Version of 11/11/13

Toward a Science of Learning Systems: The Research Challenges Underlying a National-Scale Learning Health System

Findings from a Multi-Disciplinary Workshop Supported by the National Science Foundation

Introduction

The Learning Health System (LHS) will be a "smart grid" that promises to dramatically improve the quality and efficiency of health care, empower public health, take biomedical research to a next level of achievement, and enable individuals to make better decisions about their own health. A high-functioning LHS will unlock the potential of biomedical data, convert such data to knowledge and make this knowledge available to the right decision-makers, at the right times, in the right forms. An LHS at national scale is an unprecedented, ultra-complex, socio-technical system. The LHS is best viewed as high-performing *infrastructure* supporting a wide range of users and uses, some that can be envisioned today and others that will emerge over time.

Achieving a high functioning LHS will be as challenging in this era as was the construction of the Panama Canal early in the 20th century. While the Panama Canal was a monumental achievement of engineering, the daunting problems that had to be overcome to build it went beyond engineering. The initial problems were biological, as deadly infectious diseases had to be conquered. Later challenges were social and political.

In the end the Canal led to the creation of an entirely new nation and it transformed the economics of global transportation. Before the canal was built it was possible to transit the Isthmus of Panama only by a railway completed in 1855. Moving goods in large quantity, or items too large for a railway car, required an expensive and lengthy journey around Cape Horn. When the canal opened in 1914, its effects were immediate and the world was forever changed.

So it will be with the Learning Health System. The LHS, at first blush, might be falsely conceived as an exercise, albeit a large one, in routine software engineering applied to the domain of health. But this approach is as certain to fail as one that conceived building the Panama Canal as merely digging a long ditch. Achieving an LHS will require entirely new types of systems, new technical infrastructures, new architectures, new communities of expertise, and new perspectives on what software is and how it gets executed. It will require solutions to a wide range of socio-political and economic problems. This workshop addressed the full range of these needs, and explored the research required to address them. The product of this workshop can be equally seen as an organized set of research

questions or as an emerging science of learning systems directed at the totality of these questions.

The Learning Health System

The Vision and the Imperative

The Vision: The LHS will be a fully integrated health data analysis network, a critical national infrastructure to transform healthcare and promote health. As it evolves, the LHS will allow the nation to take full advantage of the rapidly increasing amounts of health-related data available in digital and computable form.

As envisioned, the LHS will comprise an architecture and an infrastructure capable of serving purposes both planned and unforeseen. It will serve the needs of all stakeholders including patients, clinicians, researchers, public health professionals, healthcare providers and payers, and many others. A sophisticated LHS will enable health-related data from across the country to be rapidly mobilized, aggregated, continuously analyzed, converted into actionable knowledge which is then applied to shape decisions. The effects of the decisions generate new data for a next cycle of aggregation and analysis. These virtuous cycles can improve healthcare quality, safety, and efficiency; empower public health and biomedical research; and enable individuals and their caregivers, in collaboration with their clinical care teams, to make better-informed health decisions. Seen as persistent flexible infrastructure, the LHS can shorten the seventeen-year gap between knowledge generation and its routine application, while increasing the quality of and our confidence in the data and knowledge on which we rely. The LHS also has important fractal properties. Viewed at local, regional, national or global scale, the LHS will require similar structures and functions in similar ways.

It is important to emphasize that the ability merely to convert "big data to knowledge" does not make a learning health system. A complete LHS must also be concerned with packaging and curation of knowledge so it is widely accessible and actionable, and putting knowledge to use to effect change. Viewed as persistent, ultra large-scale infrastructure, the LHS is a socio-technical system requiring governance and mechanisms to ensure critical properties: functionality, stability, trustworthiness, performance, scalability, sustainability, usability, self-monitoring, self-repair, and self-improvement.

This new perspective must engage nation's health system as a complex adaptive system. Today, as one stakeholder makes a change, the others respond to protect their interest. For example, if a payer reduces reimbursement for a service, care providers are likely to react either by increasing the number of times they provide the service or by ceasing to provide the service. Neither reaction achieves the goal of the service being provided when it is needed and only when it is needed. As a multitude of stakeholders act and react to optimize their microsystems, they progressively sub-optimize the whole. One goal of the learning health system is to transform this destructive spiral into a virtuous cycle where the system iteratively improves itself as a whole, a system where all stakeholders are

equipped to optimize their microsystem in ways that also optimize the whole. Such a system will always be emergent and rapidly changing. It will be interacting with and adapting to other complex adaptive systems such as the environment, societal values, and education.

