

Mark E. Frisse, Valerie Florance, Kenneth D. Mandl,
and Isaac S. Kohane

After reading this chapter you should know the answers to these questions:

1. Which current trends will most influence the progress of biomedical informatics research and practice over the next decade?
2. What critical imperatives emerge from these trends?
3. How does the emerging massive, hyper-connected, and pervasive information technology infrastructure affect patients and other consumers?

4. What biomedical informatics “grand challenges” should be pursued over the next decade?

M.E. Frisse, MD, MS, MBA (✉)
Department of Biomedical Informatics,
Vanderbilt University Medical Center,
432 Eskind Biomedical Library,
2209 Garland Avenue,
Nashville 37232-8340, TN, USA
e-mail: mark.frisse@vanderbilt.edu

V. Florance, PhD
Division of Extramural Programs, National Library
of Medicine, National Institutes of Health, DHHS,
6705 Rockledge Drive, Rockledge 1, Suite 301,
Bethesda 20892, MD, USA
e-mail: florancev@mail.nih.gov

K.D. Mandl, MD, MPH
Children’s Hospital Informatics Program,
Boston Children’s Hospital, Harvard Medical School,
300 Longwood Avenue, Boston 02446, MA, USA
e-mail: kenneth_mandlh@arvard.edu

I.S. Kohane, MD, PhD
Harvard Medical School Center for Biomedical
Informatics and Children’s Hospital Informatics
Program, Boston 02115, MA, USA
e-mail: isaac_kohane@harvard.edu

28.1 Goals

Each of this book’s previous chapters emphasizes the state of the art while implying future courses of activity; but the primary focus of previous chapters is to summarize what has already been accomplished. This chapter addresses the future. Peter Drucker urges us to identify and prepare for the future by examining the trends and imperatives that define, liberate or constrain the actions individuals take to assure the health and well being of one another. Everyone – physicians, patients, scientists, policy makers, and citizens – must make decisions in the here and now based on the future they want (Drucker 1959). Our approach is to look toward the future by identifying present compelling trends, suggesting some challenge areas for biomedical informatics, and identifying building block areas that will lead to the future we want.

More than a decade ago, scientists at Xerox PARC described a future in which computers were pervasive – woven throughout the daily lives and activities of individuals (Weiser 1993b). Today, there is a massive, global, and ever more pervasive information network that connects people and computers all over the world, in real-time, in many commercial and non-commercial activities. Processing

power, input and output devices, bandwidth, storage, and even many common software services have become commoditized and make available to a global society an array of complex information and knowledge resources once afforded only to the elite (AAMC Task Force on Information Technology Infrastructure Requirements for Cross-Institutional Research 2010; Cairncross 2001; National Science Foundation (U.S.), Cyberinfrastructure Council (National Science Foundation) (2007)). For professionals and the public at large, internet-enabled decentralized, globalized and mobilized information sharing, when combined with the capabilities of common consumer technologies (e.g., smart phones) both broaden the availability of knowledge and expertise and lower participation costs. The proliferation of sensors and inexpensive local wireless connectivity and wearable computers creates new opportunities for ambient data collection and more thorough disease monitoring while at the same time posing threats to anonymity and personal privacy (Avancha et al. 2011; Chui et al. 2010; Nissenbaum 2010; Mandl and Kohane 2009).

Never has the need for coordinated innovation been more acute. Many nations are confronting costly increases in health care expenditures as a result of an aging populace, an increase in life expectancy, higher aggregate demand for health care services, and advanced but very expensive new approaches to caring for the very young and the critically ill. Health care costs are deemed unsustainable and pose a growing threat to the economies of nations. Facing massive costs for health care, retirement benefits, and other social programs, governments are increasingly forced to reexamine the funds they can invest in education, research, and vital technology infrastructures. Coordinated innovation is essential to find more efficient and effective means of lower health care expenditures and, at the same time, identifying new and innovative means of educating and conducting research. Without innovation, biomedical informatics may be the victim of a vicious circle in which insufficient support for research and care delivery only raises health care costs and consequently decreases informatics research support even more. This is the soil in which the seeds of the future are germinating.

28.2 Compelling Trends and Critical Imperatives

28.2.1 Trend One. Biomedical Research and Clinical Care Are Linked Through Computation

Laboratory research techniques are evolving rapidly as computational means for data collection, analysis, and simulation have become integral to traditional and novel experimental methods. The inexpensive and widespread use of genetic sequencing, candidate drug target identification, and related technologies creates data sets that exceed current capacities to analyze and interpret, and yet scientists face enormous difficulties in merging data sets to explore new hypotheses. The rate limiting step often is no longer availability of data but instead the ability to address the computational challenges around data integration, data analysis, and new means of visualizing and modeling multidimensional data that may include genotypes, phenotypes, physiological phenomena, and clinical data (Hey et al. 2009). New and innovative approaches to these challenges are emerging (Green and Guyer 2011; National Science Foundation (U.S.) and Cyberinfrastructure Council (National Science Foundation) 2007; President's Council of Advisors on Science and Technology 2010; Multiple Authors 2011b). More direct and immediate linkages among observational data captured in electronic health records and biological investigation research activities represent a new and dramatic collective research frontier. New notions of clinical trials emerge from the research laboratory. New insights derived *in silico* contribute to the prevention of adverse drug interactions and to compare the effectiveness of treatment alternatives. Biological hypotheses, in turn, are refined through more comprehensive and rapid integration of data produced in the routine delivery of health care. Collectively, these trends create, strengthen and refine real-time pipelines between observation in the laboratory and the clinic, and, in so doing, accelerate both the interpretation of data and the dissemination of new knowledge.

