

**DEPARTMENT OF HEALTH AND HUMAN SERVICES
NATIONAL INSTITUTES OF HEALTH
NATIONAL LIBRARY OF MEDICINE**

**MINUTES OF THE BOARD OF REGENTS
May 3-4, 2016**

The 172nd meeting of the Board of Regents was convened on May 3, 2016, 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 4:15 p.m., followed by a closed session for consideration of grant applications until 4:45 p.m. On May 4th, the meeting reopened from 9:00 a.m. until adjourning at 12:00 p.m.

MEMBERS PRESENT [Appendix A]:

Dr. Alessandro Acquisti, Heinz College, Carnegie Mellon University
Mr. Eric Dishman, Intel Corporation
Dr. Robert Greenes, Arizona State University
Ms. Sandra Martin, Wayne State University
Dr. Daniel Masys, University of Washington
Dr. Jill Taylor, Wadsworth Center, New York State Department of Health
Ms. Gail Yokote [Chair], University of California, Davis

EX OFFICIO AND ALTERNATE MEMBERS PRESENT:

Mr. Christopher Cole, National Agricultural Library
Dr. Joseph Francis, Veterans Health Administration
Col. Helen Hootsmans, United States Air Force
Col. Michael Nelson, United States Army
Dr. David Neri, United States Navy
Dr. James Olds, National Science Foundation
RADM Deborah Parham-Hopson, Office of the Surgeon General, PHS
Dr. Charles Rice, Uniformed Services of the Health Sciences
Dr. Dale Smith, Uniformed Services of the Health Sciences
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. Eta Berner, University of Alabama at Birmingham
Dr. Graciela Gonzalez, Arizona State University
Dr. Eliseo Perez-Stable, National Institute on Minority Health and Health Disparities, NIH

MEMBERS OF THE PUBLIC PRESENT:

Ms. Andrea Baruchin, Friends of the National Library of Medicine
Mr. Glen Campbell, Friends of the National Library of Medicine
Mr. Jason Chong, Booz Allen

May 3-4, 2016 – Board of Regents

Ms. Madeline Halpern, ICF International
Ms. Lesley Macherelli, Healthbox
Dr. Barbara Redman, Friends of the National Library of Medicine
Dr. Elliot Siegel, Consultant
Mr. Thomas West, Krasnow Institute

FEDERAL EMPLOYEES PRESENT:

Ms. Betsy Humphreys, Acting Director, NLM
Dr. Milton Corn, Deputy Director for Research and Education, NLM
Ms. Anne Altemus, Lister Hill Center, 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
Mr. Victor Cid, Division of Specialized Information Services, NLM
Ms. Deirdre Clarkin, Division of Library Operations, NLM
Ms. Kathy Cravedi, Office of Communications and Public Liaison, NLM
Mr. Todd Danielson, Office of the Director, NLM
Ms. Darlene Dodson, Office of the Director, NLM
Ms. Victoria Douglas, Division of Extramural Programs, NLM
Mr. Ivor D'Souza, Office of Computer and Communications Systems, NLM
Mr. John Doyle, Division of Library Operations, NLM
Dr. Kathel Dunn, Division of Library Operations, NLM
Ms. Gale Dutcher, Division of Specialized Information Services, NLM
Ms. Martha Fishel, Division of Library Operations, NLM
Dr. Valerie Florance, Division of Extramural Programs, NLM
Dr. Kin Wah Fung, Lister Hill Center, NLM
Ms. Katie Funk, National Center for Biotechnology Information, NLM
Dr. Dan Gerendasy, Office of Health Information Program Development, NLM
Mr. David Gillikin, Division of Library Operations, NLM
Mr. John Harrington, Lister Hill Center, NLM
Dr. Michael Huerta, Office of Health Information Program Development, NLM
Dr. Vojtech Huser, Lister Hill Center, NLM
Ms. Christine Ireland, Division of Extramural Programs, NLM
Dr. Stefan Jaeger, Lister Hill Center, NLM
Ms. Janice Kelly, Division of Specialized Information Services, NLM
Dr. John Kilbourne, Division of Library Operations, NLM
Dr. William Klimke, National Center for Biotechnology Information, NLM
Mr. Kenneth Koyle, Division of Library Operations, NLM
Dr. David Landsman, National Center for Biotechnology Information, NLM
Ms. Jennifer Marill, Division of Library Operations, NLM
Dr. Clement McDonald, Lister Hill Center, NLM
Ms. Kathy McKay, Office of the Director, NLM
Mr. James Mork, Lister Hill Center, NLM
Dr. Kim Pruitt, National Center for Biotechnology Information, NLM
Dr. Jeffrey Reznick, Division of Library Operations, NLM

May 3-4, 2016 – Board of Regents

Ms. Karen Robak, Office of Computer and Communications Systems, NLM
Dr. Hua-Chuan Sim, Division of Extramural Programs, NLM
Dr. George Thoma, Lister Hill Center, NLM
Mr. Cuong Tran, National Center for Biotechnology Information, NLM
Ms. Patricia Tuohy, Division of Library Operations, NLM
Dr. Alan Vanbiervliet, Division of Extramural Programs, NLM
Dr. Rebecca Williams, National Center for Biotechnology Information, NLM
Dr. Wanda Whitney, Division of Library Operations, NLM
Dr. Fred Wood, Office of Health Information Program Development, NLM
Dr. Jane Ye, Division of Extramural Programs, NLM

I. OPENING REMARKS

Ms. Gail Yokote, NLM Board of Regents Chair, welcomed new board ex-officio member Ms. Meg Tulloch, FEDLINK Executive Director and the new representative from the Library of Congress. She then introduced RADM Deborah Parham-Hopson, to present the report from the Office of the Surgeon General (OSG).

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

RADM Deborah Parham-Hopson said that the Surgeon General (SG) Dr. Vivek Murthy has been actively involved in meetings around the United States. His priorities are physical activity, healthy eating, emotional well-being, suicide prevention, violence prevention, and tobacco-free living.

She said that the SG continues to encourage walking in the workplace and is very engaged in working with community leaders to make environments walkable. He knows that some areas are not safe for walking, so he is encouraging walking in malls, churches, and other safe community places.

She said that the opioid problem is a priority and will be addressed in the first OSG report on addiction in the fall of 2016. The report will focus on substance abuse, addiction, and health. The Surgeon General wants to write every prescription provider in the US expressing his concern about the opioid addiction.

The OSG will continue to focus on tobacco-free living. He joined with Julian Castro, Secretary of the Department of Housing and Urban Development, in announcing a smoke-free policy in public housing.

With respect to the White House Conference on Aging, the OSG and the National Prevention Council are working with the Centers for Disease Control and Prevention to produce a healthy aging action plan later in 2016.

RADM Parham-Hopson said the OSG is frequently asked about Zika virus. The White House has been the epicenter in responding to Zika, as well as NIAID's Dr. Anthony Fauci and the CDC's Dr. Tom Frieden. HHS Secretary Sylvia Burwell has been concentrating on Puerto Rico; she was there last week. In February, the Surgeon General hosted a fireside chat with mothers to

discuss what we are doing to address their concerns. In March, at the University of Texas in Austin, he held a town hall meeting to discuss and answer questions about Zika. The OSG continues to make information about Zika widely and readily available to the public.