The Imperative: Spiraling health care costs coupled to mediocre outcomes and health status indices, the persistent latency between best practice knowledge and its application in practice, an inefficient public health and clinical research infrastructure, and many other factors, combine to make the LHS a national imperative. A series of publications from the U.S. Institute of Medicine [1,2] has described the LHS and documented the need for it. A report from the National Research Council on "Precision Medicine" strongly advocates for a "knowledge network" very well aligned with the concept of the nationally-scaled LHS [3]. A recent editorial in the *New England Journal of Medicine* cites the LHS as one of three essential steps needed to reduce healthcare's "GDP Footprint" [4]. Another has challenged all academic health centers to become learning health systems, so they can be drivers of needed change in health research and health care delivery [5]. The LHS harmonizes with the transformed health system envisioned in the Affordable Care Act.

Realizing the LHS Vision

The Foundation: The LHS requires privacy-protected access to health care and public health data in computable form at scale. This foundation for the LHS is emerging as the nation's healthcare system transforms into a system that increasingly digital. This is due in part to a $30 billion federal investment to promote meaningful use of health information technology through the HITECH provisions of the 2009 American Recovery and Reinvestment Act. Averaging several indices, the nation's health system is approximately 35% digital now but is expected to be 80% digital by 2019 [2]. Interoperability requirements built into HITECH will make significant but imperfect progress toward standardizing data stored in electronic health records. At the same time, computer science is increasingly enabling the analysis of huge volumes of rapidly changing and often unstructured data. These developments create a key foundation, without which the LHS would not be possible.

Other foundational elements of a national LHS include organizations that, within their own boundaries, have become learning health systems. While several organizations have made significant progress toward learning status, Kaiser Permanente, Geisinger Health System, and the U.S. Veterans Administration stand out as examples of institutions that are already harnessing the power of data to improve the health of the populations they serve [6, 7]. In addition, federated learning networks among independent organizations are forming to share data and harvest the benefits of learning from it, with the HMO Research Network and CancerLinQ [1,8]as key examples. These developments both document the benefits to be derived from a nationally-scaled LHS and suggest that a national-scale LHS is technically and organizationally achievable.

A Collaborative Effort: Achievement of a national-scale LHS will not be the work of a single individual, organization, stakeholder group, or even sector. Rather, it is anticipated that the

LHS will require the active participation by and cooperation among multiple and diverse stakeholders spanning the healthcare and computer science communities, nationwide and ultimately globally [9]. A number of initiatives aimed at learning are underway nationwide and around the world, and it is expected that a national-scale LHS will be realized by working collaboratively to harmonize these efforts as well as new ones into a single ultra large-scale infrastructure.

Models of multi-stakeholder collaboration that could facilitate the LHS vision on a national scale include the formation of the United Nations, the advent of the VISA network through a "chaordic" process [1], and the evolution of the Internet. Collaborative efforts to realize the LHS have already brought together multiple and diverse stakeholders. Meetings convened by the Institute of Medicine are one example, as is a grassroots effort to create a self-organizing "Learning Health Community", bound by a set of consensus LHS Core Values that were developed in 2012 and that have been formally endorsed by 54 leading organizations to date. The establishment of health-related research programs within the Computer and Information Science and Engineering Directorate of the National Science Foundation is an important initiative that promises to strengthen the links between healthcare and computer science and software engineering research that will be needed to realize the vision of the LHS. Like the Internet at its inception, today those driving the movement know some of what LHS may achieve and have an idea of how it may function, but ultimately recognize that the LHS will continually evolve and never be "finished", and that there are extraordinary applications that have yet to be envisioned.

The LHS Research Imperative

The LHS is also a consummate challenge for researchers across a range of scientific disciplines. While the LHS will inevitably evolve as a complex, adaptive, ultra large-scale system, progress at each stage of development will require the best possible answers to an enormously wide range of questions. The "Version 1.0" of the LHS that will exist upon the initial achievement of national or quasi-national scale must be sustainable and evolvable—and, in order to thrive, must realize increasing capability over time. The LHS must be able to study itself in order to demonstrate and improve its value. While the answers to this range of questions do not exist, the questions themselves can be seen with precision.

The magnitude of the LHS challenge certainly exceeds the 1960s effort to travel safely to the moon. This was primarily a challenge in top-down engineering for which the scientific problems had already largely been solved. By contrast, the LHS is more akin to a nervous system connecting to all corners of our enormously complex healthcare system. Perhaps the better analogy to the LHS challenge is the aforementioned construction of the Panama Canal, which required new technology to cross mountainous terrain, new biology to conquer deadly disease, and a new nation to ensure stable governance.

The Workshop

In this spirit, the National Science Foundation convened a two-day invitational workshop to identify the fundamental scientific and engineering research challenges (in the form of questions) to achieving a national-scale LHS. The workshop was shaped by a recognition of the expanse and depth of scientific challenges facing the creation of a high-functioning "Version 3.0" of the Learning Health System.