New formal or emergent organizational forms are evolving to capitalize more completely on the potential arising from public need and scientific capability (Collins 2011). Greater efforts are being made to strengthen the linkages between basic biomedical research activities and the health care system, within and across organizational, geographic and cultural boundaries (Olsen et al. 2011). Retaining and building upon new knowledge foster new approaches to traditional means of archiving data, knowledge artifacts, and scientific publications (Corn 2009; Luce 2008; National Science and Technology Council (U.S.). Interagency Working Group on Digital Data 2009).

28.2.2 Trend Two: Information Services Are Tailored to the Individual

Mass marketing and communication – be it in commercial or scientific arena – has been replaced by mass customization. In every information activity, users and purveyors of information benefit from decisions tailored to individual need. Internet commerce sites welcome people back by name and suggest a range of readings, music, or other services based on past buying and browsing history. Integrating information across sites allows for highly detailed demographic and preference information that is used to specify on-line advertisements highly tailored to the individual. Health care, while lagging behind the commercial sector for many reasons, has nonetheless recognized the potential of these phenomena both through the adverse impact of incomplete or impersonal information as well as through the growing number of domains in which care is improved through individualized application of highly personalized knowledge (Chan and Ginsburg 2011).

Every part in health care and biomedical research seeks to capitalize on the benefits of personalization while avoiding threats to privacy and other legitimate concerns. Enabled by online access to medical knowledge and data about health services quality and outcomes, consumers, patients, and their informal network of family and caregivers are playing a more active role in their personal

health. These same underlying technologies are enabling individuals to filter information and seek answers to questions highly specific to their information needs, scientific questions, or care delivery issues. Common informatics techniques and problem-solving approaches are uniting consumers, clinicians, and biomedical scientists to the benefit of all concerned.

Customization is no longer a luxury; it is becoming a necessity. One can no longer find answers from a data flood unfiltered by knowledge and need. Information overload impedes our individual and collective abilities to perform at our best. Finding the right answers to important questions must be simple and easily understood in context of need, language, culture, and cognitive frameworks.

The explosion of research data in genomics exemplifies the sheer magnitude of the challenge. With next generation technologies, each of the 10–20 major genome sequencing laboratories is estimated to produce at least 400 GB of data daily. The size and complexity of data sets grows when comparison and analysis is conducted. A comparative study of 629 individuals consists of 7.9 TB of data (Kahn 2011b). Clinicians share the same concerns. Clinicians, researchers, and the public also experience information overload. The size of the bibliome is growing exponentially. By 2010, there were over 20 million citations indexed through PubMed alone (Lu 2011). EHR systems are developing capabilities of incorporating data from remote sensors, personal health records, and genotype determination, but the capabilities of EHRs to translate this information into meaningful knowledge lags. Despite powerful computational infrastructure, finding answers somehow seems more difficult – not simpler – as our technologies improve.

28.2.3 Trend Three. Society's Activities Are Increasingly Online and Collaborative

The majority of Americans say they use an online social network. Worldwide, the number of individuals using social networks at the time of this

writing is near one billion. Crowd sourcing, flash mobs, blogs and virtual gatherings minimize the geographic and cultural boundaries between people and provide a sense of group action. Information is treated as a free good and a tool for social change (Shirky 2010).

This desire for free and spontaneous access runs counter to past conventions for intellectual property and privacy. Widespread deployment of sensor technologies, data collection, aggregation, and dissemination are changing the balance between open collaboration and protection of personal or proprietary interests. For the researcher, recognition and advancement could be hindered or enhanced. Commercial concerns expect that expensive investments will be recouped. For the individual, rights to privacy and desire to control access to personal information may not be well aligned with societal goals or group priorities (Nissenbaum 2010).

The potential of collaboration technology affords great new opportunities for accelerating scientific research, enabling new forms of collaborative learning at every educational level, adding to society's true health care status, and fostering new means of meeting health care and social needs. These technologies allow both informal conversation and decisive action. They provide the opportunity to broaden collective understanding and bring the right people together at the right time when important decisions or scientific interpretations must be made. The by-products of these conversations and analyses become a part of the scientific or clinical record. Captured and integrated appropriately this record clarifies the context and evidentiary base for interpretations and decisions.

28.2.4 Trend Four: Electronic Health Records Are Changing Health Care

The use of electronic medical records in health care settings is growing. The use of health information technology has improved safety and efficiency within institutions, although such benefits have not been uniformly realized as individual

seek care across institutional or state or regional boundaries. Incentives for adoption of electronic medical records in the US, for the use of health data to improve patient outcomes and for regional health information exchange were created through the HITECH Act of 2009 (Blumenthal 2010; Buntin et al. 2010). Such initiatives can improve patient outcomes and lower costs if interoperability is assured and focus is shifted from individual encounters to individual and population health.

