



**Testimony
Before the
Subcommittee on Courts, the Internet, and
Intellectual Property
Committee on the Judiciary
United States House of Representatives**

**The Public Access Policy of the
National Institutes of Health**

Statement of

Elias A. Zerhouni, M.D.

Director

National Institutes of Health

U.S. Department of Health and Human Services



**For Release on Delivery
Expected at 1:00 p.m.
Thursday, September 11, 2008**

Mr. Chairman and Members of the Subcommittee, I have been privileged to be the Director of the National Institutes of Health (NIH) for the past six years. To serve at this particular moment is a blessing, for this is truly the golden age of medical research. We know more about human biology than at any point in history. Scientists are accumulating new information at a staggering rate, and I am witness to an unprecedented explosion of knowledge.

There have been times I was informed of more discoveries in three months in such areas of research as genomics than I had in the previous five years combined – and the rapid pace continues today. These advances have illuminated previously hidden areas of the life sciences, including new and significant discoveries regarding the cellular underpinnings of disease. Our new knowledge of genes, proteins, and molecules is leading us to new areas for exploration in biomedical research.

Such progress is largely attributable to revolutionary advances in both high-throughput biology and information technology. New high-throughput technologies are resulting in exponential increases of biological data in amounts previously unattainable. New information technologies are allowing us to store, integrate, analyze and make these data accessible like never before. We are gaining an unprecedented understanding of biology, health, and disease. In this age of the

internet, the ability to share such information from one end of the globe to the other in the blink of an eye is increasing the pace and breadth of medical research. Every single week, scientists and the general public are downloading more information from NIH's databases and web-based archives of publications than exists in the entire Library of Congress. Scientists are not the only beneficiary of publicly accessible information. Students training to become the next generation of medical researchers are accessing NIH's databases. Surveys indicate that more than 60 percent of American patients consult internet medical sites prior to seeing their physicians, and they would benefit from access to the most complete and unbiased information available.

The extraordinary progress of recent years has positioned us to change the dynamics of medical treatment. In the near future, we will no longer be responding only to the acute symptoms of disease. Research advances on the horizon will enable us to identify biomarkers of illness and, in many cases, preempt disease before symptoms appear. The ability to accelerate research through innovations in information technology is leading us along this path to a new era of medicine.

Science has benefited from two revolutions. The first revolution stems from these new technologies that enable data to be generated at unprecedented rates and at dramatically reduced

costs. For example, in just its first few months, NIH's 1,000 Genome Project generated 240 billion bases of genetic information. Those data are being deposited in NIH's National Center of Biotechnology Information (NCBI) and other databases for the benefit of all scientists and the public at large.

Not long ago it was a challenge for a researcher to study the regulation of a single gene in a human cancer cell, while now it is routine for cancer researchers to measure the expression of thousands of genes and make these data available in NIH's public databases to assist discoveries by other scientists around the world.

The second revolution emerged from our ability to manage and integrate these enormous quantities of data being produced and to make them available in ways to speed research that did not exist even ten years ago. We are now capable of taking individual discoveries and integrating them with all other research findings – both publications and data. Scientists can connect the dots between discoveries instantly, an advance analogous to moving from searching for fingerprint matches manually to matching prints in a database of millions in an instant.

When viewing a report in NIH PubMed and PubMed Central databases, at the touch of a button we can link to papers that are determined to be related, as well as to papers that were actually cited. We can also link to related chemical structures, proteins, viruses, and other data, allowing us to make discoveries that advance science and even prevent deaths.

The biotechnology and IT revolutions led NIH to establish NCBI in 1988. Today, NCBI is brimming with molecular and genetic information in more than 40 free and internet-accessible databases. More than 2 million people a day are accessing these databases, seeking information to understand disease and advance research. The majority of these databases are integrated, allowing, for example, a researcher to instantaneously link from a study on a drug compound to a 3-dimensional view of the compound and then to genetic data on a gene thought to be related to the disease being studied. The linkages are copious, and this extensive integration is the great power behind these databases that drives discoveries.