The workshop was planned by a 10-member committee and ultimately engaged 45 prominent researchers spanning multiple disciplines over two full days in Washington, DC on April 11-12, 2013. Through a planned format interweaving structured presentations, panel discussions, and small group deliberations, the workshop participants collectively identified four sets of research questions organized around four system-level requirements that a high-functioning LHS must satisfy. Details regarding the workshop planning, participants, and process are provided at healthinformatics.umich.edu/lhs/nsfworkshop.

The participants were challenged to view the learning health system as a "system after next" [10]. As a "system after next", the learning health system can succeed only by creating novel combinations of role, process and technology that are unworkable without the collaborative environment provided by the learning health system. Put differently, the learning health system will be created with a new end in mind, by working back from the future, not by figuring out how to fix the various problems with the current non-learning non-system.

The research questions, that are the focal product of the workshop, developed through two rounds of breakout group deliberations. In each round, the participants divided into four groups with different composition. In the first round, each group examined one of four use cases describing the operation of a high-functioning LHS at scale. The use cases separately addressed LHS operation in support of health care quality, population health, personal health, and biomedical research. Each group identified the research challenges that must be addressed to realize a very high-functioning LHS capable of executing its assigned use case. The LHS use cases may be found at healthinformatics.umich.edu/lhs/nsfworkshop.

In the second round, each group focused on one of four broad system-level requirements for the high-functioning LHS:

- Requirement 1: An LHS that is trusted and valued by the public
- Requirement 2: An LHS that is economically sustainable
- Requirement 3: An LHS that is stable, certifiable, adaptable, and self-improving
- Requirement 4: An LHS that is capable of generating valid knowledge

The second-round groups consisted of participants with expertise directly related to each requirement; and each group identified the research challenges that must be addressed to realize a very high-functioning LHS capable of meeting the assigned requirement.

As part of the plenary session closing the workshop, each participant wrote the research challenge question stemming from the two day meeting that he/she felt was most important and that he/she could envision himself/herself being personally involved in addressing.

Following the workshop, the planning committee compiled, refined, and organized the research challenges into the form of questions. The questions were circulated back to all participants for review and comment. A modification of the four system level requirements used for the second round of the workshop's group deliberations formed the basis for organizing the key findings. The challenges and questions that follow form a broad scientific agenda for realizing a high-functioning LHS.

The Research Challenges and Questions

System Requirement 1. An LHS Trusted and Valued by All Stakeholders

Defining the Overarching Research Challenge

- There is a need to build trust and confidence in the LHS among the general public and all stakeholders.
- The LHS must generate significant actual and perceived value that benefits stakeholders involved (especially those incurring actual or perceived costs or risks to facilitate the LHS) as well as the public interest.
- Privacy, security, and transparency are key elements related to building public trust and generating value.
- Trust and confidence at all stages of the LHS operation are essential; from inputs to outputs (and outcomes).
- In order to be trusted and valued by all stakeholders, the LHS must excel at all system-level requirements.

Key Research Questions

Defining, measuring, and tracking confidence and trust

- In the context of an LHS, what are the operational definitions of confidence, trust, and trustworthiness?
- What methods will be needed to measure confidence, trust, and trustworthiness?
- What approaches will promote monitoring of confidence and trust at varying levels of scale, and thus enable improvement of the LHS over time?
- What messages and means of communication will build confidence and trust in the LHS across a broad range of stakeholders?

Building confidence and trust in the data inputs

- What properties of heterogeneous, routinely collected clinical data, including methods in place to manage and curate them, will engender confidence and trust in the knowledge generated by the LHS?
- Can the potential of a dataset to generate valid knowledge (fitness for use) be computed from the dataset itself? What metadata are required?
- What aspects and levels of data provenance, integrity and completeness are needed to enable inferences of different strengths: from descriptive portrayals to the enablement of predictive models?

Building confidence and trust in the process

- What methods and mechanisms will protect privacy and minimize the risk of breaches, and re-identification of data that has been de-identified, while still preserving its utility for learning purposes?
- What methods will be influential in developing, enabling, and applying fair information practices for the LHS stakeholders?
- What measures of transparency are needed to monitor and guide (continuous and rapid) improvement?
- How can we analyze motives and incentives for security breaches and attacks on the LHS and best structure governance and security policies and mechanisms to guard against them? What is the relationship between the real and perceived security of the LHS and stakeholders' levels of confidence and trust in it?
- What measures will build and maintain trust among individuals ("patients" and other data holders) in their control over access to and use of their data by others?
- How can the mechanism for consent for use of data be improved to enable individuals to make informed choices?
- What levels of privacy protection are desired by various stakeholders in light of known tradeoffs, risks, and benefits – as well as individual preferences? How can these be quantified and measured?
- How can the LHS be protected from subversion of all types, including breaches, re-identification, and hacking that disrupts its operation?

Generating value while building confidence and trust in the outputs

- What value metrics will assess the magnitude of improvement in health outcomes and cost-efficiency, as well as social and behavioral impacts, associated with LHS activities?
- What are the most effective means for communicating and disseminating messages that articulate the value the LHS generates and that build trust in the LHS?
- What factors affect the validity and trustworthiness of new knowledge generated by the LHS?

