina Scotto	TS-0009	Great piece in the April 3 New Yorker on medical diagnostics and machine learning. They used Thrun and Hinton's Columbia work with predicting melanoma as an example. http://www.newyorker.com/magazine/2017/04/03/ai-versus-md Anyone interested in machine learning – FAES offers python based machine learning in BIOF 509. Taught by Jonathan Street and Burke Squires on campus. https://faes.org/
14 Mar 2017	TEDxBuffalo - Become a Citizen Data Scientist	Carl McCabe	TS-0008	Here's a TEDx talk about how normal (i.e. non-academic) people can be enlisted to support certain classes of scientific research. As the talk alludes, there are many possibilities for leveraging citizen science, and it is useful to think about the technologies available and the ways people interact with technology (like gaming or idle time-wasting on a cellphone) to understand those possibilities. The value of citizen science is not just in distributed data collection or analytical tasks that are currently easier for humans than computers. It is also the engagement of more people in the purpose, goals and problems of current research efforts. https://www.youtube.com/watch?v=zgijAGkdjZc&feature=youtu.be PS. Bonus link on citizen science in cancer research: http://scienceblog.cancerresearchuk.org/2015/10/01/citizen-scientists-can-spot-cancer-cells-like-pathologists-so-what-happens-next/

08 Mar 2017	Oracle Doubling the Cost to Run in AWS	Robert Smallwood	TS-0007	Found these two articles related to Oracle's new pricing strategy which effectively doubles the cost to run Oracle products in AWS. For those of you that have worked with Oracle over the years, I'm sure this isn't a big surprise, but as we move to the cloud it may be a driver to move to other database platforms such as AWS RDS or Red Shift on AWS. Of course I suppose there's always the Oracle cloud. https://www.cloudtp.com/doppler/oracle-is-now-charging-double-to-run-on-aws/ https://www.theregister.co.uk/2017/01/30/oracle_effectively_doubles_licence_fees_to_run_in_aws/
07 Mar 2017	IBM Watson for Oncology	Joel Yobouet	TS-0006	Watson provides clinicians with evidence-based treatment options based on expert training by MSK physicians. Whether a community oncology practice or an international hospital, oncologists like all clinicians are struggling to keep up with the large volume of research, medical records, and clinical trials. Watson scales vital knowledge and helps oncologists. Now, with the collaboration between IBM and MSK, Watson for Oncology utilizes world-renowned MSK expertise to evaluate specific details of each unique patient against clinical evidence. https://www.ibm.com/watson/health/oncology/
07 Mar 2017	Data Science Job Report: R, SAS, Python	Carl McCabe	TS-0006	For anyone paying attention to technology trends in statistical analysis and data science, here's a report from an R-focused data science blog: http://r4stats.com/2017/02/28/r-passes-sas/ SQL is of course a mainstay; Python is big, R is big, and SAS is declining but big and not going anywhere soon. Java and Hadoop are also commonly used. The methodology used to track these trends involves data from job postings. Specific trends within cancer research are likely to be somewhat different.
02 Mar 2017	DNA - the Future of Data Storage?	Carl McCabe	TS-0005	You've probably heard about the concept of storing data in DNA by now. Here's a quick what/why/how/when from Yaniv Erlich's lab at Columbia U: https://www.researchgate.net/blog/post/dna-could-be-the-future-of-data-storage
05 Feb 2017	NCI Genomic Data Commons R Client	Sean Davis	TS-0004	The NCI has established the NCI Genomic Data Commons to store and distribute NCI datasets. One powerful feature of this new resource is a REStful API. The Bioconductor project is a large, open-source for the analysis and comprehension of -omics datasets. A couple of us are working on an Bioconductor package for finding and accessing data in the NCI Genomic Data Commons directly from R, exposing a community of about 4000 developers and 100,000 users to NCI genomics datasets. The GenomicDataCommons package is available on GitHub and will be making its way into the Bioconductor project in the next month or so.
26 Jan 2017	Open, Reproducible Neuroscience	Sean Davis	TS-0003	Towards open and reproducible neuroscience in the age of big data Abstract: The Stanford Center for Reproducible Neuroscience (CRN) was founded to develop a number of initiatives to help to make neuroscience more open and reproducible. These initiatives include influencing the scientific culture (e.g. The OHBM Replication Award), introducing new standards for organizing neuroimaging data (BIDS) and software (BIDS Apps) as well as building tools to assess MRI quality (MRIQC) and perform robust preprocessing of fMRI data (fMRIPREP). Tying all of those efforts together, as well as capitalizing on the experience gained from OpenfMRI and NeuroVault, is the upcoming OpenNeuro project. This novel software-as-a-service platform seeks to harness the power of high-performance computers to provide free data analysis in exchange for making data publicly available. There is a nice technology feature article in Nature on "big data" in Neuroscience, touching on many of those parallels -- data management when dealing with 100s of terabytes, attitudes toward data sharing, and the notion that in many cases "big data" is still not enough data. Neuroscience: Big brain, big data http://www.nature.com/nature/journal/v541/n7638/full/541559a.html
03 Jan 2017	AWS CloudTrail	Sean Davis	TS-002	Monitoring changes to infrastructure and data is a very important part of a robust and secure infrastructure. This service allows ongoing logging of API calls (so, changes) to infrastructure and even supports S3 events (for changes to data). https://aws.amazon.com/cloudtrail/ Other cloud providers offer similar services directly or via third-parties.