III. REPORT FROM THE DIRECTOR OF THE NATIONAL INSTITUTE ON MINORITY HEALTH AND HEALTH DISPARITIES

Dr. Eliseo Perez-Stable, Director of the National Institute on Minority Health and Health Disparities (NIMHD), discussed the roots of the NIMHD, initially created in 1990 as part of the Office of the NIH Director. It became an NIH Center in 2000, when legislation by Rep. Louis Stokes (D-Ohio) was passed. The Patient Protection and Affordable Care Act contained language by Senator Ben Cardin (D-Maryland) to make it an Institute in 2010. John Ruffin, PhD, led all the entities until his retirement in 2014, when Yvonne Maddox, PhD, became Acting Director. On September 1, 2015, Dr. Perez-Stable became the Director. Today, the Institute has a budget of about \$280 million.

The NIMHD's mission is to address the science of health as it relates to minority health and health disparities. We support research in minority health, as defined by racial/ethnic groups in the US Census. It supports research to understand the causes and reduce health disparities in specific populations, and supports the training of a diverse scientific workforce. The NIMHD also translates and disseminates research.

In defining minority health, we are talking about distinctive health characteristics and attributes of the minority, racial, and/or ethnic groups in the US. There is a common theme of social disadvantage, and historically, minorities are underrepresented in biomedical research and the scientific workforce. Minority health gives us a good focus on what leads to good or bad health.

Minority health populations are African American or Black, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, and the Latino or Hispanic, which is not a race but an ethnic group. Dr. Perez-Stable himself is originally from Cuba.

He said that health disparities affect not only racial and ethnic minorities, but also populations of low socioeconomic status, underserved rural residents, and others subject to discrimination who have poorer health outcomes often attributed to being socially disadvantaged, which results in being underserved.

In defining health disparities, we are talking about a health difference that adversely affects disadvantaged populations, based on one or more of health outcomes. Health disparities research advances scientific knowledge about the influence of health determinants to develop interventions to reduce health disparities.

The health disparity outcomes include higher incidence and/or prevalence, premature and/or excessive mortality in areas where populations differ, burden of disease measured by the disability-adjusted life years, or DALY, metric, and poorer health-related quality of life and/or daily functioning as determined by standardized measures. Some health disparity risk outcomes include risks to well-being; biological/epigenetic risks; clinical event risks; and utilization of care

risks.

NIMHD looks at many factors with respect to health disparities, including biological, behavioral, physical environment, sociocultural environment and health care system, and levels of influence including the individual, community and society. We now know how important *place* is in determining average length of life. Also, social network and community could have major effects on health outcomes.

Dr. Perez-Stable stressed the importance of inclusion and workforce diversity. He said that inclusion of minorities in clinical studies is a separate important domain, not to be confused with minority health. It is a matter of social justice and common sense given that nearly 40 percent of the US population is a racial/ethnic minority. And, he said that biomedical workforce diversity is an urgent societal issue for both clinicians (10 percent) and scientists (less than 5 percent of NIH submitted grants).

NIMHD priorities are to define the science of health disparities and minority health, to establish a health services research in clinical settings program, to support innovative, investigator-initiated research through R01 grants, to promote diversity in the workforce, and to put emphasis on population and community health.

NIMHD has three scientific branches: research in the clinical setting and health services research; integrative biological and behavioral sciences: mechanisms and etiologies; and community health and population health sciences.

The research areas we are promoting include research centers on retaining youth from health disparity populations in the HIV treatment cascade; health disparities among immigrant populations; disparities in surgical care and outcomes; and social epigenomics and health services research on minority health and health disparities.

NIMHD holds scientific workshops. Dr. Yvonne Maddox launched the Science Visioning process, and a workshop on Measurement and Methods was held in April, and on Etiologies and Interventions in May. The Institute is organizing three additional workshops: Use of Self-Identified Race and Ethnicity in Genomic and Biomedical Research (NHGRI partner), Use of IT Technologies in Minority Health and Health Disparities (National Science Foundation partner), and Structural Racism and Cultural Competence: Impact on Minority Health and Health Disparities.

NIMHD will establish an Intramural Program. Dr. Perez-Stable will recruit a scientific director with an interest in epidemiology, and a clinician with a background in social and behavioral science. NIMHD will not be doing any basic science but will fund some basic research out of some programs that we inherited.

Dr. Masys asked if research will be interventional rather than observational and what would be the best balance? Dr. Perez-Stable said that a lot of the interventions we need are mostly global, or societal interventions, and not specific to a population or a group. Training a more diverse clinical workforce is an intervention. We know that doctors who are Latino and Black see more

May 3-4, 2016 – Board of Regents

poor people and uninsured people, and it doesn't matter where they practice. So, in San Francisco for example, you could have a practice in the ritziest part of the city and if you are Latino, Chinese, or black, patients will find you. Dr. Perez-Stable said he did not know what the right balance is. Interventions are expensive.

Dr. Walker asked Dr. Perez-Stable about health care in Cuba. The NIMHD Director said he was just there in March for the first time in 26 years. He said Cuba has an excellent prevention network for the entire population, in terms of basic health care. They have a good system of health care for sick people, but have not established a level of care to provide some procedures commonly practiced here. Cubans are reasonably healthy, but there is a tobacco problem.

Dr. Francis said the VA funds research at a number of the Institutes. He asked Dr. Perez-Stable if he would be interested in working with the VA. Dr. Perez-Stable said he would be interested. The NIMHD does have a VA representative on their advisory panel.

Ms. Martin said she is from Detroit where they are turning off water. All the public schools are closed because the teachers are striking. How does education connect with NIMHD goals? Dr. Perez-Stable said when they talk about socioeconomic status, most people think of money and income. But in research, when you ask people how much money they make, about 20-30 percent won't tell you. So we use education as a more consistent predictor of socioeconomic status than income, particularly as you age. Education is a strong predictor of socioeconomic status.

IV. FEBRUARY MINUTES AND FUTURE MEETINGS

The Regents approved without change the minutes from February 9-10, 2016 meeting. The September 2016 meeting will take place on September 13-14, 2016, the 2017 winter meeting will take place February 7-8, 2017, and the Board approved holding the spring meeting May 9-10, 2017.

V. REPORT FROM THE NLM ACTING DIRECTOR

Acting Director Betsy Humphreys said that President's budget request for 2016 had included some \$4 million for the expansion of ClinicalTrials.gov, including system changes that will be required after the final rule is published. These funds were not included in the budget passed by Congress, but NIH has agreed to provide \$3 million to NLM in FY 2016 for this purpose.

On the personnel front, she announced that the new director of the NLM would be named very soon. She also noted that Dr. David Lipman will stay on as the director of the NCBI and assume an additional role as an Associate Director of NIH. This appointment by the NIH director makes official Dr. Lipman's de facto dual reporting lines, which have existed for about two decades.

