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

MINUTES OF THE BOARD OF REGENTS
May 12-13, 2015

The 169th meeting of the Board of Regents was convened on May 12, 2015, at 9:00 a.m. in the Board 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 13th, the meeting was reopened to the public from 9:00 a.m. until adjournment at 11:45 a.m.

MEMBERS PRESENT [Appendix A]**:**
Mr. Eric Dishman, Intel Corporation
Dr. Robert Greenes, Arizona State University
Dr. Trudy MacKay [Chair], North Carolina State University
Ms. Sandra Martin, Wayne State University
Dr. Ralph Roskies, University of Pittsburgh
Dr. Esther Sternberg, University of Arizona
Ms. Gail Yokote, University of California, Davis

MEMBERS NOT PRESENT:
Dr. David Fleming, University of Missouri School of Medicine
Dr. Henry Lewis, American University of Health Sciences

EX OFFICIO AND ALTERNATE MEMBERS PRESENT:
Mr. Christopher Cole, National Agricultural Library
Col. Michael Cunningham, United States Air Force
RADM Scott Giberson, Office of the Surgeon General, PHS
Ms. Kathryn Mendenhall, Library of Congress
BGEN Charles Potter, United States Air Force
Dr. Dale Smith, Uniformed Services University of the Health Sciences

CONSULTANTS TO THE BOR PRESENT:
Dr. Tenley Albright, Massachusetts Institute of Technology
Dr. Holly Buchanan, University of New Mexico
Dr. H. Kenneth Walker, Emory University School of Medicine

SPEAKERS AND INVITED GUESTS PRESENT:
Dr. Markus Covert, Stanford University
Dr. Lynn Goldman, George Washington University
Dr. Douglas Lowy, Acting Director for National Cancer Institute, NIH

MEMBERS OF THE PUBLIC PRESENT:
Mr. John Harrington, Contractor to Lister Hill Center
Mr. John Howe, Friends of the National Library of Medicine
Ms. Lesley Macherelli, Friends of the National Library of Medicine
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. Michael Ackerman, Lister Hill Center, 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
Dr. Dennis Benson, National Center for Biotechnology Information, NLM
Dr. Rodney Brister, National Center for Biotechnology Information, NLM
Ms. Kathleen Cravedi, Office of Communications and Public Liaison, NLM
Mr. Todd Danielson, Office of the Director, NLM
Ms. Darlene Dodson, Office of the Director, NLM
Mr. Ivor D'Souza, Office of Computer and Communications Systems, NLM
Ms. Mary Kate Dugan, 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. Dan Gerendasy, Office of Health Information Programs Development, NLM
Dr. Michael Huerta, Office of Health Information Programs Development, NLM
Ms. Christine Ireland, Division of Extramural Programs, NLM
Ms. Jen Jentsch, Division of Library Operations, NLM
Ms. Shannon Jordan, Division of Specialized Information Services, NLM
Ms. Janice Kelly, Division of Specialized Information Services, NLM
Mr. Paul Kiehl, Office of the Director, NLM
Ms. Lisa Lang, National Information Center for Health Services Research and Health Care Technology, NLM
Dr. Clement McDonald, Lister Hill Center, NLM
Mr. Dwight Mowery, Division of Extramural Programs, NLM
Mr. Dimpal Patel, Office of Computer and Communications Systems, NLM
Dr. Steven Phillips, Division of Specialized Information Services, NLM
Mr. Jerry Sheehan, Office of the Director, NLM
Dr. Hua-Chuan Sim, Division of Extramural Programs, NLM
Dr. George Thoma, Lister Hill Center, NLM
Dr. Alan VanBiervliet, Division of Extramural Programs, NLM
Dr. Fred Wood, Office of Health Information Programs Development, NLM
Dr. Deborah Zarin, National Center for Biotechnology Information, NLM

I. OPENING REMARKS

Dr. Trudy MacKay, NLM Board of Regents Chair, welcomed the Regents, new Regents, alternates, and guests to the 169th meeting of the Board. She introduced RADM Scott Giberson to present the report from the Office of the Surgeon General.

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

RADM Scott Giberson reported that the new Surgeon General, Vice Admiral Vivek Murthy, has been sworn in and is already actively involved in the OSG.

RADM Giberson reported that he returned from Liberia a week ago. On May 9, Liberia was declared Ebola-free. But, much more progress still needs to take place in Africa.

The new Surgeon General is interested in community-based prevention initiatives. The OSG will be reaching out to engage all its partners. Community-based prevention will be the cornerstone of their efforts. They are also going to engage and partner with sports and entertainment celebrities—knowing that they can positively influence the country, encouraging young people to avoid tobacco. The professional baseball player will be more effective in articulating this message.Mental health will be a more important focus with this Surgeon General, as will be violence against women. .

The new Surgeon General wants to build the image and the infrastructure of the Corps. None of us like the anonymity of the Corps. Hopefully, in the future you will see Admiral Murthy here. I know he wants to be here.

Given that the NLM began as the Surgeon General's library, Board member Dr. Esther Sternberg asked NLM Acting Director Betsy Humphreys if the Library should join forces with them to get their messages out.

Ms. Humphreys clarified that the NLM started in the Office of the Surgeon General of the Army, not the Surgeon General of the Public Health Service, but said that the NLM often partners with the U.S. Surgeon General to get messages out, e.g., featuring the Surgeon General in the *NIH Medline*Plus magazine.

III. FEBRUARY 2015 MINUTES AND FUTURE MEETINGS

The Regents approved without change the minutes from the February 10-11, 2015 meeting. The 2016 spring meeting will take place on May 3-4, 2016.

IV. REPORT FROM THE NLM ACTING DIRECTOR

Ms. Humphreys said this marks the first Board meeting in the Dr. Donald A.B. Lindberg Room. She said that the signs would be up the next time the Board meets. The request to rename the room in Dr. Lindberg's honor was enthusiastically approved by NIH Director Dr. Francis Collins.

With respect to the budget for 2015, for NLM's non-NCBI activities, we received a very small increase, about $700,000 over the FY 2014 level which was less than the amount required to pay for the modest salary increase. With respect to NCBI, NIH is continuing the practice begun in 2006 of transferring funds from other Institutes and Centers to allow NCBI to process the information that is coming in from various data generating projects funded by NIH and to assist in helping those that are submitting results data to ClinicalTrials.gov.

For 2016, there have some positive signs for the NIH budget, but it is far too early to say what the NIH will ultimately get.

With respect to personnel, NLM is under a partial hiring freeze pending the recruitment of a new NLM Director to ensure that the new Director will have some hiring flexibility. In the interim, the Acting Director is submitting requests to proceed with selected hiring actions to the NIH Deputy Director for Science, Outreach, and Policy. Permission has been granted in some instances.

