Fact Sheet: Data to Knowledge to Action

New Announcements

November 12, 2013

Empowering the Patient, Curing Diseases, and Saving Lives

American Society of Clinical Oncology launches formal development of a learning health system to transform cancer care and improve outcomes for patients

Our nation has achieved tremendous progress against cancer thanks to knowledge gained through clinical trials. Yet only three percent of all cancer patients participate in trials. Imagine how much faster we could improve patient care and outcomes if health care providers learned from the other 97 percent--in real time. Through CancerLinQ™, the **American Society of Clinical Oncology** aims to realize this potential. CancerLinQ, a ground-breaking learning computer network, will unlock vast quantities of information on patient experiences that are now lost to file cabinets and unconnected servers. By analyzing data from any source and providing up-to-the-minute feedback and guidance to health care providers and patients, the system will improve the quality of care for all patients with cancer. CancerLinQ is a multi-phase initiative that ASCO has undertaken with support from its Conquer Cancer Foundation. It will take an investment of $80 million over the next five years to make CancerLinQ a reality. To date, the Foundation has raised $7.8 million in major commitments, including generous contributions from **Amgen; Chan Soon-Shiong Family Foundation; Genentech BioOncology™; Helsinn Therapeutics (U.S.), Inc.; Lilly; Novartis Oncology; Susan G. Komen®;** and numerous individual supporters including: **Raj Mantena, RPh** and **Thomas G. Roberts, Jr., MD** and **Susan M. DaSilva, NP.** ASCO has committed over $20 million for the first phase. Following successful completion of a prototype with "de-identified" (i.e., anonymous) data from 170,000 patients, ASCO announced today that it has initiated development of the full CancerLinQ system. ASCO plans to make the first in a series of successively more powerful quality improvement tools available to physicians by early 2015.

NIH, IBM, Sutter Health, and Geisinger Health System to develop new methods for early detection of heart disease

The National Institutes of Health (NIH) awarded **IBM Research**, **Sutter Health** and **Geisinger Health System**, a $2 million joint research grant to develop and apply new technologies and methods to data analytics that could help doctors detect heart failure years sooner than is now possible. Demonstrating the benefits of a public-private partnership to

society, the three organizations will collaborate to develop practical and cost-effective early detection methods to use in primary care practices with an electronic health record system.

Novartis, Pfizer, and Eli Lilly partner to improve access to information about clinical trials

Developing new treatments for diseases depends on clinical research studies and 57 percent of Americans say they would be interested in participating in a clinical trial. However, almost half of all studies never reach their recruitment targets. In order to connect patients and researchers, **Novartis**, **Pfizer** and **Eli Lilly and Company**, are partnering in the U.S. to provide a new platform to improve access to information about clinical trials. The platform will enhance **clinicaltrials.gov** and will provide more detailed and patient-friendly information about the trials, including a machine readable "target health profile" to improve the ability of healthcare software to match individual health profiles to applicable clinical trials. As part of the project, patients can search for trials using their own Blue Button data. To preserve data privacy, de-identified Blue Button data, used only as an "index to search," will not be stored anywhere outside of the patient's application. The platform launch is planned for early 2014 with a starting database of about 50 clinical research studies from the participating companies. The platform will be open on both ends: other sponsors of clinical research studies may upload information about their trials, while software companies develop tools to deliver this information to interested patients.

SAP, Stanford University and the National Center for Tumor Diseases in Heidelberg accelerate real-time personalized medicine from the "bench to the bedside"

