DEPARTMENT OF HEALTH AND HUMAN SERVICES
NATIONAL INSTITUTES OF HEALTH
NATIONAL LIBRARY OF MEDICINE
MINUTES OF THE BOARD OF REGENTS
February 7-8, 2017

The 174th meeting of the Board of Regents was convened on February 7, 2017, at 9:00 a.m. in the Donald A.B. Lindberg Room, Building 38, National Library of Medicine (NLM), National Institutes of Health (NIH), in Bethesda, Maryland. The meeting was open to the public from 9:00 a.m. to 3:45 p.m., followed by a closed session for consideration of grant applications until 4:15 p.m. On February 8th, the meeting reopened from 9:00 a.m. until adjourning at 11:30 a.m.

MEMBERS PRESENT [Appendix A]**:**
Dr. Alessandro Acquisti, Heinz College, Carnegie Mellon University
Ms. Jane Blumenthal, University of Michigan
Dr. Robert Greenes (via teleconference), Arizona State University
Dr. Eric Horvitz, Microsoft Research
Ms. Sandra Martin, Wayne State University
Dr. Daniel Masys, University of Washington
Dr. Gary Puckrein, National Minority Quality Forum
Dr. Esther Sternberg [Acting Chair], University of Arizona
Dr. Jill Taylor, Wadsworth Center, New York State Department of Health

EX OFFICIO AND ALTERNATE MEMBERS PRESENT:
Col. Thomas Cantilina, United States Air Force
Dr. Wayman Cheatham, United States Navy Bureau of Medicine and Surgery
Dr. Carolyn Clancy, Veterans Health Administration
Dr. James Deshler, National Science Foundation
Mr. Stan Kosecki, National Agricultural Library
VADM Vivek Murthy, Office of the Surgeon General, PHS
Dr. Richard Thomas, Uniformed Services University of the Health Sciences
RADM Sylvia Trent-Adams, Office of the Surgeon General, PHS
Ms. Meg Tulloch, Library of Congress

CONSULTANTS TO NLM PRESENT:
Dr. H. Kenneth Walker, Emory University School of Medicine

SPEAKERS AND INVITED GUESTS PRESENT:
Dr. Elizabeth Chen, Brown University
Dr. Lawrence Tabak, Office of the Director, NIH

MEMBERS OF THE PUBLIC PRESENT:
Ms. Andrea Baruchin, Friend of the National Library of Medicine
Mr. Glen Campbell, Friends of the National Library of Medicine
Mr. Javier Crespo, National Network of Libraries of Medicine
Ms. Erica Froyd, Lewis-Burke Associates

Ms. Lesley Macherelli, Healthbox
Mr. Stavros Michailidis, Knowinnovation
Dr. Barbara Redman, New York University
Dr. Elliot Siegel, Consultant
Mr. Thomas West, Krasnow Institute

FEDERAL EMPLOYEES PRESENT:
Dr. Patricia Flatley Brennan, Director, NLM
Ms. Betsy Humphreys, Deputy Director, NLM
Dr. Milton Corn, Deputy Director for Research and Education, NLM
Dr. Sameer Antani, Lister Hill Center, NLM
Ms. Stacey Arnesen, Division of Specialized Information Services, NLM
Ms. Dianne Babski, Division of Library Operations, NLM
Ms. Joyce Backus, Division of Library Operations, NLM
Dr. Dennis Benson, National Center for Biotechnology Information, NLM
Ms. Kathi Canese, National Center for Biotechnology Information, NLM
Ms. Hua Florence Chang, Division of Specialized Information Services, NLM
Ms. Kathy Cravedi, Office of Communications and Public Liaison, NLM
Mr. Ivor D'Souza, Office of Computer and Communications Systems, NLM
Mr. Todd Danielson, Office of the Director, NLM
Ms. Victoria Douglas, Division of Extramural Programs, NLM
Dr. Kathel Dunn, Division of Library Operations, NLM
Dr. Valerie Florance, Division of Extramural Programs, NLM
Dr. Dan Gerendasy, Office of Health Information Program Development, NLM
Mr. David Gillikin, Division of Library Operations, NLM
Ms. Rebecca Goodwin, Office of Health Information Program Development, NLM
Dr. Stephen Greenberg, Division of Library Operations, NLM
Mr. Daniel Hartinger, Office of Acquisitions, NLM
Dr. Zoe Huang, Division of Extramural Programs, NLM
Dr. Michael Huerta, Office of Health Information Program Development, NLM
Mr. Nicholas Ide, Contractor, Lister Hill Center, NLM
Ms. Christine Ireland, Division of Extramural Programs, NLM
Ms. Janice Kelly, Division of Specialized Information Services, NLM
Ms. Elizabeth Kittrie, Office of Health Information Program Development, NLM
Ms. Lisa Lang, Division of Library Operations, NLM
Dr. David Lipman, National Center for Biotechnology Information, NLM
Dr. Robert Logan, Office of Communications and Public Liaison, NLM
Dr. Clement McDonald, Lister Hill Center, NLM
Mr. Dwight Mowery, Division of Extramural Programs, NLM
Mr. David Nash, Office of the Director, NLM
Ms. Deborah Ozga, Division of Library Operations, NLM
Dr. Barbara Rapp, Office of Health Information Program Development, NLM
LCDR Yvonne Santiago, Office of the Surgeon General, PHS
Lt. Jessica Scruggs, Office of the Surgeon General, PHS
Mr. Jerry Sheehan, Office of the Director, NLM
Dr. Hua-Chuan Sim, Division of Extramural Programs, NLM

Dr. George Thoma, Lister Hill Center, NLM
Ms. Amanda Wilson, Division of Library Operations, NLM
Dr. Fred Wood, Office of Health Information Program Development, NLM
Dr. Jane Ye, Division of Extramural Programs, NLM
Dr. Deborah Zarin, National Center for Biotechnology Information, NLM
Mr. Mark Ziomek, Division of Library Operations, NLM

I. OPENING REMARKS

Dr. Esther Sternberg, NLM Board of Regents Acting Chair, welcomed the Regents, alternates, and guests to the 174th meeting of the Board.

II. REPORT FROM THE OFFICE OF THE SURGEON GENERAL, PHS

Vice Admiral Vivek Murthy began with an overview of his office. He then spoke about two reports the office issued in 2016: *Facing Addiction in America*, on alcohol, drugs, and health, and *E-Cigarette Use Among Youth and Young Adults*.

Dr. Murthy said there are about 21 million people in America with a substance abuse disorder and that only one in 10 people are getting treatment. Evidence-based prevention and treatment strategies reduce the risk of overdose and death and save money. He said for evidence-based treatment strategies, for every dollar invested, they save $4 in health care costs and $7 in criminal justice system costs. Prevention programs, many of which are school-based, save up to $64 for every $1 invested. His office launched a campaign last year called "Turn the Tide Campaign" to engage clinicians on addressing the opioid epidemic, but he said we shouldn't lose sight of other substance abuse disorders.

The report on e-cigarettes and youth was issued because of growth in e-cigarette use among youth, from a rate of 1-2% to 13-19%. E-cigarettes are now used more by kids than traditional cigarettes. Board member Dr. Eric Horvitz asked if there were statistics on cigarette smoking by the same cohort. Dr. Murthy responded that smoking traditional combustible cigarettes declined among kids even before 2011, but they don't know if that was because of e-cigarettes. He said most kids who use e-cigarettes also use traditional cigarettes.

Dr. Murthy next discussed a campaign on emotional wellbeing. Emotional wellbeing or emotional fitness is an important counterpart to physical wellbeing and physical fitness, but is often ignored. We know from a growing body of scientific data that emotional wellbeing isn't just something that happens; it can be cultivated and one of the most powerful tools for cultivating emotional wellbeing is social connection. The Office of the Surgeon General (OSG) is seeking to establish emotional wellbeing as a pillar of health and prevention that stands alongside physical activity and nutrition—in particular, food insecurity. There are 42 million Americans living in food insecure households, which includes one in four children.

Public Health Reports, the official journal of the Public Health Service (PHS) and the Surgeon General since 1878, will be redesigned and modernized, and its reach among the public will be expanded.

The Surgeon General then talked about the USPHS Commission Corps, which is overseen by the OSG. A total of 6,600 Corps members address public health concerns around the country and the world. Increasingly, these officers are seen as a force that can respond to chronic issues, including the opioid crisis and the challenge of e-cigarettes. OSG will be strengthening, training, and developing officers to enhance their ability to deploy effectively and expanding ties within and outside of the US Department of Health and Human Services (HHS), with the Department of Veterans Affairs (VA), Department of Defense (DOD), and other partners. They are also looking to do joint training missions.

Board member Ms. Jane Blumenthal asked about strategies to increase social connection in older people. Dr. Murthy said that a consequence of good science and medicine is that people live longer, but it doesn't have to be a consequence that people are lonelier. He said the science around how to address loneliness in older people is growing, but we do know there are things that help. There are experiments where young people are paired with people in nursing homes, and he sees mutual benefits for both. He talked about the Silver Line, a helpline in England that assists people with social isolation. He's a fan of thinking about how to help with isolation in an intergenerational fashion. When they released the Call to Action on Walking and Walkable Communities in 2015, they had intergenerational walks that promoted physical health and helped with isolation.