The sources and magnitude of data relevant to the care of the individual are growing. Clinicians are increasingly required to integrate their traditional EHR data with information from personal health records, family health histories, personal genotypes, and an array of data from glucometers, pedometers, heart monitors, and other body sensors. Environmental hazard information and location history are also relevant to direct patient care, epidemiology, and research. To make complete and reliable decisions, health care professionals will require more sophisticated tools for visualizing and analyzing data so that the important items can be identified and acted on in a timely fashion (Stead and Lin 2009). The communication capabilities of these systems (e.g., secure messaging, health information exchange) could support the development of distributed networks of care providers and coordinators tightly linked through advanced information technologies and focused on improving both efficiency and quality while supporting research and public health if policy and economic barriers are addressed.

Biomedical informatics is playing a critical role in shaping our emerging information technology infrastructure into a learning health care system – a system that is capable both of assuring that every decision is made with complete information and ensuring that every care instance can contribute a deeper understanding of care for individuals and populations. Efforts are underway to provide individuals, families, and clinicians with decision-making support that can adapt to great differences in health literacy levels and cultural context. Collectively, these efforts should realize dramatic improvements in care access, care quality, and care costs.

28.2.5 Trend Five: Data Sets Are Massive, Heterogeneous, and Interconnected

Data once stored in and used by only a single organization are becoming more accessible for use by others. Federal and private sector initiatives are increasingly releasing data in controlled ways to foster innovation in consumer health, care delivery, and scientific investigation (Park 2011). Health information exchange – one means by which more information can be made available to support individual care – has become an essential means by which both the quality and care of individuals and populations may be improved (Frisse et al. 2011). Surveys designed to collect data about personal behavior and preferences now include biological data and connect nature with nurture. Environmental data streams are continuous and plentiful and are available in real time as a factor in public health surveillance. The potential for effective use of information often exceeds the imagination of scientists unaware of impact that linking these new data sets affords.

Traditional approaches to computer-based knowledge representation and analysis appear inadequate to handle the emerging information environment. Informaticians, engineers and computer scientists, among others, are working hard to develop new approaches for managing complexity, minimizing fragmentation, improving data accuracy and completeness, and establishing meaningful linkages among data sets. Our emerging massive data sets contain ‘chunks’ of distinct but related data, not collected from the same sources or for the same purpose, but linked by the implications for human health.

Despite stunning progress in data collection and storage, the true potential of richer and more inter-connected data sets has not yet been realized. Computable data sets have untapped potential to provide new insights into clinical care, fundamental research, public health activities, educational efforts, and health care administration – in short, reaching into every aspect of health and biomedicine. The enormous range of expression of molecular and organism-level data routinely used by the laboratory-based biomed-

ical scientist will be matched in complexity and diversity by the emergence of systems that combine genotype, phenotype, environmental data, lifestyle data, location, personal preferences and detailed analysis of inferences derived from ambient data collection.

Addressing critical imperatives will require unprecedented and consequential improvements to the efficiency and effectiveness in every aspect of research, education and care delivery. Most economically advanced nations are confronting costly increases in health care expenditures. These cost increases are the result of an aging society, an increase in life expectancy, higher aggregate demand for health care services, and new means for caring for the very young, the very old, and those at the end of life. By all accounts, these factors contribute significantly to an unsustainable increase in health care costs and a massive drain on national economies. If left unattended, the consequences will be dire. Escalating costs and weak economies force individuals and governments to make difficult choices between paying more to meet this generation’s rising health care demands or investing more in the research necessary to meet the needs of the next generation.

28.3 Biomedical Informatics Grand Challenges

The science and practice of biomedical informatics are evolving in a world where computation is ubiquitous, data are heavily interlinked, people collaborate in new ways, economic needs and constraints increase, and the nature of biomedical research and discovery has changed. Grand Challenges require innovative breakthroughs across a range of disciplines, not just in a single field. Successful solutions to a Grand Challenge will transform the way people think, they way they work, and the way they care for themselves and one another (Omenn 2006; Sittig 1994; Sittig et al. 2008; van Bommel 1997; Wild 2009). We frame five such Grand Challenges in terms of Drucker’s questions: what futurity do we have to factor into our present thinking and doing, what time spans do we have to consider, and how do we converge

them to a simultaneous decision in the present”? (Drucker 1959). Each Grand Challenge requires innovation and collaboration. Each is directed towards long-term goals but suggests identifiable immediate and intermediate milestones.

We present each Grand Challenge, describe possible outcomes if the Challenge is met, and place the Challenge in the context of critical scientific or societal imperatives. We suggest possible starting points by identifying some “building blocks” that, when developed, will contribute to successful outcomes. By virtue of the interdependencies among informatics foundations, technology, policy, and action, a building block elaborated under one Grand Challenge may very well be a critical component of other Challenges described.

28.3.1 Grand Challenge One. Comprehensive and Dynamic Information Resources for Research, Education, and Health Care Delivery

When this challenge is addressed, our ability to characterize and understand complex biomedical phenomena will increase as a result of new capacity to analyze and interpret data from ambient information, from deftly engineered vehicles for collecting research data or patient-reported observation, and from social networks and an array of intelligent medical devices.

Biomedical researchers, clinicians, and patients will benefit from extensive and tightly-linked data resources that improve understanding of disease mechanisms through deeper insights into relationships among genes, behavior and environment. New opportunities to improve health care delivery and biomedical research arise when more integrated and manageable systems are brought to bear at the very moment clinical or personal health decisions should be made.