Show and Tell Archive

Date	Topic	Who	Presentation
09 Oct 2018	Introduction to Blockchain	Manish Malhotra - <i>Chairman & CEO, Unnisant, Inc.</i>	 The presentation slide for 'Introduction to Blockchain & Cryptocurrency Perspective' by Manish Malhotra. It includes a speaker profile, seminar details, learning outcomes, and a PDF icon.

Attachments

File	Modified
Microsoft Powerpoint Presentation CBIIT Sync Meeting_Globus_MK.pptx	Apr 11, 2017 by kimbroughmg
PDF File Introduction to Blockchain.pdf	Sep 24, 2018 by kimbroughmg
PNG File HPC-Blockchain-intro-image.png	Dec 03, 2018 by Frost, Ruth (NIH/NCI) [C]

[Download All](#)

Welcome to CBIIT TechScouts!

We are a forum for promoting continuous improvement across CBIIT through the cross-fertilization of ideas, experiences and recommendations.

The goal of CBIIT TechScouts is to collectively improve scientific productivity through the sharing of knowledge.

What We Do

CBIIT TechScouts is designed to foster new collaborations, learn about opportunities to better serve our customers, engage CBIIT more broadly, and raise awareness of new techniques and technologies that promote innovation.

Central to the CBIIT Tech Scouts is the focus on gathering insight from the community with ideas and experiences on how emerging information, data-oriented, and software technologies can be used to improve scientific productivity and accelerate cancer research.

Our Service Includes:

1. **An email distribution list**, allowing members to stay updated on the latest topics and trends
 - a. To join: email [Miles Kimbrough](#), subject 'TechScouts Access Request'
2. **An annotated archive of topics and presentations**, organized for future reference
3. **A monthly summary found on CBIIT Central**, providing a lightweight overview of recent trends
4. **And finally, Monthly Show and Tell Sessions**, opening the door for new ideas to be presented both in person and virtually
 - a. To present: email [Miles Kimbrough](#) with topic and availability

Questions?

General Support : Miles

Kimbrough | [miles.](#)

kimbrough@nih.gov | 240.276.

5251

Consultation and Guidance :

Eric Stahlberg | [eric.](#)

stahlberg@nih.gov | 240.276.6

729

Technical Support : George
Zaki | george.zaki@nih.gov |
240.276.5171