III. REPORT FROM THE NIH OFFICE OF THE DIRECTOR

Dr. Lawrence Tabak, principal deputy director of the National Institutes of Health, focused his talk on four areas: the 21st Century Cures Act; the NIH Clinical Center; the Associate Director for Data Science (ADDS) Office and the Big Data to Knowledge Program (BD2K); and transition.

The 21st Century Cures Act, landmark legislation passed overwhelmingly in a bipartisan fashion in December 2016, established the "so-called NIH Innovation Account" that provides $4.8 billion for 10 years. That funding has to be appropriated on a year-by-year basis and is distributed among specific initiatives. Dr. Tabak gave a broad overview of: the BRAIN (Brain Research through Advancing Innovation Neurotechnologies) Initiative; the Precision Medicine Initiative, now called the All of Us Research Program; the Cancer Moonshot; the Regenerative Medicine Initiative, which is being done in coordination with the Food and Drug Administration to support clinical research using adult stem cells, including autologous stem cells to further the field of regenerative medicine; and the Next Generation Researchers Initiative. The 21st Century Cures Act also authorizes the NIH Director to require awardees to share data, which Dr. Tabak called "a potential game changer."

At the Clinical Center, the largest research hospital in the world, all the patients are on some protocol. After contamination of a sterile product that turned out to be indicative of broader challenges, an advisory committee to the NIH director was established "to fortify our culture and practice of safety and quality, to strengthen our leadership for clinical care quality, oversight, and compliance," said Dr. Tabak. He explained how the Clinical Center used to have decentralized management like NIH. Since the contamination, they strengthened leadership by appointing the first-ever chief executive officer, an associate director for clinical research and chief scientific

officer, and a new chief of pharmacy, and they changed the reporting structure so all clinical directors report directly to an Institute Director. In addition, 70 focus groups with over 800 employees resulted in creation of an anonymous toll-free hotline, a newsletter about safety, and more.

Before briefly talking about ADDS office and BD2K, Dr. Tabak emphasized that biomedical research is now conducted in the Big Data space and that data science is integral to everything that NIH does. He said that support for the BD2K program is derived from the NIH Common Fund and all the Institutes and Centers. He said that Dr. Brennan has "graciously agreed" to serve as the interim director of ADDS and BD2K, which will be managed as a Common Fund program by the Division of Program Coordination, Planning, and Strategic Initiatives (DPCPSI).

Then Dr. Tabak addressed transitioning in a new administration and how there are regular meetings with the new leaders, which he described as professional and cordial. They are reviewing NIH processes and priorities. *[NOTE: At the time of this meeting, Dr. Tom Price had been nominated to serve as HHS secretary; Mr. Norris Cochran was acting secretary.]*

After Dr. Tabak's talk, there was a short discussion with board members. Dr. Daniel Masys asked whether the NIH strategic plan should be considered a model for NLM. Dr. Tabak replied, "No, so please don't be constrained" and Let the creative juices flow." Ex-officio member Dr. Carolyn Clancy from the Veterans Health Administration asked about the 21st Century Cures Act sounding like a new direction for research, more analogous to the Human Genome Project, where there are large initiatives and multiple components, but all collaborate and work together in a different way. Dr. Tabak replied that, indeed, many of the programs will require large team science efforts, and that the BRAIN Initiative and the Cancer Moonshot are of the scale and the complexity where team science will be at the forefront.

IV. SEPTEMBER MINUTES AND FUTURE MEETINGS

The Regents approved without change the minutes from September 13-14, 2016 meeting. The May 2017 meeting will take place on May 9-10, 2017, the 2017 Fall meeting will take place September 12-13, 2017, and the Board approved holding the Winter meeting February 13-14, 2018.

V. REPORT ON DATA SCIENCE

NLM Director Dr. Patricia Brennan announced that, effective January 8, she was appointed Interim Associate Director for NIH for Data Science. In this capacity, she will work with a team, including new and existing NLM staff, to provide input to the overall NIH vision and actions undertaken by each of the 27 Institutes and Centers in support of biomedical research as a digital enterprise.

NIH became interested in data science following the initial completion of the Human Genome Project, when it was left with volumes of data and needed to figure out how to analyze, organize, and keep them safe. Following the 2012 recommendations of the Advisory Committee to the NIH Director Data and Informatics Work Group, the starting point for trans-NIH data science was the Big Data to Knowledge (BD2K) initiative. Five years after its creation, the NIH data science

community has conducted a wide range of data science-related activities, including training, community building, standards setting, and coordination. But we're now at the point of a pivot process.

Dr. Brennan encouraged the Board to think about data science in the strategic planning process. The Advisory Committee to the Director of NIH regarding the NLM (2015) left NLM with a daunting challenge: to become the epicenter of data science at NIH. Why are we concerned about data science at NIH? Because we can accelerate biomedical discoveries by leveraging the large amount of data available. Of course, being a platform for discovery has always been a central mission of NIH and NLM.

With commitments from the Common Fund and the NIH Institutes and Centers, BD2K will now see a commitment of $250 million in resources over the next four years. This down payment for the future, which will grow, will benefit NLM.

Dr. Brennan then gave an overview of scientific discovery through time, asking, where do discoveries originate? At first, scientific discovery was experiential. The ancients could only discern medical cause through observation, and therapy through trial and error. The experimental era, from the late 17th through late 20th century, used controlled strategies and was a period of robust discovery.

In the 1950s and 60s, the products of research included both answers to questions and the data generated to support the answer. With the dawn of the 21st century, the world entered a new period—the data-driven era. Now, an increasing number of NIH extramural and intramural projects are generating large amounts of data, which prompts the question, what should we preserve?

Dr. Brennan then briefly discussed five data science approaches to accelerate discovery in Alzheimer's disease: machine learning (highlighting the role of white matter hyperintensities); data discovery (locating high-value data sets for exploring new hot spots in the brain); data wrangling (exploring the impact of lifestyle and other factors); treatment planning (determining which of millions of voxels should be the target of sequential observations); and stochastic models (selecting the most likely disease progress from thousands of trajectories).

So, the fundamental substrate for discovery today is data. She mentioned a key aspect of NLM's commitment to data science—ensuring that new and existing data sets are Findable, Accessible, Interoperable, and Reusable (FAIR). These FAIR principles, agreed upon by the wider data science community, are fundamental to what library scientists already know and have long been doing. She described each element in greater detail, along with its parallels to ongoing information management efforts at NLM and new functions required for the future. For example, it's important to create mechanism to determine high-value data sets, forecast their costs, and anticipate their utilization. We also need to develop preservation strategies for data sets and consider provenance and sustainability.

As her next theme, Dr. Brennan discussed computation. This critical area of data science includes analytics, visualization, and management. There are challenges because of the volume,

size, and distribution of data. There is also a concern to create privacy-preserving analytics. There are emerging opportunities in the areas of machine learning and optimization, which can serve as a complement to current biostatistical methods. Also, visualization will take on new forms, because the type of data being produced is so large and complex that it can't be scanned by the human eye. NLM must also look closely at data management—what needs to be preserved, in what format, and at what cost. Considerations include preserving the provenance of analytical strategies and maintaining version control.

Another important consideration of data science is infrastructure. Elements here may include: a commons, which is a physical storage space and a suite of utilities, currently gaining wider acceptance across several disciplines; identity management and authentication; planning and forecasting tools; and business analytics.

So how does this pivot to data science relate to NLM's history? It's a logical progression. Since our founding, we've provided services to the public, with changes in infrastructure allowing us to expand our reach using digitization and networks. In the current era, the Library will be attuned to the public's need for information, as always, and the new focus will be on data.

Some things are already in place at NLM, such as the free archive, PubMed Central, which will be accepting data deposits beginning in fall of 2017, and ClinicalTrials.gov. Current needs on campus include improving the safety and accessibility of high-value data sets and accelerating observational trials. And what's needed in the future to make data FAIR? Robust, sustainable storage solutions. Again, all of this is designed to facilitate discovery. Other key needs are analytical tools and building new relationships, both in the public and private sectors.

NLM has a "secret sauce" in all of this, which makes it unique at NIH: our many years of experience with standards, which will help make data FAIR and usable. NIH is beginning to see how helpful NLM can be to them, in discovering relevant data resources and avoiding duplication of research efforts.

NLM is also very active already in developing tools for discovery and analysis, at the Lister Hill Center and NCBI. The Library will continue promoting data science training at universities and on the NIH campus. We will partner with our sister Institutes and Centers in promoting a greater level of sophistication and expertise in data science across the board. NLM will also work to train more sophisticated, data-informed medical librarians, through continuing education, on-the-spot training and by other means. And, most importantly, for the health of the country, NLM will work to build data-informed clinicians. Not all need to be in this category, but their experiential research will no longer be sufficient, in many cases.

The Library won't do this alone. It's engaging with and learning from government, national, and global collaboratives in data and open science. Partnerships are essentials, with clinicians, patients, and scientists, as well as with government agencies and others. Going forward, Dr. Brennan noted, she expects the Library to ensure integration of its own planning and efforts with BD2K and NIH Cloud Pilot learnings. Another important goal, as NLM continues to act on its commitment to data science, is to establish a viable organizational structure at the Library to support it.

Dr. Brennan next laid out a list of critical tasks for DataScience@NIH, formerly a function of the NIH Office of the Director, as it pivots to its new home at NLM. These include: creating mechanisms to determine high-value data sets, locate them, and forecast their cost and utilization; implementing efficient, secure preservation strategies that facilitate access and reuse; re-engaging and stimulating intramural and extramural efforts in standards; developing new methods for data management and data-driven discovery; growing a talented workforce; engaging with government, national, and global collaboratives; fostering open science; and ensuring the integration of lessons learned from BD2K. She then opened the floor for questions and comments.