SAP's HANA Healthcare Platform provides researchers, hospitals, pharmaceutical and insurance companies with biological, lifestyle and clinical data to optimize patient health by personalizing prevention, treatment and health maintenance; and mapping drug development more precisely to the biology of disease. In an SAP-funded collaboration with Carlos D. Bustamante's lab at the **Stanford School of Medicine**, researchers are using the SAP HANA Platform for healthcare for real-time analytics to uncover genetic variants that contribute to population health and disease. Discovering patterns of variation within and among different populations can help clarify how genes contribute to disease susceptibility in humans. To date, Stanford has seen from 17-600X faster computations in analyzing their genomics data. Stanford's work will ultimately lead to new treatments targeted for autism and cardiovascular disease, both of which are critical public health concerns. In clinical care, SAP and the **National Center for Tumor Diseases (NCT)** in Heidelberg, Germany are piloting a Medical Explorer tool based on SAP HANA. Physicians and researchers can securely analyze clinical and genomic data in real-time for patient breakthroughs in cancer diagnostics and treatment options. Phase 2 was launched in September 2013 and will emphasize how to predict disease risk and better match patients to clinical trials.

Rutgers University and the State University of New York, Stony Brook work with industry to implement new methods for medical imaging and managing streaming data

Through the **National Science Foundation**'s Industry/University Cooperative Research Center (I/UCRC) program, the Center for Dynamic Data Analytics (CDDA)--established between **Rutgers University** and the **Stony Brook University**--conducts multi-stakeholder projects that integrate research efforts between industry and academia. Their Quantitative Tissue Assessment project is developing methods for evaluating liver and spleen disease, as well as sarcopenia and other related debilitating illnesses. Recent estimates indicate that approximately 45 percent of the older U.S. population is affected by sarcopenia (18 million people in 2010) and the numbers are rising. Through a partnership with **BioClinica**, which specializes in medical imaging for clinical trials, the CDDA has developed novel algorithms to aid in the assessment of new therapies which will have a major impact in the health of the aging American population. The Scalable Indexing Project is developing techniques for storing and organizing high-bandwidth streams of data on disk drives so that they can be queried in real time. The result is that some big-data workloads can be accelerated by ten to 100 times, or even more. **Tokutek**, a rapidly growing enterprise database company with over two dozen corporate and academic customers, is commercializing this research.

New patient-owned cooperative to help citizens take ownership of their health data

Our Health Data Cooperative (ODHC) is a new cooperative developed to help patients better manage and directly benefit from the use of their health data while providing it anonymously for clinical research uses. Patients participating in the cooperative will have their health data anonymously collected using an OHDC stock number. All collected data will be aggregated to help health researchers and clinicians. Fees paid to OHDC and knowledge gained by subscribing health organizations for access and use of the data will be shared among the patient shareholders. The School of Information and Library Science at the **University of North Carolina at Chapel Hill** and the **Center for the Advancement of Health Information Technology at RTI International (UNC/RTI)** are partnering with ODHC to help build and design the database access as well as the evidence-based educational infrastructure for members that will help them understand the research uses of their data. ODHC is also partnering with **Touchstone Energy Cooperative** to enlist its more than 30 million existing members. Funding of up to $12 million is being planned to support the effort and enrollment is open to all U.S. residents with a goal to have 10 million members enrolled in the next 24 months.

DC-NET and CAAREN partner with George Washington University Medical Center on next generation genomic sequencing

The **George Washington University (GW)** Medical Center's Division of Genomic Medicine, GW's Capital Area Advanced Research & Education Network (CAAREN) in cooperation with

the District of Columbia's citywide network **DC-Net** have established a partnership to further GW's initiatives in next-generation genomic sequencing and other advanced research initiatives. Together, the partners will facilitate the compilation, analysis, and transmission of massive data sets of DNA and RNA sequences to identify biomarkers and new targets for therapies to treat diseases like cancer and heart disease. Next-generation sequencing machines produce terabytes of raw data and require analyzing data from hundreds to thousands of patients. Privacy issues are controlled at multiple levels from informed consent, anonymous and encrypted data files, and secured networks. To reduce the transfer and analytic turnaround from days to just hours, GW has invested in a massive new research High Performance Computing cluster and partners are helping to create a wide-area network between the cluster, GW's multiple campuses, the St. Laurent Institute, and at the researchers' homes in the DC area.

