2009 Atlanta Conference on Science, Technology and Innovation Policy

The Role of Information Technology in Medical Research

Daniel Castro

October 2009

The Role of Information Technology in Medical Research

Daniel Castro
The Information Technology and Innovation Foundation
dcastro@itif.org

***Abstract*-Using information technology (IT) to modernize our health care system will lead to improvements in medical research. Health informatics will allow medical researchers to determine the effectiveness of a particular treatment for a given population or to discover the harmful side-effects of a drug. While some of this research will occur in the private sector, public investment in this area will play a major role. This report finds that both the United States and the United Kingdom commit roughly the same percentage of total public medical research funds to health informatics. However, the United Kingdom is uniquely positioned to benefit from advancements in health informatics research because it is significantly ahead of the United States in its transition to electronic health records among primary care providers. More importantly, the National Health Service (NHS) has made an important strategic decision to emphasize medical research as one of its core missions. Thus, as the NHS continues to develop its IT infrastructure, it will be able to make technical upgrades and policy changes to improve information sharing and its information base for research. The United States currently lacks the capacity being developed by the NHS to turn its existing or future electronic health records into a usable database for medical research. To benefit from the full potential of health informatics, the United States should develop the capability to share medical data for authorized research in a timely and efficient manner.**

I. Introduction

Many developed countries have announced initiatives to modernize their health care systems with investments in health information technology (IT). The goal of these initiatives is to use technology to improve the health care system by reducing costs, increasing patient safety and improving quality of care. Improving health care is a common goal for these countries, but there are wide disparities in the success with which nations have pursued this goal [1].

In particular, countries such as the United States have lagged behind some European nations in the adoption of health IT, such as electronic health records. Interoperable electronic health records are a prerequisite for a modern health care system and the key to delivering a number of benefits to health care patients and payers. For example, the computerized decision support systems used in hospitals provide patients the most benefit when they use a complete and accurate set of patient data. These systems can help ensure a return to the core principle of evidence-based medicine—that patients and doctors have the best evidence available when making a decision about treatment.

While much attention has been paid to the degree to which nations have made progress with investment in health IT, less attention has been paid to the level of investment in health IT research. Yet evidence-based medicine relies on high quality medical research. Moreover, as we enter an increasingly digital world, the amount of health data that will be available to medical researchers will be increasing substantially. While past medical researchers had only a few limited data points recorded on paper on which to base their hypotheses, in the future researchers will have massive online databases containing terabytes of data for their analysis.

Some of the major benefits from modernizing our health care system are expected to come from the improvements in medical research that it will enable. For example, medical researchers will be able to use rapid-learning health networks to determine the effectiveness of a particular treatment for a certain population or to discover harmful side-effects of a drug. While some of this research will occur in the private sector, for example through private pharmaceutical research, public investment in this area will also be important.

Already a variety of projects offer a glimpse into the possibilities that IT will allow for future medical research. But achieving this vision will require substantial leadership and effort on the part of nations to overcome the technical and social hurdles ahead.

Some of the questions this paper will look at are as follows: How are the United States and the United Kingdom integrating health informatics into their overall commitment to improving health care? To what degree are these nations investing in the technology that will provide the platform for this research? How have national research institutes addressed medical research as not simply a domestic issue, but as an international challenge that must be answered with international partnerships?

This paper will look at the degree to which the United States and the United Kingdom are pursuing data intensive,

IT-based medical research. The paper will review public programs and efforts in this field in each country. In addition, the report will quantify public investment in these programs—both past investment and projected investments. Finally, the report will make a qualitative assessment of the effectiveness of policies and initiatives in each country to advance this type of research.

II. Background

A. Informatics in Health Care

Health care is becoming an increasingly data-intensive field as doctors and researchers generate gigabytes of medical data on patients and their illnesses. While a patient visiting the doctor 20 years ago may have only generated a few data points—basic information such as weight, blood pressure, and symptoms—a medical encounter today may leave a long trail of digital data from the use of high-definition medical imaging to implantable or wearable medical devices such as heart monitors. More importantly, as doctors and hospitals transition away from paper medical records, this data is increasingly being collected and made available in an electronic format. The availability of large data sets of digital medical information has made possible the use of informatics to improve health care and medical research. Often referred to as "in silico" research, informatics offers a new pathway for medical discovery and investigation. Informatics focuses on developing new and better ways of using technology to process information. Today, informatics is being applied at every stage of health care from basic research to care delivery and includes many specializations such as bioinformatics, medical informatics, and biomedical informatics.