Ms. Humphreys updated the Board on Dr. Lindberg's retirement, noting that in the days preceding his departure, he was honored by numerous friends of the Library, including this Board, the Office of Science and Technology Policy, publishers, CENDI, the Congress, among many others. On March 26, 2015, the NLM staff paid tribute to Dr. Lindberg and his wife, Mary, at an afternoon ice cream social reception in the Lister Hill Center lobby. Native partners from Alaska (Dr. Ted Mala), Hawaii (Mr. Tay Perry), and the lower 48 (Mr. Ralph Forquera) also spoke at this event, which featured a montage of video tributes from staff throughout the NLM and presentations of mementos to the Lindbergs from each NLM Division. Dr. Barbara Redman, President of the Friends of the National Library of Medicine (FNLM), announced that the FNLM had commissioned a joint portrait of Dr. and Mrs. Lindberg that will be hung at the Library.

The public was invited to "Fair Winds and Following Seas," a late afternoon event on March 30, 2015 in the NIH Natcher Center, followed by a reception sponsored by the FNLM in the Lister Hill Center Lobby. A program from the ceremony was included in individual Board Member books. Dr. Collins spoke at the retirement event and commented on the fabulous videos which highlighted Dr. Lindberg's career. For those not in attendance, Ms. Humphreys showed the opening video from that day's program, which included excerpts from the speech Dr. Lindberg gave at his swearing-in ceremony in October 1984. The event included an announcement that the FNLM and the AMIA would be cosponsoring a periodic lecture in honor of Dr. Lindberg and former FNLM Chairman Dr. Donald West King. Ex-officio Board member Dr. Cathy Nace presented Dr. Lindberg with the Order of Military Medical Merit, a very high honor rarely given to a civilian.

Ex-officio member Dr. Dale Smith informed Board Members that Dr. Lindberg would also be awarded the first honorary doctorate of Humane Letters ever awarded by Uniformed Services University at its upcoming commencement. .

To help guide the selection of Dr. Lindberg's successor, Dr. Collins has named a Working Group under the Advisory Committee to the NIH Director (ACD) to prepare a report on the needs of the Library going forward. The ACD received more than 600 responses from NLM users and

supporters regarding the NLM and their hopes for its future. Board Chair Dr. MacKay, who sits on the ACD, mentioned that the Board would be pleased with the final report of the Working Group which will be presented to the full ACD and the NIH Director on June 11 at a public meeting that will be videocast.

With respect to legislation, Ms. Humphreys said that the Congress has been introducing numerous bills to increase public access to results of federally funded research. The NIH statutory language goes back a number of years. Last year, legislation was enacted in the Labor-HHS Appropriations bill expanding this requirement to all agencies that Committee's purview. It would require relevant agencies with extramural research expenditures of more than $100 million per year to develop and implement policies for making articles resulting from agency-funded research freely available to the public no later than 12 months after publication. The public access requirements outlined by the President's Office of Science and Technology Policy apply to all federal agencies with great than $100 million in research expenditures annually and apply not only to publications but to data as well. It appears that the arguments against public access to government-funded publications and data have lost force.

Rep. Fred Upton (R-MI), chairman of the House Energy and Commerce Committee, released a discussion draft of the 21st century Cures Act to accelerate the discovery, development, and deployment of cures through a set of reforms aimed primarily at NIH and the Food and Drug Administration (FDA). It suggests an increase in NIH funding. NLM's immediate concern is language in the bill that talks about a level of standardization of eligibility criteria for clinical trials that sounds good, but may not actually be feasible or desirable. We have been suggesting clarifying language for this.

Not mentioned in the Board book, but interesting, is the Medicare Access and CHIP (Children's Health Insurance Program) Reauthorization Act of 2015. It will change the landscape in 2019 in terms of how eligible professionals might be rewarded or docked in their Medicare payments for meaningful use of their electronic health records.

Ms. Humphreys then showed the Board the draft Office of Management and Budget (OMB) guidance on the Federal Information Technology Acquisition Reform Act (FITARA), which is part of the Defense Reauthorization bill. NLM hopes that OMB and HHS implementation of this law will not be too problematic.

The next topic was the Notice of Proposed Rulemaking (NPRM) regarding clinical trials. The public comment period on the NPRM closed on March 23, 2015, and a total of 892 comments were received on the Federal policy and more than 200 on the draft NIH policy. Comments ranged from, "This is wonderful," to "This doesn't go far enough" to concerns about the burden associated with some of the data submission requirements. Some were concerned that the proposed rule did not cover more trials (beyond those meeting the definition of an "applicable clinical trial"), and others wanted to increase the types of data made available on ClinicalTrials.gov. The comments came from pharmaceutical and medical device companies, professional societies, biomedical journal editors, nongovernmental organizations and others. We are sorting through them now so NIH can consider them for the development of the final regulations. They will be developed in consultation with FDA and reviewed by HHS and OMB

before publication in the *Federal Register*.

Also, since the last Board meeting, a new funding opportunity for the National Network of Libraries of Medicine (NN/LM) was issued. NLM changed the award mechanism for the 2016-2021 from contracts to cooperative agreements. On April 2, 2015, the funding opportunity announcement for the cooperative agreements was published on the NIH Guide, titled, *Regional Medical Libraries for the National Network of Libraries of Medicine*, at http://grants.nih.gov/grants/guide/rfa-files/RFA-LM-15-003.html.

Ms. Humphreys reported that on March 17, 2015, the NLM held "A Tribute to Marshall Nirenberg," the first of a three collaborative NIH events that will celebrate the legacy of Nobel Laureate Dr. Marshall Nirenberg and the 50th anniversary of his deciphering of the genetic code. Future NIH events will be announced by the NIH Office of Intramural Research. This first program focused on NLM's recent acquisition of Marshall Nirenberg's Nobel Prize and certificate, through a contribution from his wife, Dr. Myrna Weissman, who gave the keynote presentation.

Betsy noted that the *Native Voices* exhibition would be closing at NLM on July 17. She encouraged Board members to visit the exhibition and urge others to do so before it closes. The traveling exhibition is currently making stops at each of the RMLs. NLM has reached an agreement with the American Library Association to tour the exhibition more broadly in 2016 and beyond.

V. MEDLINEPLUS RESPONSIVE INCREASES USE AND CUSTOMER SATISFACTION

Ms. Jen Jentsch reported that in October 2014 NLM released a responsively designed, one-column version of MedlinePlus for mobile users. She gave a brief overview of the MedlinePlus web site, noting that it first launched 16 years ago, with the Spanish version following shortly thereafter. MedlinePlus provides reliable, authoritative, consumer-friendly information, contains no advertising and is updated daily. It is in plain language with almost 1,000 health topic pages with summaries and links to other organizations and interactive content.