Ms. Humphreys acknowledged Martha Fishel, chief of the Public Services Division (PSD), who will retire on June 3, 2016, 42 years after beginning her federal career; Michael North, head of the Rare Books & Early Manuscripts Section of the HMD, who recently left NLM after 15 years of service to become head of Reader Services and Reference in the Rare Books and Special Collections Division of the Library of Congress; and other recent staff departures, including

May 3-4, 2016 – Board of Regents

Mary Kate Dugan, John Butler, Patti Stathopoulos, and Dr. Mike Sappol.

In the legislation area, little has happened on any of the items outlined in the Board book. The Senate is working on companion legislation to the 21st Century Cures Act. Items of interest to the NIH and NLM are summarized in your book.

On the public access front, the Patient Centered Outcomes Research Institute (PCORI) recently announced a requirement that publications arising from their funded research be deposited in PubMed Central.

Recently, five Senators wrote to NIH Director Dr. Francis Collins and Ms. Humphreys, asking that NLM consider providing access to conflict of interest information in PubMed. A copy of the letter from the Senators and our response is included in the packets at the Board members' places. We received a similar letter signed by many scientists and another one from a concerned citizen, all saying similar things. It is good editorial practice for journal editors to require that competing interest information be submitted along with papers and that this information be linked to published articles. The Senators and scientists are concerned that conflict of interest information is not evident to people reading abstracts in PubMed and that news articles are written without mentioning the competing interests of some article authors. The conflict of interest statements included in articles may only be available behind a pay wall so journalists don't have immediate access. NLM believes that it is probably feasible to provide access to such statements in PubMed, and we indicated that we are looking into this with publishers.

There is something in your book about the Federal Information Technology Acquisition Reform Act, FITARA. Its purpose is to improve transparency and accountability in acquisitions made by federal agencies. One of the mechanisms for achieving this goal is by requiring agency-level CIOs to approve all agency acquisitions.

In May, NLM transitioned from supporting the NN/LM via contracts to a grant mechanism: cooperative agreements. The new Regional Medical Libraries are listed in the Board Book. There are two new RMLs: the University of Iowa for the Midwest Region and the University of North Texas Health Science Center for the Southcentral Region. The NLM greatly appreciates the service of the institutions that were previously the RMLs in these regions and their staff members.

Ms. Humphreys noted that NLM uses the American Customer Satisfaction Index to track customer satisfaction with its health information websites. It has been renamed the E-Government Satisfaction Index. NLM continues to do very well in the surveys and is moving many services to responsive design, which automatically adjusts output to the size of the display on a user's device. Most recently, Genetics Home Reference was moved into a responsive design format launched on DNA Day, April 25.

She also mentioned the series of well-received Genomics Hackathons that Dr. Ben Busby is orchestrating for NCBI, here at the NLM and elsewhere.

Ms. Humphreys had previously reported how NLM's People Locator helped save an individual

May 3-4, 2016 – Board of Regents

from a collapsed building in Pakistan. More recently, after a very severe earthquake in Ecuador, NLM was asked by the Ecuadorean government if they could use People Locator as the official site for their citizens to consult and publicized it in local papers. This was the first time a government had done this.

Ms. Humphreys noted that NLM loans items from its collection for exhibitions by outside organizations. Most recently, NLM lent items to a new exhibition, *Beyond Chicken Soup: Jews and Medicine in America*, which opened on March 13, 2016 and will run through January 16, 2017 at the Jewish Museum of Maryland in Baltimore.

Dr. Greenes asked Ms. Humphreys two questions about funding for the NLM. First, he asked how BD2K funds would come into play under NLM. Also, would NLM be responsible for any informatics-related funds of the Precision Medicine Initiative portfolio?

Ms. Humphreys said these were future possibilities that would be on the top of the agenda for the new NLM Director. NLM has been working closely with Dr. Josephine Briggs, the interim director of the PMI, and Dr. Bill Riley, her deputy. She and Lister Hill Center Director Dr. Clem McDonald came up with some suggestions for PMI funding of standards-related activities that they thought would be helpful to the PMI, as well as to health care in general. They will fund these suggestions. Mr. Dishman is quite interested in how he can work with NLM.

Following up on Dr. Greenes' question, Ms. Yokote asked whether there are some other key areas that NLM should be involved with. With respect to PMI, Ms. Humphreys said that NLM has focused on trying to move forward some of the standards-related activities that will help the PMI, as well as health information exchange in general. It will be up to Mr. Dishman and the new NLM Director to determine new initiatives to move forward.

Dr. Masys asked about the output and input side. On the input side, the standards that would support this novel direct participant engagement, and on the output side, this question of participant engagement, of making it alive and a useful and vibrant activity seems to fit NLM like a glove. This isn't the kind of clinical research that you sign a piece of paper and don't know what happened. They want to create a personal value proposition, and it sounds like it is so close to the motivations like MedlinePlus and Genetics Home Reference.

VI. HEALTH FORMS ON-THE-FLY

Following Ms. Humphreys' announcement that Dr. Eugene Koonin, an NCBI scientist, had just been elected to the National Academy of Sciences, Ms. Yokote introduced Dr. Clem McDonald to report on health forms on-the-fly. Dr. McDonald said the focus of his report would be on forms that capture clinical data. NLM-Forms is a data capture widget designed in partnership with the Regenstrief Institute that generates an executable web form on the fly from a stored form description. NLM-Forms supports HL7 data types, questions and question groups, and repeats; scores weighted surveys; and can set default values, answer lists, and other question attributes as a function of responses to other questions. NLM-Forms can render an input form "widget" for any LOINC panel.

Dr. McDonald discussed the auto-completer, the “secret sauce” of NLM-Forms. It is a service that provides look up functions to tables (master files/coding systems). They provide “answer lists” for fields in many forms. The tables are connected to fields in the form by a URL with parameters to control fields in the table searched. The implementation provides autocomplete service for LOINC, RxTerms, ICD-10-CM, and many NCBI genomics resources, including RefSeq, and more.

When a user enters text into a table connected form, the table connector delivers an auto complete service showing the rows of the table which the text matches. The user then picks the row they want. This same tooling could be used to connect to any table. Once a user picks a record from a connected table, the NLM-Forms can use other fields in that record to provide answer lists, help messages defaults and other controls to subsequent fields in the form. This is what gives special power to NLM-Forms. He then demonstrated the service showing how the form autofills.

Dr. McDonald noted that there are other form builders. Most of them are commercial and not separable from a specific system and usable across systems. Also, other form builders do not have special support for clinical content or clinical data standards. Everything is written in JavaScript (the universal web language that runs on all computing devices). The browser turns form descriptions into live forms via a rendering program. The forms can be described in a spreadsheet or in JSON (JavaScript Object Notation), and authored with the NLM form builder. Examples of forms include the meaningful use vital signs form, PHQ-4 form, the Surgeon General’s Family Health History, and a form for reporting genetic results.

Lastly, Dr. McDonald explained a JavaScript validator and translator for the Universal Code for Reporting Units of Measure (UCUM), a computable standard for units of measure that has been adopted by HL7, DICOM, IEEE, and ISO-11240:2012—an international standard for units of measure in the pharmaceutical industry. In a large sample of HL7 messages, we found 60 different ways to express the units for a red cell count. UCUM was developed to solve this problem and has been adopted by most clinical standards groups.