Board member Dr. Alessandro Acquisti commented that he had previously considered data science just a fancy name for advanced statistics. Now, following Dr. Brennan's talk, he's convinced of its merits as a field. He expressed concern about data privacy, even in the face of the great opportunity afforded NIH with an expansion of data science research. Dr. Brennan mentioned that privacy is a choice made by the patient. Any data stored at or by NLM should hew to government privacy policies, however—for not only the patient but also the investigator. This conversation is just beginning, but NLM is the right place to have it.

Dr. Horvitz said he supports the concept of NLM becoming the data science lead for NIH and for health data worldwide. However, NLM can build on the efforts of many other entities that have been working in the field, including academia, industry, the National Academy of Sciences, etc. He suggested an 18-month exploratory phase for the Library, to not only review the work by BD2K centers but also these other institutions, including Microsoft and Amazon. This is important work that seems to be in line with the long-term dream of true evidence-based health care. The Library should take the role seriously and ask for appropriate investments to carry out the work.

Board member Dr. Daniel Masys quoted a guiding principle of former NLM Director Dr. Donald Lindberg: "Systems that get used get better." Perhaps a corollary to that rule would apply here: "Data that gets used is saved." Dr. Brennan agreed, but noted that NLM needs to purchase space before it can store any data. About 40% of the cost of data produced at NIH goes into storage!

Board member Dr. Gary Puckrein expressed concern about the term "high-value data set." Might that be viewed as a subjective determination? One would hope that the cost of date storage will match the amount of data created, and that the outside world could use the maximum access to data. Dr. Brennan agreed. It's important to strike a balance between the usefulness of unique data and the value of data that's used frequently. Dr. Puckrein noted a convergence of health care, science, and everyday life in the possible benefits of data science. He hoped there might be a public outreach component to NLM's new efforts, as there was for the Human Genome Project. Dr. Brennan said that the Library considers that an important element.

VI. NLM STRATEGIC PLAN: PROGRESS REPORT

Dr. Mike Huerta started with background information on the Strategic Plan. When Dr. Lindberg announced his retirement, Dr. Collins established a working group of his advisory committee to

form a strategic vision for NLM. The group issued six recommendations: the first was basically to continue leadership of existing NLM efforts at NLM; the second was to lead and innovate Open Science worldwide; the third was to lead Data Science at and for NIH; the fourth was to strengthen the role of NLM in workforce development, which is important because NLM has a long tradition of supporting informatics and Data Science; the fifth was to continue as a custodian of historic collections; and the sixth was that NLM identify how to best pursue our mission and vision. Dr. Collins accepted the recommendations.

Dr Brennan then initiated the Strategic Planning Process in September 2016 with a goal of identifying bold ideas to inform an NLM Strategic Plan with a vision for the next 10 years and action priorities for five years. A Board Planning Committee, co-chaired by Dr. Daniel Masys and Dr. Jill Taylor, is overseeing the planning effort, which will involve extensive input from internal and external stakeholders.

Multi-disciplinary panels of external experts are being organized around four themes— advancing biomedical discovery and translational science; advancing data science, open science and biomedical informatics; supporting the public's health, including clinical systems, public health systems and services, and personal health; and building collections to support discovery and health in the 21st century. In addition to these themes, all four panels with consider other topics, including research needs and funding; infrastructure, both physical and computational; standards; workforce development; partnerships and user communities; user engagement and educational outreach; international engagement; and health disparities. Each panel will generate a report, which will feed into the development of the strategic plan to be reviewed and approved by the NLM Board of Regents for presentation to NIH and the Secretary of Health and Human Services. The goal is to have the Board consider the draft plan at its September 2017 meeting.

In terms of broader stakeholder engagement, Dr. Huerta talked about Dr. Brennan's blog, social media, presentations at NIH and across the country, visits to extramural training sites, a request for information in the NIH Guide, a staff survey, a staff town hall meeting, and more.

After Dr. Huerta spoke, Board members Dr. Jill Taylor and Dr. Daniel Masys, co-chairs of the strategic planning committee, offered a few words. Dr. Taylor spoke of being honored to serve and how we're at a pivotal point in health care and science, biomedical discovery; the importance of data; and partnerships with industry where there is commonality in the mission. Dr. Daniel Masys commended Dr. Huerta for his synopsis and said that they're in the midst of planning how the agenda and sequence of work will be presented to the panels. He said that the enterprise would follow the first rule of Frisbee, which is to never precede any maneuver by an explanation more predictive than "watch this."

VII. NLM STRATEGIC PLAN: 10 YEAR VISION

Dr. Dan Masys reviewed highlights from the 20-year strategic vision included in the NLM long range plan formulated in 2005-2006. This earlier planning group tried to imagine what the world was going to look like in 2025. We're now halfway there, and that previous document gives us the opportunity to revisit their vision.

So, how did they do in terms of their predictions? In the area of genomic research, there would be an enormous pool of shared data on genetic variation, humans, animals, and viruses. (However large you think it is, it is going to be much larger than that.) That genomic data would yield new opportunities as well as new ethical dilemmas all the way out to health care and disease prevention. Remember that in 2005 the Human Genome Project had just been completed and it was viewed as a collection of data ripe for opportunities of discovery.

That earlier planning group also envisioned how high-throughput bioassay data would challenge existing NLM data, and that the Library would link these to the literature—this is the data wrangling that Dr. Brennan referred to this morning. In their view, the data would be so voluminous that it no longer could it be browsed by researchers; instead, there would be a shift in emphasis to analytical systems whose purpose was to find patterns of interest for researchers. The previous panel also believed that, by 2025 or 2026, the public would be demanding their individual genetic fingerprints and access to that information. There would be a shift in emphasis to genome sequencing, which was very much the technical topic of 2005. There would be an increasing focus on gene-environment interactions, giving credence to the notion that genetics loads the gun and environments pulls the trigger. The future would also see new classifications of disease based on molecular characterization, and traditional organ-based characterization.

With respect to clinical research, they saw a relentless convergence of clinical and genomic research, involving traditional kinds of research models but supplementing that with high through-put molecular assessments. Clinical trials would be carried out in partnership with patient communities, community-based provider organizations, and academic researchers without the need for physical co-location or even the traditional bricks-and-mortar setting. The research would increasingly leverage electronic records, both for health care and for public health surveillance. And there would be a standardized health research data infrastructure in place for the sharing and making available of that data, certainly in line with the vision the NLM Director expressed this morning.

As to the health care system, a fundamental premise of the group was that our traditional model of publication—reading and remembering by practitioners—would be necessary but completely insufficient to turn knowledge into effective action. Going forward, system approaches to care would be dependent on executable knowledge in the form of computerized logic—shared, made available, and evaluated by healthcare organizations. Increasingly, that care would be informed by activated consumers who play an important role in quality control and their own care. The panel predicted that intelligent devices would be the dominant way of accessing NLM services—we might have crossed that one already.

The earlier planners envisioned NLM being the leader for a worldwide cyber-infrastructure of health knowledge management provided in real time and mediated by expert systems, that there would be ubiquitous telemedicine capabilities connecting community health aides, globally, to specialists for the interpretation of data generated by affordable wireless diagnostic tools. Most of these ideas were futuristic in 2005 and are more commonplace today, especially regarding personal health records, which were at that time largely disconnected from other systems.

Regarding personal health knowledge bases, these would interact continuously with provider

systems (something yet to be achieved), and automatic personal knowledge updates would be pushed by authoritative sources such as the NLM. In addition, patients' personal imaging data CT scans, MRIs, digital radiography, and other nanoscale molecular characteristics would be part of that personal health knowledge base.

In the area of lifelong learning, professional learning would replace discipline-specific training. The type of education which has been the model for health care for that last several hundred years would be transformed. Computer science simulations would lessen the trial and error of clinical practice done on patients, and these simulations would increasingly be used to test practitioners' skills and to document their limits.

With respect to community health, they predicted the digital divide, access to the technology would disappear but there would still be residual last-mile bandwidth issues. A sizable fraction of the population would continue to lack the education and training necessary to make most effective use of the information technology. As a partial solution, they pointed to the enduring importance of the National Network of Libraries of Medicine as an essential link between NLM's core services and local community needs.

Regarding global health, the panel predicted that whole-genome analysis would routinely be applied globally to emerging diseases. Also, NLM would facilitate a worldwide system, in part enabled by wireless communication and portable, wearable devices, making the delivery of information and the acquisition of clinical data possible in the remote and resource-poor parts of the world. They forecast that NLM's knowledge resources would continue to provide new opportunities for partnerships with international organizations and foreign governments, to improve health and prevent the spread of emerging infections.

With respect to the core activities of publishing and the Library's information delivery systems, they predicted that there would be increased understanding of visual reasoning, and increased understanding of how multi-media data sets and the methods of interaction of information would affect identification, understanding, and retention of information. They even predicted that, as new items are added to NLM's digital collections, they would automatically establish connections already related to the present, "understanding" information users' needs and delivering concise and appropriate responses.