Biomedical informatics professionals must ensure that the potential of a hyper-connected informatics infrastructure is realized. Data from disparate sources must be linked and integrated with sufficient precision to assure safe and effective use across a range of scientific, clinical, and

administrative actions that are increasingly reliant on large data sets and accurate machine interpretation. Data standards, ontologies, representation formalisms, pattern recognition, systems and many other techniques must be coordinated seamlessly to ensure efforts will scale and be applicable broadly to a wider array of researchers, teachers, clinicians, and individuals seeking answers to health questions. Standardized, open, and modular software architectures will be essential to ensure that development of and access to systems is affordable to all with needs.

28.3.1.1 Intelligent, Learning Machines

Both data mining and text mining can be used to markup data sets so scientists can identify trends and make discoveries. Real-time collection of reliable data must be more tightly linked with resources that present evidence and enable better clinical decision support. Linkages among full-text biomedical journals can guide ontology development and experimentation (Bada and Hunter 2011). Machine learning techniques should also be directed towards systems that can determine when and how new data suggest reconsideration of published research and care guidelines. The impact of quantifying data and embedded reasoning aids must be more thoroughly studied and understood.

28.3.1.2 Linkages, Multi-Scale Models, and Testing Environments

Integrating disparate systems will require both multidisciplinary collaboration and new data environments in which novel approaches may be tested. In clinical care, progress can be accelerated through more pilot programs designed to evaluate and improve large-scale computable phenotypes from disparate clinical data sources (Kullo et al. 2010). A wide array of open-source software systems, communications standards, natural language processing technologies, data mapping systems, and ontology resources must remain available to foster the broad collaborations necessary for substantive work (Multiple Authors 2011a). Using similar approaches, technologies, test beds, complex multidisciplinary collaborations, and other

novel methods should be employed to transform increasingly voluminous bio molecular data into proactive, predictive, preventive and participatory health techniques (Kohane et al. 2011; Multiple Authors 2010).

28.3.1.3 Informative Interfaces

Sensors and other devices will deliver live data streams from homes, test labs and patients for analysis and knowledge management. Better approaches must be developed to present data to clinicians, researchers and consumers in ways that foster and enhance both understanding and use of the data. Each tool, from the most sophisticated systems used to plan surgery to the consumer-oriented gaming systems for health care promotion, must be tailored to the capabilities and needs of the individual using it at the moment. Each must evolve as the skills of the user grow. Systems are needed that can adapt to various needs and levels of literacy without requiring customized programming for each encounter.

28.3.1.4 New Models for Biology and Health

Collectively, linkage and analysis techniques can be used to formalize interdependencies between genotypic and phenotype data, drawing data from populations with multiple conditions to complement current clinical trials. Analysis of the value, investment and impact of different components in this cycle requires new statistical models as well as systems that can track or predict the elements taken into account in a decision. New forms of simulations must be developed to accelerate understanding in scientific and clinical settings.

28.3.2 Grand Challenge Two. Decision Support Is Fully Informed by Evidence and Personalized to the Individual

When this challenge is addressed, it will be far simpler to gather and personalize many types of evidence about an individual's health and information needs in ways that support personal decisions made alone or with the assistance of family

members, clinicians or friends. Personal health information, medical records data, and published clinical trial results would be complemented by broader array of genetic information, environmental context, and ongoing clinical trial data that are presented at an appropriate level for the user.

The sophistication and reach of intelligent tools already embedded in the health information infrastructure makes this scenario possible, even likely. The seeds for receiving guidance this way are already planted – in software 'wizards', in e-commerce recommender systems, and in retrieval aids supporting consumer health information sites (Ricci et al. 2011). New efforts focused on health care and biomedical research should tailor plentiful and rich sources of data, information and knowledge to meet the individual needs in any context at any moment in time. Efforts may create more helpful dashboards summarizing health status summaries, recommending consumer health actions or even monitoring critical care delivery. They can also be directed to highly sophisticated treatment planning systems supporting the complex treatment protocols in modern oncology, critical care, and other disciplines in which lives depend on the results (Stead 2010). There is an opportunity to ensure that every appropriate health data item and every entry in a person's care record contribute appropriately to a sound health decision.

Once personalized diagnoses can be reached through integration of comprehensive information, the next step can be highly personalized care plans that are shaped by social, economic and cultural knowledge. Every health decision can be based on information that is made meaningful for the decision maker, delivered at the right time and place, in a usable form. For example, every decision to prescribe a medication could be made on the basis of a computer-generated synthesis (of current medications, past medication history, allergies, health status, phenotype, physiology, and full genetic profiles) that is delivered quickly to the clinician, saving hours of valuable time.

28.3.2.1 Personalized Choices

Each decision support application, from the most sophisticated systems used to plan surgery to the consumer-oriented gaming systems for health

promotion, must be able to tailor itself to the capabilities and needs of the individual using it at the moment. Decision support applications based on ever-improving user modeling systems will afford greater adaptation to changes in needs and user sophistication. Decision support application research and practice can lead to systems that assess levels of literacy without needing specialized programming for each encounter. Through contextual analysis, these new systems can allow their users to focus on the decision at hand rather than the mechanics of software interfaces.

