NYU Data Science Community features journalism, research papers, events, tools/software, and jobs for April 15, 2016

Data Science News

Eric Dishman exits Intel to head National Institutes of Health precision medicine research

Healthcare IT News from April 13, 2016
The longtime **Intel** fellow will be responsible for creating a longitudinal study to more effectively treat disease and ultimately improve health. **Eric Dishman** also brings experience using precision medicine tactics to beat cancer he fought for 23 years.

Inside Equinix's NY4 Data Center Where Wall Street Trades

Bloomberg Markets from April 13, 2016
Six miles northwest of the **New York Stock Exchange** as the microwave flies, across the Hudson River and within earshot of Interstate 95, is a building with no name. Only three numbers mark its address, and, like much of its surroundings, it's nondescript, encircled by windblown trash and lonely semitrailers waiting to be hauled away somewhere. It's a part of New Jersey that's, well, ugly.

It's also a critical node in the U.S. financial system: The 49 different exchanges that lease space at this data center sent a record 9.6 million messages per second through its fiber-optic cables in February. Every day, electronic trades representing trillions of dollars' worth of equities, derivatives, currencies, and fixed-income assets pass under this roof. This is NY4. This is where Wall Street actually transacts. [video autoplays, pre-roll + 3:43]

The academic, economic and societal impacts of Open Access: an evidence-based review

F1000Research from April 11, 2016
Ongoing debates surrounding Open Access to the scholarly literature are multifaceted and complicated by disparate and often polarised viewpoints from engaged stakeholders. At the current stage, Open Access has become such a global issue that it is critical for all involved in scholarly publishing, including policymakers, publishers, research funders, governments, learned societies, librarians, and academic communities, to be well-informed on the history, benefits, and pitfalls of Open Access. In spite of this, there is a general lack of consensus regarding the advantages or disadvantages of Open Access at multiple levels. This review aims to to be a resource for current knowledge on the impacts of Open Access by synthesizing important research in three major areas of impact: academic, economic and societal.

Also:

- OpenTrials: towards a collaborative open database of all available information on all clinical trials (April 8, BioMed Central)
- Sharing all types of clinical data and harmonizing journal standards (April 03, BMC Medicine)

Ex-Darpa Head Regina Dugan Leaves Google for Facebook

WIRED, Business from April 13, 2016
Regina Dugan—the former head of Darpa who was running the Advanced Technologies and Projects lab at **Google**—has left the internet giant to take up a similar post at rival **Facebook**.

Also:
- How to win friends and influence people — The social network has turned itself into one of the world's most influential technology giants, and wants to become ever more powerful (April 9, The Economist)
- Facebook's Massive New Antennas Can Beam Internet for Miles (April 13, WIRED Business)
- Everything Facebook announced at F8 2016 (April 13, The Next Web)
- Powering Facebook experiences with AI (April 13, Facebook Code, Engineering Blog; Joaquin Quiñonero Candela)

12 Inspiring Women In Data Science, Big Data

InformationWeek from April 11, 2016
Women make up half the population, yet it's been well documented that they don't come close to parity in STEM fields. Could the rise of big data and data science offer women a clearer path to success in technology? Here's a list of 12 inspiring women who work in big data and data science.

Revolutionizing research communication through a new academic publishing platform

Gordon and Betty Moore Foundation from April 13, 2016
The way that researchers communicate their work has not changed significantly in the last few centuries; academic publishing still relies on journal articles and has not kept up with technological advances, new analytical tools, or the globalization of research. To help address this discrepancy, the **Gordon and Betty Moore Foundation** has granted Collaborative Knowledge Foundation, or CKF, $1.3M to develop a new open source, modular platform for scholarly knowledge production and academic publishing.

Zika Virus and Birth Defects — Reviewing the Evidence for Causality

New England Journal of Medicine from April 13, 2016
The Zika virus has spread rapidly in the Americas since its first identification in Brazil in early 2015. Prenatal Zika virus infection has been linked to adverse pregnancy and birth outcomes, most notably microcephaly and other serious brain anomalies. To determine whether Zika virus infection during pregnancy causes these adverse outcomes, we evaluated available data using criteria that have been proposed for the assessment of potential teratogens. On the basis of this review,

we conclude that a causal relationship exists between prenatal Zika virus infection and microcephaly and other serious brain anomalies.

