Toward an Information Infrastructure for Global Health Improvement

C. P. Friedman¹, J. C. Rubin¹, K. J. Sullivan²

¹ University of Michigan
² University of Virginia

Introduction

We live in an era of profound challenges and significant opportunities for global health. Among them, the rate of health data generation is increasing almost exponentially. For example, experts worldwide believe that, by 2020, people and health systems in the U.S. and around the world will generate data 50 times more rapidly than they did at the start of the decade [1, 2, 3]. Biomedical knowledge generation worldwide has been accelerating at a correspondingly profound rate [4, 5]. One estimate suggests that the doubling time of biomedical knowledge—which stood at 50 years in 1950 and 7 years in 1980—will decrease to an estimated 73 days by 2020 [6]. The year addressed by this IMIA Yearbook, 2017, coincidentally marks the 17th anniversary of a study documenting the 17-year lag between the generation of new biomedical knowledge and its widespread application to health practice [7], and also the 17th anniversary of a landmark study describing the prevalence and impact of patient safety issues that take a major toll in human lives [8]. These developments occur against a backdrop of declining individual and public health indicators coupled with pressures to decrease public and private health expenditures, but also a backdrop of potentially disruptive technological advancements [9, 10, 11, 12, 13].

In response to the challenges, as well as to seize the opportunities, the past decade has witnessed the introduction and evolution of the concept of Learning Health Systems (LHSs) [9, 14, 15], capable of continuous self-study and improvement. LHSs embrace but also transcend the secondary use of health data, the theme of the 2017 issue of the IMIA Yearbook, pointing to what may be possible through complete cyclical processes that mobilize health data, analyze it to create new knowledge, and apply that new knowledge to improve the health of individuals and populations. LHSs are supported by infrastructures that enable these processes to take place routinely and with efficiency of scale and scope. Augmenting their potential to improve health by analyzing data and acting on it, LHSs also hold the potential at a meta-level to improve their own capacity to learn.

In this essay, we describe our vision of LHSs, connect this vision to the concept of infrastructure, and argue for the development of LHSs that would occur primarily through coordinated composition of organizational, network, and regional initiatives as opposed to centralized, top-down, initiatives. While initiatives that are imposed top-down generate immediate attention and may lead to localized and short-term benefits, we take the position that widespread enduring benefits to health will result from a more deliberate and evolutionary process of infrastructure co-production in which the full spectrum of stakeholders are directly engaged. Above all, this essay is an expression of hope and optimism for a future of better health for all people, suggesting a pathway—albeit a pathway fraught with challenges—to this future.

Learning Health Systems and Learning Cycles

The concept of LHS was first advanced by the U.S. Institute of Medicine (now the National Academy of Medicine) in 2007 [16, 17]. In the ensuing 10 years, the concept has gained increasing attention, initially in the U.S., but currently and progressively around the world. This is seen, in part, in a substantial literature with 1,940 Google Scholar citations retrieved from a search on the term...
“Learning Health System” conducted by the authors in January of 2017.

In the U.S., several recent developments point to increasing interest in and development of LHSs. A national LHS Summit, held in 2012, established consensus around 10 LHS Core Values that have been formally endorsed by 109 organizations [18] (see Table 1). In 2015, the U.S. Office of the National Coordinator for Health IT, citing these LHS Core Values, established a nationwide LHS as the pinnacle goal of its 10-year Interoperability Roadmap [19], and other federal agencies have expressed strong endorsement of the concept [20, 21]. The U.S. National Science Foundation (NSF) has supported workshops to identify the research challenges to achieving a high functioning LHS [22, 23, 24, 25]. More recently, the Computing Community Consortium (CCC) [26], which advises and works with national research policy makers to understand and shape the future of computing and computing research, has funded a series of three workshops on the broader topic of Cyber-Social Learning Systems (CSLSs), of which LHSs are a special case [25]. The Patient-Centered Outcomes Research Institute (PCORI) has developed a health data and analytics network at national scale [27, 28]. The literature reveals a panoply of reports of individual organizations seeking to achieve the capabilities associated with LHSs [29, 30, 31, 32, 33, 34].