The field of bioinformatics has exploded within the past decade to keep pace with advancements in molecular biology and genomics research. Researchers use bioinformatics to gain a better understanding of complex biological processes by, for example, analyzing DNA sequences or modeling protein structures. The most famous example of this is the Human Genome Project which relied on informatics to correctly analyze and sequence the 3 billion chemical base pairs that make up human DNA [1]. Much progress in basic research has been made possible by advancements in information technology, including the computing power, storage technology and software algorithms needed to collect, store and analyze the large data sets involved in genetic research.

Informatics has also had a major impact on the field of systems biology. Systems biology uses computer modeling and mathematical simulations to predict how complex biological systems will behave. For example, researchers have created models to simulate tumor growths. Through the application of computer models researchers can gain a better and more comprehensive understanding of how diseases affect an entire biological system in addition to the effects on individual components [2].

Medical informatics, or clinical informatics, focuses on using information processing to improve health care delivery. It covers various applications including using information technology within the clinical setting for medical billing, patient and resource scheduling, and patient care. An example of medical informatics is the use of clinical decision support systems (CDSS) which provide feedback and instruction to health care workers at the point of care. Such a system may, for example, provide warnings of potential drug interactions to a prescribing doctor based on a patient's existing medical history and known allergies. By integrating patient information with clinical guidelines, health care providers can help reduce medical errors. Adverse drug events alone account for an estimated 19 percent of injuries in hospitalized patients in the United States and cost hospitals over $2 billion per year, excluding medical malpractice expenses [5].

Biomedical informatics is a unique discipline that bridges multiple fields including medical research, clinical care and informatics. At its core, the objective of biomedical informatics is to develop new tools and technology to better collect, display, retrieve and analyze biomedical data. Such research can lead to new treatments, diagnostic tests, personalized medicine and better understanding of illnesses.

B. Benefits of Health Informatics

Bringing together large data sets of medical data and tools to analyze this data offers the potential to expand the research capabilities of doctors and scientists. Medical researchers can use this vast source of biological and clinical data to discover new treatments and better understand illnesses. Pharmaceutical companies can use the biomedical data to create drugs targeted at specific populations. Health care providers can use the data to better inform their treatments and diagnoses.

Applying informatics to health care creates the possibility of enabling "rapid learning" health applications to aid in biomedical research, effectiveness research and drug safety studies [7]. For example, using this technology, the side-effects from drugs newly introduced to the market can be monitored in real-time, and problems, such as those found with the recently withdrawn prescription drug Vioxx, can be identified more quickly. Moreover, the risks and benefits of drugs can be studied for specific populations yielding more effective and safer treatment regimens for patients.

As Reference [7] has noted, using rapid learning techniques can not only improve patient safety, it can also lead to substantial improvements in the quality and cost of care. By turn-

ing all of this raw digital data into knowledge, these rapid learning health networks can enable doctors to better practice evidence-based medicine. Evidence-based medicine is the use of treatments judged to be the best practice for a certain population on the basis of scientific evidence of expected benefits and risks. Cost savings in health care is a growing priority in both the United States and the United Kingdom as their health care costs continue to rise and populations get older. By using rapid learning networks, health care workers can identify not only the most effective treatments, but also the most cost-effective treatments given a patient's specific medical profile.

C. Building the Digital Platform for Medical Research

Achieving this vision of an intelligent and fully-connected health care research infrastructure has not yet been realized. While various pilot projects have shown success and have demonstrated the potential benefits that can emerge from a ubiquitous deployment of informatics in health research, many technical obstacles still need to be overcome. These obstacles include making data accessible, connecting existing data sources, and building better tools to analyze medical data and draw meaningful conclusions.

Much medical research data is not accessible electronically. For example, one challenge for the United States and the United Kingdom are the low rates of adoption of electronic health records among primary care providers and in hospitals. Electronic health records provide a complete medical history for a patient, including a full account of the patient's illnesses, treatments, laboratory results, medication history and known allergies. Among primary care providers, approximately one quarter use an EHR system in the United States and 89 percent use them in the United Kingdom. At hospitals, the rate of use is much lower with only about 10 percent or fewer of the hospitals in the United States and the United Kingdom having adopted EHR systems [19]. Achieving the widespread use of electronic health records is a necessary requirement for creating the underlying data sets needed for biomedical informatics research. Access to the electronic health records of large populations will help researchers apply informatics to various problems including clinical trial research, comparative effectiveness studies, and drug safety monitoring.