UCUM includes all metric units, every kind of “conventional” unit, and a formal definition of the syntax and tables with coefficients and controls for conversion. Lister Hill converted this to a JavaScript widget—a service that can be used in any browser that will be an integral part of the NLM-Forms tool kit.

Dr. Greenes asked whether the forms were downloadable. Dr. McDonald said most were. Dr. Greenes asked if you are looking up diseases or medications, you might have a different way to say it. Have you thought about this? Dr. McDonald said that the forms do have synonyms which should address this concern.

Dr. Francis said an app developer could go to this site and start building patient generated data forms for a smart phone or a handheld. Dr. McDonald said we have three ways that the forms can be displayed. We encourage other users to add to it or change it. Dr. Francis said he has a whole team that would do so. Dr. Walker asked whether it would be good for patients to fill out such a form on a smart card so it was available whenever needed. Dr. McDonald said that the

May 3-4, 2016 – Board of Regents

number of forms is so infinite that he didn't know if that would be possible. Drs. Masys and McDonald agreed that these forms would be a nice lens into common data elements for researchers. Dr. Masys said that the world standard right now is Google Forms, because it gives you the completer, data sets, and analysis and the back end to do things with. He suggested linking to Google Apps.

VII. PRESENTATION OF OUTGOING BOR CERTIFICATE, NLM DIRECTOR'S AND FRANK B. ROGERS AWARDS

Ms. Humphreys presented outgoing board chair Ms. Gail Yokote with a Board of Regents certificate and a gavel for her service. Dr. Charles Rice, president of the Uniformed Services University of the Health Sciences, also received a certificate.

The Frank B. Rogers Award, established by an anonymous gift from an NLM employee to recognize important contributions of NLM staff members, was presented to three recipients: John Doyle for his work on the management and development of the NLM digital collections repository which complements PubMed Central's *Bookshelf* and allows us to have digital copies of many rare books and films in our collection; Dr. William Klimke for his work on developing the pathogen pipeline project; and Mr. Jim Mork for his contributions to the Medical Text Indexer.

NLM Director's awards were presented to: Dr. Wanda Whitney for her contributions to NLM and NIH effort to improve access for persons with limited English proficiency; Mr. Cuong Tran for managing NCBI's Title 42 appointment process for scientists and fellows; Mr. Todd Danielson for preventing and solving problems and providing wise advice for all NLM staff at all levels; and Ms. Karen "Janie" Robak, for her exceptional contributions to maintaining NLM's computer operation services.

Ms. Gail Yokote then adjourned the Board, inviting them to lunch and a group photo.

VIII. LHC BOARD OF SCIENTIFIC COUNSELORS REPORT AND NCBI BOARD OF SCIENTIFIC COUNSELORS REPORT

Lister Hill Center (LHC) Director Dr. Clem McDonald introduced Dr. Eta Berner, a professor at the University of Alabama at Birmingham and chair of the LHC Board of Scientific Counselors (BSC) to provide highlights from two years of BSC meetings, at which specific projects have been reviewed and advice provided to the LHC Director.

Dr. Berner showed a photo of the BSC, noting that its members' diversity of backgrounds and professions are a good fit for the diversity of the projects they evaluate. She then summarized the BSC's reviews of several Lister Hill research areas: clinical text de-identification; 3D informatics; drug terminology and ontology research; automated detection of lung diseases in chest X-Rays; consumer health question answering; and medical text indexing.

The other thing we've done each April since 2012 is to interview participants in the LHC Medical Informatics Training Program. This augments the written training reports; we try to get

a good sampling of fellows with different areas of expertise and different mentors to get a picture of the training program. Each time we've met with them, the Board has been very impressed with the diversity and caliber of the fellows, and their accomplishments. The fellows also praise the mentors for their instruction and inspiration. In addition, the fellows always want more information on the range of opportunities that are available to them. Our recommendations for the Training Program is that it should include a formal goal-setting planning for the experience, and an intensive weeklong "boot camp," to identify who needs what. We also suggested the creation of a fellows' handbook and that the fellows themselves develop it, to save on mentor time and give fellows more of a say to that process. .

The themes in the BSC comments and recommendations about Lister Hill's research are quite consistent across the projects. One is to engage the users— focusing on their needs and obtaining feedback on how well the tool is meeting their needs, and working with users who are outside NLM, which can provide a clinical context outside the basic research. The second is further collaboration and expansion within Lister Hill Center, and basically to clarify some of LHC's research objectives. The third is to increase the evaluation effort as part of that collaboration; we recognize that this could require more staff involvement. Fourth, we would urge LHC continue to dissemination and outreach, but publish more broadly in order to get the word out through different media and make the tools available and more accessible and to a larger audience. We see that that's beginning to happen. So, in general, the Board felt that the teams that we've reviewed had done a very good job, consisting of top-notch scientists doing work that is incredibly useful.

Dr. Masys asked how LHC research projects are selected in the first place. Are they constantly evaluated and, depending on how they're going, are they scored for whether or not they should continue, or whether their budgets should be adjusted? Because one of the questions is, is this work intramural or extramural? LHC Director Dr. Clem McDonald replied that he has not seen a growth in overall research funding during his term, but there have been no complaints about priorities, although there are of course discussions about what we're going to do next.

Ms. Humphreys commented that a review of all those issues should be part of the strategic planning process. There should be a discussion about what kinds of things NLM may be uniquely positioned to do or what kinds of problems and issues that other people have, like improved efficiency, we might tackle.

Dr. Masys noted that four out of six of these common themes that Dr. Berner presented, and the associated projects, came to the BSC relatively late in their lifespan. That suggests maybe an earlier process review that is more structured, with checkpoints along the way to say, are you collaborating? Are you connected to your customers? But it could be done internally and, in the life cycle of any project, it may help create a trajectory that external reviewers see downstream.

Ms. Humphreys agreed, but added that, when the BSC urges expanded collaboration, at the same time, there are almost no projects they would review where there is not a fair amount of interaction with some set of users. You also have a number of cases with external reviewers so, while it's probably true that you can't have too much collaboration and interaction, it's not that it doesn't exist.

May 3-4, 2016 – Board of Regents

Ms. Yokote next introduced Dr. David Landsman, to report on the NCBI Board of the Scientific Counselors. Today, he noted, he will be talking as the Deputy Scientific Director, standing in for the BSC chair, who normally doesn't attend this meeting.

For those unfamiliar with the NCBI organization structure, Dr. David Lipman is the Director, and there are three branches. Most of the research is done in the Computational Biology Branch, with Dr. Landsman as the Branch Chief. NCBI senior research staff are NIH appointees, and their research programs are reviewed by the BSC. In addition, the BSC does a separate review of NCBI projects, such as PubMed, GenBank, etc. The NCBI has two other Branches, the Information Resources Branch that maintains hardware, software, the network, etc. The third branch is the Information Engineering Branch with James Ostell as the Branch Chief. This branch is the largest at NCBI and prepares all the database resources that NCBI provides to the scientific community.