New OSTP initiative formed to leverage big data to predict pandemics

Globalization, ecological pressures and industrial practices are increasing the risk of a pandemic. **The Office of Science and Technology Policy's (OSTP) Predicting the Next Pandemic Initiative** will establish a consortium of communities of interest from government departments and agencies at all levels, non-governmental organizations, academic institutions, industry partners and others, to enhance multi-sector collaboration and the use of big data to predict pandemics before they occur. Closer collaboration with remote sensing, sensor technologies and observation networks will enhance identification and fusion of new and existing data sets, to enable development of big data analytics. Participants will begin with a pilot project that will explore the drivers of biological events and determine how big data can be used to predict pandemic potential of novel infectious agents. Existing data collection capabilities and requirements will be examined based on historical and recent disease emergence, data fusion, barriers to collaboration, and big data analytics to allow the future development of a predictive capability.

NIH's Big Data to Knowledge (BD2K) enhances access and use of biomedical data to researchers

The **National Institutes of Health (NIH)** Big Data to Knowledge (BD2K) initiative is a multiyear, multimillion dollar investment to develop new approaches, standards, methods, tools, software, and competencies necessary for biomedical scientists to capitalize on the big data being generated by the research community. It aims to enhance access to available biomedical data through technologies, approaches, and policies that enable and facilitate widespread data sharing, discoverability, management, duration, and meaningful re-use. It will provide for protections for individual privacy, confidentiality, and intellectual property; and develop a cadre of researchers skilled in the science of big data, in addition to having elevated general competencies in the

usage and analysis of big data across the biomedical research workforce. NIH plans to devote nearly $27 million in funding to BD2K in FY2014.

CMS creates virtual data center for lower cost and more security sharing of health data

The **Centers for Medicare and Medicaid Services (CMS)** is launching the **CMS Virtual Research Data Center (VRDC)**, a $13 million a year investment in developing a secure and efficient mechanism through which researchers will virtually access CMS data. Using the VRDC, researchers will access CMS data from their own workstation and will be able to analyze and manipulate the data within the virtual environment, affording numerous advantages in terms of efficiency, cost, privacy and data accuracy. VRDC is a more efficient mechanism for CMS to share data, lowering data costs for researchers. It enhances the overall privacy and security of protected beneficiary information by ensuring identifiable data does not leave the VRDC. And it provides researchers with access to timelier data by allowing them to refresh their study cohorts on a regular basis.

MedRed BT Health Cloud works for better healthcare in The U.S., U.K. and around the world.

MedRed and **BT** have collaborated to create the MedRed BT Health Cloud (MBHC), a new international project, currently in beta release, aimed at enhancing the integration and dissemination of open health data. This multiyear, transatlantic effort makes available one of the largest open health data repositories anywhere in the world. Based on the BT Cloud Compute platform, it currently includes several years of de-identified population health data from England, Scotland and Wales. The parties are committed to continually updating and refreshing the data in partnership with academic researchers and industry. U.S. public data sets, such as the FDA adverse event reporting data, recently released Medicare data from the Center for Medicaid Services, and data from other healthcare systems are currently being added. The MedRed BT Health Cloud is specifically designed for the life sciences and healthcare industries and we expect the collaboration to lead to new and actionable insights that can help change the way healthcare is delivered in the U.S., U.K., and around the world.

New York Mayor's Office uses city data to help new businesses get off the ground

In New York, the Mayor's **Office of Data Analytics (MODA)** focuses on projects that improve daily operations, help to prepare for and respond to disasters, and support economic growth. MODA is currently working with the **New Business Acceleration Team (NBAT)** to analyze the City's performance in helping new restaurants cut through red tape and open their doors to

customers. MODA calculated "time-to-open" by connecting information on construction permits (**Department of Buildings**), restaurant inspections (**Department of Health and Mental Hygiene**), and NBAT counseling notes. MODA found that NBAT's free services cut an average of 45 days off a new business' time-to-open; that's 45 more days of happy customers, revenue for business owners and jobs for New Yorkers. The analysis performed by MODA will help the city as it considers how to help even more businesses get off the ground quickly.