However, collecting medical data in electronic format is only the first step. Interoperability poses a substantial challenge for biomedical research. The vast amount of electronic medical data cannot fully be utilized by researchers because the data resides in different databases. Even when the organizations that collect and distribute biomedical data are willing to share data, incompatible data formats or data interfaces can create challenges for analyzing data across multiple data sets. As a result, researchers wishing to use multiple data sets must devote significant resources simply to managing the differences between the data and, as a result, have fewer resources available for working with the data [6].

For many years individuals in the research community have called for increased coordination and interoperability among data repositories to advance the use of informatics in health care. They have proposed various options to address interoperability although, to date, no proposal has achieved universal acceptance [6, 7]. One interim solution has been the development of online communities to share programming code to reduce the burden of working with diverse data sets. The most notable, Bio*, is a collection of open-source biomedical informatics projects that provide re-usable code for researchers to use that automate common computing tasks. For example, the project includes modular programming code to manipulate DNA sequences or combine data sets from different data sources [6].

III. National Priorities in Health Informatics

Both the United States and the United Kingdom have made significant investments in health informatics. Measuring the level of public investment in health informatics at the national level is an imprecise science as many forms of medical research involve an IT or informatics component. In addition, funding for this research comes from various government sources. However, one trend is clear: the proportion of medical research that relies on mathematical modeling or data intensive, high-speed computing is on the rise.

This section will enumerate the current investments being made by the national governments in the United States and the United Kingdom in research to develop IT tools for medical research. Although the distinction is not always perfect, this section makes an effort to not highlight investments in medical research that simply uses IT, but rather to focus on those that rely on IT as the principal method for investigation.

A. United States

In the United States, the public funding for biomedical informatics research has come principally through the U.S. Department of Health and Human Services (HHS). Within HHS, the primary funding agency is the National Institutes of Health (NIH), although additional funds come from the Centers for Disease Control and Prevention (CDC), the Agency for Health Research and Quality (AHRQ), and the Food and Drug Administration (FDA). Some additional financial support for biomedical informatics research may come from the National Science Foundation, although its intended mission is to fund science and technology research outside of medicine.

NIH invests $30.5 billion annually in medical research. The growing importance of information technology at NIH can be

seen in the increasing level of investment NIH makes in IT-related grants. Funding for "Network and Information Technology R&D", while not strictly funding for biomedical informatics, has shown a significant increase over the past 5 years. Whereas funding for FY 2005 totaled only $509 million, the NIH estimates funding will reach $950 million in FY 2010 [11]. The NIH also runs the High Performance Computing and Informatics Office within its Center for Information Technology. The mission of this office is to provide the high performance computing resources and tools needed to allow the NIH scientific community to conduct its biomedical research. This office provides the software applications needed by researchers for bioinformatics, structural biology and proteomics.

Investment in biomedical informatics research at NIH comes principally from three sources of funding: the NIH National Centers for Biomedical Computing (NCBC), the National Cancer Institute Center for Bioinformatics (NCICB), and the National Center for Biotechnology Information (NCBI).

The NCBC is an important strategic investment for NIH. In 2004, the NIH created a Roadmap for Medical Research to "address roadblocks to research and to transform the way biomedical research is conducted by overcoming specific hurdles or filling defined knowledge gaps" [12]. One objective of creating the Roadmap was to ensure that these programs would get funded since many of these initiatives might otherwise fall outside the domain of existing centers within NIH or appear too risky. The Roadmap was initially funded by a 1 percent contribution from each center within NIH, but since 2006 has been funded directly by Congress.

One of the three principal themes of the Roadmap is to develop a better toolbox for medical research that will enable scientists to better understand diseases at the molecular level. A key initiative in this theme is to have a Bioinformatics and Computational Biology initiative that will allow researchers to share, analyze, integrate and visualize large data sets.

The major project within the Bioinformatics and Computational Biology initiative is the NCBC. NIH created the NCBC through a two-stage funding process with the goal of creating specialized biomedical computing centers at educational institutions in the United States. As described by NIH, the NCBC is "devoted to all facets of biomedical computing, from basic research in computational science to providing the tools and resources that biomedical and behavioral researchers need to do their work" [12]. In addition, the NCBC serves as a center of learning to educate and train additional biomedical informatics researchers. In FY 2004 NIH devoted between $14 and $17 million to fund the NCBC, and in FY 2005 NIH committed $12-14 million. A total of seven grants have been awarded to different institutions [12].