Medical librarians and medical libraries would continue to be essential, although libraries will increasingly become a place of collaboration and information analysis. The panel also predicted that new research results would be freely available in digital format after initial production of publication. Regarding preservation, they predicted that some but not all digital works would be released in standard-based formats that would facilitate permanent access and transformation overtime, remaining accessible as technology formats came and went. NLM would be part of an international network of trusted digital repositories supported by a variety of institutions. They envisioned that this shared digital content would be archived redundantly in multiple independent sites to provide resilience and disaster recovery in the face of major disasters. There would also be better technology for preserving and providing digital access to historically important physical papers, records, and images.

The bottom-line vision of the 2005-2006 planning process was that of a 21st century health system in which NLM had in their view a catalytic role, building on its reputation as a credible source of reliable information and a leader in biomedical informatics.

There were of course unforeseen developments such as widespread adoption of smart devices and social media, and global health threats such as the Ebola virus. There have been some technology failures, wars, and unpredictable elections. Dr. Masys invited the Board to comment on what the previous strategic planners got right and perhaps what they got wrong, and how these elements might inform the upcoming planning process.

Board member Dr. Jill Taylor thought the panel underestimated the amount of change in technologies for biomedical engineering, such as synthetic technology. That field has implications not only in synthesizing new forms of DNA but also putting them up to make a new organism. This has implications for health, in production of therapeutic material, but also security and biodefense.

Dr. Puckrein hailed the 2005 vision, but said that, after working closely with patient advocacy groups, a missing element is the healthcare marketplace. There is an enormous debate about health care in the US, and the assumption is that healthcare is going to be a commonplace and everyone is going to have access. However, what can we realistically cover and how do we best deliver the care? Going forward, we should introduce the marketplace into this discussion because it impacts the movement of medical information as well as the way care is delivered.

Ex-officio member Dr. Clancy from the VHA added that part of that marketplace is how much of that data is being sold. What might it cost to acquire data needed for medical research.

Board member Ms. Jane Blumenthal noted that the group overestimated the disappearance of the digital divide. It may require another ten years. Even when the infrastructure is in place, you still have to pay for high-speed Internet in your home, sufficient data on your cell phone, etc. It's not always easy for people to afford and acquire, which is another marketplace issue.

Dr. Horvitz mentioned a new notion, around 2005, when Amazon and others promoted the idea of selling computation. But now it's a commonplace to acquire these resources.

VADM Murthy supported much of the report, but thought the planners overestimated experts' ability to understand how people learn about health and medicine, and to explore that question in terms of actual products. Also, they didn't comment on what would drive change in health information—a few large institutions, perhaps, or a larger grassroots demand? If there were a demand from the populace, that has implications for sustainability and for the support that we can build for increased funding.

NLM Director Brennan recalled that, when the Board spoke about privacy and security earlier—specifically, the ethics with gene editing—it didn't consider the perverse threats of data disruption, stealing data, denying access to data sources, etc. These can disrupt fundamental operations, and we should ponder them in our planning, going forward.

Ex-officio member Dr. Wayman Cheatham of the US Navy Bureau of Medicine and Surgery, echoed Surgeon General Murthy, accenting the importance of helping consumers discern truth and non-truth in medical information.

Ex-officio members Ms. Meg Tulloch of the Library of Congress called for a continuing examination of the role of libraries and librarians in the 21st century.

NLM Deputy Director Humphreys stressed the importance of capturing the record of what occurs in healthcare delivery, without excessive burden on providers or patients. A reasonable record would show successes and failures in patient care, and that data could likely be reused, going forward. This would be a huge boon for everybody. She noted that the future of NLM might depend largely on the public's trust in government. NLM must continue to prove itself a reliable source of valid information. Increasingly, people are the generators of their own information, thanks to the Internet and, much more broadly, social media.

Ex-officio member Col. Richard Thomas from the Uniformed Services University of the Health Sciences, concurred with Ms. Humphreys that a core strength of NLM is developing medical records and standard formats for the data associated with them. He described similar efforts at the Department of Defense and the questions posed, including, what information do you truly need to make good decisions and how accurate it is? The altruism of the NLM puts it in a unique position of knowing the value of health data, not related to forces driving the market, but rather its value to the health of the individual.

NLM Associate Director for Health Information Programs Development Dr. Mike Huerta, said he didn't hear mention of the challenge with metadata and software, and how that relates to data sets. In the future, digital research objects will need systems of unique identifiers, so that a given data set would be linked to the author, the patient, and other data elements. It would be kind of a fingerprint.

Dr. Horvitz talked about the interesting work underway in the artificial intelligence community, and what it might mean to have formal methods for augmenting someone's cognitive abilities. Possibilities for AI go beyond just in time provision of the right documents to include providing specific information based on inferences about what a decision maker already knows and identification and display of additional information that would be of ideal assistance to an informed decision.

Dr. Cheatham noted that most of the major medical societies have created patient registries that are being used for macro research. He thought that, if NLM brought in some of those medical societies, to discuss the benefits and challenges of such registries, there would be a tremendous reaction.

Lister Hill Center researcher Dr. Terry Yoo thought a major consideration should be the aging of the US population. The typical 2026 patient is going to be far older than the patient of 2017.

Board member Ms. Sandra Martin said she missed hearing about getting back to the basics of health, including a greater focus on prevention. Board member Dr. Sternberg agreed.

Dr. Masys thanked the group for a lively discussion and assured members that it would continue, in different forums, going forward.

VIII. NCBI STRATEGIC PRIORITIES

National Center for Biotechnology Information Director Dr. David Lipman noted that NCBI has been around for almost 30 years, developing many new resources, seeing a growing number of users, and experiencing dramatic growth as a staff. It's a good time to step back and look at what NCBI needs to do to maximize its impact, given the gradual plateauing of staff size and program resources.

Obviously one thing is to make sure operations are as efficient as possible. Toward this end, NCBI has started a strategic review of its information resources and tools, to maximize the value of these services for its diverse users. Possibly the most important thing in this process is to make sure you're working on projects to which you can add the most value.

One high-priority area with the greatest potential to increase impact is search and access to the biomedical literature. That category includes PubMed, PubMed Central, BookShelf, and PubMed Health. Altogether, these sites see millions of users a day, so that is one scale of use.

One tier down are some of NCBI's sequence resources, chemistry resources, and ClinicalTrials.gov, which collectively get several hundred thousand users daily. Still, to those who consult these more specialized resources, they are important. Also, they are still expensive to maintain, and a fair amount of our effort to goes into them. (As an aside, Dr. Lipman mentioned that an infrastructure and a culture that is good at handling lots and lots of users isn't necessarily adept at handling these strategic ones on the lower level of use.)

In focusing on maximizing impact, one measure is the number of users you have and how much they use the resource. The nice thing about usage is that you can drill into those figures in detailed ways and explore many aspects. The key thing to consider is the ratio of usage to cost.

Information quality is another aspect of making an impact—even with the same number of users, improving the quality of the resource can make a significant difference. In the case of the literature sites, that's hard to do in a systematic way. However, if someone is looking at an arbitrary clinical trial vs. a recent systematic review that's relevant to that area, one could argue that in some contexts it's important to direct them to the systematic review. In some of the other database areas, you could objectively get into the notion of information quality and how to assess it—for example, if it's been contaminated in some way, if it's annotated structurally in the right way, and so forth.

The last element of impact is consequence. If an NCBI resource didn't exist, would that make a difference in the world. The pathogens work that we're doing in food safety is an area in which use by a relatively small number of users of data is able to impact the health and safety of thousands of people.

Another consideration in terms of maximizing impact is cost. If two resources have the same

impact and one is cheaper than the other, an organization might do well to let the more expensive one go. Users, usage, and cost are factors that are relatively easy to manage, whereas in some cases information quality is not.

Dr. Lipman next focused on this ratio of usage to full-time employees (FTEs). This FTE equivalence is for NCBI an expression of the main cost of work performed, because the ratio of that figure to users and usage is something that can be obtained objectively; in measuring impact, NCBI needs to assess and use that data.

Looking at possible return on investment, if you're contemplating the creation of a new resource, you might do some market research, to find out how useful it could be. You don't have any users, initially, but you have some idea of what you're looking for to justify the cost of this undertaking.

You can take an existing project and reduce the costs, which makes that FTE-users cost-benefit ratio more favorable. Another way to improve your impact is to come up with ways to increase the number of users. With some NCBI resources, they've been around for quite a few years, and if we don't have many users, we may need to reconsider their purpose and worth.

One concept that NCBI is pushing internally is a "minimal viable product"—you want to do the most minimal thing that gets enough usage that you can learn from it, and then build on that knowledge. In setting its strategic priorities, NCBI is also looking at projects underway which are lower-tier in terms of usage, but the consequences of which are hard to predict. In the past, these may have been deemed strategic and impactful, but over time that may have changed.

One area that we are clearly going to invest more in is the literature. In the biomedical field, most knowledge is in text. (Other scientific fields, like mathematics and physics, are more grounded in models.) Another obvious reason to ramp up efforts is that NCBI already has a vast user base. If it can improve its resources for them, that would have a greater impact on research and discovery, and probably drive the user and usage figures even higher.

Another area NCBI has examined closely is DNA in protein sequences. Part of the reason for this focus is there are already many scientists submitting data to these databases, and many who consult these resources. NCBI is confident it can improve the information quality and make existing websites easier to use. It can also spark discoveries by cleaning up the data and by integrating them into higher level knowledge structures.