Center for Technology in Government (CTG) and the Institute for Financial Market Regulation (IFMR) engage financial sector stakeholders and academia on market stability

Bringing together industry representatives, regulatory bodies and academia with a grant awarded by the **National Science Foundation**, **CTG and IFMR** will collaboratively establish a research agenda focusing on the stability, openness and fairness of financial markets. Scheduled for November 14, 2013, this timely workshop will focus on the information sharing and collaboration challenges created through data complexity and volume, the dynamic and interconnected nature of markets, financial instruments, technologies, and institutions working within regulated markets and the regulatory process. CTG is a research center within the University at Albany, State University of New York (SUNY); IFMR is collaboration between the University at Albany and Albany School of Law.

Supporting the Earth, Energy Use, and the Environment

Amazon Web Services and NASA bring access to data about the Earth to the public

Amazon Web Services (AWS) and the **National Aeronautics and Space Administration (NASA)** will bring access to data about the Earth to the public through the NASA Earth eXchange (NEX), a collaborative sharing network for researchers in Earth Science. In a new NASA Space Act Agreement, AWS is working with NASA NEX to host a significant amount of NASA's Earth-observing data as an AWS Public Data Set. In exchange, NASA will create machine images, workshops, tutorials and other resources that help earth sciences researchers leverage the cloud in their daily work. For NASA, having this data in AWS will enable the calculation of the next National Climate Assessment on the AWS cloud. For the general public, open data access enables projects like **Citizen Science Alliance's** Zooniverse.org, which allows researchers to leverage the power of the crowd to quickly analyze massive data sets and work on problems that cannot be efficiently solved by computers. Currently, hundreds of thousands of "citizen scientists" regularly contribute to the Zooniverse projects running on AWS—including efforts to find exoplanets and gravitational lenses, model the Earth's climate, classify life on the ocean floor and classify animals photographed on the Serengeti. The Galaxy Zoo projects at

Zooniverse have classified over a million different galaxies in the Sloan Digital Sky Survey using contributions from citizen scientists.

The DOE to sponsor new data challenge to build tools to use electricity more efficiently

The U.S. Department of Energy (DOE), dedicated to helping citizens play a role in the future of how energy resources are used throughout the country is building platforms that enable the diverse energy resources within the United States. DOE's **Office of the Chief Information Officer** and **the Office of Electricity and Energy Delivery** is sponsoring the American Energy Data Challenge, which encourages the public to build tools that empower American consumers to use energy more efficiently and incentivizes families and businesses to make informed decisions.

Google Earth Engine to power new project to monitor changes in the world's forests

Google is working with the **World Resources Institute** on Global Forest Watch 2.0 (GFW 2.0), a new forest monitoring tool that uses data from satellite imagery, monitoring systems, and mobile technology to provide near real time information on forests all over the world. It will also include crowd-sourcing capability so people on the ground can report deforestation when it takes place. Google's Earth Engine is hosting the underlying satellite data and the scientific analysis to detect changes in forest cover. This project also receives support from multiple government, industry, and civil society partners including the **Government of Norway, Staples,** the **University of Maryland** and **United States Agency for International Development (USAID)**. **USAID** has already contributed $5.5 million to the project which will be launched later this year.

UCSD, Clean Tech San Diego, OSIsoft and San Diego Gas & Electric to create a "Sustainable San Diego"

Clean Tech San Diego, along with the **Predictive Analytics Center of Excellence (PACE), the San Diego Supercomputer Center** (SDSC) at the University of California, San Diego (UCSD) and **OSIsoft** are working to develop a "Sustainable Communities" project for downtown San Diego. The project will deploy a data infrastructure that connects physical systems such as those managing electricity, gas, water, waste, buildings, transportation and traffic; and enable the city of San Diego to drive city-scale applications that lower electricity consumption and cost, discover and anticipate grid instabilities, educate the public, and improve both quality of life and economic development. The OSIsoft software system will connect to and acquire significant volumes of highly detailed data streams which will be published in a cyber-secure, private cloud that is accessible based on signed and approved access mechanism agreements. Presently, **San Diego Gas & Electric** (SDGE) and **UCSD** are beta-testing the OSIsoft software and UCSD

researchers are using UCSD's Microgrid to analyze the data on the main SDGE grid and the UCSD Smart Grid. Processes developed and results will be published lead other communities on their own path to sustainability.