Ms. Jentsch provided some history of MedlinePlus for mobile devices. In 2010, a mobile-optimized MedlinePlus web site was a launched, seeking to allow access to users on any device. It had less content than the desktop site and different architecture. In 2013, after seeing increased traffic to the desktop version from smartphones, the MedlinePlus team decided to adopt responsive design to create an optimal viewing experience across most devices -- from desktop monitors to mobile phones -- with the goal of phasing out the separate mobile site and retaining one Web site.

With responsive design, the user has access to the same content but the presentation is tailored to the screen size of the user's device. She demonstrated responsive design using *The Boston Globe* website, showing how it looked on a desktop, laptop and a smart phone, and noting how the site content stays the same but is rearranged as display environment changes.

Ms. Jentsch noted that MedlinePlus is not the first site at NLM to use responsive design. Early

adopters were PHPartners, DailyMed, AIDS*info*, NLM Digital Collections, the Environmental Health Student Portal, and Health Hotlines. All have experienced significant increase in traffic since redesigning using responsive design.

Ms. Jentsch reported on the use and user feedback for the responsively designed, one-column version of MedlinePlus for mobile users, released in October 2014:

- Comparing the six months before and following the one-column mobile release, the number of visitors to the mobile site increased from 800,000 to about 99 million.

- The responsive MedlinePlus is more visible to Google than the former mobile site. In the month following its release, referrals from Google increased 29 percent, with mobile referrals growing from 34,000 to over 11 million in the 30 days after the release.

- There was also an increase in user satisfaction, reported by visitors who completed the Foresee ACSI survey. In just four months following the release, the overall satisfaction score grew by six points, from an already excellent score of 83 to an unprecedented high score of 89.

- Eighty-one percent of users said their experience on the new mobile MedlinePlus was "better" when asked, "How does your experience on this mobile site compare to your experience using other mobile sites?"

Board consultant Dr. Kenneth Walker asked what the most popular topics searched on MedlinePlus were. Ms. Jentsch said that disease information and diagnosis reports were most popular. [NOTE: The 100 specific terms most frequently searched during the previous week can be viewed in the search clouds at http://www.nlm.nih.gov/medlineplus/cloud.html and http://www.nlm.nih.gov/medlineplus/spanish/cloud.html] Dr. Sternberg wondered how Google works, for those who go there first for health information. Ms. Jentsch replied that there are things you can do to optimize your Web site so it appears on the first page of a google search. This is called Search Engine Optimization—to rank as high as possible in the Google search engine. Ms. Humphreys said that many talk about how users are going to Google, not to MedlinePlus. Google attempts to refer people to the most credible sources of information so often Google users end up at NLM.

Board member Ms. Gail Yokote asked whether the search experience is good for those who have visual problems. Ms. Jentsch replied that NLM is working on that issue as well.

A board member asked what happens if you are on a mobile phone and click on one of those sites that are *not* compatible with mobile apps? Ms. Jentsch explained that they will just take you to that Web site, which would need to be optimized as well to provide a good user experience.

Board consultant Dr. Tenley Albright said that responsive design was an exciting new way of searching MedlinePlus. Board member Eric Dishman asked whether the MedlinePlus Web site was also optimized for desktops. Yes, replied Ms. Jentsch. Mr. Dishman then asked if there

were issues with videos, in terms of lag time. Ms. Jentsch said that videos work fine. "What's next?" asked Mr. Dishman. Ms. Jentsch said that NLM will continue to work on the platform and to respond to criticism. She also indicated users accessing content from EHRs via MedinePlus Connect also get the optimization.

VI. PRESENTATION OF FRANK B. ROGERS AWARDS, NLM DIRECTOR'S AWARDS, AND OUTGOING BOR CERTIFICATES

Drs. David Fleming and Henry Lewis were unable to attend this meeting, Ms. Humphreys noted, but each will be sent a Board of Regents certificate. Outgoing Board chair Dr. Trudy MacKay was presented with a certificate and gavel for her service. Dr. Ralph Roskies also received a certificate, with thanks.

NLM Director's Awards were presented to Dwight Mowery, NLM's Chief Grants Management Officer, for improvements he made to the grants process at the NLM that have been adopted by other entities at the NIH, and Dimpal Patel, head of System Services Section of the Office of Computer and Communications Systems, for his leadership in implementing virtualization technology at the NLM.

The Frank B. Rogers Award was established by an anonymous gift from an NLM employee to recognize important contributions of NLM staff members to the Library's operations and services. This year's recipient was Mary Kate Dugan, head of the NLM Collection Unit in the Preservation and Collection Management Section within Library Operations. Ms. Dugan's difficult task has been to ensure that NLM collections were both well protected and available for use during a multi-year project to strengthen the floor of one of NLM's underground levels and install compact shelving to increase available collection storage space. This required planning and implementing multiple moves of virtually all components of NLM's physical collections.

Dr. Trudy MacKay then adjourned the Board, inviting them to lunch and a group photo.

VII. PRECISION MEDICINE AND CANCER RESEARCH

Dr. Douglas Lowy, Acting Director of NCI, and the Chief of the Laboratory of Cellular Oncology in NCI's Center for Cancer Research, discussed NCI's role in President Obama's recently unveiled Precision Medicine Initiative (PMI).

Dr. Lowy noted that mortally rates from cancer have declined—they are about 20% lower than 20 years ago. Today, there are many more cancer survivors in the United States than previously.

The Precision Medicine Initiative has four parts: (1) NCI expansion of the application of precision medicine in cancer, with an FY 2016 President's Budget request of $70 million; (2) NIH coordination of the establishment of a cohort of 1 million or more volunteers to help bring precision medicine to a wide range of diseases, with a, $130 million budget request; (3) increased FDA capacity to support the advance of precision medicine in regulatory issues will be increased for which $10 million is proposed; and (4) development of related interoperability standards, across systems, while protecting privacy, need to be developed. The Office of the National Coordinator for Health IT has lead responsibility for the standards aspect, for which $5

million is included in the President's Budget request.

According to the National Research Council's 2011 report, *Toward Precision Medicine; Building a Knowledge Network for Biomedical Research Towards and a Taxonomy of Disease*, "precision medicine refers to the tailoring of medical treatment to the individual characteristics of each patient…Preventive or therapeutic interventions can then be concentrated on those who will benefit, sparing expense and side effects for those who will not." This definition inadvertently appears to exclude population-wide public health recommendations, even those based on a mechanistic understanding of disease. A preferable definition is that precision medicine involves "interventions to prevent, diagnose, or treat a disease (e.g., cancer), based on a molecular and/or mechanistic understanding of the causes, pathogenesis, and/or pathology of the disease. Where the individual characteristics of the patient are sufficiently distinct, interventions can be concentrated on those who will benefit, sparing expense and side effects for those who will not."

NCI's part of the PMI is focused on cancer treatment. The goal is to expand our genomics-based clinical and preclinical studies. It has four parts. There is the genomics master protocol trials of adult and pediatric cancers. Also proposed is a repository of patient-derived preclinical models for evaluating targeted therapeutics to improve our predictive oncology. The long-term goal is to take tumors and predict how people are going to respond to what drugs—analogous to what's already done in infectious disease and antimicrobial agents.