Building confidence and trust in the system as a whole

- What are the effects on confidence and trust in the LHS attributable to the key legal and ethical issues related to secondary use of data: issues such as dissemination of knowledge that is inevitably imperfect and the potential for misuse of results?

- What processes should be used to collaboratively develop an ethical framework for the LHS that instills trust and will prove adaptable as the LHS evolves over time?
- What special ethical, legal, social considerations are raised by the inclusion of "-omic" (e.g., genomic, proteomic) data in the LHS and how should they be addressed?
- How can the LHS be developed to ensure that it provides learning experiences for all stakeholders, but especially for the general public?
- What aspects of LHS governance will build confidence and trust in the system? Where is the balance point between the control needed to ensure compliance with whatever rules exist and the flexibility needed to enable evolution of a complex adaptive system? How can legitimate conflicting interests be resolved in a manner perceived by all stakeholders as fair?
- How can the LHS be designed, engineered, and operated as a self-defending and self-repairing system for purposes of protecting individual and institutional privacy and the integrity of data knowledge against malicious attack and accidental disclosure?

System Requirement 2. An Economically Sustainable and Governable LHS

Defining the Overarching Research Challenge

- There is a need to incentivize key activities constituting the inner workings of the LHS, including the sharing of data, the work of transforming data into information and knowledge, the sharing and dissemination of lessons learned, the underlying LHS governance activities, and more.
- The LHS will require revenue sources that must be governed in a way to ensure preservation of public trust and equitable treatment of success-critical stakeholders.
- Misalignment of incentives may pose challenges to a sustainable LHS; benefits may not necessarily automatically flow to the stakeholders whose efforts are requisite to creating the foundation for the LHS.
- At the same time, there is significant optimism about the potential value the LHS will generate – for the overall public good and for individual stakeholders.
- The value of the LHS should primarily be seen, ultimately, as the promotion of individual and population health at greatly reduced cost to our nation and people.
- Some types of value creation within the LHS will be dominated by the generation of benefits that flow to stakeholders beyond those involved in directly in data collection or analysis activities, requiring appropriate incentive structures and policies to be developed and applied.
- Developing sustainable business models that maximize the value created and appropriately incentivize and compensate those involved at various stages of its creation will be a key challenge to be addressed – getting to win-win-win outcomes.

Key Research Questions

Private sector incentives and markets
Predictive models and empirical studies will be needed to address the following:

- What is the relationship between incentive structures and data sharing and other key activities essential to the function of the LHS that will advance the public interest?
- How can we understand and predict the evolution of a marketplace for the LHS.
- What is the relationship between LHS governance and its ability to drive efficiency in this market?
- What characteristics of the marketplace will enable it to serve as a mechanism that promotes learning in the context of the LHS?
- How can certain costs associated with activities of the LHS, such as governance, be sustained and funded from private sources in ways that are efficient, equitable, and sustainable?
- To what extent will releasing public data spur LHS business models and how will alternative ways of selling, accessing, licensing these data would affect LHS operation?
- What is the relationship between: a) property rights structures and incentives attaching to new information and knowledge generated by the LHS and b) rates of new knowledge creation and rapid innovation in the context of the LHS?
- How can the science of networks, markets, game theory, and mechanism design inform the design and operation of the LHS?

The role of the public sector in the LHS

Predictive models and empirical studies will be needed to address the following:

- Which LHS activities have significant positive externalities and characteristics of public goods, thereby meriting significant government funding and involvement?
- What ingredients essential to standing up and sustaining the LHS have no private rationale for funding or are unlikely to be funded privately, and therefore should be considered for catalysis stemming from public funding?
- In what areas of the LHS are commons likely to be created and what government oversight and regulation would best serve the public interest and the interest of the LHS in such a context?
- What organizational dynamics would result from government playing a varying set of convener roles and would these serve to advance the development of the LHS?
- What are potential unintended consequences of public sector involvement and how could their risks best be mitigated?
- What lessons for the development of the LHS can be learned from other examples of public-private partnerships catalyzed by government leadership?

Conceptualizing value in the context of the LHS

- What measurement methods exist and what new methods need to be created and validated to assess:
  - LHS value propositions for various stakeholder groups.
  - The public good derived from the LHS.
- What value metrics can be assigned to privacy and measure the tradeoffs involved with policy choices concerning privacy?

- What value is the LHS uniquely capable of producing and how can it be identified, measured, and incentivized?

The LHS and the healthcare/health system as a whole

- What is the potential relationship between healthcare system reforms and financing strategies, as a function of how they reward value, and alternative business models for the LHS?
- What might be the expected effects of changes in medical education and other foundational aspects of the healthcare system on rapid learning and new knowledge adoption/application?
- What is the relationship between health care delivery innovations, such as team practice and patient engagement, and the extent and quality of learning in an LHS?
- In assessing the LHS, what methods will differentiate effects attributable to the LHS from changes in the health ecosystem?
- How can lessons from other rapid learning systems and Ultra Large Scale Systems in other sectors be applied to thinking about incentives and markets vis-à-vis the LHS?