USGS, NOAA, and USDA launch new Big Earth Data Initative

The **United States Geological Survey (USGS)** will provide $9 million (as indicated in the President's 2014 budget request) for a Big Earth Data Initiative in collaboration with the **National Oceanic and Atmospheric Administration (NOAA)**, **NASA**, and the **United States Department of Agriculture (USDA)**. Building on lessons from USGS' Powell Center for Analysis and Synthesis that engages in cross-disciplinary scientific collaboration on complex natural science problems, this initiative will dramatically improve data discovery, accessibility, and usability for scientific discovery and problem solving.

DataKind works with industry and non-profits to bring data analytics talent to society's most challenging problems

DataKind, a non-profit that matches data scientists with non-profit and non-governmental organizations, is partnering with **Pivotal** to bring some of industry's top data analytics talent to bear on some of society's greatest challenges. Many high impact social organizations have huge troves of data but lack the resources to analyze them; through Pivotal's new data philanthropy initiatives, their data scientists can volunteer their unique skills and engage data scientists around the world. DataKind is also exploring similar partnership opportunities with **Teradata**, a sponsor of DataKind's efforts and another industry leader committed to using their data services for the greater good. One of DataKind's newest supported projects is a partnership with **The Mission Continues** which aims to better understand the effects of their volunteer programs on improving veterans' lives. DataKind is also starting a new project with **Medic Mobile**, a non-profit that uses communication technologies to improve the health of under-served and disconnected communities. The global health community estimates that a frontline health worker prevents a child death every three seconds. DataKind is working with Medic Mobile to quantifiably measure the impact of their various health initiatives. By scaling up access to the world's standard for impact measurement, Medic Mobile and DataKind could improve health for millions of people by making the case for large scale mobile-based interventions.

The Kamusi Project aligns universities around the world and "citizen linguists" to create a dictionary of every word in every language on earth

Documenting the millions of words in the 7000 languages spoken around the world seems almost impossible – and yet, this is the data mission of the **Kamusi Project**. Kamusi, a Swahili word for dictionary, with support from the **National Endowment for the Humanities**, has designed

the Global Online Living Dictionary (GOLD), a system to capture the full range of human linguistic data and make it available for education and for technologies such as speech recognition and automatic translation. Kamusi GOLD will begin working with such partners as the **Long Now Foundation** of California and the **Global WordNet** based at Princeton to unite existing data sets and a partnership network of U.S. and global linguistic experts such as the **University of Ngozi** in Burundi to collect new data. With EPFL, one of the two **Swiss Federal Institutes of Technology**, Kamusi is designing a unique crowdsourcing system to gather and validate a vast array of data now only in human brains, including the many form changes that words go through when exposed to grammatical rules, the historical evolution of terms, and geo-tagging to discover the ranges of particular words and pronunciations. Users who share their knowledge will earn free access to the data, a "play-to-pay" system that will include social networking and competitive gaming elements. The data will be accessible through apps for most devices, including a lightweight offline system under development for One Laptop per Child.

MIT and the city of Boston launch big data challenge around urban transportation

The big data initiative at the **MIT Computer Science and Artificial Intelligence Laboratory (CSAIL)** is organizing competitions designed to spur innovation in how scientists think about and use data to address major societal issues. Working with a number of partners, the MIT Big Data Challenge will define real-world challenges in different areas such as transportation, health, finance and education, and make available data sets for the competition. The first MIT Big Data Challenge, which launched in October 2013, in partnership with the **City of Boston,** focuses on transportation in downtown Boston. The Challenge makes available multiple data sets, including data from more than 2.3 million taxi rides, local events, social media and weather records, with the goal of predicting demand for taxis in downtown Boston and creating visualizations that provide new ways to understand public transportation patterns in the city.