The third part represents a very big problem —drug resistance, especially to the targeted treatment, which we need to understand and overcome with combination therapy. Fourth, the NCI database needs to integrate the genomic information with clinical response and outcome. This is building on our cloud pilot models and is clearly important for housing the information.

Genomic analysis of tumors through the Cancer Genome Atlas, which is a cooperative research endeavor of the NCI and the National Human Genome Research Institute (NHGRI), has identified simultaneously heterogeneity and vulnerabilities in tumors. There are important genetic and epigenetic differences between tumors even within the same tumor type. Some changes are common to several tumor types. There are "driver and "passenger" changes. (Targeted treatment typically aims to inhibit the "driver" changes.) The changes for which the NCI has experimental drugs, or FDA approved drugs, are so-called actionable changes—those molecular changes for which there are candidate or FDA-approved drugs.

Under the best of circumstances, we would have treatments that are able to target as high of proportion of potentially lethal cancers as possible. For example, cancers that carry mutations in a Ras gene, the most common oncogene in human cancer, tend to respond poorly to standard chemotherapy and carry with them a poor prognosis. It would therefore be extremely beneficial if there were effective drugs against these. Dr. Lowy said the public and private sector research communities have not been successful at developing effective inhibitors against cancers with mutant Ras. The NCI has a Ras initiative whose goal is just that. Efforts are underway to identify as many driver mutations as possible and to develop inhibitors within the public or private sector.

To test these in the Precision Medicine Initiative, NCI will change the way clinical trials done at the NCI. In typical cancer trials, you evaluate the impact of drugs on a particular cancer and organ type. By contrast, the MATCH (Molecular Analysis for Therapy Choice) trial follows a different paradigm based on targeted treatment of specific actionable molecular abnormalities in cancers independent of the organ site. This is a public/private partnership with more than 20 different pharmaceutical companies, testing a range of targeted treatments in a single trial. And the schema here is that you sequence a tumor and if you identify an actionable mutation for which we have a study agent, the patient then goes on it. If there is progressive disease, then you look for other actionable mutations. If yes, then they can go back on another treatment. If no, then they go off study. With stable or complete or partial response, they continue on the agent until there is progression. What we hope to find relatively quickly is those cancers for which there's some suggestion that specific treatments are going to be effective. Then we will quickly start a phase 2 trial within the context of the MATCH trial to see how the agent performs in larger number of patients with that particular cancer. It's flexible, but managed centrally. We have a central institutional review board (IRB) so it's just one IRB that gives the approval for the entire country. A trial arm can be added or deleted without affecting the other arms. We have more than 2,500 US sites.

Dr. Lowy said that a biopsy sample is received and within less than two weeks it is processed and evaluated to determine whether a patient is eligible for the trial and what arm of the ten-arm trial he or she should go on with. The Adult MATCH trial has been in the planning for more than a year, but it has not yet opened.

To summarize, Dr. Lowy said that precision medicine is a rational approach to cancer prevention, screening, and treatment. Genomic analysis has identified substantial molecular heterogeneity within the same tumor type, however the same molecular changes occur in several forms of cancer. The precision medicine initiative in oncology can use this understanding to improve treatment outcomes for patients. The Adult MATCH trial, the most advanced component of the precision medicine initiative, will target a wide range of actionable molecular abnormalities.

Mr. Dishman asked about the detail of the analysis and what data model would be used to scale precision medicine. Dr. Lowy said that they are looking at about 200 different genes that are pre-specified for which there is a CLIA (Clinical Laboratory Improvement Amendments of 1988)-approved protocol. They will be analyzed in one of three different US centers. He also stated that they are looking at the issue while sequencing a large number of lung adenocarcinoma, to see if we can identify genes that are mutated even at the 1 percent level. If there is an actionable change, then we can intervene and help a small number of people. What will be rate limiting is developing effective interventions to intervene in a larger number of patients who have a particular type of cancer.

Lister Hill Center Director Dr. Clem MacDonald observed that there has been a lot of work in recent years trying drugs on every cancer. The breakpoint blocker has been used on a couple things. But is this a different approach? Dr. Lowy said that both approaches are based on basic research. The immune checkpoint blockers are the outgrowth of research from cellular immunologists who are trying to understand the basis for the regulation of the immune system

particularly for antibody production. What is proposed here is very analogous except it targets a specific abnormality in cancers.

Board member Dr. Ralph Roskies noted that, in Dr. Lowy's formulation, the word "personalized" is left out. Dr. Lowy agreed, saying it was an error of omission. At the NCI, the distinction exists in part because Dr. Harold Varmus, its former Director, used to say is that what his father did was "personalized medicine," and what we are trying to do is "precision medicine." But many do characterize PMI as personalized medicine, too.

Board member Dr. Robert Greenes asked whether other NIH Institutes will focus on their specific diseases and do trials on those with the same model. Dr. Lowy said it is probably going to be a broader approach. He said that there is a Working Group of the Advisory Committee to the NIH Director that will make recommendations along those lines.

Dr. Greenes observed that a million person cohort or other large ones, like Apple's research kit, that enables its many users with smartphones and connected devices to become research participants, could obtain very individualized information about people beyond the genome. He asked whether broader questions like that were part of the initiative. Dr. Lowy said that the NIH-coordinated million+ cohort part of the PMI should take advantage of multiple components. One would be electronic health records (EHRs). Another would be smartphones and mobile health. A third would be environmental exposures and lifestyle. He said that your health risk is much more determined by your ZIP code than by your genetic code. The next PMI workshop will look at EHRs and the follow-up meetings will focus on populations with health disparities and mobile health. The composition of the cohort should represent the US population.

Dr. Greenes said that the data pipeline for different sources is important. Dr. Lowy said that there are at least three challenges in terms of the data. First, EHRs have not been used for research. How do we modify EHRs so that the data can be useful? Second, whatever the technology is today, the cohort has to last for some period of time. Third, some populations aren't connected to EHRs. How do you engage them?

Dr. Sternberg asked Dr. Lowy if he envisioned precision medicine as measuring the impact of the physical environment on health prevention. Dr. Lowy said that was the intention. National Institute of Environmental Health Sciences Director Dr. Linda Birnbaum has been an advocate for the need to evaluate environmental exposure. Dr. Lowy said that he is not in a position to say what the million person cohort will look like. But high blood pressure is clearly an enormous problem for which the cohort might be used as kind of a testbed to firm up what should be the standard of care.

VIII. SCREENING CHEST X-RAYS FOR TUBERCULOSIS IN RURAL AFRICA

Dr. Sameer Antani from the Lister Hill National Center for Biomedical Communications (LHNCBC) described how his team is helping to solve a global health problem by improving methods of screening people for tuberculosis (TB). This is critical because a person infected with both HIV and TB requires a different treatment regimen than someone HIV-positive who does not have TB. Dr. Antani and his team have developed a tool that is being used by an HIV

treatment program in rural Kenya that aims to screen the 2.5 million people for HIV and then to check those who are HIV positive for signs of TB. Global health is a key NIH objective.