System Requirement 3. An Adaptable, Self-improving, Stable, Certifiable, and Responsive LHS

<u>Shaping the Overarching Research challenge</u>

- The LHS will be a massive complex-adaptive socio-technical system, perhaps unlike any system built in this nation to date.
- Data sources and standards will evolve, changing dramatically over the time horizon in question.
- Some aspects of the LHS will function automatically, without human intervention in real time. Other aspects will require human intervention.
- The LHS is person-centric, not patient-centric. Data will flow, with appropriate consents and privacy protections, from all aspects of personal health, many coming from individuals and organizations other than those that provide healthcare.
- In order to be stable, the LHS must inspire confidence and develop trust. It must be able to learn and improve itself through its own operation.
- Appropriate methods of certification of critical system properties will increase trust in the system and help all stakeholders know what to expect from it; building such trust will itself contribute to the responsiveness and stability of the system.
- In order for certification to serve an optimal role in building trust, certification requirements and processes must accommodate and leverage system properties of the LHS itself.

<u>Key Research Questions</u>

An Adaptable LHS

- How does a socio-technical ecosystem integrate people and technology so as to address the inherent dangers of an autonomous turnkey system?

- How do we design information systems that employ digital computers, individuals, and institutions to carry out large-scale computations in ways that are evolvable and adaptable to ongoing changes in technology, knowledge, policies, and values?
- How can the system adapt to specific types of changes, including general changes in technology but also specific changes in requirements, specifications, data types and granularity, standards, devices, and data flows?
- What features of the system will best position it to incorporate innovation in data sources that include real-time monitors, sensors, and devices intended to augment bodily functions and directly improve a person's health?
- In the context of an adaptable system, how do we determine what adapts?
- How can a system adaptably ingest, manage, refine, and emit data from a rapidly growing source environment?
- How can the adaptation be controlled so that powerful stakeholders cannot exploit the system?
- How can adaptation arise from individual unplanned innovation at the fringes of the system, as distinguished from institutional innovation within the system?
- How can the system automatically infer and reconfigure itself to meet emerging and changing human and system requirements?

A Self-improving LHS (that builds trust in the system)

- How can the LHS be certified to be safe, effective, and adaptable at the same time?
- How are damages to the system (e.g. security breaches or data corruption) traced, tracked, and remediated?
- What processes are needed to promote self-improvement?
- How do we enable the system to analyze previous output and improve upon it?
- How do we enable the requisite human intervention to make the system (continuously and rapidly) self-improving?
- What are the self-improvement targets to measure and what are the appropriate metrics?
- What are the reporting requirements associated with and necessary for self-improvement?
- From where do standards for system performance (an LHS "score card") emanate? (If a system is to be self-improving, it needs to know what it is improving against.)
- How do we combine data generated in learning-from-doing and from observational data with careful experiments to be sure knowledge is scientifically solid and valid for stated purposes but also as widely useful as possible?

A Certifiable LHS

- Viewing the LHS as a socio-technical system, what are the roles of information, actors, equipment, processes, and organizations in a certification process?
- What should be certified: the system, the process, or the output or some combination of these?
- The LHS places unique demands on certification. What are the appropriate legal, ethical, policy, and technology frameworks for certification of this type of large-scale system?

- What evidence must be gathered about the development, design, and operation of the system and about the environment in which it operates to enable certification? How is this evidence gathered, organized, interpreted, maintained?

A Stable LHS (through trust and confidence)

- How can we build trust in the data as well as in the processes for correcting it (and how can we best develop these correction processes)?
- How can an infrastructure-enabled system that is continuously evolving, and over which there is limited central control, inspire the trust of all stakeholders: both in its dependability (security, reliability) and the validity of the knowledge it generates?
- How do we make data sufficiently self-describing so that, for example, the system might be able to identify, without human intervention, data that are most relevant to addressing a question?
- What architecture is needed to accommodate ongoing data representation and metadata standards?
- What are the key indices of trustworthiness that will be sufficiently compelling to generate public confidence in the LHS?
- How can the LHS explain its reasoning to stakeholders faced with major decisions?

A Responsive LHS

- How can the LHS deliver output appropriate to the context, incorporating notions of urgency and risk?
- How can we understand and develop mechanisms to balance the tradeoff between speed (rapidity) and accuracy?
- How can users of all kinds be educated to properly interpret and account for risks of inaccuracy?
- How can we create an agile capability for various stakeholders to ask questions to the system and receive timely and understandable answers?