Splunk, NoticeandComment, and Sunlight Foundation innovate to facilitate federal and local civic engagement using Regulations.gov

Regulations.gov, the centralized portal for all federal regulatory dockets, including proposed rules, and publicly submitted comments—is enabling innovative new ways to improve civic engagement by making the millions of regulatory documents available on the site more accessible, searchable, and usable for the general public via their data API**. Splunk4Good**, the social responsibility initiative of **Splunk**, will analyze large and complex data and create a new, public interface that enables users to explore the federal regulatory data through real-time graphs, dashboards, and visualizations. **The Sunlight Foundation** uses Regulations.gov to power Docket Wrench, which lets users analyze more than four million documents by using visualizations that group textually similar documents together to illuminate how rulemakings are influenced by outside groups. Sunlight is now expanding Docket Wrench using natural-language

and machine-learning techniques to visualize authorship, sentiment and content. In addition, **NoticeandComment.com** has developed a new publishing platform using the framework of Regulations.gov as their model for all local governments to help publish and manage the 17 million public notices generated by over 89 thousand municipal and special district governments in the U.S., as well as the hundreds of millions of public comments and responses. Over $500 million annually is spent by local governments on advertising paper notices; use of the new publishing platform is free to all local governments, and the public can access the database via the web and through a new customizable mobile app.

NSF investments in a diverse portfolio of research to advance data analytics

Since the launch the National Big Data Research and Development Initiative, the **National Science Foundation** (NSF) has invested tens of millions of dollars in foundational research, infrastructure and education for data innovation. Nearly $62 million was awarded through the Core Techniques and Technologies in Big Data and Data Infrastructure Building Blocks programs, funding projects to advance and build data analytics capabilities, with potential impacts ranging from improved genome analysis to smarter search engines. NSF also sponsored an intensive, 5-day interdisciplinary workshop that generated transformative ideas for using large datasets to enhance the effectiveness of teaching and learning. Other investments include a $3.75 million award for data analytics research to improve the usability of online privacy policies, and a $25 million investment to launch a **Center for Minds, Brains and Machines at Harvard** that will both use and build new capabilities in data analytics while unraveling the mysteries of the human brain and building smarter machines.

Berkeley's AMPLab releases open source license on software for lightning fast cluster computing used in major industrial collaborations

UC Berkeley's Algorithms, Machines, and People Laboratory (AMPLab) has released its popular Spark big data analytics system under an Apache Open Source license, making the software freely available to data scientists, developers, researchers and other users. A key component of the Berkeley Data Analytics Stack (BDAS), Spark has been shown to be 10 to 100 times faster than the industry standard Hadoop Map Reduce for many important types of data analytics. In addition to Spark, BDAS includes support for SQL queries, real-time streaming, approximate query processing, large graph databases, and declarative machine learning. Spark and BDAS are seeing increasing adoption by companies of all sizes. For example, **Yahoo!** is using Spark and its Shark SQL interface in its advertising and data platforms to personalize user experiences. **ClearStory Data** uses Spark to provide highly interactive access to big data, and **DataBricks**, is a newly formed and funded company focused on furthering the Spark ecosystem.

The AMPLab is supported by a combination of public and industrial sponsors, including a $10 million **National Science Foundation** Expeditions in Computing award, **DARPA** XData, and 21 industrial partners including founding sponsors **Amazon Web Services, Google,** and **SAP**.

SGI and Fedcentric introduce cutting edge supercomputing power to the biggest and fastest data challenges

In data-intensive industries such as manufacturing, media, life sciences and earth sciences, the volume and speed of streaming data that must be analyzed are pushing the boundaries of hardware capabilities. **SGI** is working on bringing the power of cutting edge supercomputing technology to the toughest data challenges that these industries face every day. Working with **Fedcentric Technologies**, SGI has created a cost efficient commodity high density computing systems to manage high velocity data, in particular to address high-speed detection, processing, and analysis for fraud prevention. This year SGI and Fedcentric are working with the **U.S. Postal Service** to leverage the system to manage the over 8 billion transactions and 275 terabytes of data they see every day and create new capabilities designed to dynamically route mail, enabling expedited shipping, delivery and service.