It's feasible to do these things now because of the Pathogen Detection Project; the challenge now is figuring out how we can generalize that effort and do it to scale across more resources. NCBI plans to start focusing mostly on bacteria and viruses, areas in which it has lots of sequence data. NCBI can also do a lot of comparative analysis across the DNA levels, because there are consequences. As sequencing improves, NCBI can also do more of this type of analysis for eukaryotes. These pathogen efforts and resulting data are already consequential for public health.

Amplifying that point, Dr. Lipman cited a 2016 book by Robert Gordon, *The Rise and Fall of American Growth*, which examines the US standard of living since the Civil War. What struck

Dr. Lipman was the table of death rates, showing a peak with the 1918 Spanish flu epidemic and, not so far back in time, the millions of people have been impacted by infectious diseases—HIV/AIDS, bird flu, and Ebola, for example.

Obviously, infectious disease remains a threat to public health. Anti-microbial resistance, which itself peaks every eight to nine years, complicates matters and increases the challenge. For all the forms of bacteria in a hospital, patients have twice the risk of death if the microbe shows anti-microbial resistance.

Taking a broader view, Dr. Lipman said that NCBI remains hopeful about the role it can play in helping researchers and clinicians prevent and treat cancer and other chronic diseases. There are a lot of exciting new discoveries, but it's a long slog.

Public health measures, vaccines, and drugs are crucial in the fight against infectious disease, but it's also critical to catch these outbreaks as early as possible. In the case of foodborne pathogens, whole-genome sequencing is revolutionizing that process.

Foodborne illness is a major public health problem in the US, with almost 50 million cases annually. There aren't many deaths but the number of hospitalizations has meant considerable costs across the US for the last few years. NCBI has been working with Centers for Disease Control and Prevention, the FDA, the US Department of Agriculture, and state laboratories to transform how surveillance and monitoring is done for food safety. For example, all islets for Listeria collected in the standard network in the states are being sequenced locally, the raw data being sent to NCBI. NCBI then does the sequence analysis and sends them back out. Listeria outbreaks are common, but the number of cases per outbreak has gone down because this partnership is catching them earlier.

This successful method is being extended to all foodborne pathogens and, in the coming year, NCBI will be building up its existing infrastructure to collect data on antimicrobial resistance, in addition to surveillance and monitoring. It's a very exciting project and has resulted in techniques NCBI can do across the board for viruses, bacteria, etc.

Another promising new direction is using whole genome sequencing to guide therapy. There's a lot of enthusiasm around using it for tuberculosis, because characterizing the microbial resistance profile for TB is difficult and time-consuming. If you have the whole genome sequence, however, you can use certain standard tools to identify the variants that make the bugs resistant to anti-microbial agents.

In closing, Dr. Lipman said that, to make a bigger impact, NCBI must understand the quality and the consequence of its resources, know and monitor closely its products (and understand their "usage over cost" ratios). Two areas that we want to increase our focus are search and access to the literature and the comparative analysis of DNA sequences, particularly with respect to public health and infectious disease.

Ms. Blumenthal expressed appreciation for Dr. Lipman's "real world" view of the consequences of the work of NLM, and how the Library allots its resources. Often, the questions of how we

can do something better or more cost-effectively seem very abstract, but those processes become concrete with these examples. She acknowledged NCBI's reputation for being innovative and said she hoped that cost-effectiveness would not inhibit things in that community or squash new ventures. No, Dr. Lipman assured her. But, as in organization, priorities must be set. Perhaps NCBI isn't going to be as innovative in some areas but, in the areas where they think they can make the most impact, new ideas and approaches will keep coming.

Dr. Deshler agreed with Dr. Lipman—when you have high usage of an asset and you pull it away, people get upset. But it's important, as mentioned, to take a hard look at low-usage things, too. Might they be useful in the future? Are they yielding important discoveries? And how do you measure those things?

Dr. Lipman concurred with that approach. You can look at the market, literature, grants, etc. to gauge value. But you have to look at the risks, too, of scaling up or scaling back a resource. NCBI wants to be as explicit as possible in the process that it's using, so that it can understand and defend why it's making these strategic decisions.

Board consultant Dr. Kenneth Walker said that, with Dr. Lipman's vast knowledge of the collection and curation of data, does he think that there's a place at NCBI for clinical research data, with its possibilities for discovery—or does that fit better into the vast data set expected to result from the Precision Medicine Initiative?

Dr. Lipman, who trained in medicine, said he was no expert in big data, but that he thought it would be challenging to create a database of clinical information at NCBI until it can devise a way to answer simple questions about treatment and practice, and make the answers comprehensible to clinicians.

Dr. Taylor said she appreciated NCBI's digital immune system, but a big gap in it is biomarkers of human infection. What if a respiratory virus is identified but it's dormant in a patient? How do we go about getting that data in the database, along with the pathogens genome? Is there any hope of using data in ClinicalTrials.gov or a vaccine trial to study such questions?

Dr. Lipman admitted that there is a big gap in our knowledge, period. Several years ago, the National Institute of Allergy and Infectious Diseases convened a panel to look at the spread of flu. The scientists and clinicians discussed matters thoroughly, and their top recommendation was to consult clinical information on what happens when you get the flu. There's a huge need to do exactly what Dr. Taylor suggested, and a real opportunity for research and investment.

IX. LEVERAGING EHRS TO COLLECT AND ANALYZE SOCIAL, BEHAVIORIAL AND FAMILIAL FACTORS

Dr. Elizabeth Chen, an Associate Professor at Brown University, addressed the Board.

For the past decade, there has been increased emphasis on understanding the influence of social, behavioral, and familial factors on health and health outcomes. There is a need for enhanced data methods for research about these factors, to enhance evidence-based tools for diagnosis,

management, and prevention.

The report of a 2009 conference focused on family history and improving health described a need to better understand collection and analysis of family history information, and to study the linkages between social, behavioral, environmental, and genetic factors. More recently, a two-phase IOM report on capturing social and behavioral domains in measures of electronic health records emphasized the need for specialized data collection and measurement of social and behavioral determinants of health in electronic health records (EHRs).

Collectively, these developments motivated the creation of the SFHERE Project—Social and Family History Extraction Representation and Evaluation. The focus has been on looking on EHR data as a rich longitudinal source of data, capturing information about diseases and conditions, family history, social history, and a wide range of patient data. A challenge is that some of this information may be captured or structured in a discreet format, or in an unstructured or free text format, as clinical text and clinical notes. The overall goal of SFHERE has been to develop computational approaches to enhance existing knowledge of interactions among social, behavioral and familial factors, and diseases. So, while the focus is on a few specific conditions, the hope is that the work, methods, and processes developed are generalizable to any condition, and any factor. The team also wants to establish systematic processes for collecting and analyzing data, developing open source tools, and providing these resulting knowledge-based resources to the broader community.

This project started while Dr. Chen was on the faculty of the University of Vermont. The University of Minnesota was a partner and now Brown University is, too. Where in the EHR is family and social history documented? Not surprisingly, one of the first challenges for training our computational techniques is to find the sections we want to study. In its work, the team spent time looking at EHR data and also at existing standards for defining a family history. (They came up with 20-30 different elements.)

They did a similar study of social history, which might include elements like substance use and occupational information. They've developed natural language processing tools to pull out family and social history, and worked with the University of Minnesota's BioMedICUS, a natural language processing (NLP) system. Among other functions, it does automated extraction of family and observation predications from unstructured text, making use of NLM's Unified Medical Language System (UMLS).

In preliminary studies done in 2011, SFHERE looked at three different sources of notes to see what defines social history. They finally identified 35 different statement types that are captured within the social history. Some that have been of special interest recently are alcohol use, drug use, residence information, details about things in the house, living situation, number of people in the home, and family. If that's the data, how do we get from data to knowledge?

They've defined the disease/knowledge discovery process. There are many health data sources—registries, public health databases, literature, and biobanks, among others—so they're looking at using individual data sources but also integrating these data sources with each other, and the steps to get through to the knowledge discovery. So the data selection—defining your inclusion and exclusion criteria, and your target data—and transforming that data so that you can use it to

pull that history from notes, is key. There is also the matter of dealing with integration challenges. Finally, after that, you get to the data mining steps.

Getting the data and cleaning it to the format needed is 80 to 90 percent of the work. Then you can start doing the data mining, to discover the patterns and then interpret that knowledge, whether is it already known or new. It's this knowledge that helps to generate the hypotheses, and foundational to all of this work are standards. They're essential to our efforts to integrate and analyze information from these different sections and formats. Of course, there are data privacy and security needs to be taken into account in each of these steps, too.

They've developed the family history NLP model and also have one for substance use. The team has recently focused its attention on pediatric asthma and adult epilepsy. In the early days, they focused on tobacco use, which continues to be the leading preventable cause of morbidity and mortality in the United States. There has been increasing emphasis on how EHRs can help with tobacco use assessment and cessation. One of the team's studies looked at how EHRs could potentially be used to inform improved collection and subsequent analysis of details associated with tobacco use. To give context, Dr. Chen shared examples of tobacco use entries that reflect the fields within the EHR. In some, you select from a specified set of values, but relevant (and sometimes conflicting) information may also be embedded in free text notes. The information may also change over time. Throughout this study, the team wanted to get a sense of the related content present in EHRs to inform natural language processing techniques to pull out the information. Depending on how you define your inclusion and exclusion criteria and what fields you use, a patient may be categorized differently on the variable of interest. Dr. Chen described different levels of granularity that could be applied, and also a challenge for natural language processing is the use of acronyms and how they disambiguate. Misspelling poses problems, too.