The second major biomedical informatics program at NIH is NCICB. NCICB is a division of the National Cancer Institute, an organization within NIH. Created in 2001, NCICB has been a pioneer in advancing the use and development of biomedical informatics infrastructure, tools, and data to improve medical research. In 2006, NCICB was reorganized into the Center for Biomedical Informatics and Information Technology (CBIIT). Funding for the Center has increased steadily over the years from $71.7 million in FY 2005 to $101.2 million in FY 2009, an increase of approximately 40 percent over 4 years [4].

The major investment in this area by NCI has been for the cancer Biomedical Informatics Grid (caBIG), a program targeted at harnessing bioinformatics to advance cancer research. While caBIG has been funded by the NCI, it is a collaborative program involving over 80 organizations. Described as "an Internet for cancer research", the caBIG project is intended to make the vast amount of medical data being generated by patients available, accessible and usable to medical researchers by connecting cancer centers, cancer researchers, and participants involved in clinical trials [2].

As explained in Reference [2] by Dr. Kenneth Beutow, the director of the caBIG program, the purpose of caBIG is as follows: "Personalized medicine is all about information. But for information to be useful, it has to be accessible. What we are doing with caBIG is facilitating accessibility through interoperability, essentially creating an environment where information can be exchanged, integrated, and acted upon." In addition, efforts from caBIG include developing software applications to manage clinical trials and share research data. caBIG has also worked to establish a common lexicon for data exchange and develop best practices for the use of electronic health records.

caBIG has already yielded important applications for medical research. For example, one tool developed by caBIG is the Biological Pathway Exchange which is used to model the signal transduction pathways, or biological pathways, used for communication between and within cells. These communication pathways help determine cell behavior, such as whether they thrive or perish and if they spreading to other parts of the body. Such research is especially useful to scientists studying proteomics as it helps them to better understand how proteins interact with each other [2]. The success of the caBIG program has led to the development of the BIG Health Consortium, a public-private partnership that seeks to bring together previously unconnected sectors of the life sciences and health care using the caBIG model to pursue research in personalized medicine.

Between 2004 and 2006, NCICB (now CBIIT) provided $20 million annually in funding to caBIG during its pilot phase [3]. The pilot phase concluded in 2007 and funding for the enterprise phase of caBIG has increased. For FY 2008 funding increased to $45.8 million and estimated funding for FY 2009 is $43.1 million [4].

The 2010 Professional Judgment Budget Request from the National Cancer Institute (NCI) shows the importance that the National Health Institutes places on an increase in investment in bioinformatics. The purpose of the budget request is to describe "what a financial infusion could make possible and how NCI would spend those monies" [2]. For 2010, the NCI has proposed an increase of $2.1 billion in total funding, a substantial increase over its 2009 budget of approximately $5.0 billion. Of these funds, NCI would direct a substantial portion of the increased investment into bioinformatics including $40 million to increase research in systems biology, $45 million for increasing biomedical computing capabilities and $100 million to expand caBIG and help support the BIG Health Consortium. In addition, NCI proposed a substantial investment in other research priorities with a heavy bioinformatics component such as committing $200 million to expanding The Cancer Genome Atlas (TCGA). The Cancer Genome Atlas is a collaborative effort with the National Human Genome Research Institute to collect large data sets on the genetic makeup of various forms of cancer and develop the technology needed to sequence and analyze DNA from tumors. With the additional funds, NCI predicts it could record the genome of up to six tumor types per year.

The third significant source of investment in bioinformatics from NIH comes from the National Center for Biotechnology Information (NCBI). NCBI is a division of the National Library of Medicine (NLM), an organization within NIH. NCBI is a significant program within NLM, accounting for $73.5 million in the FY 2006 budget out of a total NLM budget of $329.5 million. In addition, NLM supports biomedical informatics through its extramural programs division. This division awards grants to support basic and applied research in biomedical informatics, training and education for informaticians, resources for medical libraries, and scientific conferences. In FY 2006 the budget for extramural programs totaled $69.2 million.

Congress established NCBI in 1988 to create a national repository for molecular biology information. The importance of the NCBI mission has expanded with the flood of genomic data and the increasing reliance on bioinformatics by medical researchers. NCBI supports its mission by developing and supporting the information systems and software applications needed to store and analyze molecular biology and genetic information.

One of the signature projects of NCBI is GenBank, an annotated online database of all publicly available DNA sequences. NCBI acts as a central repository for genetic sequence data, exchanging data with multiple international partners on a daily basis, collecting sequence information directly from researchers, and receiving data submitted to the U.S. Patent and Trade Office. NCBI offers integrated search tools such as Entrez which searches NCBI's vast collection of biomedical databases and BLAST which allows researchers to find similar nucleotide or protein sequences in sequence databases. These tools link sequence information with related publications in databases such as PubMed thereby helping accelerate gene-based discoveries and research.