Interest in LHSs has spread across the globe [35, 36, 37, 38]. Specifically, in the European Community, the TRANSFoRm project has addressed some of the challenges of achieving a robust infrastructure for LHSs [39, 40]. The European Institute for Innovation through Health Data seeks “to tackle areas of challenge in the successful scaling up of innovations that critically rely on high-quality and interoperable health data” [41]. In the U.K., the LHS concept has become a beacon for health improvement [36, 42, 43, 44]. The Swiss government has recently announced a national LHS initiative [45]; and in Asia, collaborative efforts joining Japan to Taiwan have resulted in an incipient Consortium for Asia Pacific Learning Health Systems [46, 47].

We frequently refer to LHSs in the plural because there is no single accepted definition of the term, and it is almost certain that there will not, in the end, be a single monolithic LHS spanning the globe, but rather a global system of more and less tightly integrated LHSs. LHS proponents also differ in their interpretation of the “H” appearing in the abbreviation. Some interpret the “H” restrictively to stand for “health care”, focusing the LHS concept on improving the delivery of care to individuals and aligning it almost exclusively with health care quality, safety, and research. Others interpret the “H” more broadly to connote “health” in recognition that an LHS can promote public and population health.

Information is central to the LHS concept, and an LHS can improve health through successive iterations of “learning cycles”. The concept of iterative improvements achieved prominence through the efforts of W. Edwards Deming and has subsequently been elaborated in a robust literature [48, 49, 50]. At the highest level of abstraction, learning cycles convert data to knowledge (D2K), apply that knowledge to influence performance (K2P), and document changes in performance to generate new data that seeds the next iteration of the cycle (P2D) (see Figure 1). Early visions of Learning Healthcare Systems focused exclusively on the reuse of digital data generated by care delivery to drive the D2K component of the learning cycle [16]. The broader vision of Learning Health Systems recognizes that a much wider array of data sources—including purposefully collected data outside of care experience, such as geospatial data or data addressing the social determinants of health—can be important components of the learning process.

As an illustrative example, a learning cycle might be organized to reduce the incidence of falls by residents of long-term care facilities. Participating facilities—forming a learning community—could be those in a given region, those owned by a particular company, or a self-organized group of volunteers. The D2K segment of the cycle would engage participating facilities in describing what practices they currently employ to prevent falls. These data, aggregated across the community, would be analyzed in relation to each facility’s fall rate. The results of the analysis would likely identify some practices that could potentially reduce falls. The community would review these results and decide which have sufficient credibility to generate recommendations for change. This engages the K2P segment of the

<table>
<thead>
<tr>
<th>Table 1 Learning Health System Consensus Core Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.) Person-Focused</td>
</tr>
<tr>
<td>2.) Privacy</td>
</tr>
<tr>
<td>3.) Inclusiveness</td>
</tr>
<tr>
<td>4.) Transparency</td>
</tr>
<tr>
<td>5.) Accessibility</td>
</tr>
<tr>
<td>6.) Adaptability</td>
</tr>
<tr>
<td>7.) Governance</td>
</tr>
<tr>
<td>8.) Cooperative and Participatory Leadership</td>
</tr>
<tr>
<td>9.) Scientific Integrity</td>
</tr>
<tr>
<td>10.) Value</td>
</tr>
</tbody>
</table>
cycle, wherein each facility would receive recommendations, ideally customized to its own specific circumstances. Each facility would then respond to the recommendations, in whole or part, and perhaps make some additional changes stimulated by the recommendations. This engages the P2D segment of the cycle that would document the changes made at each facility, and resulting effects on fall rates, which initiates the next iteration of the cycle.

Characteristics of a Fully Functional Learning System

For the purposes of the discussion to follow, we will describe an LHS as a socio-technical system with the primary goal of significantly and safely improving health while reducing costs and other harms. An LHS is a cyber-social system composed of people and technology [51]. A fully functional LHS will exhibit all of the following five attributes [52]:

1. The health-related characteristics and experiences of very large numbers of persons, along with other relevant data types, are securely available as data to learn from. The D2K component of learning cycles requires data in sufficient quantity and of sufficiently high quality to generate credible findings. Efficient and scalable LHSs require all or a significant fraction of this data to be routinely collected and persisted in data marts that can be centralized or federated, institutionally- or patient-controlled. Data can describe individuals who play a variety of roles; it can also describe processes, organizations, and environments. Control over access to and use of personal data is governed in ways that protect the equities of all stakeholders including especially the individuals to whom the data pertain.