28.3.2.2 Personalized Health Libraries

In the current information-rich culture, information seeking and management are routine for everyone. Individuals increasingly expect all useful information to be available on the Internet in real time. Individuals expect continuing improvements in access to and management of personal health information collections drawn from health records, personal notes, correspondence, Internet sites, and health publications. Efforts to transform data collections into useful, self-managing, and usable personal libraries are promising research and application directions. Equally valuable will be health-library applications capable of searching through the universe of health information for possibly pertinent new data, assessing it, asking advice and making recommendations to the library owner about updates. Research is also needed to foster patient discussion with knowledgeable peers or professionals when advice and recommendations are sought about the relevance of a new piece of information.

28.3.2.3 Personalized Translators

Inevitably, clinicians, scientists and patients use technical jargon or culturally determined language that is clear to the members of their group, but not to others. They speak in phrases and not sentences, inject humor, scramble many ideas into a single thought, and make mistakes. Efforts that improve processing the spoken and written language of health care and biomedical research will increasingly be foundational components of far more sophisticated knowledge management

systems. Language applications capable of ‘translating’ across and among different cultural or technical domains may be as valued as those that translate among foreign languages.

28.3.2.4 Personalized Health Curricula

Efforts focusing on patient education are often situational, such as instructions for wound care. But longitudinal health records support a new approach, the ability to formalize a “learning curriculum” tailored to the needs of individual who receives care and to professionals and informal caregivers who support this care. Pilot programs should be developed linking the educational agenda of patients and doctors with the care plan. Social/behavioral research on health literacy, language barriers, culture and demographic factors that influence learning should be exploited in delivery of this curriculum (Woods and Kemper 2009). Similar efforts could apply to providing clinicians with continuous education about new types of information (e.g., genetic tests) feeding into the electronic health record and their utility in health decisions.

These efforts complement the use of collaborative models, individualization, and simulation in more traditional health science education. With few exceptions, our professional schools remain fixed on nineteenth century models of education, segregated by professional class, and certainly ill-suited to the collaborative nature of care essential for today’s economy. Personalized education services are a part of a more systematic study of the transformation of traditional education through distance learning, computer-supported collaboration, team-based education, experiential knowledge gathering, and simulation. Much can be learned remotely – whether from the other side of a state or the other side of the world. Teams can be brought together sooner and exposed both to real world clinical and laboratory settings or simulations of critical complex problems. Teaching of surgical techniques must be developed to support widespread and routine pre-operative simulation and robotic surgery when such techniques yield cost-effective improvements. These same techniques can be

applied to create far more customized case-based teaching technologies foster a greater reliance on three-dimensional interactive exploration over traditional anatomical dissection (Satava 2007).

28.3.2.5 Personalized Health Records

Clinicians, patients and their families need access to similar information whether in their homes, in care settings, in offices or other public settings. Information must flow between care-center based electronic health records and personalized health records maintained by patients and their families. Techniques for self-monitoring by an electronic health record, whether corporate or personal, can be improved through the addition of synchronized updates, gap identification, error avoidance, and a range of alerting applications. The integration of two (or more) views of a person's health history can help foster dialogue efficiently about new research findings, relevant changes in family health, or possible changes to a care plan suggested by data from sensor readings. Opportunities abound to create systems that allow individuals to use health information to link data, make personal discoveries, find others like themselves, and promote positive collaborative dialogue to extend the notion of the self-managing individual to the notion of the self-managing group (Kohane and Altman 2005).

28.3.3 Grand Challenge Three. Learning from Every Experiment and Health Care Decision

When this challenge is addressed, individuals and groups will be able to more readily exploit the virtuous circle of evidence creation, critical analysis, rapid application, rigorous evaluation and effective use (Frenk 2010). Research and practice become continually informed by new findings and the evidence supporting fairly limited situations becomes relevant to a broader and more complex population. Through this virtuous circle, research is transformed from a static set of activities to a continuous and never-ending process of advancement and transformation.

To optimize care, practitioners of evidence-based medicine use scientific information largely gleaned from the published literature, often based on randomized clinical trials. The challenge is to bring about a new paradigm in which evidence-based medical decisions include not only traditional published medical knowledge, but also evidence from highly processed clinical, environmental and epidemiological information.

Electronic medical record systems contain both structured and unstructured health data. These systems may hold more longitudinal data than a paper patient record held, but identifying all pertinent information for the problem at hand is not yet routine. The promise of this large data repository lies in the clinician's ability to quickly see outcomes for a given patient based on the experience of many similar patients. To use stored health data in this care paradigm requires novel probabilistic modeling paradigms and machine learning techniques.

The Institute of Medicine has called for a learning health care system that generates and applies "the best evidence for the collaborative health care choices of each patient and provider; to drive the process of discovery as a natural outgrowth of patient care; and to ensure innovation, quality, safety, and value in health care" (Etheredge 2007; Olsen et al. 2011). A recently shifting emphasis from clinical trials to post marketing surveillance creates one clear example. Withdrawals of major drugs such as cerivastatin and rofecoxib have highlighted the importance of post-marketing surveillance where pre-launch clinical trials often fail to detect adverse drug effects due to limits in sample size and demographic diversity, and in the diversity of comorbidities and drug interactions represented in trials (McGettigan and Henry 2006; Staffa et al. 2002; Strom 2006; Wadman 2007). Hence, the data to support a complementary system of evidence, based on observational data about large numbers of patients, can be delivered. While there are inherent selection and confounding biases in evaluating outcomes in observational datasets, there are also advantages to measuring the impact of therapeutic trajectories in real world contexts.