In addition to the funding activities already mentioned, the American Recovery and Reinvestment Act of 2009 allocated $1.1 billion for comparative effectiveness research. Of these funds, $400 million has been directed to the Office of the Secretary in HHS. Comparative effectiveness research helps provide information on the benefits and drawbacks of different treatment options and offers much potential for informatics research. Thus a significant portion of these funds will likely go to support informatics-related research. The Federal Coordinating Council for Comparative Effective Research tasked with prioritizing spending for these funds has recommended the primary investment be for data infrastructure. As described in the Council's report to the President, "Data infrastructure could include linking current data sources to enable answering CER questions, development of distributed electronic data networks and patient registries, and partnerships with the private sector" [21].

Many other federally-funded health care programs have also made important efforts to bring together useful data sets and analyze this data to improve health care. For example, NIH has also created the National Electronic Clinical Trials and Research (NECTAR) network to share clinical research data between researchers and institutions. Better access to clinical trial data will help eliminate duplication between different trials and help doctors find and apply the most effective treatments. CDC runs the National Electronic Disease Surveillance System (NEDSS), a program to monitor public health for disease trends and outbreaks. Using public health, laboratory and clinical data, each state implements its own electronic surveillance system for communicable disease surveillance, either using NEDSS or its own custom information system. A recent survey found that 40 states have a fully-functional electronic surveillance system in use [13]. CDC also operates BioSense, a program designed to rapidly identify and monitor bioterrorism and disease outbreaks. The FDA launched the Sentinel Initiative in 2008 with the goal of developing a system to monitor the safety of drugs and other medical products regu-

lated by the FDA. While still in its early states, the goal of the Sentinel Initiative is to allow the FDA to query external databases—such as electronic health record systems, insurance databases and other medical registries—and more rapidly detect potential threats.

B. United Kingdom

National-level funding for basic medical research in United Kingdom has historically come from two principal sources: the Medical Research Council (MRC) and the National Institute for Health Research (NIHR). In 2006, in response to a report published by Sir David Cooksey, the British government decided to recalibrate spending on medical research by creating the Office for Strategic Coordination of Health Research (OSCHR). The purpose of OSCHR is to better coordinate funding for medical research at the national level to make research more effective, maximize the clinical benefits for patients, and better utilize limited resources. Following the 2007 Comprehensive Spending Review, the budgets of both departments were combined into a single research fund. The last stage of consolidation occurred in 2008 when Scotland and Wales both decided to fully commit their share of financial resources to the research fund. By 2010, total annual research funding will be approximately £1.7 billion.

As the October 2008 OSCHR report details, "The OSCHR Board has identified E-health records research, and particularly the research potential of large electronic patient record databases, as a major opportunity for UK biomedical science, patient safety and public health" [15]. OSCHR has made a substantial financial commitment to health informatics in its planned budget for 2010-2011, the period at the end of the current three-year Comprehensive Spending Review. Funding for this research will come both from MRC and NIHR sources. The NIHR budget includes £18 million for the Research Capabilities Programme of Connecting for Health in England. The MRC budget includes £0.6 million to support collaborative programs that will study how to use electronic data sets to improve medical research [15].

The Research Capability Programme is part of the National Health Service (NHS), the publicly-funded health care system that serves all residents in the United Kingdom. In 1998, the Department of Health created a national initiative, NHS Connecting for Health, to modernize its health care system through the use of IT. The Research Capability Programme is an initiative of NHS Connecting for Health. Originally conceived in 2005 as an initiative to gather population data for epidemiological and comparative effectiveness studies, the mission of the program has evolved into a broad effort to transform the NHS so that health care research is a core area of focus. The objective is to tap into the vast potential supply of NHS data for the purpose of improving health care quality and safety for patients through improved medical research.

The Research Capability Programme was created in response to a 2007 report from the UK Clinical Research Collaboration's (UKCRC) Advisory Group. The UKCRC report identified six specific recommendations for the UK to improve its research capabilities. These recommendations included: mandating the use of a unique identifier in all patient records; making research a core objective of the NHS Care Records Service; making available databases of complete, longitudinal medical records that cover the entire population; improving data completeness and data quality; addressing regulatory and governance issues regarding the use of data; engaging with all relevant stakeholders [18]. In direct response to the UKCRC report, the Research Capabilities Programme is currently developing the technical architecture, functional requirements, data standards, information governance, infrastructure, and stakeholder engagement needed to improve the UK's clinical research capabilities.