2. Best practice knowledge derived from these data is available to support health-related decisions and actions by individuals, care providers, as well as health service managers, planners, policy-makers, and payers. To enable the K2P function, the knowledge generated by LHSs must itself be represented and persisted in machine-executable forms. An LHS must know what it knows, and be able to compute from that knowledge advice to decision makers that is: specific to the decision setting, associated with a level of uncertainty, and, ideally, customized to the recipient of the advice [53]. The advice generated by the K2P function transcends the traditional clinical decision support paradigm focused on recommending care for a given patient. K2P can offer recommendations to all participants of the health ecosystem and address changes that may be implemented at the individual, group, organizational, or system levels.

3. Learning and health improvement are routine and continuous processes. Every health system is capable of self-study and improvement when faced with an externally imposed stimulus such as potential loss of accreditation, or an internal crisis. By contrast, in a learning health system, there is a shared recognition of the need for continuous improvement in the absence of crisis or external imposition, driven by a shared belief that the LHS approach is the best means to achieve health improvement at lower cost with greater safety. As such, an LHS will be characterized by multiple learning cycles ongoing simultaneously [54, 55].

4. Infrastructures enable the routine execution of multiple learning cycles. Even though each of the simultaneous learning cycles in an LHS addresses a unique health problem, all learning cycles depend on a common set of supportive services. The technologies, policies, and standards comprising these services constitute the infrastructure for the LHS. If these infrastructural services are shared across learning cycles, LHSs can function with economies of scale and scope, as discussed below.

5. Stakeholders within the system see value in the above activities and view them as part of their culture. Continuous learning is an intensely human process. Learning cycles will be undertaken by diverse groups of individuals, forming learning communities, who are driven by the motivation to solve problems that are barriers to improving health while reducing costs, harms, and other externalities. External motivations can incentivize participation in these learning communities, but the majority of the motivation must originate within the culture of the organization. If asked...
why he/she participates in the activities of an LHS, perhaps the best answer from a community member would be: “This is who we are and this is what we do.”

The above five characteristics can serve many purposes. They can serve as important guideposts to inform the strategy for realizing LHSs, for identifying key success factors, for quantifying and measuring progress and developing relevant success metrics, and for enabling LHSs to learn from one another. They can also be considered to comprise a set of continua suggestive of a capability-maturity model [56].

Infrastructures, Scale, and Informatics

The fundamental process of an LHS is the learning cycle as depicted in Figure 1. From an informatics perspective, a set of learning cycles, each addressing a specific health problem, can function as a learning system when they are supported by an infrastructure serving multiple simultaneous learning cycles, as depicted in Figure 2. Some learning cycles will progress more rapidly, others more slowly, as a function of the health problem being addressed. An infrastructure provides a set of integrated services, each service supporting a component of the learning cycle and all services together establishing a smoothly articulated workflow from each stage to the next in the learning process.

As previously noted, infrastructure endows the LHS with economies of scale and scope. Without infrastructure, each learning cycle is figuratively a “tub on its own bottom” requiring its own concepts, methods, tools, and support systems to undertake the necessary activities of D2K, K2P, and P2D. Without infrastructure, implementing each new cycle requires people specific to that cycle and the deployment of cycle-specific methods, tools, and processes. Under that circumstance, implementation of the next cycle will cost almost as much as did its predecessor. Also, methods developed from an initial cycle may not transfer to successive ones because the methods of the initial cycle may be over-fitted to the health problem it addresses.

Figure 3 illustrates the services required by an LHS infrastructure and provides a schematic representation of a complete infrastructure. The inner circumference of the figure displays the fundamental D2K, K2P, and P2D components of the learning cycle. The boxes on the outer circumference depict the infrastructure services required to support each stage of the complete cycle. The circle that connects the boxes illustrates the requirement that the services articulate into a smooth workflow.