This project involves a combination of LHNCBC research and software development that utilize advances in image processing and machine learning, which have been part of LHNCBC research portfolio for decades. LHNCBC researchers are developing machine-learning algorithms to automatically segment the lungs; detect and remove ribs, heart, aorta, and other structures from the images; and then detect texture features characteristic of abnormalities, which allows us to discriminate abnormal from normal cases. Developing these machine-learning algorithms, which allow computers to learn so they can do a task without being programmed to do it, requires large sets of example X-rays. LHNCBC acquired chest X-ray training sets from TB control programs and hospitals in the United States, India, China, and Japan. The LHNCBC's automated lung disease detection system performs at near 100% accuracy on these chest X-ray training sets. Research continues in lung segmentation and texture detection by techniques such as PHOG (pyramidal histogram of gradients) and others, with the objective of speeding up the detection performance. LHNCBC's lung-segmentation method can now segment lungs in tens of seconds rather than minutes. LHNCBC also created a valuable biomedical research resource by adding the chest X-ray data sets from China and Montgomery County to Open-i, an LHNCBC-developed open access biomedical image search engine (http://openi.nlm.nih.gov/).

Chest X-rays are the quickest, most cost-effective way to diagnose TB; however, challenges limiting its use as a screening tool in an isolated rural area include limited access to trained radiologists, poorly-maintained roads, and limited technology and electricity. LHNCBC is collaborating with an international non-governmental organization in Kenya, the Academic Model Providing Access to Healthcare (AMPATH), which runs an HIV treatment program in western Kenya that is based at Moi University. The international team designed and deployed a chest X-ray imaging truck that can travel to multiple locations in western Kenya to improve access. Each truck contains radiology equipment provided by funding from NLM. The x-ray generator is rugged, inexpensive and lightweight—only 65 kilograms. A cassette reader converts the X-ray image into a digital image. A radiologist sees and reads the images. The first chest x-ray truck was deployed in Fall 2014, makes about two trips per week and is seeing about 50-70 patients on each trip. The First Lady of Kenya, Margaret Kenyatta, came to inaugurate the truck and pronounced it "the coolest thing" she has ever seen.

This project was selected for HHS Ignite, an initiative of the Department of Health and Human Services (HHS) Innovation Council, and received a cash award that was used to help acquire field-deployable equipment. Dr. Antani thanked the members of the LHNCBC team, including post-doctoral fellows, visiting scientists, researchers and developers. Dr. Clem McDonald, director of LHNCBC, and Dr. George Thoma, director of LHNCBC's Communications Engineering Branch, have led the effort.

Board member Sandra Martin asked how many such trucks there are and what NLM's role is regarding them. There are two trucks, Dr. Antani replied, and one is in service already. NLM doesn't do anything with the trucks—that's left to AMPATH— but NLM provided the two x-ray units and a special box for x-ray power. Ms. Martin asked whether there were problems with wireless connectivity in remote places that could affect how the truck was deployed. There's poor connectivity in general across Kenya, Dr. Antani replied, but cell phone connection is very

good however. AMPATH has made agreements with the local telephone carriers to lend part of their towers to allow direct wireless connection.

Mr. Dishman asked about acceptance of the project by radiologists in the field who aren't directly involved. How do they view it? Actually, replied Dr. Antani, they wanted it badly because there is such a heavy workload.

Ms. Yokote asked whether LHC was capturing the images on their own, so they can refine and also study them. Yes, Dr. Antani responded. The images are being stored. They initially get the unrefined version, and then they are delivered to us in final, once the whole process with AMPATH is complete.

Dr. Greenes asked whether the goal right now is detection rather than diagnosis. Dr. Antani replied yes, the truck starts the process by pronouncing patients positive or negative. However, clinical officers and others on the ground can do a lot more. They have patient histories and other things we don't have. All we have is an image.

IX. THE FUTURE OF WHOLE-CELL MODELING

Dr. Markus Covert described a climactic moment in Tolstoy's *War and Peace*, in which Pierre, the main character, is facing a firing squad. He asks himself, what is killing me? It isn't the teenagers with guns pointed at him; instead, it is a system—what Tolstoy calls a "concurrence of circumstances"— things that are happening in Paris, in Moscow and in between. All of these are coming together, and that's what's killing him**.**

Dr. Covert suggested that patients with a disease must feel the same way. For example, researchers may discover a genetic factor in cancer or an Alzheimer's gene, but that is only one step toward a cure. Cancer, for example, is a complex interaction of hundreds or thousands of genetic, epigenetic, environmental factors, and the way that they come together. We'll never be able to address the critical problem of complex diseases, he said, until we learn to embrace their multifactorial complexity.

After reading a 2008 *New York Times* article on how Dr. Craig Venter had sequenced the bacterium *Mycoplasma genitalium*, Dr. Covert formulated an ambitious goal: to try and build and synthetically construct a model of a simple cell, to capture its complexity and how its many structures and functions work together. Since that article, a number of innovations have occurred in modeling and in collecting and making machine-readable data compilations. A lot has been happening on the experimental side, too—successes and failures suggesting how this whole-cell model could be created, using the vast knowledge that's being generated in the world.

Dr. Covert presented a schematic of what he and his team tried to do: to look at all of the genes and functionality in a cell, and then try to portion that out into modules. It was helpful to identify groups of genes that act together, like carbon energy metabolism genes, or the genes that are responsible for RNA decay. Next, they combed through about a thousand research papers, pulling out about 1,700 parameters on this simple cell. What they were left with was a collection of sub-modules that they tested and parameterized, each of which uses one method. They might

seem separate, he noted, but they all interact. You can run them serially or in parallel for one second of simulation time, studying their extensive output. Dr. Covert chose *Mycoplasma genitalium* because it's the smallest culturable organism and lends itself to experimentation.

He presented two stories: one of a complicated cell behavior that would have never been identified without a model, in a very complex phenotype; and another, in which the model can be used to test very specific predictions, experimentally. These two are instructive of the value of this computational modeling activity.

The team is currently focusing on making a computational model of *E. coli*, which has 10 times the genes and 100 times the molecules of mycoplasma. Dr. Covert said their first simulations took about 10 hours each to run, but, with technical optimizations, they've reduced that time to about five minutes in some cases. Some are hoping the whole-cell modeling could be done for mammalian cells. Live-cell imaging may be in the future, too.

Dr. Covert thanked his lab team and again thanked NLM and NIH for their support for what could have been seen as a total long shot project, giving them the resources and the encouragement to go in and do our best in this area.