Industry, academia, and government come together to build big data performance benchmarking standards

The Big Data Top100 List is a new open, community-based big data benchmarking initiative coordinated by a board of directors including representation from the **San Diego Supercomputer Center, Pivotal, Cisco, Oracle, Intel, Brocade, Seagate, NetApp, Mellanox, Facebook, IBM, Google**, and the **University of Toronto**. The initiative is announcing the start of the Big Data Benchmark Challenge to seek community input in defining big data benchmarks and metrics. The creation of objective standards for application-level performance and price/performance fosters competition and innovation in the marketplace. Over the past 20 years, for example, the benchmark performance of commercial database software has improved by about a million times, while the price/performance has improved by a factor of a couple of hundred thousand. With support from the **National Science Foundation**, the initiative has been hosting a series of community workshops, with a fourth workshop to be held in October of 2013. As a part of this effort, the **National Institute for Standards and Technology (NIST)** has recently funded researchers at the San Diego Supercomputer Center to study different strategies for data generation for big data.

NIST launches new community based big data effort

The **National Institute of Standards and Technology (NIST)** invested $750K in a community-based effort to shed light on the question, "What is "Big Data" and how does it differ from

traditional data environments?" Answers to these and other basic questions are required to address the big challenges of big data comprehensively. NIST is pleased to announce that initial answers to these questions have emerged from a community-based effort by dedicated, expert participants from the academic and commercial sectors in the U.S. and abroad. They encourage members of the public to listen to their ideas and learn how to participate in the continuing conversation.

University of Illinois to develop big data as an engineering discipline with a Grainger Foundation endowment

Engineering at Illinois combines rigor and a culture that embraces interdisciplinary collaboration to tackle society's most challenging problems. They are announcing that the new Grainger Engineering Breakthroughs Initiative, built on a $100 million gift from **The Grainger Foundation**, will focus on two topics-- big data and bioengineering, fields that grow from collaborative and multidisciplinary roots. Illinois' big data efforts will serve as a broad foundation for knowledge and innovation, creating new and original ways to improve the world through cross-cutting exploration, education, and research. Using the power of big data to challenge assumptions and break barriers, thought-leading teams from highly ranked departments across the college, and across the Illinois campus will make unprecedented advances in every field. Just as importantly, these teams will define the curriculum and preparation for future generations.

NYU, Berkeley, and the University of Washington launch a bold new partnership to harness the potential of data scientists and big data for basic research and scientific discovery, supported by the Moore and Sloan Foundations

New York University, the **University of California, Berkeley** and the **University of Washington** have launched a five-year, $37.8 million, cross-institutional effort with support from the **Gordon and Betty Moore Foundation** and **Alfred P. Sloan Foundation**. This ambitious collaboration seeks to harness the full potential of the data-rich world that characterizes all fields of science and discovery. Success depends on the individuals and teams that combine subject-matter expertise with computational, statistical and mathematical skills-sometimes called data science. While data science is already contributing to scientific discovery, substantial systemic challenges need to be overcome to maximize its impact on academic research. This new partnership--a coordinated, distributed experiment involving researchers at these leading universities--hopes to establish models that will dramatically accelerate this data science revolution.