To date, the term personalized medicine has denoted use of a patient's genotypes to define whom the patient most closely resembles, in order to target and tailor therapeutic options. In fact, genotype is only one variable that can be used to personalize therapies or to predict trajectories and outcomes. Another approach would be personalization based on clinical and quality outcomes across a similar population. For example, a given patient or her doctor considering a new medication might predict rates of adverse events and hospitalizations based on the experience of a cohort of individuals similar to the patient with respect to age, gender, genetic makeup, and underlying conditions (Polifka and Friedman 2002). The predictions may differ from those made by extrapolating the results of often small clinical studies and applying them to the patient at hand. More often than not, a patient differs from the population studied in traditional clinical trials in terms of age, race, ethnicity, co-morbidities, duration of exposure to the therapeutic and even indication.

With personalization of therapy based on contextualization of a patient among those with similar characteristics, clinical decisions are made not only on generic knowledge and the traditional medical history but also on a worldwide collective knowledge of how such decisions have affected the care of others with similar highly specific characteristics (Feero et al. 2010; Hamburg and Collins 2010; Kemper et al. 2010; Kho et al. 2011).

This contextualized information has a role in supporting decisions by patients and doctors. There is already a grassroots movement to share information among patients with similar diseases through online social networks. This movement is predicated on the belief that patients should be able to use their own information to find appropriate others with whom they can exchange ideas, share data, make discoveries on their own, and develop new collaboratively created knowledge resources (Kohane and Altman 2005; Weitzman et al. 2011a; Wicks et al. 2011).

28.3.3.1 Comprehensive, Self-Updating Data Sets

The generation of self-updating pan-health care datasets will be advanced through additional research and policy efforts directed towards data governance, de-identification, aggregation, and use policies. Data may come from EHRs, from public health documents, or from patients themselves (Fine et al. 2011). Meta-level representation is essential for simplifying integration, supporting new uses of data, and enforcing privacy. The 2010 report from the President's Council of Advisors on Science and Technology represents one general approach (President's Council of Advisors on Science and Technology 2010). Efforts to apply on a systematic basis meta-level representations to all forms of health care information will be vital both to support tractable machine-interpretation and to ensure data privacy and security policies are maintained.

28.3.3.2 Personalization Information Meeting Individual Needs

The precise synergy and complementarity of observational studies and traditional study designs (e.g., randomized controlled trials) must be established. Providing personalized decision-support services will require more systematic research and evaluation of methods that will find the cohort of patients similar to a given patient, identify the treatment paths taken for that cohort, and quantify outcomes as a function of treatment paths.

28.3.3.3 New Channels for Distribution and Discovery

Health information exchanges should allow processed clinical information to flow across boundaries among various clinical and public health environments, allowing a new awareness of epidemiologic context. But many technical, cultural, and policy challenges impede realizing the full potential of data exchange. Additional research, implementation, and evaluation in operational settings of means by which EHRs and PHRs can communicate with one another afford promising opportunities. Fruitful research directions include

the adaptation of these principles to formal care coordination, virtual medical homes, and custom-tailored information services.

28.3.3.4 Personalization Information Meeting Group Needs

Through more widespread access to a range of information, new opportunities arise for understanding the role of the individual within their environment. By “lifting one’s head up” and looking at broader social needs, one can gain a broader perspective of the context in which systems are used. Research that helps prioritize what must be done and for whom will help assure that developers do not so much create software as they create societal value. These activities require a range of skills to assess needs and markets, to implement rapid and iterative user-centered systems, and to evaluate critically the extent to which their efforts help others do their work more effectively and efficiently.

28.3.3.5 Effective Support for Cognitive Work

In clinical settings, far greater attention must be paid to systems that will allow emphasis to shift from the administrative coordination of clinical and financial transactions to support for the cognitive work of teams (Gooch and Roudsari 2011; Stead et al. 2008). Research activities identifying more effective means of supporting collaborative work will be highly valued; systems that accomplish this goal while also supporting financial transaction management and other administrative tasks will be valued to an even greater degree.

28.3.4 Grand Challenge Four. Access to Information is Ubiquitous

Responding to this challenge will advance the coordination of care delivery and decision-making across the health care ecosystem. Individuals making health care decisions will no longer be burdened by seeking information but

instead can concentrate on effectively managing and applying biomedical knowledge. This will require the advancement of a comprehensive, modular, secure, sustainable, and transparent information infrastructure that automatically integrates data from disparate systems, maintains appropriate policy, privacy, and intellectual property controls, and ensures reliable and standardized access by individuals and organizations authorized to access and use the data and knowledge within.