The NCBI BSC consists of eight members, each serving a maximum term of five years. All NCBI Senior Investigators and Tenure track Investigators get reviewed by the BSC every four years. Tenure is awarded by NIH Central Tenure Committee to qualified Tenure track Investigators. BSC reviews retrospective performance, as opposed to the grant review system which is prospective. Tenure track Investigators are converted to tenure after five to seven years. They are reviewed mid-term, after two to three years. Conversion to tenure means you get support for your research. Besides the BSC there are two committees that are involved in the tenure process, the NCBI Promotions and Tenure Committee and the NIH Central Tenure Committee. The NCBI Committee consists of two-thirds NCBI scientists and about one-third scientists from other NIH Institutes. We frequently invite ad hoc members to the BSC meetings because each BSC does not always have members who know a lot about the particular field under review. For example, this coming Tuesday, we don't have anybody in text mining so we added an ad hoc board member who is an expert in that field. The BSC meets twice a year. The spring meeting reviews NCBI resource projects, while fall meetings are on research programs.

NCBI has 11 Senior Investigators. The preparation for a BSC review entails a written summary of the research performed over the past four years as well as two to three pages of future goals. They include a copy of their CV and five recent manuscripts demonstrating their best research. The investigators also give a 30-minute talk to the BSC on their past four years of work. In addition, they get to talk to the BSC alone. The BSC evaluates each investigator, writes a report addressed to the director of NCBI, and then they rank them based on the significance of their research. They also consider such factors as collaborations with outside and NIH scientists.

BSC members are also required to be assess mentors and trainees relationships with postdoctoral fellows and students. There is a poster session where the BSC speaks to the fellows and students. The BSC members enjoy the poster sessions.

Dr. Landsman presented a brief summary of some of the NCBI research projects.

Board member Dr. Jill Taylor said she could see how the NCBI BSC plays a crucial role in maintaining the quality of the research that you do. There appears to be a mixture of basic research and much more applied research that has implications within the larger community of

science. What role does the Board play in deciding next steps?

That is a good question, Dr. Landsman replied, because, in the research realm, they could make recommendations for additional projects. Scientists could discuss such topics with them, gaining suggestions for additions to ongoing projects and science.

IX. NLM'S RESPONSE TO PUBLIC HEALTH EMERGENCIES

Ms. Yokote next introduced Ms. Stacey Arnesen, to talk about public health emergencies and how NLM responds to them. Actually, Ms. Arnesen noted, the Library has provided health information for disasters and public health emergencies since the 1980s, in response to several chemical spills, and teamed with the Pan American Health Organization to develop disaster information centers in Latin America. In its 2006-2016 Long Range Plan, *Charting the Course for the 21st Century*, NLM solidified its place in this area with the creation of the Disaster Information Management Research Center or DIMRC. The plan recommended that NLM be a partner in the federal disaster preparedness, response and recovery efforts, to ensure access to health information and effective use of libraries. It also stated that NLM should demonstrate how libraries and librarians can provide critical information services, and foster a culture of community resiliency.

DIMRC is the branch at NLM/NIH tasked to coordinate the collection, organization and dissemination of information on disasters and public health emergencies. We rely not only upon colleagues throughout the Division of Specialized Information Services, but also on staff throughout NLM. Even more broadly, DIMRC collaborates with federal, state and local agencies, and national and international organizations.

Ms. Arnesen then showed the DIMRC home page and its many resources. One section features WHO definitions of public health emergencies, including an occurrence or imminent threat of an illness or health condition caused by bioterrorism, epidemic or pandemic disease, or a novel and highly fatal agent or biological toxin that poses a substantial risk. As for disasters, there are several definitions. One is a situation or event that overwhelms the local capacity, necessitating a request to the national or international level for external assistance. NLM's scope includes emerging infectious diseases such Zika and Ebola, acts of terrorism, natural disasters, and major accidents.

In this country, response to disasters is local until local authorities ask for assistance from the state or federal level. She described what DIMRC and NLM do, between disasters, when it's quiet. On a daily basis, staff are maintaining and updating resources and infrastructure to better prepare for incidents. They're providing wider access to the disaster health literature by developing tools for emergency preparedness and emergency response, and by conducting research in disaster informatics and communications. NLM is also committed to helping libraries in disaster-stricken areas to maintain their operations, because if you can't maintain or take care of yourself, you can't take care of others.

We work closely with HHS Assistant Secretary for Preparedness and Response as well as with other NIH Institutes, and the Department of Transportation and other federal agencies for major

incidents. Among numerous incidents, NLM has been responsive to the earthquake in Japan, which led to a tsunami and then a radioactive leak; REMM, our radiation emergency medical management tool, developed in conjunction with HHS, was used after that incident and some of its information translated into Japanese.

NLM has also responded to hurricanes, earthquakes, a number of chemical emergencies and, more recently, some of the emerging infectious diseases. NLM has numerous resources that can provide relevant and critical information in times of public health emergencies. Ms. Arnesen and SIS senior computer scientist Victor Cid then presented a virtual tour of NLM's Emergency Operations Center or EOC in Second Life, an open virtual world where you're allowed to purchase islands, like this one representing NLM, and build upon them.

In the US, most EOCs follow the incident command system or ICS, and the structure and people play clearly defined roles in that room. NLM's virtual room was developed to conduct disaster drills and disaster exercises, and to practice the incident command system with hospitals, public health departments and related health organizations. We've done about five different drills using this room with groups across the country. Mr. Cid then took the Board on a brief tour of the room, highlighting its capabilities.

Using a timely example, the Zika virus, we can see how we put all of our information resources together in the ICS. We have the Zika Health Information Resource Guide, which links to all relevant resources from NLM, as well as to the CDC and many other key players. This also links to pre-formulated searches on Zika in PubMed. Next, we'll go over to Disaster Lit, which compiles the "grey literature"—information quickly coming out of emergencies and disasters but not yet appearing in the published literature, or in PubMed. SIS does get reports of these emerging government documents, and we already have over 200 Zika-related items that have been added to this database.

We take seriously the fact that, following a major incident, there may be a disruption to library services or buildings. In selected cases, we offer free access to books, articles and databases through the Emergency Access Initiative, a partnership of NLM and publishers.

Other NLM resources include an archive of Web-based global health events information on Zika and Ebola being produced by the History of Medicine Division, virus and genetic information that is being put together by NCBI, and information on disaster and public health emergency topics on MedlinePlus.

Several NLM tools are specifically designed for disaster and emergency response, including WISER, the Wireless Information System for Emergency Responders, for first responders in hazardous material incidents, and People Locator which helps reunite family, friends and other loved ones in disasters.

DIMRC is always responding via daily activities, such as updating our databases, and we are always looking for more information on public health emergencies to add to these resources, so in effect we're always monitoring for new emerging incidents. We also collaborate with and are in contact with other agencies to see what they're working on, what's coming up, and what they

may have of interest. Of course, if you're going to create new resources or enhance existing ones, you have to make sure that those who need them have access to them. The Zika Virus, for example, set in motion a variety of activities, including basic research, genetic sequencing, development of countermeasures against the virus, etc. What we do is try to compile all of this information in one place, to get people started. In addition, we syndicate this content, and other organizations then have access to our information and can easily take our information, post it on their websites, and redistribute it.