System Requirement 4. An LHS Capable of Engendering a Virtuous Cycle of Health Improvement

<u>Defining the Overarching Research Challenge</u>

- The LHS should engender a virtuous cycle of continuous health improvement that touches the domains of health care quality, public health practice, decision making by individuals, and biomedical research.
- Achieving rapid learning requires sharing and integration of new knowledge in a timely and responsible manner.
- The LHS may require a new way of thinking about validity and credibility of findings. Learning requires drawing the strongest inference possible based on the data available, and questions the randomized controlled trial (RCT) as the gold standard of creating knowledge.

- Many different types of learning by many different actors can occur in an LHS, and such types of learning as well as the methods by which we learn will continue to evolve over time.

Key Research Questions

Rapidly creating knowledge that engenders learning

- Knowledge that engenders learning must be valid, trustworthy, actionable, and fit for purpose. How are these attributes defined and measured? To what extent do these measurement methods depend on domain (e.g. care quality vs. public health) and context?
- What existing analytical and inferential methods serve the needs of the LHS, and what needs of the LHS require new methods? In particular, in the context of an LHS at scale, what methods are needed to address bias, confounding, propagation of erroneous information, and other potential sources of mis-learning in the LHS?
- What rigorous methods for data description and new data quality metrics will capture the semantics that enable rapid learning from (big) heterogeneous data streams? What approaches will describe fitness for use within the context of a specific purpose? For example, what are the semantics for characterizing data and identifying and describing bias?
- What approaches can illuminate the question of how strong inferences need to be in order to achieve the status of "learning" in the context of the LHS?
- Knowledge that drives a virtuous cycle of improvement must be packaged with meta-knowledge such as where and how it should be applied and how its effect should be measured. What existing syntactic and semantic structures are applicable to knowledge management in this context and what new methods are needed?

Communication within the LHS: What mechanisms will enable communication of methods used and results obtained, in actionable forms, to all stakeholders with interest in the results?

- Results will be generated by the LHS through sophisticated models and analytical methods. What methods are required to generate messages appropriate to and understandable by different stakeholders? How can defensible explanations of inferences be generated?
- How do we develop ways to communicate results/information/knowledge generated to others who may wish to replicate (or build upon) the work done, as well as to the general public?
- How does a research team describe what they have done in ways that others can understand?
- How can the computational procedures employed in the system be documented in ways that are assuredly consistent, understandable, checkable, and repeatable, and how can the computational provenance of derived data be tracked from its points of production through consumption and use?

Building a smart system. How can the LHS become smart enough:

- To detect attempts to answer a question that is not answerable with the resources in the system?
- To detect the use of sub-optimal research/analytical approaches and possibly even correct them?
- To cross-validate putative new results with the extant body of knowledge
- To detect attempts to subvert appropriate use, and to report those attempts?
- Given the methods used, to quantify the uncertainty and epistemological gaps (through confidence limits and other methods) around the results?
- To be able to look for something and subsequently communicate to the system whether or not it found what was sought, so that the system itself can learn from such feedback?
- To ensure, through simulation and other methods, that we are not building things or collecting data that the system does not need?
- To explain its own inferences: how it learns, what it learns, and what it has already learned?
- To decide what knowledge to produce and how to select the learning methods that will produce actionable knowledge as quickly and efficiently as possible?
- To know when and how to employ, deductive, inductive, and abductive forms of reasoning, for both the testing and the generation of important new hypotheses; in particular to notice weak signals beyond human detection that might nevertheless point to important but previously unrecognized or just-emerging phenomena (such as unexpected side effects of new drugs, or weak but systematic health impacts of environmental exposures or socio-economic conditions)

Learning about learning, research about research

- What are the best approaches to measuring and understanding the predictive value of models resulting from big data?
- What new methods are needed to determine the value of knowledge generated by the LHS and resulting actions?
- What methods are required to measure the impact of the LHS on affecting behavior change at an individual and population levels?
- To what extent can well-designed research registries emulate the results of, complement, or replace traditional randomized clinical trials? With much refined diagnostic coding and ability to sharply reduce unmeasured confounders, can future research registries become increasingly useful, perhaps a new "gold standard" as they come to include much larger and closely matched, samples, with clinically rich, individualized, longitudinal data?

Key questions specific to the health domain

- To support health care quality improvement, how can the LHS develop a complex multi-dimensional model of a person's future health status and communicate it to them or their trusted healthcare provider to promote their health and wellbeing?
- To support public health practice, how can the LHS embrace essentially all aspects of the health of all individuals, including assessing environmental, occupational,

lifestyle, and healthcare determinants of health and evaluation of the impact of a wide array of intended and unintended changes and interventions?

- To support decision making by individuals, how can the LHS optimally define "patients like me" in the context of answering questions such as "Based on the experiences of other patients like me, what are my treatment options and tradeoffs, and which treatment is likely best for me, given my values and preferences?"
- To support biomedical research, how can the LHS enable us to go from myriad point data in EHRs and other data sources on the health experiences of individuals over time to descriptions of clinically meaningful events/processes/constructs?