What kind of computing do you use in order to do these simulations, asked Dr. Walker. Those original ten-hour simulations were performed on one Intel processor. Now, thankfully, people have seen their work and been inspired by it, so the lab has been invited to use computer resources in several places. That, combined with the fact that they have been very mindful of how to improve the runtime, has made it possible for that single processor to complete a simulation in five minutes.

NLM Extramural Programs Director Dr. Valerie Florance asked about Dr. Covert's slide showing whole-cell models and techniques. She said she was looking at those and thinking about the multi-disciplinary teams of people who do this. Dr. Covert said collaborations are happening that have never been seen before. For example, he recently co-hosted an NIH-sponsored event that brought together video game designers and scientists. Video game designers are experts in complexity and visualization, but have absolutely no idea what they are doing in biology. It was interesting to see them walk in the other's worlds, solving problems. He also had a friend who had received a $1 million bonus from Google for making Gmail profitable; the friend decided to take six months off to help Dr. Covert's lab do computer programming. He didn't know biology, but his work had a dramatic impact on the whole-cell modeling project.

X. EXTRAMURAL PROGRAMS REPORT

EP Director Dr. Valerie Florance reviewed grant outcomes for NLM ARRA (American Recovery and Reinvestment Act) grants awarded in 2009 and 2010. In 2012, she gave the Board a general report on grants given, amount spent, papers published, etc.

The purpose of the ARRA grants was to stimulate discovery, spur economic activity, and create and save jobs. Here at NLM, in that two-year period, we spent about $82 million in addition to our appropriated funds for those years. In addition to its own allocation, NLM received several million dollars of ARRA money from the Office of the NIH Director for awards in our portfolio

of interest.

NLM awarded several different types of grants with ARRA funds, including traditional research grants, small business grants and new ARRA-specific "Challenge" and "Opportunity" and Summer research experience grants. Overall, in the two-year period, NLM issued 128 new ARRA grants. There were also supplements and competitive revisions awarded to existing grants.

Before discussing outcomes, Dr. Florance provided some additional background on the variety of research grants funded during this period. NLM's ARRA Challenge grant topics were largely clinical and information science-oriented. Due to the short turn around for defining Challenge grant topics, NLM used topics that it had solicited through a Request for Information (RFI) several years before. Knowing that the ARRA-funded research was skewed in that direction, when new research grants were selected for awards using appropriated funds, emphasis was placed on bioinformatics and translational bioinformatics, for purposes of portfolio balance.

Dr. Florance next showed slides comparing the 50 ARRA-funded and appropriation-funded research grants (R01s) awarded during the period. Under non-ARRA R01's, people have the usual span of three or four years in which to complete their work. With ARRA grants, they had to compress it into two years, even though they were originally conceived as four-year projects.

Publication rates for ARRA and non-ARRA grants are nearly the same, with the 27 4-year awards producing an average of 20 publications per grant, compared to 10 publications per grant for the two-year awards. However, the citation rates were higher (average 14 per publication) for the ARRA 2-year grants than for the non-ARRA 4-year awards (average 7 citations per article) There are highly cited publications in each group. Dr. Florance noted that 14 R21 Exploratory Developmental grants were awarded, some in each category, but only 9 of those grants produced publications.

Overall, for all grants awarded in 2009-2010, many resulted in publications—858 total, with an average of about nine articles per award, with a range of 0 to 88 publications. She noted that when you see that someone has produced 88 publications in this period, you're reminded that there is a model in which PIs continue to cite grants after they've ended. There is probably a conceptual link between the work they did before, so they cite the grant that actually got the work started. Twenty percent of our ARRA awards and 10 percent of the non-ARRA regular grants had no publications at all. It's not necessarily a failure to have no publications, but it does make one want to explore further, Dr Florance observed. At a future Board meeting, Dr. Florance would like to bring examples from the community that didn't publish, asking questions like, did they go on to receive different grants? Did they produce some kind of resource but didn't bother to publish it and put it up on a Web site somewhere or made the data available? The people who had no publications were experienced investigators, so it did surprise her to see that nothing was published.

Dr. Sternberg said she had been an intramural research scientist who got one of those NIH Director's Challenge Grants. They published one paper and immediately got a call from the Air Force Research Laboratory, saying that they wanted to start working with us. So you have not

seen any more papers resulting from that grant, totally on purpose. They received several different contracts afterwards, from the military and from industry.

Something similar might have happened with NLM grantees with low or no publication rates. The National Science Foundation or some other entity might have supported their work afterwards, Dr. Florance observed.

Dr. Sternberg suggested there would be merit to having the Extramural Programs office go back to investigators who didn't publish, asking them what happened with their research. We can actually do that, said Ms. Humphreys.

A board member then suggested that internal faculty evaluations would be a better measurement of researchers and their impact. Look at the paperwork they have to file for their promotions and you'll get a good sense of their influence.

IX. HAZARDOUS SUBSTANCES DATABANK

Ms. Shannon Jordan of the Division of Specialized Information Services (SIS) discussed the Hazardous Substances DataBank (HSDB), created in the early 1970s as a result of Environmental Protection Agency (EPA) Superfund legislation and an interagency agreement between NLM, the Department of Health and Human Services (HHS), and the Agency for Toxic Substances and Disease Registry (ATSDR).

HSDB includes information on potentially hazardous chemicals such as drugs, industrial reagents, consumer product ingredients, nanomaterials, venoms, and more. The database currently contains over, 5,700 substances, with over 150 data fields covering human exposure, industrial hygiene, pharmacology, and more.

Ms. Jordan explained the HSDB content development process. On a four-month cycle, staff members select about 100 chemicals that are brand new or in need of updating. HSDB staff then conducts literature searches using PubMed, TOXLINE, and other relevant databases and drafts the content of a new or updated HSDB record. An SIS senior toxicology advisor performs a final review and then the entire package goes to 16 experts on a scientific review panel.

Toxicity and environmental impact are among the key considerations in adding substances to HSDB. The 2014 Elk River chemical spill in West Virginia involved SIS staff in responding to a disaster as well as a news media/public information situation. The HSDB panel analyzed data on a chemical that was not well known, compiling and analyzing technical information to allow timely decision making. That spill involved extensive collaboration among federal agencies, including the Centers for Disease Control and Prevention, the National Institute of Environmental Health Sciences, the National Toxicology Program, and NLM.

Ms. Jordan then introduced Dr. Lynn Goldman, dean of the Milken Institute School of Public Health at George Washington University. Dr. Goldman described a movement among Members of Congress to legislate that raw data from every study that EPA uses in a regulation are available to the public. However, the matter is not simple. There are many types of

environmental studies. Some are quite old, but still have bearing on present situations. Often, no human studies can be done in connection with the regulation of chemicals, because harm might be done to people. Also, much of the data on chemical and environmental health are difficult to locate. Dr. Goldman supports the concept of getting whatever data we have out there to the people, promoting systematic reviews, and making it easier for people to update systematic reviews when new studies are done. .That would be a positive thing in terms of transparency, showing how agencies—in this case EPA—make regulatory decisions.