Outside of the NHS, the United Kingdom has seven research councils that fund research in various fields. Since 2001, all of the research councils have participated in the UK e-Science Programme, a coordinated effort to give researchers across all domains access to the large data sets, computing resources, and software tools need to exploit informatics research. Launched as a joint program between the research councils and the now defunct Department of Trade and Industry, the initiative received £118 million in initial funding. It has experienced some notable successes, for example, using grid computing to identify three drugs that can be used to treat antibiotic-resistant bacterial infections. Funding for the e-Science Programme also supported the development of CancerGrid, an initiative to develop software tools to reduce the cost of clinical research and make data sharing more efficient.

While most of the health informatics research funding comes from the MRC, for some of this research the MRC partners with other research councils including the Biotechnology and Biological Sciences Research Council (BBSRC), the Engineering and Physical Sciences Research Council (EPSRC) and the Economic and Social Research Council (ESRC). For example in 2008, the BBSRC, which has an annual budget of approximately £450 million, established a £6.5 million bioinformatics and biological resources fund to support health informatics research.

Past MRC activity has shown its commitment to health informatics research in previous years through its funding and strategic decisions. Between 2005 and 2006, MRC increased its funding for informatics fellowships from £0.9 to £1.4 million. In 2006, the MRC provided £1.1 million in funding to the National Cancer Research Institute bioinformatics initiative. In

2007, the MRC granted funds totaling £2.2 million to support workforce development in bioinformatics. In 2008, the MRC allocated £1.5 million to award grants supporting the use of electronic databases for medical research. EPSRC, ESRC and the Wellcome Trust contributed additional funds to this project resulting in a total funding of £10 million. MRC also worked with EPSRC to jointly fund a £2.3 million initiative to study how information systems can be used to drive better quality diagnosis and treatments in health care [17].

The MRC has set out a new strategic plan for 2009 to 2014 that continues to emphasize health informatics research. This plan includes four strategic goals, one of which includes the objective of fully exploiting the potential benefits of population-based data by developing tools to use existing data sets, sharing and linking future data sets, and developing a national framework to support this type of research.

Various nongovernmental organizations in the United Kingdom also focus on health informatics and operate, at least in part, with public funding. For example, the European Bioinformatics Institute (EBI) at Cambridge is one of the primary sites for biomedical informatics research in the United Kingdom. EBI is one of the five centers of the European Molecular Biology Laboratory (EMBL), a major basic research institute supported by public funds from 20 European nations. In addition to providing free public access to online biological databases, EBI pursues bioinformatics research and offers bioinformatics training to students and scientists. In 2008, EBI operated with a budget of approximately €43.2 million with approximately 45 percent of the budget from EMBL. EBI receives additional funding from various sources, including close to €3 million from the NIH in the United States and a similar combined amount from two of the research councils (MRC and BBSRC) in the United Kingdom. The remaining funds for EBI come from the European Commission (€8.8 million) and the Wellcome Trust (€7.5 million). The Wellcome Trust is the largest charity in the United Kingdom, spending approximately £600 million annually on research domestically and abroad.

Another important nongovernmental organization contributing to this research is the National Cancer Research Institute (NCRI). NCRI is a public-private partnership between various stakeholders supporting cancer research in the United Kingdom. NCRI began the Informatics Initiative in 2003 with the goal of maximizing the impact of cancer research through the application of informatics. It has focused on improving data sharing within the cancer research community by developing internationally-accepted data standards, databases, and data tools. The primary interface to this data is the NCRI Oncology Information Exchange (ONIX). ONIX is a portal to various data sources and provides researchers specialized tools to search biomedical databases.

One of NCRI's key accomplishments is promoting data sharing of all publicly funded research both through creating technical standards and enacting cultural change in the research community. The MRC has been one of the leading partners in this effort having established a data sharing initiative in 2001. Other cancer research partners, including the Wellcome Trust, BBSRC, and Cancer Research UK have followed with their own policies to support better data sharing. For example, in April 2007, BBSRC established a new data sharing policy that states that it will fund efforts to make data available with as few restrictions as possible for further scientific use by its researchers.