Also:
- Proving Zika guilty: A long and painstaking task (February 24, Reuters)
- On Zika, Congress Is Failing to Do Its Job (April 14, The New York Times, The Opinion Pages)
- SAS: Understanding Zika with Zetabytes (audio 41:07, April 11, FederalNewsRadio.com)

The dark side of Guardian comments

The Guardian from April 12, 2016

As part of a series on the rising global phenomenon of online harassment, the **Guardian** commissioned research into the 70m comments left on its site since 2006 and discovered that of the 10 most abused writers eight are women, and the two men are black. Hear from three of those writers, explore the data and help us host better conversations online.

Also:
- Investigating the algorithms that govern our lives (April 14, Columbia Journalism Review; Chava Gourarie)

Events

Reading and Talk by Alexander Pschera about Nature and the Digital Revolution – April 18

Deutsches Haus at NYU presents a talk with **Alexander Pschera**, the author of Animal Internet: Nature and the Digital Revolution (New Vessel Press), and the book's translator, Elisabeth Lauffer.

Monday, April 18 at 6:30 p.m., Deutsches Haus – located (42 Washington Mews at University Place)

Moore-Sloan Data Science Lunch Seminar Series

Wednesday, April 20 — **Boris Leistedt** from **NYU Center of Cosmology and Particle Physics**

Seminar meets from 12:30 - 1:30 p.m.

The Data Science Lunch Seminar Series is an informal weekly gathering of NYU Data Science affiliated persons to discuss data science related topics. Each week there is a 30 minute presentation, over lunch (provided), with additional time for conversation and questions.

Genes, Germs & GMOs: Social Media Effects on Scientific Controversy

Social media have dramatically increased the public's voice in debates over the social consequences of scientific developments. Millions of heretofore uninvolved

citizens have an opportunity to weigh-in on often complex and uncertain scientific questions in ways that are meaningful to themselves and their peers, and that affect the trajectory of debate as well as governmental support and regulations. To what extent has this development changed levels of social media users' empowerment and participation, and how has it affected scientists who feel its effects?

Wednesday, April 20, at 1 p.m., **Boston University**, Hillel Student Lounge (213 Bay State Road Room 227)

Text as Data Speaker Series

The NYU 'Text-as-Data' speaker series takes place on Thursdays from 4 – 5:30 pm in room 217, 19 West 4th St (unless otherwise noted). The series provides an opportunity for attendees to see cutting edge text-as-data work from the fields of social science, computer science and other related disciplines.

Thursday, April 21, will be **Noémie Elhadad** (**Columbia**) on *Summarizing the Patient Record.*

NYU Tech Meetup: Tech for the Public Good

WagnerTech, in partnership with **NYU Tandon School of Engineering**, **The GovLab** and the **NYU Center for Data Science**, invites you to hear first-hand from tech enthusiasts playing lead roles in their field, about how they are leveraging technology to advance high-impact innovation for the public good.

Lightning talks by:
* **Noelle Francois**, Executive Director, **Heat Seek NYC**
* **Chris Whong**, Adjunct Assistant Professor of Urban Planning at Wagner
* **Justin Salamon**, Post Doctoral Associate at the **Center for Urban Science and Progress**
* **Milena Berry**, Co-Founder and CEO at **PowerToFly.com**
* **Mark Swarek**, Artist & Lecturer on Integrated Digital Media

Friday, April 22, starting at 6 p.m., Leslie eLab (16 Washington Pl.)

Save the Date: WBDIH All Hands Meeting May 12, 2016

Join us for the first **West Big Data Innovation Hub** All Hands Meeting in Berkeley, CA, on May 12, 2016!

Deadlines

Neurohackweek -- September 5-9, 2016

Summer school for neuroimaging and data science ... Neurohackweek is a 5-day hands-on workshop in neuroimaging and data science, held at the **University of Washington eScience Institute**. Participants will learn about technologies used to analyze human neuroscience data, and to make analysis and results shareable and reproducible.

Deadline to apply: Monday, April 18

1st Workshop on Representation Learning for NLP

The 1st Workshop on Representation Learning for NLP (RepL4NLP) will be held on August 11, 2016 and hosted by the 54rd Annual Meeting of the Association for Computational Linguistics (ACL) in Berlin, Germany. The workshop is being organised by **Phil Blunsom**, **Kyunghyun Cho**, **Shay Cohen**, **Edward Grefenstette**, **Karl Moritz Hermann**, **Laura Rimell**, **Jason Weston**, and Scott Wen-tau Yih, and will be sponsored by **Google DeepMind**, **Facebook AI Research**, and **Microsoft Research**.

Deadline for paper submission: Sunday, May 8

2016 ACM SIGKDD Innovation and Service Awards

ACM SIGKDD invites your nominations for its 2016 Innovation and Service Awards.