She described the Toxic Substances Control Act (TSCA) of 1976, which gives EPA the authority to require reporting, record-keeping and testing requirements, and restrictions relating to chemical substances and/or mixtures. An estimated 65,000 chemicals that existed when TSCA was adopted were grandfathered onto the market—basically, the industry came in and said we're registering this chemical, even if it was not on the market, but just existed.

TSCA definitely needs to be changed, to open up that data to the public. Too often, "confidential business information" (CBI) has been claimed as a reason not to file reports of effects, so EPA collects adverse effects in another database in which a lot of information has been redacted, such as the compound, the name, etc. If the pending EPA bill goes through, it would strip CBI protection from hundreds of thousands of data items that are in that vault. A lot of data would then become available.

At NIEHS's National Toxicology Program, there is a movement toward adopting principles of systematic review of the chemicals. That's another opportunity for collaboration for NIH. These kinds of questions and concerns lead to the fact that an organization like NLM, experienced in data collection and analysis, might be the logical seat for data on chemicals, rather than placing that burden on the individual investigator. Some of us in the field aren't going to be very good at doing that.

Board consultant Dr. Holly Buchanan asked how many substances *should* be in the HSDB. A realistic estimate, said Dr. Goldman, is 20,000-25,000. Actually, if you look at how many have ever been registered by EPA and FDA, the number jumps to 85,000.

Dr. Walker asked whether there are international chemicals that we have no recordings of because they are not used in the US. Also, what about all of the degradation products that go into our sewer and water systems? Dr. Goldman replied that, believe it or not, scientific references and other articles on Wikipedia have improved through the years and are a pretty reliable source on these matters.

XII. KEY QUESTIONS IN THE ERA OF BIG DATA & DATA-CENTRIC SCIENCE

Dr. Michael Huerta, NLM Associate Director for Health Information Programs Development, noted that the previous presentation was the perfect introduction to this topic. Some domains of biomedicine, like genetics and genomics, have a long tradition of sharing data. The major public products of these domains are not only scientific papers that interpret the data, but also the underlying data themselves. However, other biomedical research domains, which comprise the

majority of research supported by NIH, are not data-centric. Papers will be published about the data, but the underlying data are never seen beyond the labs that collect or produce them.

Now, new forces are in play: societal expectations (prompted in part by the social media culture, in which everything is shared), policy directives from the Executive Branch and Congress that urge greater access to research results, increasingly sophisticated technical capabilities and scientific opportunities emerging from the digital nature of modern data are converging to assure that data from all biomedical research domains—including those not currently data-centric—will be made broadly available very soon.

We now have an imminent flood of data for which there is little or no data infrastructure. What can be done to make the most of it? And, what is an appropriate role for NLM, especially in light of the increasing NIH-wide interest in data-related issues?

If all domains of biomedical research will be making data available, a fundamental question is, how much data will this be? Kevin Read, a former NLM Associate Fellow now at New York University, worked with other NLM staff to analyze the 2011 NIH-supported articles published in 2011 and found that the research projects reported involved over 200,000 datasets. Of these, only 12 percent were deposited in recognized repositories, meaning that the vast majority of datasets would have been hard to discover and were essentially invisible.

What is a dataset? In the same study, 385 papers were scrutinized by 30 NIH staff members. Each was looked at by two staff members, who were asked how many datasets the paper contained. The numbers varied widely between the two staff members looking at the same paper. Journals like *Scientific Data* and others will probably help to clarify this as time goes on.

Another question is what is the current amount of investment in data infrastructure, specifically data repositories? Dr. Tony Chu of NLM/OHIPD and Dr. Susan Gregurick, Director of the Division of Biomedical Technology, Bioinformatics, and Computational Biology at the National Institute of General Medical Sciences, did separate analyses of this question and found that NIH invests at least $117 million in data repositories annually. As far as who owns and operates the current data infrastructure, the vast majority of this figure pertains to extramurally-run repositories and reflects only a few internal NIH repositories. Another question, of course, is what other questions should we be asking about the current landscape? And what role might NLM play?

Everyone knows that all data are not created equal and, given that, which data should be made available “as is”, which should be indexed, which should be deposited in a repository and curated, and which should be preserved for the long term? The cost rises dramatically if we go from simply posting the data on a Web site versus preserving it long-term; we have to ask which data should be treated which way.

Another important decision to be made is when, where and how datasets should be cited. In November of 2014, the International Committee of Medical Journal Editors decided to consistently list at the end of abstracts: the name of the dataset; the name of the data repository in which it resides; and the dataset unique identifier. NLM is currently developing ways to cite

these items and data journals.

What role might NLM play in supporting and advancing data sharing? The flood of data from non-data-centric domains is a threat, but it is also an opportunity for discovery. If we're looking at that flood as an opportunity to change the paradigm of all biomedical research so that data-centricity is the norm, we can reap the benefits that we've already seen in genomics and genetics. Of course, incentivizing would speed this change; these incentives would best be considered systematically across the biomedical enterprise, not one component of the enterprise at a time. Such components that could each offer incentives include (but are not limited to) funders, research institutions, patient advocacy groups, and professional societies.

So, how might NLM contribute to understanding the current landscape, identifying key decisions to be made, examining how those decisions might best be made, and how we might facilitate emerging opportunities, especially in light of the increased interest in data across the NIH? Dr. Huerta invited the Board's views.

Dr. Roskies expressed fear that, with so many questions, people might throw up their hands up and say, "This is too much." In that case, nothing would get done. Which immediate questions NLM should pay attention to? Dr. Huerta pointed to "what is manageable to study," then suggested reaching out to other organizations to look more closely. Stressing importance of answering these key questions to groups outside the NIH campus would be useful.

Ms. Humphreys thought a good starting point would be to figure out how big the "big data" is at NIH. Without a clear understanding of the size of the data, we could be off by a factor of 50 or 5000. We may not be able to organize it all, but if we can identify five types of data that we do *really* care about, we can focus on those five.

Ms. Yokote asked how the shift toward data sharing could be encouraged among academic researchers. Decisions would have to be made—am I going to share this and, if so, with whom, do I own this, does the university own this? We can have conversations with our own colleagues, encouraging some of them at academic institutions funded by NIH to answer some of these questions. This might constitute a core group that would be willing to tackle this as a pilot.

Dr. Huerta said that NIH plans to require data management and sharing plans of everybody asking for research money from NIH, regardless of the size of the budget or the mechanism of the award. He added that NLM Assistant Director for Policy Development Jerry Sheehan is leading a group under BD2K (Big Data to Knowledge) that is looking at these issues for clinical and clinical research data.

Because this challenge is so daunting, Dr. Roskies suggested that first steps be taken, examining data not related to humans. This approach offers a better chance to make progress adopting a data management plan, and clinical data can be subsequently added.