28.3.4.1 All Relevant Information Available

Access ideally should conform to the A⁷ principle: An^ywhere An^ytime, Affordable Access to An^ything by An^yone Authorized (Wing and Barr 2011). Whether viewed as part of an electronic health record or located by searching Google, MedlinePlus or another online health information source, optimal decisions can only be made efficiently if new means are developed to integrate disparate data sources in standardized ways. More consistent means can be developed to transform and integrate into decision-making contexts a wider range of audio, graphic and printed information. Many research avenues can be explored to develop support software that can be embedded in ways that enhance ability to mark-up and explain any item in the information at hand. These systems can confer a greater “intelligence” to information by demonstrably incorporating in appropriate ways and on a “real time” basis all relevant new information.

28.3.4.2 Modular, Open-Source Software

Novel methods for developing and evaluating the creation and distribution of open-source software are essential to afford interoperability, economy, and scalability. Affordable access to software systems that improve productivity and growth is every bit as important as the equally open and affordable “big data” systems that drive innovation. Modular and open-source systems are vital to every topic discussed in this chapter. Through systematic study of foundation

systems, developer intuition can be complemented by far-reaching and more formal needs assessments.

28.3.4.3 Leveraging the Promise of Emerging Technologies

In some contexts, sophisticated software systems can complement or replace human decision-making. As has been the case in the computer interpretation of electrocardiograms, efforts should be made to determine the circumstances in which pattern recognition and other techniques can confer the same accuracy and efficiencies in other areas such as radiology and histopathology (Burnside et al. 2006; Gabril and Yousef 2010; Noble et al. 2009; Rubin and Napel 2010). More broadly, new means of analysis can be applied to leverage genome-wide association studies and to personalize medical care. The former requires particular attention to the curation of massive data sets; the latter requires study of how mobile devices can be integrated into systems of care.

28.3.4.4 Compelling Demonstration of Value

Further work is needed to study the effectiveness and efficiency of interconnected systems and of the decision tools and other intelligent aids built into them. First, automated approaches associated with incorporating and integrating data from outside sources (e.g., environmental data, monitoring device information) must be studied. Second, widespread data exchange among mobile devices, personal health systems, electronic health records, and other technologies necessitates new analytic approaches derived from the perspective of individuals and systems (Friedman and Wyatt 2006). Of necessity, these analyses will include measuring impact on patients, families, providers, researchers, and those responsible for the financing of research and care. These analyses will require collaborative approaches across a range of technical, social, and economic disciplines. Finally, these same systems will provide new capabilities for knowledge discovery, post-market surveillance of drugs and medical devices,

emergency preparedness, Biosurveillance, and public health.

28.3.5 Challenge Five. Policies that Ensure an Accessible, Secure, Reliable, Effective, and Sustainable Global Computational Infrastructure

Our hyper-connected global information infrastructure continually changes science, commerce, collaboration, education, migration, politics, government, and almost every other aspect of work and life (Friedman 2007). As demands for this infrastructure grow, it will in the aggregate become increasingly costly and will require some degree of continuous management and governance. The impact of this infrastructure upon biomedical informatics, scientific collaboration, education, clinical care and related disciplines is apparent (American Medical Informatics Association 2011; Breslow 2007; Christensen et al. 2009; National Research Council of the National Academies 2011). But at both a global and a local level, disagreements arise on the benefits to individuals and groups of both specific short- and long-term investments and, correspondingly, differences in opinions concerning financial support, governance, and management. Like transportation, utilities, education, and other public goods, the debate about global infrastructure is not about whether it is needed, it is about the value models, regulations for use, governance, and the equitable payment essential for long-term sustainability.

Through effective governance, individuals contributing information to a large and heavily connected system must be assured that their expectations for use and for recognition are reasonably met. Infrastructure governance structures – at any level from local to global – must be studied to validate their ability to maintain data provenance, protect intellectual property, and to ensure data are not disclosed or used in ways contrary to the consent provisions of the individuals or researchers who served as the primary

contributors to any data item or collection. Starting from classic management theory, these new systems afford new and necessary opportunities to study of relationships between strategy, governance, operations, sustainability and growth (Evans 2011).

28.3.5.1 Data Privacy and Security

Data use must be consistent with policies and the expectations of those who contribute data (Avancha et al. 2011; Evans 2011; Markle Foundation 2006; Nissenbaum 2010). As data become more integrated, the challenge of enforcing policies and preferences becomes more problematic. Fruitful and significant efforts focus around the modeling of privacy policies and preferences, the discovery of policies and violations through audit log analysis, methods to anonymize data, and the creation of methods to ensure the protection of everything from massive distributed databases to individual wearable medical devices. As is the case with other biomedical corpora, privacy and security policies defy comprehensive translation into computable forms. Practical means of representing policies and developing policy enforcement mechanisms present major opportunities made even more challenging by the need to ensure these policies are operable in systems composed of mobile devices and other large-scale computing systems, multiple users, and multiple intentions for use (Avancha et al. 2011; Markle Foundation 2006; Nissenbaum 2010).

28.3.5.2 Sustainable Preservation

Pilots designed to study the preservation of new data sources and collections are essential components of a health care technology infrastructure. These collections may originate from machines, scientific research, care provision, patients, family members, social networks, ambient environmental sensors, formal and informal scientific or social publications, and many other sources. Derived artifacts may represent complex biological models, disease processes, health status of individuals or populations, or evidence for advancing research and clinical care. They may be relatively static or they may be generated

dynamically as needs arise. Approaches can be evaluated by the extent to which the collection and support of such artifacts are sustainable and by the extent to which they facilitate subsequent inquiry through exploration across a broad topic at a high level or through deep immersion into detail. Methods can be developed that are applicable across organizations, regions, and countries. Ideally, these methods can encompass the various representations and use intentions for clinical and administrative data. Current approaches to assessing sustainability models for digital data will benefit from a fusion of relevant techniques applicable to large-scale challenges in other domains.