IV. Recommendations and Conclusions

Both the United States and the United Kingdom make significant investments in medical research. Thus it is little surprise that both nations have made investments in informatics research as its importance to medical research continues to grow. The total public investment by the United States in health informatics is substantially greater than in the United Kingdom. However, both countries commit roughly the same percentage of total public medical research funds to health informatics. Because this report has only looked at publicly-funded research, it cannot be used to infer which country possesses more technical capacity for biomedical informatics research.

Qualitative differences exist between the approaches taken by these two nations on health informatics. An important development in the United Kingdom has been the creation of the Research Capability Programme. The United Kingdom is uniquely positioned to benefit from advancements in health informatics research because it is significantly ahead of the United States in its transition to electronic health records among primary care providers. More importantly, the NHS has made an important strategic decision to emphasize medical research as a core activity that it must support. Thus, as the NHS continues to develop its IT infrastructure, it will be able to make technical upgrades to improve information sharing and its information base for research. Moreover, because of its unique role, the NHS can more directly impact issues critical to researchers such as improving data quality through its own policy directives.

Researchers in the United Kingdom can take advantage of the national electronic health record system with projects like UK Biobank. UK Biobank is a large-scale medical research project to study how an individual's health is affected by lifestyle, environment and genes. Using both public and private funds, the project intends to enroll 500,000 participants in the United Kingdom who will participate in an initial health as-

sessment, provide medical samples and consent to allow researchers to monitor their medical records indefinitely. This project is made possible by the data sharing capabilities of the NHS that will allow researchers to track the health of participants over the next few decades. Even initiating the project required NHS data—NHS medical records were used to identify potential participants and invite them to join the study [20].

The United States currently lacks the capacity being developed by the NHS to turn its existing or future electronic health records into a usable database for medical research. This is not too surprising given the decentralized approach of the current efforts to increase adoption of electronic health record systems. In the United States, the closest alternative to the information base the NHS is building is the HMO research network, a consortium of 16 health maintenance organizations (HMO) in the United States that provide researchers access to health data for a large population.

To address this deficiency, future efforts in the United States to speed adoption of electronic health records systems should include functional requirements to allow the secondary-use of medical data for research. For example, HHS should consider the importance of secondary use of medical data as it develops interoperability requirements and other standards in its evolving definition of "meaningful use" that will determine how funds are spent from the 2009 stimulus package. The goal should be to develop a national data-sharing infrastructure to support health informatics research, rather than to create isolated, project-specific research databases. Many current or proposed projects focus on adding an additional layer of reporting requirements to health care providers to gain access to important patient data rather than simply making all patient data accessible for research. For example, Rep. Dingell (D-MI) recently introduced the "America's Affordable Health Choices Act of 2009" (H.R. 3200) which included a provision that all health care centers receiving federal dollars from Medicare or Medicaid programs be required to report hospital-acquired infections to CDC's National Healthcare Safety Network [10].

Challenges remain for both nations, especially in regards to data sharing. Continued funding is necessary to develop the technical infrastructure and data standards needed to improve data sharing between existing systems. In addition, a mechanism is needed to allow relevant medical data to be shared for authorized medical research in a timely and efficient manner. Safeguards must be put in place to protect patient privacy, but these individual protections must be balanced against the potential benefits from research. As Reference [9] has shown, privacy regulations can have a substantial impact on technology diffusion. As a result, policymakers should be cautious of implementing privacy regulations that impede technology adoption in health care, as it could have a significant impact on health care quality and medical research.

The United Kingdom appears more prepared to address these challenges. As part of its Information Governance and Threat Assessment agenda, the Research Capabilities Programme has produced work documents that address many of these issues relating to data sharing. For example, it has identified options and next steps to develop a pseudonymization service for the de-identification of patient data, and it has analyzed the legal issues that need to be clarified in the United Kingdom to use patient data for research. Issues such as patient consent must be resolved before data from electronic health records can be used extensively for medical research. Patient consent may be required to either use patient medical data directly in research studies or to identify patients for potential inclusion in research studies.

The United States should also form a comprehensive review of these data- sharing challenges. For example, it should consider the current legal framework for sharing research data. Like the MRC, the NIH has made a clear commitment to data sharing. As of October 2003, NIH has required that all grant applicants seeking funds of $500,000 or more include a data sharing plan as part of the proposal or explain why data sharing is not possible. However, NIH has acknowledged that state and federal laws, including the HIPAA Privacy Rule, may interfere with data sharing [16].