Data Science for Social Good Fellowship will connect students with non-profits and government partners on high impact problems

The **University of Chicago** will soon begin accepting applications for fellows, mentors, and project partners for the 2014 class of The Eric & Wendy Schmidt Data Science for Social Good Summer Fellowship. Funded by **The Schmidt Family Foundation**, the fellowship trains aspiring data scientists with computer science, statistics, and quantitative social science backgrounds to solve real-world problems with high social impact. In 2013, the fellowship's first year, 40 graduate and advanced undergraduate students worked with more than 20 government agencies and non-profit organizations on 12 projects tackling problems in health care, energy, education, disaster relief, transportation, and city services. Fellows worked in small teams led by experienced mentors from academic and industry, learning and applying technical data analytics skills in order to find solutions to high impact social problems. These problems were collaboratively scoped with government and non-profit partners--including the **City of Chicago, Cook County Land Bank,** the **Cook County Sheriff, Ushahidi, Qatar Computing Research Institute, Lawrence Berkeley National Laboratory, Environmental Defense Fund, NorthShore University Health System, Nurse-Family Partnership**, and the **Case Foundation**--who contributed data and expertise to the projects. The teams developed prototypes for reducing bus crowding, identifying students at risk of missing college opportunities, predicting cardiac arrests in hospital patients, and other applications that will help project partners do more with their data in the future. For institutions interested in founding similar programs, the fellowship will hold a workshop on November 15, 2013 in Chicago.

IBM partners with universities and industry to create a tool to help position students for the most in-demand data jobs

IBM is unveiling the IBM Analytics Talent Assessment, a first-of-its-kind online tool that allows university students to gauge their readiness for public and private sector big data and analytics jobs. The assessment tool will also help students gain guidance on ways to further develop themselves as in-demand data crunching job candidates. The assessment focuses on essential competencies and traits that indicate whether a student not only has the ability to analyze data, but can parlay it into effective strategies that will fuel the Smarter Workforce in the public and private sectors. The eight universities collaborating with IBM on the project are among the 1,000 plus universities that partner with IBM through IBM's Academic Initiative to offer big data and analytics curriculum.

TechAmerica Foundation to host Big Data Road Shows: Big Data, Bigger Results

Bringing together industry representatives, senior federal and state government officials and academia, the **TechAmerica Foundation** have recently concluded a series of Big Data

Roadshows this Fall in Silicon Valley, California; Austin, Texas; and Boston, Massachusetts. The roadshows focused on the impact that big data is having on both the public health and energy sectors through insightful analysis and case studies from the private sector and government. The roadshows follow the 2012 release of the TechAmerica Foundation report "Demystifying Big Data: A Practical Guide to Transforming the Business of Government," which provides a road map for the federal government to better leverage and utilize big data and its necessary technologies.

MIT forms new working group on big data and privacy

The MIT Big Data Initiative is launching the Big Data Privacy Working Group, which will officially kick off November 2013, and bring together stakeholders from academia, industry, government and non-profits to focus on the future of big data and the unique issues and challenges surrounding privacy. The group's goal is to think long term to better understand and help define the role of technology in protecting and managing privacy, in particular when large and diverse data sets are collected and combined. The group will work toward collectively articulating major privacy challenges and developing a roadmap for future research needs. While there are a wide variety of technical approaches to privacy protections, what is lacking is a sense of how they might actually work at scale or if new technical tools are needed. This working group aims to close that gap so that large scale analysis of data can proceed in a manner that is respectful of privacy values. Technology and policy will play a role in defining how we collect and manage personal data in the future, but the path forward will be a mix of technology and public policy approaches, which is why bringing together key stakeholders, across disciplines, to discuss and better understand these issues is so important.

New Council for Big Data, Ethics, and Society formed to provide critical social and cultural perspectives on big data initiatives

In collaboration with the **National Science Foundation**, the **Council for Big Data, Ethics, and Society** will launch in early 2014 to provide critical social and cultural perspectives on big data initiatives. The council will bring together researchers from diverse disciplines--from anthropology and philosophy to economics and law--to address issues such as security, privacy, equality and access in order to help guard against the repetition of known mistakes and inadequate preparation. Through public commentary, events, white papers and direct engagement with data analytics projects, the council will develop frameworks to help researchers, practitioners, and the public understand the social, ethical, legal and policy issues that underpin the big data phenomenon. The council is being co-directed by danah boyd, Geoffrey C. Bowker, Kate Crawford and Helen Nissenbaum.