What the team learned was that, within EHR systems, there are limitations in capturing and accessing details related to tobacco use and secondhand smoke exposure. What is needed is flexibility in describing amount, frequency, and start and quit dates. Today, too, there will need to be a broader "nicotine use history" option, for capturing information on nicotine replacement therapies and e-cigarettes. Improving data quality and currency improved quality and is a common challenge in any part of the EHR. However, free text probably won't go away, so they're going to have to have robust natural language processing methods to extract information.

With asthma, the leading chronic condition in children globally, there is existing evidence of familial and environmental risk factors for pediatric asthma, but there is a need for further studies to explore and understand the interactions among these factors. There is a wide range of data on these factors in the EHR and the goal is to enable use of these data to support both patient care and quality research in public health, using all the steps they've performed, including the development of data mining techniques to transform the EHR data into actual knowledge.

Dr. Gary Puckrein hailed Dr. Chen for capturing the many challenges and opportunities that lie ahead, working with real world data. He suggested her analysis include geography, as that factor can play an important role in asthma prevalence.

Dr. Sternberg praised her expert analysis of complex data, and asked how the researchers balance

the provider and patient burden in answering these questions. Earlier in her career, when she was seeing patients, there was limited time to take a social history. If that information has to go through the EHR, it may be ignored altogether.

Dr. Chen says she has heard that question in several discussions. Who is responsible for documenting this information? Personal health records may be helpful, having the patient specify it. With wearable devices, too, people are collecting data every day and it can be documented.

VADM Murphy commented that we are moving increasingly toward mining the data that medical professionals are documenting. Has the team thought about how we predict behavioral change and how we manage it so that we don't end up creating more obstacles? Also, how we should train people to document, which is a difficult thing to do for which little training is provided in clinical services?

Dr. Chen said there should always be a balance between clinical needs and research. If you document to support clinical care, are the data useful for research? We need to inform training of clinicians regarding documentation We can't just say of their recordkeeping, "You need to do it better." We need to tell them how.

X. EXTRAMURAL PROGRAMS REPORT

Dr. Valerie Florance, director of Extramural Programs, brought two administrative issues to the board members for their review and vote. The first concerned the approval of NLM Operating Procedures. These procedures are reviewed annually every February at the Board of Regents (BOR) meeting. Annually, the BOR approves operating procedures for grant adjustments, downward or upward revisions, and NLM staff report any adjustments that result in an increase of more than $40,000 direct costs at the next BOR meeting. The last amendment reported was in 2012. The second issue requires reporting, as part of a "Special Council Review," of any Scientific Principal Investigator on a NIH-funded grant that receives greater than $1 million per year in funding support. Currently, NLM has never supported an investigator greater than the targeted level set by the NIH. The members approved the procedures.

Dr. Florance next provided an overview of a new grant concept that is instantiated in a 2017 NLM Funding Opportunity entitled, "Data Science Research: Personal Health Libraries for Consumers and Patients (R01)" (https://grants.nih.gov/grants/guide/pa-files/PAR-17-159.html).This new research grant funding opportunity will receive applications in the spring, with the goal of issuing the first awards by September 2017. The applications submitted to NLM and NIH under this announcement will be reviewed by NLM's standing review study section, the Biomedical Library Informatics Review Committee (BLIRC).

Dr. Florance provided some examples of past research in related areas and commented that much of the data science research related to consumer health and patients was aimed at clinical decision support and basic tools and not directly for consumer-patient supported research. NLM had in the recent past added an additional bullet to its list of areas for biomedical informatics research, consumer-patient oriented research. The research goal of the new announcement is to "bring benefit of data science research to consumers and patients." The development goal of a

Personal Health Library, as described by Dr. Florance, involves building, managing, and using all aspects of this library in support of the consumer or patient.

Dr. Horvitz asked whether there currently was any coordination with other agencies in respect to the grant announcement and funding. Dr. Florance responded, "Not at this time," due to the short interval between the announcement and funding.

Dr. Walker asked Dr. Florance if she believes that an application (app) will one day be available for smartphones, to help people manage their personal health information. Dr. Florance's response was that she hopes NLM can provide options and choices so that everyone can store and manage such important records as they are most comfortable.

NLM Director Dr. Patricia Brennan explained that this announcement is a R01 grant for research and hopes that the applications submitted will be more focused on advancing knowledge as opposed to developing a health app for consumers. The real question, she believes is, "What are the research problems that help advance bioinformatics knowledge?" She suggested further that it is "research problems that exist in personal health information integration." She hopes to see grant applications come in that support advances in informatics and data science knowledge, not applications requesting support to create mobile apps.

A lengthy discussion was held by many around the table about the use of mobile apps for health management for individuals.

NLM Deputy Director Betsy Humphreys commented on research that looks at the impact apps have for people with certain conditions. Most studies found that the apps alone did not work unless they were integrated into another program that helps people change their behavior.

Lisa Lang, head of the National Information Center on Health Services Research and Health Care Technology (NICHSR), NLM, proposed that it might be helpful to develop common data elements and common metrics that could be used by the researchers that NLM funds and could help contribute to future projects and the knowledge base.

Finally, Dr. Florance said that an important real issue for consumers is having a health information resource designed for the person, that incorporates information from family history and activities of daily life as well as from clinical encounters.

XI. REPORT FROM THE NLM DIRECTOR

NLM Director Dr. Patricia Flatley Brennan began her report by discussing NLM's participation in an art contest as part of the 2016 Combined Federal Campaign. She showed images of entries prepared by 20 NLM staff members, from which a winner was selected. NLM's winning entry was displayed at the NIH Clinical Center.

Dr. Brennan gave an update on selected NLM FY 2016 accomplishments, organized under the four goals in the Library's 2006-2016 Strategic Plan, thanking NLM division heads for providing input.

Dr. Brennan said that the strategic plan developed at NLM ten years ago worked and NLM continues to be guided by this process.

She told the Board that NLM is operating under a continuing resolution until at least April 28. She continues to work with the NIH Office of the Director on data science initiatives which is a way for NLM to grow. These funds will not supplant current funds.

Dr. Brennan said NLM's goals for 2017 are to:

- Provide leadership to the pivot of data science, open science, NLM, and NIH to the future
- Reorganize how NCBI creates, develops and delivers data resources
- Decrease the time to complete Medline indexing by streamlining processes
- Modernize and adopt and cloud technology
- Promote the *All of Us* research initiative to our partners
- Issue a new research grant funding opportunity announcement in the area of data science research for personal health information
- Implement at least one more imaging data science project – deploy the computable 3D anatomic/organ model and crowdsourcing of consumer pill images
- Advance the mission of NLM through effective financial, acquisitions, personnel, facilities, and general administrative management

Dr. Brennan talked about the Strategic Plan, noting that NLM would conduct several functional audits. One will measure the impact of MeSH on indexing and literature retrieval. Another audit on outreach will determine how effective we are in ensuring that people are making use of NLM products and services.

Dr. Brennan said there are four strategic planning panels scheduled. Board Members who would like to attend by webinar or as observers are encouraged to do so. The four panels will discuss NLM's future role in:

- Advancing biomedical discovery and translation. March 1-2, Chair: Arthur Levine, MD, University of Pittsburgh;
- Advancing data science, open science, and informatics. March 14 and 15. Chair: Russ Altman, MD, PhD, Stanford University
- Supporting the public's health: clinical systems, public health systems and services, and personal health. April 4-5, Chair: Suzanne Bakken, RN, PhD, Columbia University, Building 21st Century collections for discover and health. April 19-20. Chair: Patricia Thibodeau, MLS, MBA, Duke University.

Dr. Brennan said that, on September 16, 2016, NLM celebrated the issuance of the final rule with requirements for registering certain clinical trials and submitting summary results information to ClinicalTrials.gov. On January 18, 2017, the final rule became effective. The ClinicalTrials.gov team made sure that researchers could input their information to the system correctly by that date.

Dr. Brennan asked Joyce Backus, Associate Director of Library Operations, to introduce newly appointed LO staff. Ms. Backus said that Stephen Greenberg, PhD, had been appointed head of HMD's Rare Books and Early Manuscripts Section. Dr. Greenberg, who joined NLM in 1992,

was most recently the Coordinator of HMD's Public Services. Ms. Amanda Wilson, MLS, was selected to head NLM's National Coordinating Office for the National Network of Libraries of Medicine. Ms. Wilson most recently served as the director of the Department of Transportation library. Lastly, Backus introduced Mark Ziomek, MSLS, who has been appointed chief of NLM's Public Services Division. Mr. Ziomek joined NLM from the US Government Accountability Office, where he was the manager of library services.

Dr. Brennan also welcomed two new Lister Hill Center Fellows, Catherine Blake, PhD, and Sivarama Rajaraman, PhD.

Earlier this year, the NLM received a gift from the DeBakey Medical Foundation to support enhanced access to HMD's Michael E. DeBakey Archives and to establish HMD's Michael E. DeBakey Fellowship. Dr. Brennan announced the first set of five 2017 Michael E. DeBakey Fellows and their diverse set of research projects that will draw on the DeBakey Archives at NLM.

NLM also bade farewell to Arthur Petrosian, PhD, who retired from his position as Chief Scientific Review Officer after 11 years of service to the NLM.