Another interesting thing we are working on is archiving Web content surrounding public health emergencies. HMD is tracking the changes that occur about a particular incident such as Zika overtime because as we learn new information that gets updated on the webpage. HMD's efforts to chronicle and understand the evolution of knowledge about and response to the disease may help to drive future research in these areas.

Finally, on Zika, NLM is working with SNOMED, LOINC and others, to develop standard terminology that can be used in electronic health records, and in turn link to the consumer health information that is provided through MedlinePlus Connect.

Ms. Arnesen mentioned DIMRC's work related to the Flint, Michigan water crisis, creating a page on lead in the water supply as well as links to NLM databases. She also described the 2014 Elk River chemical spill in West Virginia. The chemical in question is MCHM, which now has more than 25 citations in PubMed but, at the time of the incident, had none. She described SIS's quick and effective collaboration with several HHS agencies, which led to the discovery of unpublished studies on that chemical which formed the basis for a new record in the Hazardous Substances Databank (HSDB). Ms. Arnesen says she envisions a stronger role for NLM in the future, regarding emerging public health issues. DIMRC is already working with the National Institute of Environmental Health Sciences, for example, figuring out ways of doing research on disasters much faster. Typically, in most disasters, the time lag to start collecting any data is about nine months.

Dr. Jill Taylor said that she has been through many public health emergencies in New York but has never considered NLM as a resource in addressing them. What can we do about that? Ms. Arnesen said she is open to any suggestions for promoting our public health emergencies resources.

Dr. Charles Rice said he was very impressed by the great work of NLM, but he wonders about duplication of effort within the federal government when the NLM, the Department of Homeland Security, HHS, CDC, etc. which are focused on the same goals. He is hopeful that when the new NLM Director arrives, across Rockville Pike, we can expand even further and coordinate more efforts because each center will have a different focus.

Yes, said Ms. Humphreys. Going forward with strategic planning, this is an area where we need to determine where NLM can add the most value and focus our efforts on some of the things that libraries do best, pulling together and organizing all of the relevant information.

X. USING SOCIAL MEDIA FOR PUBLIC HEALTH MONITORING AND SURVEILLANCE

Ms. Yokote introduced Dr. Graciela Gonzalez Hernandez, an Associate Professor in Bioinformatics at Arizona State University.

Dr. Gonzalez Hernandez said her focus today would be on her research on mining social media for public health surveillance and monitoring. Health dialogue is very lively online. We found some references that say 26 percent of Internet users discuss health information and, of that group, 30 percent reported having changed their behavior based on information that they got on the Internet. Forty-two percent discuss their current medical conditions and, contrary to popular belief, there is more personal health information on Twitter than people share with their doctor. We know that there are 310 million active users on Twitter, so perhaps 24 million people might change their behavior after reading something online.

We have been following social media throughout the years and interest in analysis of social media is growing. Last September, one of our papers, “Utilizing Social Media Data for Pharmacovigilance,” was nominated as one of the ten articles with the greatest potential social impact from the more than 2,500 journals published by Elsevier/Atlas.

Social media mining involves many steps. The pipeline begins with data collection and then goes to annotation and resource adoption. The next steps are a classification process, information extraction, and normalization—this takes something that is said in that Tweet or post and turns it into something standard—and, finally, validation where we begin our case studies. The pharmacovigilance study we are working on is about adverse drug reaction, that is, an unintended, harmful response suspected to be caused by a drug taken under normal circumstances. There are over 2 million serious reactions in the US annually, and it is the fourth leading cause of death, ahead of pulmonary disease, diabetes, AIDS, pneumonia, accidents and automobile deaths.

Dr. Gonzalez Hernandez presented a sampling of tweets that her team has seen, all dealing with on drugs. She and her team have amassed over 15 million tweets about over 100 drugs. Their data collection process takes into consideration phonetic spelling variants and alternative drug names. To assist in this process, they have multiple trained annotators to try and interpret exactly what a person is saying.

From this vast collection, we take a sample. I usually take a two-step approach: combine data and form short text nuggets. Then we start doing a handbook with the rules for annotation, so that when other annotators do this type of classification we have a standard. Some text nuggets are said to be stronger based on the language around them or even how the words are represented. For example, if people put something like “VERY” in all capital letters that obviously changes the strength of the effect, as opposed to “very.” Unfortunately, the downside of auto-corrected text is that it removes the emotion behind what people wrote. We’re trying to have features that detect the weight of this text, not autocorrected, so that we can detect the true meaning. Actually, this idea of text classification is paying off, the same classifier was retrained with a quick annotation, and this means, when prescription drugs are being abused, we can now

also detect that.

In the area of drug safety classification, we are trying to get a predictive system going just by looking at these comments on Twitter. Adverse drug reaction (or ADR) extraction is the next hard step—finding mentions of symptoms and other phrases. We need to either add them to the lexicon or learn to find them on their own. We've tried doing this with traditional approaches, but unfortunately they perform poorly on social media. So we developed an ADR mine system, which is deep learning. It uses conditional random fields, and outperforms lexicon-based approaches and all others currently out there. It looks at clusters of words and also at topics, and what different words will mean depending on their placement and order.

We look at the groups of clusters first and then give that collection a title. It will start identifying words and thoughts within that cluster. We train on the many millions of tweets that we have, so we can train unsupervised. This process is working really well, and every time we do more, it gives it a better edge. Another ongoing concept for the team is normalization, so that we can do an exact text match, then a definition match and then a semantic-related match; we go through these different levels and try to find the best match. How do we validate this? We take the drug, the documented adverse effects of the drug, and then we compare these to the things we are finding in social media.

Going forward, we will be exploring health timelines. The idea is, we are not looking at only one message but we can create cohorts of people that we are interested in following. For example, we created a cohort of pregnant women, starting at the beginning of the pregnancy and following their progress until delivery of their children. In this way, we can figure out what happens throughout pregnancy; we can really tell whether there is a diagnosis given the mother during pregnancy, and whether she took the drug or not. We look at outcomes, miscarriage, babies who are born or babies who passed away. But we are still basically finding phrases. We highlight the drug mentioned throughout and try to see if we can create a correlation or not.

Finally, we are very interested in prescription drug abuse monitoring. When we observed that common drugs were mentioned together, we asked a professional, who said that the top two drugs taken together on social media are not generally prescribed together. So we stumbled on this prescription drug abuse idea. We have other specific medication case studies going on, and we seek to combine our findings with data the FDA is getting and have a better view. I think many interesting things have come out of this; automatic monitoring and language processing are vital.

Ms. Martin asked whether this same information and mining and other techniques could be used on social media platforms other than Twitter. Yes, Dr. Gonzalez Hernandez replied. There is a whole list of social media outlets out right now and her teenagers keep her posted on them. Some of those platforms are naturally closed, meaning you can't freely access them, unlike Twitter. Facebook used to be closed but then they added hashtags, which turns anything with a hashtag public and makes it searchable. When we have a health-related website, our techniques have to change. Most of this pipeline will work across platforms once we have more a developed algorithm.

May 3-4, 2016 – Board of Regents

Dr. Olds asked whether Twitter and other social media platforms were hackable. Like could someone essentially trash a medication or a drug?