Toward a Science of Learning Systems

The challenges and specific research questions--organized into four areas corresponding to requirements a high-functioning LHS must meet--constitute one level of expression of the workshop's findings. At a higher level, the findings in their totality and their multiple interactions give rise to a novel scientific perspective. Rising to the challenge of the LHS may require a new science of large-scale learning systems best seen as an evolution from the science of information systems, through a science of cyber-physical, and ultimately to a science of cyber-physical-social ecosystems. The applicability of this new perspective may transcend the health domain that was a focus of this workshop.

Information Systems: The revolution in computing over many decades has transformed the *information* sectors of our society. Commerce, finance, entertainment, communications, science, advertising, media, and other fields, have been profoundly reconfigured by the digital representation, processing, storage, transmission, and use of information enabled by advances in computer and information science and engineering.

Cyber-Physical Systems: Computer science and engineering are now revolutionizing the *physical* systems our society, driven by advances in sensors, the ongoing miniaturization of digital computers, the wireless networking of edge devices to massive cloud computing systems, computer control of robotic machinery, and advances in such areas as machine learning and artificial intelligence. Advances in the production of cyber-physical systems are becoming visible to the public. They include sensor-laden cell phones, self-driving vehicles, and smart power grids, smart buildings and cities, and a dramatic revitalization of manufacturing in the United States driven by advanced manufacturing methods.

Cyber-Social Ecosystems: With scientific research and applications of information and cyber-physical systems now firmly established, the next major dimension of our society to be revolutionized by computer and information science and engineering is that of large-scale, decentralized, *human*-intensive, information processing systems.

In the realm of cyber-social ecosystems, scientific research will seek to develop and validate fundamental principles to enable the effective design, analysis, construction, governance, operation and evolution of ultra-large systems that deeply integrate information and physical phenomena, in which complex information processes are effected

by evolving networks of people, institutions, and digital computing machines.

A key challenge will be to find ways to "engineer" such systems to carry out desired large-scale computations—for example, providing proprietary or sensitive data as inputs to the LHS so as to obtain desired knowledge--when the system components--including individuals, private and public institutions, and computing devices--generally operate in their own self-interests, or as required by social norms, laws, or regulatory frameworks. The notion that we might design social and institutional networks, enabled by digital infrastructures, to carry out large-scale evolving computations in the interest of the nation and its citizens is new and compelling.

A national-scale Learning Health System will have to be understood and designed as such a cyber-social eco-system. The system as a whole, not just the digital infrastructure, but also networks of people and institutions, will have to be understood not just as users of a technological infrastructure, but as parts of an information processing ecosystem. So, for example, trust in the digital infrastructure is necessary to induce privacy-sensitive citizens to agree to permit information flows that are critical enablers of the desired computations. Institutions with commercial interests in the clinical data that they generate will have to be incentivized or compelled to transmit such data under certain conditions.

Furthermore, the LHS infrastructure will be understood not just as a machine carrying out operations for a certain institution, whether government or private, but also as a platform for innovation across the healthcare ecosystem. It will provide *data as a service* enabling a host of unforeseeable innovations in data-, knowledge- and evidence-driven healthcare, bio-surveillance in the public interest, and health-related research and development. Just as no one involved in the creation of the internet in the 1970s could have reliably predicted the emergence of Google, Amazon, Facebook, or Twitter, no one today can predict what the LHS will catalyze in the decades to come. But if it is properly designed not only as a system but as a platform architecture, then it can reliably be predicted to have profoundly catalytic effects on future innovation for health, healthcare and the economy.

Achieving a learning system will require discovery of methods to loosely couple the parts of a complex ecosystem in ways that enable them to align with one another over time. Just as the internet easily survives the loss of component systems due to loose coupling, and just as it fosters the emergence of completely unforeseen functions, so an architecture enabling loose but influential coupling in the health ecosystem will allow each component of the system to participate in the evolving network while maintaining its own autonomy, pursuing its own interests, and optimizing its own performance, while also being able to rely on trusted other parts of the system doing their parts in concert.

The broad areas of research that emerged during the workshop provide examples of methods to enable loose coupling and iterative alignment. These include methods to estimate the magnitude of the value proposition of the learning health system. According to AHRQ's National Quality Report [11], healthcare quality is improving at about 2% per year in this country. What is the upper bound given a learning health system, i.e. a Moore's Law for the learning health system? Could quality double every five years as the NIH budget did

at the turn of the century? Or is a doubling every ten years a structural upper limit? What are the key determinants such as the cycle time from discovery to translation into practice, or the percent of data from practice that is harnessed for discovery? What methods could be used to forecast a collective upper bound based on assumptions about various determinants?