Dr. Brennan also announced the pending retirement of NLM's Deputy Director Betsy Humphreys. NLM will celebrate her career on June 19, from 1:30 to 3:00 p.m., with a program in Natcher Auditorium with a reception in Lister Hill Center Lobby to follow. Ms. Humphreys said she started a four-month appointment at the NLM in 1973. It turned into 45 years. NLM, she said, is a great place with a great staff and mission. Dr. Masys said that the NLM has a wonderful combination of career loyalty and altruism making it, as Paul Harvey once said, "the crown jewel of the federal government." Ms. Martin asked for Ms. Humphreys' celebration proceedings to be broadcast for those unable to attend. Dr. Brennan agreed to do so.

The NLM Director then asked the Board how the NLM, in a flat budgetary environment, might manage its public service and scientific responsibilities. Dr. Masys said the Congress has shown itself to be a willing set of investors. The creation of the NCBI and the legislation introduced by Claude Pepper was an example of a good idea resulting in funding that did not have to come out of NLM's operating budget. He recalled giving a talk before an advisory panel on the Human Genome Project in 1984, at which Dr. Ruth Kirschstein, then director of National Institute of General Medical Sciences, said, "Well, you don't want librarians taking care of gene sequences, do you?" That meeting resulted in a supplemental appropriation of about $25 million. If the strategic planning process recognized a new opportunity, funding will follow. Dr. Sternberg agreed recalling that it was former NIH Director James Shannon who built the Institutes on the basis of disease. He did that in order to get the support of the Congress.

Ex-officio member Col. Thomas Cantilina of the Office of the Surgeon General, US Air Force, said that if you focus on priorities and allocate funds accordingly, you will get the important things done. Ms. Blumenthal said, don't expect it to be easy to get funding for new programs. You are going to have to have political astuteness in your skill set to do so. Ms. Humphreys said it is easier to stop doing things when you are in a budget crunch. It is harder to stop doing things when people see your budget going up.

Dr. Walker said that NLM does not have champions in Congress as it did in the past.

Dr. Brennan said that the NLM has an important place at every table at NIH where decision-making happens. Because of the emphasis on data science and NCBI, NLM is considered central to the future of the NIH. She had a meeting recently with Appropriations Committee members and found them to be concerned about the health of their communities. The NNLM is an important pathway into communities.

Dr. Taylor said the New York Department of Health is in a similar situation. You want to avoid across the board cuts. She said NLM has its incredibly committed and creative staff that will come up with solutions that will help.

Dr. Brennan mentioned that NLM will have a town hall meeting with staff soon. She will communicate to the staff that NLM's core mission is not going to change.

Dr. Taylor said that it is all about honesty and that Dr. Brennan is the right person to do that. It is a different time and it is respectful to help them be part of the solution.

Dr. Masys said that NLM has shown over the years that you can convert someone who is sympathetic into a champion by educating their staff. If they learn what the issues and resources are—they will help you.

RADM Trent-Adams said that as you look at the increase in demand for NLM, you look not only for the champions but to building nontraditional partnerships. Look at how those connections can be made in the community, so every American can view him- or herself as part of the NLM community. Then they become your voice and the demand has a face. As you leverage those relationships with nontraditional partners and NLM will have a presence in communities where NLM may not be right now, including across government, departments, and HHS.

Dr. Brennan said that she and Acting SIS Director Florence Chang visited the Children's Inn here at the NIH. She said that NLM sees this as a pathway for how we might engage differently with children. K-12 education and outreach has always been part of our mission.

Ms. Tulloch said that staff is your biggest cost and you will need to focus on them. How do you get people trained? How do you reposition people? How do you reorganize the NLM to make it more effective? In about a year you will see how you can realign your staff and save money.

Dr. Brennan said that NLM faces a problem similar to other federal environments: how to compensate staff in a way that reaches the market level of salaries for similar workers. It has a great influence on our research capacity and its ability to expand.

Ms. Tulloch said that innovation is key. Dr. Brennan needs to be seen as an innovator. People will rally around you. The American Library Association is getting into communities, pushing out policy programs—the Washington office is a good contact. The National Library Services for the Blind and the Disabled pushes out to the state libraries. They might be a good contact for

the NLM Director. And there are university libraries that are facing similar constraints. They might be good partners.

Dr. Gary Puckrein said spends a good part of his day trying to get folks access to health. One of the unique opportunities that is presented to him is to hear about what people are thinking about in the future. His clear takeaway is that we are in the middle of trying to build a health care communication information infrastructure in our country. It is complicated by the structure of the existing health care system, the various players. Clearly, the Library has an important role to play in this area. Looking at what NLM is thinking about doing with respect to data science, that is all about the future. How you communicate—that is really the trick of the matter. Sustainability is going to be the issue of the 21st century. How do we keep our populations healthy? By 2030, 75 million Americans will be on Medicare—one-third of the US population. When you add Medicaid to it, over 100 million Americans will be on the two programs. Right now, about 60 percent of those on Medicare go to the hospital with diabetes for emergency room visits or inpatient stays. This is not sustainable. We do not have enough hospitals to care for that population. That is the challenge. If you can carry on what you are doing into the broader conversation, it would be very important.

Ms. Blumenthal said that we have a very activist population – on both sides of the political spectrum and the way you get champions in Congress in this environment is through their constituents, not just librarians, but those who are influential friends of congressional members, including disease associations, and online communities who support new ways of processing and making information available.

Ms. Martin said that one of NLM's secret weapons its exhibits. They help to tell your story. It makes it clear and creates a picture. You might have some other secret weapons to take advantage of to speak to the community like traveling exhibits.

XII. LEGISLATIVE UPDATE, INCLUDING 21ST CENTURY CURES ACT

Mr. Sheehan began by directing the Board to their board books for a legislative update prepared in January. He indicated that his report would focus on the 21st Century Cures Act.

Mr. Sheehan explained that there are many areas within the 21st Century Cures Act that are relevant to NLM, including some that are directed generally toward NIH. There are also activities that are of interest to NLM, but that might not rise to the level of interest or be a priority for the NIH. Specifically, activities that are managed by the Office of the National Coordinator for Health IT (ONC) and that involve collaboration with the Food and Drug Administration (FDA).

The 21st Century Cures Act was developed a bit differently than other bills. In addition to the typical types of hearings that bring in witnesses to provide testimony, there were town hall meetings and hearings held across the country, drafts of the bill were posted on public websites, and written comments were requested and accepted. The bill was introduced into the House in May 2015 and passed by a wide margin in July 2015.

Typically, the House Bill would be sent over to the Senate to be taken up and debated, but

instead the Senate launched its own initiative around medical innovation, and 19 individual bills were introduced on topics ranging from interoperability of electronic health records (EHR) to medical device regulation to research priorities for NIH. These bills were considered between February and April 2016.

Further negotiation between the House and Senate resulted in a broad-based law signed in December 2016 that affects not only NIH, but the FDA, ONC, Centers for Medicaid and Medicare Services, SAMHSA, and others. The first part of the 21st Century Cures Act, Division A, is of most interest to the NLM and the NIH.

One of the key elements of the bill was a novel approach to increase the NIH budget. Congress created a separate fund, outside of the typical appropriations process, by which treasury funds are put in an NIH Innovation Fund and on an annual basis, Congress appropriates the funds to NIH. Division A of the Act provides almost $5 billion to the NIH Innovation Fund over ten years. It must be used for the Precision Medicine Initiative, the BRAIN Initiative (Brain Research through Advancing Innovative Neurotechnologies), Cancer Research, and Regenerative Medicine.

The law also requires that NIH complete a strategic plan at least every six years and instructs individual ICs to prepare strategic plans "on a regular basis" and following a common template. It includes instructions that IC Directors must consult with the Director of the National Institute for Minority Health and Health Disparities (NIMHD), as well as the NIH Office of Research on Women's Health (ORWH), to identify objectives for the inclusion of minorities and women in strategic plans.

In addition, the Secretary is supposed to look at financial reporting requirements that impose burdens on researchers and find ways to adjust and lessen the burden while achieving the intended goals. There is also a new research policy board that will be established within the White House Office of Management and Budget to address concerns about the regulatory burden on the research community. This provision follows several recent reports by the National Academies and General Accountability Office (GAO).

One of the topics that has come up in these academy studies and in GAO reports is how to reduce the burden of reporting requirements associated with data management plans that many agencies are now requiring for new funded research. NLM has been trying to work with other agencies so when NIH puts a data management plan requirement in to place across the board, it is consistent with other Federal agencies.

Another provision expands the Senior Biomedical Research Service which is a tool that HHS agencies can use to provide more competitive salary packages or retention plans for PhD holders in biosciences. The Act expands the service in three different ways. First, the total number of people across HHS who can qualify for this service has increased from 500 to 2000. Second, the salary cap has been increased, and third, it expands beyond just those with PhDs in biological sciences to include Ph.D. and master's degree holders in engineering and bioinformatics or related emerging fields. This can give potentially more flexibility to NLM in the kinds of packages it could offer to Federal staff with those qualifications.

The bill also authorizes the NIH Director to require funded investigators to share their research data, consistent with laws governing human subjects protections and proprietary information. It also puts in place certain requirements to protect privacy. One exempts human subjects data collected in biomedical research from disclosure under the Freedom of Information Act if there is "at least a very small risk" of reidentification. It also requires the Secretary to issue certificates of confidentiality, which provide certain protections from disclosure of information to participants in research studies, like the Precision Medicine Initiative. And it requires the Secretary to establish a working group to report on uses and disclosures of protected health information under HIPAA.