Dr. Gonzalez Hernandez said she has received additional funding to identify or create software to identify bots that may carry out those kinds of activities. Her team is also trying to identify fake postings about vitamins and supplements. “Good fake” will fly under the radar, but “traditional fake,” like multiple repetitions of the same false messages, is easily caught. The harder ones to identify are when companies have humans write things. We can detect them with spikes, and filter them from contained areas.

XI. EXTRAMURAL PROGRAMS REPORT

EP Director Dr. Valerie Florance noted that last year NLM made 135 awards, totaling \$42.131 million. Most of our money goes into research project grants. Last year, we made 45 new awards with 90 continuing awards; 77 percent of our budget was spent on continuing awards. Typically 50 percent of our awards go into health care and public health research. This distribution depends on the applications and funding we receive.

Using RePORTER, NIH’s Research Portfolio Online Reporting Tools, Dr. Florance presented a context for NLM awards within all of NIH. Generally, for research projects, our success rate is 21.9 percent, which is even more successful than NIH. Congress asked us to be more transparent, so that people would know how we make decisions. Her slide showed that, for a given percentile score, what we did and didn’t fund. There is always an area of overlap where there are some skips and some awards. Looking at the R01 or Research Project Grant, Dr. Florance compared all of NIH to NLM.

Dr. Florance highlighted some of the awards made in 2015 on interesting topics. Health care informatics and public health informatics are two areas where we are trying to do more. In our translational bioinformatics portfolio, text mining is a hot topic, as is modeling larger, multiscale models. Consumer informatics is also popular, as is precision medicine. In the information sciences and methods area, work continues to focus on improved NLP for mining literature and electronic health records

NLM also has a resource grant programs, which currently centers upon information resources to help reduce health disparities, and a unique program to support scholarly works on history, policy and scientific trends in biomedicine and health. In 2015, we made 11 new Informationist supplements awards to research grants funded by NIH institutes and centers, to bring librarians onto their grants to assist in data management. We manage two Common Fund grants, which are for innovators and early-stage investigators.

As part of the BD2K initiative, NLM manages several programs focused on training for data management, annotation and curation of scientific data. Three of grants awarded are led by librarians. The point of these resources is that they can be used by librarians and others in training scientists to better plan and manage their data, with an eye toward sharing it.

Currently, there are 147 active NLM grants and a US map was presented that shows where the

May 3-4, 2016 – Board of Regents

money goes; 53 percent of our grants are in Pennsylvania, California, Massachusetts, New York and Texas. In 2015, there were 696 publications that acknowledged NLM research grant support. Dr. Florance showed a list of the top titles in which NLM grantees were published. Our 2015 NLM Informatics Training Conference was held at NIH in July 2015, with 264 attendees. The agenda included a variety of presentations and posters by trainees and had a showcase of NLM resources of interest to informatics researchers, presented by staff from LHC and NCBI. Extramural Programs staff gave two lunchtime webinars. Dr. Florance showed pictures of trainees giving their talks; these are available on the EP web site at <https://www.nlm.nih.gov/ep/trainingconf2015.html/>. In closing, for our grants operation and grantees, 2015 was a good year.

Ms. Yokote asked about measuring the impact of the grants, and being able to leverage things like My Bibliography from NCBI.

WEDNESDAY, MAY 4, 2016

Board Chair Gail Yokote opened the May 4, 2016 meeting of the Board by recognizing Mr. Eric Dishman, former Board Member and recently named Director of the NIH Precision Medicine Initiative (PMI) Cohort Program. Mr. Dishman said he will be starting officially on June 13, 2016 and that the Health Provider Organizations (HPOs) should be announced soon. The HPOs will recruit the bulk of the million or more U.S. volunteers who will make up the required diverse landmark longitudinal research study.

Other announcements about the Program will come a little later. He said he will meet this week at the White House to determine what will be accomplished before the President is out of office. He hopes to work with the NLM because NLM builds tools that the entire research community uses. While this program is at NIH, we are building an entity very much like industry would build.

XII. DOCUMENT DELIVERY IN THE E-JOURNAL ERA

Dee Clarkin, Deputy Chief of the Public Services Division, explained how document delivery is integral to NLM, how it has influenced some important developments in U.S. copyright law, and how it operates in today's information ecosystem.

Ms. Clarkin traced the origins of document delivery at NLM to John Shaw Billings' loan system in the 1880's. She discussed how expanding technology and increasing demand for medical literature forced a legal show-down between medical publishers and the NLM which led to the Supreme Court copyright case, *Williams & Wilkins v United States* (1975) which affirmed that it was a fair use for libraries to photocopy materials for scientific research. However, document delivery in the e-journal era is different. Electronic items are usually governed by contracts law - - by the license that is negotiated between the user and the publisher -- lessening the role for fair use in the e-journal era.

Ms. Clarkin provided an overview of how NLM's electronic resource sharing routing system, DOCLINE, was developed in the 1980s and how it revolutionized document delivery for medical librarians. Now, DOCLINE requests have diminished with the growth of open source and open

access articles. NLM has studied the reduction in requests and determined the reason for this is the overwhelming desire of users to have instantaneous access to information, which precludes even the shortest wait for an item on document delivery. There are also alternative software systems supporting document delivery that are in use by some health sciences libraries.

Continuing to describe the current ecosystem for STEM literature and information, Ms. Clarkin mentioned a shift from purchases of large bundles of journal titles from publishers to alternative models like “pay per view” where users only pay for limited use of a single article at the point of need. She also discussed current trends such as the growth of the illegal downloading site, SciHub, which garners up to possibly 400,000 downloads a day. Even more interesting is that many of the SciHub users who have access to library resources prefer to use SciHub due to its simplicity and ease of use. Many in the document delivery and library community believe we should make our services easier to use.

Ms. Clarkin described how document delivery actually works at NLM -- including a variety of streams of requests organized by membership status in DOCLINE and emphasized that NLM only receives document requests that cannot be filled by local resources. She also discussed how NLM decides whether license terms are acceptable or not, providing adequate document delivery rights as well as preservation rights. Clarkin wrapped up her talk by stating that regardless of continuing changes, NLM will need to provide ILL to provide access to its uniquely held materials that are not available electronically.

Christopher Cole contributed information about the differences and similarities between NAL and NLM, including the need to keep materials accessible. Betsy Humphreys stressed the importance of the NLM collection as a last resort for borrowing libraries.

Daniel Masys asked about archival copies of scanned literature. Clarkin explained that the Library doesn't keep scanned copies produced for document delivery. Robert Greenes asked when materials are no longer under copyright. Clarkin said that in the U.S. materials published before 1923 were in the public domain -- or out of copyright. Betsy Humphreys said that when Newt Gingrich was on the NLM Board he explained that Congress' expansion of the copyright term was to “protect the Mouse” -- Mickey Mouse -- because the Disney property was about to go into the public domain. Ken Walker talked about the importance of practicing physicians downloading papers and asked about the recent Google decision about scanning books. Ms. Clarkin explained that the Google Books decision is about a different aspect of copyright.