The NCBI databases are critical tools for the discovery of gene function and the identification and cures for many diseases. For example, about three years ago, a child was hospitalized with an undiagnosed illness in Minnesota. The state health laboratory had isolated an unknown virus. After determining the DNA code of the virus, laboratory staff used the internet to access the 55

million DNA sequences at NCBI and immediately found a match. The virus turned out to be the first polio case in the United States since 1999.

Following the Hurricane Katrina disaster in New Orleans, local officials were unable to identify thousands of bodies because of their poor condition. NIH responded with software that analyzed 10,000 DNA samples in two minutes, as compared to the full day of work required by an analyst to examine 14 samples by hand.

The biology and IT revolutions have enabled NIH to launch genome-wide association studies to identify genetic variations that are common with various diseases. Such studies have identified multiple genetic variants common to type 2 diabetes, information that will be vital as we seek to curtail this epidemic. Through a relatively new NCBI database called dbGaP, the data from these NIH genome-wide association studies are being made available to researchers across the world, in order to accelerate the discovery of cures and prevention strategies.

Recently, NIH's data bases were used to identify a virus that had caused the mass death of honeybees in the United States. Scientists scanned the DNA code against all known viruses and pathogens and linked it to a new virus known as the Israeli acute paralytic virus.

With these new life-saving tools, the main limitation on their use is the capacity to store and retrieve the data, given the extent to which data is being submitted. While today we are storing and retrieving only a fraction of the data and findings that could be available, the mandatory public access policy enacted last year will increase the scale of information that will be available from the library. Under the law, scientists who receive taxpayer dollars to conduct research will post their findings in PubMed Central, a public archival database at NIH.

From May 2005 to December 31, 2007, 14,397 research-articles supported by NIH – out of a total 189,000 – were made publicly available through PubMed Central through a voluntary policy. Since the establishment of the mandatory policy, well over half of NIH-funded articles are being submitted to PubMed Central, and the percentage is growing every day. During this early period of policy implementation: 400,000 users are accessing 700,000 articles every day.

Congress applied the mandatory public access policy to manuscripts resulting from NIH-supported research. The policy has two basic premises: 1) the integration and accessibility of biomedical research will speed discoveries, resulting in the prevention of death and disability;

and 2) the public has a right to have full access, without charge, to research findings supported by taxpayer dollars, after a reasonable period of embargo.

The House Committee on Appropriations first expressed concerns about lack of access to NIH-supported research reports and data in July 2003. A year later, that Committee recommended that NIH develop a mandatory public access policy, with reports becoming available six months after publication.

NIH responded with caution. The Agency proposed a voluntary public access policy in September 2004 and published it for public comment. After public debate and comment, NIH started a voluntary policy, with a 12-month embargo period, in May 2005. As part of the Consolidated Appropriations Act for FY 2008, Congress enacted mandatory deposition in PubMed Central of published manuscripts from NIH-supported research. Throughout this process, continuing to this very day, NIH is engaged in public discussions about the mandatory policy, and is being responsive to concerns about implementation.

NIH began a formal process to engage its stakeholders in enhancing the effectiveness of the NIH Public Access Policy implementation. NIH held an open meeting on the Public Access Policy on

March 20, 2008, and conducted a Request for Information (RFI) from March 31 to May 31, 2008. NIH is considering all the comments and suggestions it received from the RFI. Among other issues, the NIH was particularly interested in information about the following:

- Do you have recommendations for alternative implementation approaches to those already reflected in the NIH Public Access Policy?
- In light of the change in law that makes NIH's Public Access Policy mandatory, do you have recommendations for monitoring and ensuring compliance with the NIH Public Access Policy?
- In addition to the information already posted at <http://publicaccess.nih.gov/communications.htm>, what additional information, training or communications related to the NIH Public Access Policy would be helpful to you?

The NIH is in the process of analyzing all submissions collected through this RFI, along with comments collected before and during the March 20 meeting, and will report its analysis by September 30, 2008.

We understand that a bill has been introduced on this matter. The Administration is reviewing this bill and will get back to you with our views on it.

Thank you for the opportunity to present this information to you. I would be happy to answer any questions you may have.