Another area includes novel frameworks and methods to align policy, governance and ethics. Take management of use and abuse of the learning health system as an example. What safeguards can be architected into the enabling technology? Which aspects can be managed through shared governance? Which aspects are best handled through professional norms and ethics? What is the role of policy? How do research advances in one of these areas affect what is required of the others. What attributes (e.g. provenance) can be included with the data to reduce what needs to be managed. What methods might "immunize" the learning health system from subversion by enabling it to detect and reject suspicious behavior? What methods would enable stakeholders to explore, understand, judge and improve the learning health system?

A third area includes novel approaches to combining "big data" analytics, research study designs and modeling methods. For example, what combinations of data fusion and machine learning methods, study designs and statistical methods, and analytic frameworks could support rapid cycle hypothesis generation, hypothesis testing and dynamic inferencing? As inferences emerge, how can they be reflected in models to test whether they fit together? How can these models then be used for prediction at scales ranging from an individual or organization to society?

In sum, the LHS raises a host of deeply important, difficult, and unresolved research issues at the intersection of computer science, software and systems engineering, the behavioral and social sciences, healthcare, and public health. How can the diverse, often conflicting, and generally evolving value propositions of all success-critical stakeholders be understood and factored into the negotiation of (eco)system requirements and specifications? How do we extend the current science of computing, which remains very focused on digital machines and logic-based digital computations as agents of information processing, to a science of information processing carried out in, and by, ultra-large scale, digitally enabled but also intensely human, cyber-social ecosystems? How do we integrate social mechanisms, such as rules, regulations, laws, standards, governance mechanisms, incentives, mechanism designs, and mechanisms for educating and training people and into a theory of computing systems, architecture, design, implementation, evaluation, certification, monitoring, repair, economics, and system evolution?

Addressing the deep scientific questions related to the LHS will clearly require not only collaboration among multiple and diverse stakeholder types, but will also require research, learning, and innovation to occur at the intersection and fusion of multiple and diverse disciplines. No one discipline alone has the tools or knowledge requisite to attack this multi-faceted sociotechnical challenge. Crucially, neither the research community currently organized around the National Science Foundation, *nor* that around the National Institutes of Health, has the requisite intellectual capabilities to produce the required innovations.

The LHS requires a new and significant crossing of capabilities that today are present only in these largely disconnected communities. To realize this goal, the NSF will have to continue to expand its embrace of health and healthcare as a driving domain for fundamental research. The NIH will have to dramatically expand its embrace of computing as a transformative power for biomedical research and healthcare innovation and delivery in this century. Additional mechanisms, of which this workshop was one starting point, and resources to bring together the largely disjoint research communities will be needed.

The intellectual merit and potential broad impacts of the innovations that will be driven by investments in a Learning Health System are enormous. Whether this combination of challenges and the methods to address them nucleate to form a "new science of learning systems" or whether these problems are addressed through an emergent network reflective of existing disciplines, the participants in this workshop were deeply excited by the vision and the possibilities that emerged from it.

References

[1] (IOM) Medicine, I. of. (2011). Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care. In C. Grossman, B. Powers, & J. M. McGinnis (Eds.), *The Foundation for Continuous Improvement in Health and Health Care Workshop Serier Summary* (pp. 1–311). National Academies Press.

[2] Smith, M., Saunders, R., Stuckhardt, L., & McGinnis, J. (2012). *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*. (M. Smith, R. Saunders, L. Stuckhardt, & J. McGinnis, Eds.) (1st editio., pp. 1–450). Washington, D.C.: National Academies Press.

[3] National Research Council. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. Washington, DC: The National Academies Press, 2011.

[4] Milstein, A. (2012). Code Red and Blue—Safely Limiting Health Care's GDP Footprint. *New England Journal of Medicine*, *368*(1), 1–3.

[5] Dzau, V. J., Cho, A., ElLaissi, W., Yoediono, Z., Sangvai, D., Shah, B., … Udayakumar, K. (2013). Transforming Academic Health Centers for an Uncertain Future. *New England Journal of Medicine*, *369*, 991–993.

[6] Technology, P. C. of A. on S. and. (2010). *Report to the President Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward* (pp. 1–108). Washington, D.C.

[7] Etheredge, L. M. (2007). A Rapid-Learning Health System. *Health Affairs*, *26*(2), w107–18. doi:10.1377/hlthaff.26.2.w107

[8] "CancerLinQ." *ASCO.org*. American Society of Clinical Oncology, n.d. Web. 06 Nov. 2013.

[9] Summit, D. at the M. 2012 L. H. S. (2012). Core Values Underlying a National-Scale Person-Centered Continuous Learning Health System (LHS). In *Core Values Underlying a National-Scale Person-Centered Continuous Learning Health System (LHS)* (pp. 1–2).

[10] Gorry GA. Envisioning the system after next. EDUCOM Rev. 1997: 30-3

[11] *National Healthcare Quality Report 2012*. (2013). Rockville, MD.