ACM SIGKDD, ACM's Special Interest Group on Knowledge Discovery and Data Mining (KDD), is the premier global professional organization for researchers and professionals dedicated to the advancement of the science and practice of knowledge discovery and data mining. It established the Innovation and Service Awards to recognize outstanding technical and service contributions to the KDD field.

Deadline for nominations is Friday, May 20.

Open Data Research Symposium 2016 Call for Abstracts

The Open Data Research Symposium program committee is pleased to announce that the second Open Data Research Symposium (#ODRS16) will be held on October 5, 2016 prior to the International Open Data Conference 2016 in Madrid, Spain.

The deadline for abstracts submissions is Monday, May 30.

Tools & Resources

After 2.5 years, my co-authors and I have completed the deep learning textbook.

Facebook, Ian Goodfellow from April 06, 2016
www.deeplearningbook.org

Many thanks to the **Google Brain** team for encouraging me to devote several months of concentrated effort to this project!

New Tools to Summarize Text

Fast Forward Labs from April 11, 2016
We're excited to introduce the latest report and prototype from our machine

intelligence R&D group! In this iteration, we explore summarization, or neural network techniques for making unstructured text data computable.

Making language computable has been a goal of computer science research for decades. Historically, it has been a challenge to merely collect and store data. But it's now so cheap to store data that we often have the opposite problem: once we've data, how should we analyze it to find meaning and insights?

Accessing a Neo4j Graph Database Server from RStudio and Jupyter R Notebooks Using Docker Containers

AJ Hirst, OUseful.Info blog from April 12, 2016
In *Getting Started With the Neo4j Graph Database – Linking Neo4j and Jupyter SciPy Docker Containers Using Docker Compose* I posted a recipe demonstrating how to link a Jupyter notebook container with a neo4j container to provide a quick way to get up an running with neo4j from a Python environment.

It struck me that it should be just as easy to launch an R environment, so here's a docker-compose.yml file that will do just that.

Bing's New Live Code Editor Is Surprisingly Useful

Gizmodo from April 10, 2016
Google is famous for integrating tiny, handy features directly into its search results. With a new in-line live code editor that pops up when you search for programming questions, Bing may just have beaten Google at its own game.

If you search for something related to a major programming language in Bing, you'll now get a live code editor with examples of how to solve your problem. You can edit the code and run it all in-page, with instant feedback.

CRAN - Package urltools

Oliver Keyes from April 12, 2016
A toolkit for handling URLs that so far includes functions for URL encoding and decoding, parsing, and parameter extraction. All functions are designed to be both fast and entirely vectorised. It is intended to be useful for people dealing with web-related datasets, such as server-side logs, although may be useful for other situations involving large sets of URLs.

Careers

Data on PhDs Across the Disciplines

Daily Nous from April 04, 2016
The **National Science Foundation** (NSF) has released a new edition of its Survey of Earned Doctorates, according *Inside Higher Ed*, including data through 2014. According to the survey, 54,070 PhDs were bestowed in 2014,with 5,486 of them in humanities fields. Of those, only 54.3% had "job or post-doc commitments." For 2004, these figures were 5,210 and 63.4%, respectively.

Also noteworthy is the average amount of debt humanities PhDs accumulate during their studies (including their undergraduate years): $29,953, with 17% of all humanities PhDs having more than $70,000 in debt.

Research Associate in Natural Language Processing (Fixed Term) - University of Cambridge

University of Cambridge, Department of Theoretical and Applied Linguistics from June 06, 2016
Applications are invited for a three-year Research Associate post on the project LEXICAL: Lexical Acquisition across Languages. The project is funded by the **European Research Council** (ERC) in the form of a Consolidator Grant awarded to **Anna Korhonen**. The aim is to develop a novel computational modeling framework for learning and transferring lexical and semantic information across languages without the need for parallel resources. The project will cover a variety of typologically diverse languages and language domains and will demonstrate the usefulness for NLP applications such as machine translation.

Deadline to apply is Monday, June 6.

University of Delaware Postdoctoral Researcher in Special Collections and Digital Humanities

University of Delaware, Interdisciplinary Humanities Research Center from May 15, 2016
The **University of Delaware** Library and the College of Arts and Sciences invite applications for a Postdoctoral Fellow of Special Collections and Digital Humanities. We seek an untenured scholar in the humanities (PhD received January 2010 – June 2016). The mission of the Fellow is to promote primary sources related to African American culture found in Special Collections at the University of Delaware Library through collaborative instruction, programming, creative outreach, and project development.

Deadline to apply is Sunday, May 15.