There are several provisions in the bill that relate to ClinicalTrials.gov. The first makes technical corrections to the law governing clinical trial registration and results reporting. Previously, if someone registered a trial in ClinicalTrials.Gov for a device that was not yet approved or cleared, that information could not be made public, even if the person submitting the information wished to make it public. This was called Device Lockbox. The Act now permits such information to be made public at the request of the registrant. In addition, the Act requires that clinical trials of combination products like drug-eluting stents, be classified by their primary mode of action, rather than as clinical trials of drugs, as previously proposed. Both requirements have already been incorporated into the final rules for clinical trial registration and results reporting that were issued in September.

A second provision related to ClinicalTrials.gov requires two kinds of reports that will be prepared on outreach and education, as it pertains to clinical trials registration and results reporting, as well as on compliance efforts. One report is due in two years; another with numbers of registrations and results summaries is due every two years for six years. A third provision calls for increased NIH efforts to promote the submission of the results of "valid analyses" to ClinicalTrials.gov. This would mean providing the results of a trial separated out by the sex of the participant, age, race, or ethnicity. When submitting to ClinicalTrials.gov, researchers would be instructed to include these types of analyses, if valid for the research they are conducting. Fourth, a consultation needs to take place within 90 days with the NIH, FDA, and ONC and other stakeholders about enhancements to be made to ClinicalTrials.gov as it relates to usability, functionality, and search capabilities.

There also is a requirement for a NIH to develop a report with recommendations for policies for improving rigor and reproducibility of pre-clinical and clinical research.

Given NLM's role in creating standards for electronic health records (EHR), NLM also has interest in provisions that aim to enhance the interoperability among EHRs and other types of health IT. Cures has provisions to advance interoperability and trusted exchange of electronic health information; prohibit information blocking and other harmful business practices; expand the availability and use of certified health information technology; and enhance transparency and the usability, accessibility, and privacy and security of health IT.

These are just a few of the provisions in the bill that relate to issues at NLM. NLM will continue to review the bill to identify other provisions of interest.

NLM is in involved in a NIH effort to develop a Legislative Implementation Action Plan (LIAP) for the 21st Century Cures Act. This plan will describe how NIH will implement every provision that is directed to the NIH director. The effort is led by the NIH Office of Legislation and Policy Analysis and managed by a working group consisting of select IC directors. NLM has been given the responsibility for three out of the four items related to ClinicalTrials.gov (the NIH Office of Science Policy has responsibility for the fourth item related to consultation).

In response to a question from the Board, Mr. Sheehan explained that NLM needs to start working with the other parts of HHS and other Departments that are implementing different pieces of the Cures Act. There are many places in which the Act touches on NLM interests and can be informed by NLM's experience with information systems.

Mr. Sheehan summed up his presentation by saying that as NLM goes through this implementation process he will share highlights with the Board of how NLM is working with other departments and agencies.

XIII. APPOINTMENT OF NOMINATION COMMITTEE FOR NEXT BOR CHAIR

Dr. Sternberg said three ex-officio members will serve on the Nominating Committee for the next BOR Chair: Dr. Nelson, Dr. Francis, and Ms. Tulloch. Dr. Francis will serve as chair.

XIV. NEW PUBMED MAINTENANCE AND DATA MANAGEMENT SYSTEM

Ms. Kathi Canese of the NCBI discussed the redesign of PubMed's citation data management system (PMDM), the cornerstone of a tremendous effort to improve how PubMed citation data is managed.

They had to rebuild the backend system, a set of APIs, as well as the user interface (UI). We spent a fair amount of time on the UI, she said to make users (in this case publishers or their data providers) happy and to make the system easy to use. Previously, NLM had two separate systems where the data were managed: a PubMed repository and a Data Creation and Management System (DCMS) used by the Index Section. So NLM had content that was similar, but not exactly the same, in two systems. The bulk of the data is submitted by data providers, or publishers, in XML to NCBI, and then the citations are loaded, indexed, and made available online in PubMed. Until recently, the data were also shipped by NCBI to the DCMS. A group of data reviewers would review author names and abstracts, make sure everything was correct, and ship it back to PubMed. Finally, the MeSH indexers would add MeSH terms and value added content to the citation in the DCNS and then it would be shipped back to PubMed again.

In addition to that stream, there was another stream for OCR (optical character recognition)-readable data. Most data comes from data providers or publishers in XML, but there are still journals that are only in print (about 4% of citations). In order to include that content in PubMed, we have a group of citations which are OCRed and then submitted to PubMed. It is easier and less costly to get the data in XML.

Having two discreet databases, and having information travel between two systems on a daily

basis, created a number of pain points. Large volumes of citation data were in flux between two systems, incurring a high risk of data loss and corruption. Once a citation was in the DCMS, only NLM staff could edit it.

When a PubMed user, author, or publisher noticed an error in a citation, they would send an email message to NLM customer service. If the citation was in process for Medline, they were told to wait for indexing to complete. Users could wait months for the citation's typographical error to be corrected. The biggest complaint from PubMed users was that an author name was misspelled. This was a huge problem.

With that in mind, we had three main goals for the redesign of the PubMed citation data management system:1) centralize citation data storage and maintenance so it was not in two discreet systems; 2) minimize reliance on manual NLM data review by automating corrections, revising data correction policies, and accepting additional data from publishers; and 3) allow publishers to correct their own citation data.

We began by tracking the changes made to citations while in the DCMS. We found that 87% of citations were corrected during data review. We brought that number down by changing data review policies that were no longer necessary. One of these was capitalization in journal article titles. Instead of standardizing them, they are now displayed exactly as submitted and as represented on the publisher's website. We also identified data corrections that could be automated when data were loaded from publishers, like adding periods to the end of titles, stripping extra spaces in abstracts, and removing duplicate colons from headings. This significantly reduced the number of citations changed manually.

Not only were we trying to automate corrections, we let publishers submit data that had previously been added by the data reviewer, like grant data, databank numbers, associated citation links. We found publishers did a good job.

Because we had a lot of different user groups, we had to create an elaborate permission system for the new single citation management system. We wanted to give NLM staff access to all 26 million citations. We only gave publishers access to edit citations to their own publications. We have data providers that submit XML for a number of publishers and needed access to edit data for all of the publications they manage. The permission system was improved and corrections can be made to appropriate citation data by all of these players.

On October 4, 2016, the PubMed Data Management System (PMDM) was launched. We have a single system and simplified workflow for managing citation data. All citations are loaded by XML submissions, including OCR data. And, once loaded, citations are available in PMDM for editing by both NLM data reviewers and providers. We received overwhelmingly positive responses from our publishers.

Since the release, we have seen more data providers correcting citation data in PMDM. Ninety percent of publisher-supplied journals have data providers who are actively using PMDM. Data providers are correcting author lists and names.

In summary, PMDM has drastically changed the systems and processes used to manage PubMed citation data, by: developing a single centralized system for citation data; minimizing the need for manual corrections; and empowering publishers to correct their citation data.

Index Section head Ms. Deborah Ozga discussed NLM's Indexing Management System (IMS). Ms. Ozga said that the Index Section likes the new PMDM because it has improved workflows for data review, reduced data review customer service queries by 72 percent, and enabled development of the new IMS.

We no longer need to do the data review on every citation, looking for metadata to add and data to correct. In addition to applying the MeSH terms, our focus now is on value-added metadata, like clinical trials and grant numbers, retracted article links to retraction notices, corrected article links to erratum notices and comments linked to original articles.

We identified which journals are most likely to publish value-added metadata by analyzing citations from every MEDLINE journal and added a processing code to each journal to define the level of data review needed.

But, we still have a challenge. Data review staff must still sift through hundreds of thousands of articles to manually extract the value-added metadata. The Index Section is working with NCBI to encourage publishers to: 1) submit more metadata, such as comments links, to PubMed; 2) follow ICMJE recommendations when publishing journal articles; and 3) provide NLM with access to the full text of journal articles for data mining. If we can do that, we can change our data review process to emphasize verifying the accuracy of metadata extracted and reviewing problem metadata identified through automated processing.

Ms. Martin asked about error reporting. Do you want to train us not to call you when we have an error? Should we go directly to publishers now? Ms. Canese said yes, and Ms. Martin said it certainly makes good sense.

Ms. Humphreys said that the changes made to this system have allowed NLM to encourage publishers to submit other types of information that multiple Senators and scientists are interested in having. Ms. Canese agreed, noting that next month, conflict of interest statements would be coming to PubMed.

Dr. Sternberg said that adding the metadata will distinguish PubMed in the marketplace. Ms. Humphreys agreed noting that what they have done is to make adding the metadata cheaper, quicker, and better. And, Dr. Sternberg added, and more accessible too.

XV. ADJOURNMENT

Dr. Sternberg adjourned the Board of Regents meeting at 11:30 a.m. on February 8, 2017.

ACTIONS TAKEN BY THE BOARD OF REGENTS:

- Approval of the September 13-14, 2016 Board Minutes
- Approval of the February 13-14, 2018 Future Meeting Dates
- Approval of the Grant Operating Procedures

Appendix A - Roster - Board of Regents

I certify that, to the best of my knowledge, the foregoing minutes and attachment are accurate and complete.

Patricia Flatley Brennan, RN, Ph.D.
Director, National Library of Medicine

Esther M. Sternberg, M.D.
Acting Chair, NLM Board of Regents