XIII. REPORT FROM THE NOMINATING COMMITTEE FOR THE NEXT BOR CHAIR

Ex-officio alternate member Ms. Kathryn Mendenhall gave the report from the nominating committee for the next BOR chair. They are pleased to place in nomination the name of Gail Yokote, associate librarian at the University of California, Davis. Dr. MacKay called for a vote and the nomination was unanimously approved. She congratulated the incoming chair.

XIV. PICKING UP THE PIECES FROM THE GENOME EXPLOSION

Dr. Rodney Brister, staff scientist with NCBI's Information Engineering Branch, said he would talk about viruses. They come in all shapes and sizes, can be double stranded, single stranded, etc. What all viruses have in common is their dependence on a host cell for the replication and production of progeny.

You could say that our world is coated in viruses, with estimates of 10 to the 30^{th} power viruses existing on the planet today. They contribute to a large number of human fatalities each year and can have a devastating economic impact. Given the havoc they can wreak, you would expect there have been a lot of viral sequencing efforts in the last decade or so, and this indeed has been the case. A virus genome sequencing explosion means that there are currently nearly 2 million virus sequences in GenBank.

In the era of Next Generation Sequencing, transforming these data into usable form includes several steps, that is, the assembly of short reads into contigs, identification of the contigs, the alignment, and alignment to reference genomes. All of this has to come together to get that final sequence that you find in GenBank or other sequence databases. All of this is dependent on referenced genomes, and that is where Dr. Brister's group comes in; their job is to help facilitate the transformation of sequences and data into the usable things that you find in the world's public databases.

As part of this work, they provide reference infrastructure for the identification and assembly and annotation of viral sequences. How do we do this? First we go out into the public databases, such as GenBank, and aggregate viral sequence data. Next, we divide it into groups and validate viral genomes in those groups. For viruses, taxonomy offers a pretty effective framework for the grouping of sequences; at NCBI, we use the taxonomy provided by the International Committee on Taxonomy of Viruses (ICTV). They currently recognize over 3,000 viral species. We then validate each genome on the basis of that taxonomy and the rules of taxonomy they provide.

Despite their best efforts, the ICTV can't keep up with the proliferation of information, which means NCBI is often left organizing things on the fly, using the rules ICTV developed and then applying them to situations they haven't considered yet. Not only are we trying to create reference sequences, but we are also trying to ensure that these reference sequences contain the best annotation possible. To do this, we have reached out to various stakeholders for sequence data, including sequencing centers and individual scientists, and tried to aggregate their knowledge.

In addition to aggregating all of our genome sequences and validating them, NCBI is also interested in adding value to these sequences. Basically, what we're doing is assigning a host type to each species that has a reference genome, so if something infects a human, horse, or mosquito, we aggregate this kind of information and allow users to search by the information that's relevant to those specific studies. Viruses come from all over the place not just humans, but also other animals, plus bacteria and plants. Where can you find all of this information? If you search "viral genomes" in Google, it will bring you to this page, http://www.ncbi.nlm.nih.gov/genome/browse/. We have aggregated a number of resources including a browser we make available which includes every single validated viral genome as well as some taxonomy and host information.

We have created several value-added resources, including our Retrorvirus resource page (http://www.ncbi.nlm.nih.gov/genome/viruses/retroviruses/), which includes an HIV human interaction database, a collaboration of the Southern Research Institute and Dr. Kim Pruitt's group at NCBI. This is a great example of outreach, and how public resources can come together in some common interest. Within the community here, they are aggregating all known, published HIV and human protein interactions, and we are organizing these data and making them available to the public.

NCBI has developed some approaches that deal with variable and out-of-date information, inconsistent feature annotation, and protein, gene, and source vocabulary. We have a general idea that if we take raw data data from the publicly available databases and apply computational pipelines to it, we can create better data. .We are lucky enough to have an innovation platform called Virus Variation Resource (ViV), developed by staff who have been working on some of these topics. We are using references to standardize annotation so that every gene has the same name. We standardize metadata vocabulary, too, so that, rather than "pig" being "porcine" or "swine," users can easily search by one word and find what they are looking for.

This project is influenced by the Influenza Resource, and as we began to learn more about what we were doing, we wanted to make this a compact structure so that we could quickly create new resources. When there is an ongoing viral outbreak and suddenly people start sequencing—can we respond to that in a timely way? Can we put together a resource that serves the community in just a few weeks, when people are putting together sequences? Just as we were figuring out what kind of process would serve us best and what would make it possible, along came the Ebolavirus outbreak. Within a matter of a week, we were able to construct an Ebolavirus-specific database. It literally was in place after the first Ebola sequences from the last outbreak were in the database.

As we anticipate the world of tomorrow including many, many more sequences, these efforts to standardize things and make the data more usable become increasingly important.

Ex-officio member Mr. Christopher Cole asked whether NCBI had other partners in this effort. If NCBI has been standardizing and upgrading its metadata, how does that interact with that of other groups? Are they moving on the same path or are you going to have multiple search mechanisms to look at all of the databases? Dr. Brister noted that there are stakeholders of different varieties. You have the primary archival databases and then you have the value-added

databases that are supported by NIAID. NCBI identified the stakeholders very early on and some of their policy publications include those stakeholders on them. So we try to tackle these problems as a group. There are other efforts out there that have matched up with our efforts for outreach.

Dr. MacKay said she noticed that in one of the slides that environment was almost a negligible source of viruses and mutations? Dr. Brister replied that this approach represents how they've been changing the data model. Initially they wanted isolated things, studied in the laboratory, but the truth is that that just doesn't happen in viruses to the same extent that it used to. They're becoming adaptable; part of NCBI's job is to represent what is out there and part aggregating really well annotated references—documented with experimental evidence. The idea of indexing sequences as "environmental" helps separate them and is often different from the misleading isolation sources included on that record.

Ms. Humphreys asked whether some of this effort related to foodborne pathogens. Dr. Brister said that he has had some interaction with the FDA; to date most of their most of their work in this vein has been focused on bacteria, but he sees a vital role for his virus group. Their goals are simply to create a prototype of a variety of pipelines, infrastructure and, collaborations, and that effort is doable. The use cases that have come up are different; there may be foodborne pathogen use cases, but thus far our use cases have been about contamination and biologicals.

XV. ADJOURNMENT

Dr. MacKay adjourned the Board of Regents meeting at 11:45 a.m. on May 13, 2015.

ACTIONS TAKEN BY THE BOARD OF REGENTS:

- Approval of the February 10-11, 2015 Board Minutes
- Approval of the May 3-4, 2016 Future Meeting Dates
- Selection and Approval of new Board of Regents Chair, Ms. Gail Yokote

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	Trudy MacKay, Ph.D.
Acting Director, National Library of Medicine	Chair, NLM Board of Regents