Gail Yokote asked about negotiations among large academic research and health sciences libraries. Betsy Humphreys noted the proliferation of open access materials due to the requirements of research funders. Robert Greens inquired about those without access to libraries. Betsy said open access materials are an alternate for those users.

Daniel Masys referred to the quote of Pat Brown of Stanford: “...scientific publishing is the only enterprise that after a long and difficult gestation the midwife keeps the baby.” He said empowering users who can self-publish is a destabilizing force. Betsy Humphreys mentioned the variety of models in open access and also why various journals are using the open access model.

Ms. Clarkin commented on a variety of possible paths including different types of business models, researcher self-publishing in blogs that completely skip the publication step and others. Publishing is changing and having a dramatic impact on libraries and document delivery.

XIII. NCBI REFERENCE SEQUENCES COLLECTION

Dr. Kim Pruitt said sequence data collection is biology's big moon shot and is a massive investment. She explained that sequence data doesn't mean anything in isolation, that you need to add analysis, metadata, and annotation to make it useful. She said the reference sequence (RefSeq) collection—which provides sequence records connected to functional information—is an important resource to address major world issues like food security, health, disease, and treatment plans.

Dr. Pruitt said the RefSeq project has multiple goals: accuracy in sequence and annotation representation; transparency in the GenBank source of sequence data; staying current; using consistent methods across the dataset; providing rich feature annotation; and connecting sequence to function.

She gave an overview of GenBank and explained that RefSeq is based on public data in GenBank (and RNAseq) but is a vetted, selected data set. Having a sequence available in GenBank doesn't mean it's in RefSeq. She also said that RefSeq records are updated more often than GenBank and are supported by manual curation. She said it's important to understand that RefSeq provides some information that is not otherwise publically available in GenBank – this includes alternative transcript variants (based on RNAseq analysis) and more complete transcript and protein sequence representation.

Dr. Pruitt provided examples of custom data sets, such as the human RefSeqGene project and the fungal internal transcribed spacer (ITS) project. She also explained the benefit of using both a manual curation approach and a computational annotation pipeline. She said one of the unique aspects of RefSeq is that it can provide a transcript even when the genome sequence isn't complete.

RefSeq staff are engaged in a collaboration with the Genome Reference Consortium (who maintain the human genome sequence) and they're the leader in an international collaboration to harmonize the human and mouse genome annotation that's emitted by the NCBI annotation pipeline and the Ensembl genome annotation pipeline.

She talked about how RefSeq supports clinical needs by providing sequence records that are widely used to report the location of sequence variations that impact phenotype—both transcript records and also genomic gene-centric records, provided by the RefSeqGene Project, are used for this, e.g., in ClinVar. She gave an example of how RefSeq provides a custom resource of short sequence regions (fungal ITS project) that can be used to identify fungal organisms, at the request of an International collaboration with mycologists. To support this project, RefSeq curators comb the literature to find fungal organisms that have been correctly identified and analyze available GenBank sequence data.

She also talked about prokaryotic RefSeqs, where they annotate all genomes that pass quality metrics and curate functional names, and said that annotation of eukaryotic genomes is more selective. She said last August, they re-annotated the full complement of RefSeq prokaryotic genomes (over 42,000) and they plan to do this regularly when their software has improved accuracy. A small number of bacterial RefSeq genomes are annotated by collaboration including the Reference Sequence for E. coli, the E. coli K12 MG1655 gene strain, which is the most important bacterial reference. This sequence has been annotated, reviewed and curated by the E. coli research community.

After Dr. Pruitt's presentation, there was a discussion about scaling the eukaryotic genome annotation pipeline to meet the annotation needs of population based sequencing, which Dr. Pruitt said they have been having conversations about but haven't committed to yet. Dr. Walker asked about the curating ancient genomes and wondered if he had this new technology if he could use the database. Dr. Pruitt said the database could be used to identify the contextual information on other genomes to identify gene locations and possible alternative splice variants. There was also a discussion about the infant metabolic panel, and RefSeq accession number format (different prefixes) and machine learning. During this discussion, Dr. Pruitt said they're using RNAseq data heavily in the annotation pipeline because they look for regions where they have evidence that's trustworthy and that part of her team's curation focus has been responding to new genes that are being called by the annotation pipeline.

XIV. REPORT FROM THE NOMINATING COMMITTEE FOR NEXT BOR CHAIR

Board Chair Gail Yokote called upon Board Member Mr. Christopher Cole to report on the Nominating Committee's decision for the next BOR Chair. Mr. Cole said that the Committee had met and named Dr. Robert Greenes to serve as the next BOR Chair. The report of the Nominating Committee was approved unanimously by the full Board.

XV. PROACTIVE APPLICATION AND DATA SECURITY AT THE NLM

Mr. Ivor D'Souza, director of the Office of Computer and Communications Systems, began by emphasizing that NLM's mission revolves around being a trusted provider of medical information so security is a very important aspect of it.

Mr. D'Souza's talk covered three themes: 1) application security versus network security and other forms of security, 2) the importance of application security and challenges with the reactive model that is prevalent across this industry and the government; and 3) NLM's journey to proactive application security and how that informs and helps the Library.

Mr. D'Souza referred to application security as the next frontier for security, because the Library started off with operating systems security. He said you could have physical security, network security, and firewalls, but that you can't block people out from using applications and that's why he is focusing on application security. He said most reported vulnerabilities turn out to be application oriented.

Quoting an expert in the field of security, Mr. D'Souza said, "If you think technology can solve

May 3-4, 2016 – Board of Regents

your security problems, then you don't understand the technology." He emphasized the importance of security practices that were repeatable and sustainable over time.

He discussed the Open Web Applications Security Project or OWASP, which looks at security practices and breaks them into three levels: opportunistic, standard, and advanced. The opportunistic level is the lowest standard and the advanced is the highest. The Library uses the dynamic scanning tool, the source code scanning tool, and a security checklist. He discussed mandatory security training for employees and the security group.

He emphasized people, technology and process, and the importance of each in the application security journey, which he said, is a journey that never ends. Mr. D'Souza finished by talking about the importance of having adequate licenses for source code scanning tools to find vulnerabilities in software prior to deploying them. He said the Library is beginning to see benefit from its early investment in a small set of licenses, and the Library is looking into other ways to acquire additional licenses.

After Mr. D'Souza's presentation, there was a conversation about the importance of NLM's proactive approach. The discussion also included positive news about costs associated with software security tools, which are going down and how software developers are writing code that is both better and more secure. Mr. D'Souza also mentioned that they are minimizing investments up front until tools are proven to be worth the money, and that investments must be outcome and results focused.

XVI. ADJOURNMENT

Ms. Yokote adjourned the Board of Regents meeting at 12:00 p.m. on May 4, 2016.

ACTIONS TAKEN BY THE BOARD OF REGENTS:

- Approval of the February 9-10, 2016 Board Minutes
- Approval of the May 9-10, 2017 Future Meeting Dates
- Approval of Dr. Robert Greenes as the new BOR Chair

Appendix A - Roster - Board of Regents

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

Betsy L. Humpheys, M.L.S.
Acting Director, National Library of Medicine

Gail A. Yokote, M.S.
Chair, NLM Board of Regents