28.4 Summary

This chapter summarizes some of the major, compelling trends in technology, society, economics, and science that serve as the context in which biomedical informatics research is conducted. Collectively these resources constitute an infrastructure that is pervasive, computationally powerful, and capable of providing dramatic scientific and clinical progress. These technologies are changing the way every person lives, works, and learns. Self-evident and irrevocable trends suggest a number of imperatives critical to social and scientific progress. Each imperative challenges the biomedical informatics community to build upon the momentum of technical and social change to address unresolved problems in the creation, management, mining, dissemination and use of biomedical data, information, and knowledge. Some future possibilities are apparent today; some are suggested in the Grand Challenges. Substantive and lasting engagement of the biomedical informatics community can make this possible future a reality.

Suggested Readings

Chan, I. S., & Ginsburg, G. S. (2011). Personalized medicine: progress and promise. *Annual Review of Genomics and Human Genetics*, 12, 217–244. This

paper provides a comprehensive review of the full spectrum of personalized medicine, from laboratory discoveries to their application at the bedside.

- Christensen, C. M., Grossman, J. H., & Hwang, J. (2009). *The innovator's prescription: a disruptive solution for health care*. New York: McGraw-Hill. This book provides background information on the notion of “disruptive technology” across a wide range of application and policy areas in health care. Based on observations in many industries, it describes the various ways in which innovation can emerge.
- Chui, M., Loffler, M., & Roberts, R. (2010). The internet of things. *McKinsey Quarterly*, 2, 1–9. This book looks at the emerging technology of network-enabled sensors being embedded in everyday objects, from roadways to pacemakers. The pervasive nature of computing blurs boundaries between health care and broader consumer software systems.
- Friedman, T. L. (2007). *The world is flat: a brief history of the twenty-first century* (1st further updated and expanded hardcover ed.). New York: Farrar, Straus and Giroux. This book provides a perspective on globalization of economies and technology and how they are “leveling the playing field” for international business competition. The combination of inexpensive computing, global networks, and liquid data change the very nature of collaboration in every setting.
- Green, E. D., & Guyer, M. S. (2011). Charting a course for genomic medicine from base pairs to bedside. *Nature*, 470(7333), 204–213. This paper by the Director of the National Human Genome Research Institute offers predictions about how our increasing understanding of the genetic contributions to human health and disease will alter the way medical care is practiced.
- Kohane, I. S. (2011). Using electronic health records to drive discovery in disease genomics. *Nature Reviews Genetics*, 12(6), 417–428. This paper provides a review of methods used to entrain EHR data into genomics research, makes comparisons with analogous aspects of conventional cohort-driven research, and summarizes current major projects.
- National Science Foundation (U.S.), & Cyberinfrastructure Council (National Science Foundation). (2007). *Cyberinfrastructure vision for 21st century discovery* (pp. iii, 58 p.). Retrieved from <http://www.nsf.gov/pubs/2007/nsf0728/index.jsp>. This publication from the NSF lays out its vision of the future of a culture of peer-to-peer collaboration that is emerging to support distributed communities focused on education, science and development.
- President's Council of Advisors on Science and Technology. (2010). *Realizing the full potential of health information technology to improve healthcare for Americans: The path forward*. Washington, DC: Executive Office of the President. The “PCAST Report” reviews the state of health information technology, in particular, those aspects influenced by US

federal efforts. The report argues for meta-data standards, mapping of semantic taxonomies onto tagged data elements connected through a common infrastructure. It describes how these features can promote privacy, increase data security, and broaden use of health care information.

- Stead, W. W., & Lin, H. S. (Eds.). (2009). *Computational technology for effective health care: Intermediate steps and strategic directions*. Washington, DC: The National Academies Press. This report from the Computer Science and Telecommunications Board of the National Science Foundation describes the current status of transaction-based health care information systems and promotes the adoption of systems more focused on knowledge management and effective clinical use. It serves as a foundation for future research directions in health care informatics.

Questions for Discussion

1. Describe a hypothetical 3-year informatics research proposal that would enable you to address a specific grand challenge. What critical imperatives would your effort address?
2. Identify three organizations that support such research. Provide links and a rationale.
3. Identify research funded by these organizations that provides a methodological foundation for your proposed research.
4. Describe your specific aims, your methods, environmental dependencies, timelines, and a budget represented in terms of your time and financial resources required.
5. If working as a team, specify the specific roles and responsibilities each team member would assume.
6. If working independently, create a list of three informatics researchers from other institutions who pursue similar lines of inquiry. Explain the rationale for your choices by describing their hypothetical contributions to your effort and how you would contribute to their ongoing research.

7. If your interest is in health care delivery or public health, identify a hypothetical project that recognizes a critical imperative, responds to a grand challenge, and advances your organization's mission. Describe your approach, your requirements, and your resource

commitments, and your expected results. Whenever possible, extend your personal efforts to encompass teams across the wide spectrum of health care activities from research, administration, care delivery, public health, individuals, and society.