Finally, both nations need to ensure they have strong research communities. Advanced biomedical informatics research will not only require having the technical infrastructure in place, it will also require having a talented pool of researchers trained in biomedical informatics and related fields. In the United Kingdom, the MRC has funded workforce training and fellowships. In the United States, the NCBC has been used to expand the population of trained researchers in biomedical informatics, bioinformatics and computational biology [14]. In addition, both the United States and the United Kingdom recognize the need to work collaboratively on this research and partner with the private sector. As previously discussed, research communities in both countries have formed working partnerships, such as the collaboration seen between the NCI and the NCRI.

The need for pursuing informatics in health care has been recognized at the local, national and international levels. In 2005, the World Health Organization adopted Resolution WHA58.28 to establish an eHealth Strategy that noted the "potential impact that advances in information and communication technologies" could have on medical research and urged member states to implement "national electronic public-health information systems and to improve, by means of in-

formation, the capacity for surveillance of, and rapid response to, disease and public-health emergencies" [8]. Both the United States and the United Kingdom have responded to this call, and both have made substantial commitments to continue to improve the use of IT in medical research in the following years.

REFERENCES

[1] D. Castro, "Explaining International Health IT Leadership," Information Technology and Innovation Foundation, Washington (2009).

[2] "Connecting the Nation's Cancer Community," National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services, NIH Publication No. 08-6363, January 2009, http://plan.cancer.gov/pdf/nci_2010_plan.pdf.

[3] "The caBIG Pilot Phase, Report: 2003-2007," National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services, November 2007.

[4] K. Colbert, "Funding data for the NCICB," Personal communication with author to Office of Budget and Finance, National Cancer Institute. (July 2009).

[5] D.W. Bates et al., "Effect of Computerized Physician Order Entry and a Team Intervention on Prevention of Serious Medication Errors," Journal of the American Medical Association 280 (1998): 1311-1316.

[6] L. Stein Creating a bioinformatics nation. Nature 2002; 417(6885):119–20.

[7] L.M. Etheredge, "A Rapid-Learning Health System" Health Affairs, 26, no. 2 (2007): w107-w118.

[8] "Resolution WHA58.28 on eHealth", World Health Organization (May 2005).

[9] A. Miller and C. Tucker, "Privacy Protection and Technology Diffusion: The Case of Electronic Medical Records," Management Science, 55 (July 10, 2009): 1077-1093.

[10] United States Congress, America's Affordable Health Choices Act of 2009, 111th Congress, 1st session, H.R. 3200. Washington: GPO, 2009.

[11] "Estimates of Funding for Various Research, Condition, and Disease Categories (RCDC)" National Institutes of Health (May 7, 2009), http://report.nih.gov/rcdc/categories/.

[12] "Bioinformatics and Computational Biology", NIH Roadmap for Medical Research. Accessed July 31, 2009. http://nihroadmap.nih.gov/bioinformatics/index.asp.

[13] "Status of State Electronic Disease Surveillance Systems --- United States, 2007" Morbidity and Mortality Weekly Report, Vol. 58, No. 29 (July 21, 2009): 804-807, http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5829a3.htm.

[14] "Programs and Services: Fiscal Year 2006, No. 07-256" National Library of Medicine, National Institutes of Health, U.S. Department of Health and Human Services (2006), http://www.nlm.nih.gov/ocpl/anreports/fy2006.pdf.

[15] "OSCHR Chairman's first progress report," Office for Strategic Coordination of Health Research (November 2008), http://www.nihr.ac.uk/files/pdfs/OSCHR_Progress_Report_18.11.08.pdf.

[16] "NOT-OD-03-032: Final NIH Statement on Sharing Research Data" National Institutes of Health (February 26, 2003), http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html.

[17] "MRC Delivery Plan: 2008/09–2010/11," Medical Research Council (December 11, 2007), http://www.mrc.ac.uk.

[18] "Research Capability Programme: Background, Vision, Overview and Glossary (PD00)" NHS Connecting for Health (August 13, 2008), http://www.connectingforhealth.nhs.uk/systemsandservices/research/docs/pd00.pdf.

[19] A. Jha, D. Doolan, D. Grandt, T. Scott, D. Bate, "The use of health information technology in seven nations," International Journal of Medical Informatics, Volume 77, Issue 12, Pages 848-854.

[20] "UK Biobank: Information Leaflet" UK Biobank (2007), http://www.ukbiobank.ac.uk/docs/Informationleaflet130608.pdf.

[21] "Report to the President and the Congress," Federal Coordinating Council for Comparative Effectiveness Research, U.S. Department of Health and Human Services (June 30, 2009), http://www.hhs.gov/recovery/programs/cer/cerannualrpt.pdf.