The services depicted in Figure 3 support and engage people, process, policy, and technology. In their totality, they emphasize that D2K infrastructure services are necessary to, but not sufficient for, supporting an LHS. Of special significance to a complete infrastructure are K2P services required to persist and manage knowledge so the system “knows what it knows”. K2P services enable the system to represent knowledge in machine-executable forms so it can be rapidly updated as the system learns, it rapidly generates advice to inform decisions, and it tailors the advice to the needs and characteristics of the recipients [53]. Methods drawn from the behavioral and implementation sciences to promote practice change complete the services necessary to support K2P [57, 58].

Learning cycles can occur at varying levels of scale. They can be undertaken by a single organization, by networks of otherwise independent organizations, by specialized disciplines that span organizational, legal, and geographic boundaries, and across geographical regions varying in size, from counties to states/provinces to entire nations. Because the actions necessary to execute learning cycles are, to a significant degree of approximation, invariant across these levels of scale, LHSs exhibit important fractal-like properties of self-similarity [59, 60]. The infrastructure supporting LHSs, displayed in Figure 3, is capable of delivering the same services at any level of scale. This self-similarity property has strong implications for the evolution of LHSs at large scale, as will be discussed below.

Fully functional, integrated LHS infrastructures, as suggested by Figure 3, do not yet exist, nor do architectures that would
underpin such infrastructures. Nonetheless, pieces of LHS infrastructure that provide specific services shown in Figure 3 are emerging. For example, PopMedNet and I2B2 are widely deployed D2K infrastructure components [61, 62]. Apervita [63], Semedy [64], and the Knowledge Grid [53], provide K2P services by representing and curating biomedical knowledge in machine executable forms and generating from that knowledge messages tailored to support decisions taken by care providers, patients, and/or managers. Systems such as Deliberative Dialogue can support the work of learning communities [65]. The European Community’s TRANSFoRm project [39] provides scalable infrastructure that supports both D2K and K2P. Further maturation of these existing services, development of infrastructure components supporting other needed services, and their integration into a coherent workflow are the fundamental challenges of an LHS, comprising some of the most interesting and important challenges faced by informatics in the coming decade. Integration of these services will require solutions to many additional problems in software and system design, including the formation of basic architectural design rules, the standardization of representation and especially the interpretation of data and knowledge that flow across the services, the distribution of local control and system-wide learning that both rely on and inform local control, the need to accommodate and account properly for data of varying quality, and the need to assure the trustworthiness of system operation in the face of daunting challenges, ranging from natural disasters to human error and attacks on data and system function coming from capable adversaries. Among other things, the LHS Core Value of Scientific Integrity (see Table 1) hinges on assaying the quality of data and the analytic methods used to learn from them; and societal trust in LHSs will in turn hinge on the presence of correct and compelling arguments that these systems have been constructed according to the highly demanding standards of quality, resilience, and integrity.

**Fig. 3** Prototypic LHS Infrastructure Services

---

**Toward a Health Improvement Ecosystem**

We sit collectively at a propitious moment for the development of LHSs and the realization of their potential to improve individual and population health. As discussed previously, there is an intensifying collective interest in and enthusiasm for the LHS concept around the world. The vision of LHSs that are enabled by shared sets of socio-technical architectural and infrastructure elements points to how a global progression toward increasingly functional LHSs may occur.

To understand this, it is important to recognize that the LHS is much more than a “moonshot”, although it resembles the 1960’s moonshot in its ambitiousness and, perhaps also, the enthusiasm it is generating. Paraphrasing John F. Kennedy, the widespread achievement of LHSs will not be one organization “going to the moon”, it will be the entire world [66]. But unlike the original “moonshot”, LHS is not a direct build-out toward a single sharply defined end goal that we already largely know how to achieve modulo certain engineering details. LHSs are architectural, infrastructural, organizational, and ultimately societal means to achieve a wide range of ends that themselves comprise a moving and evolving set of targets. LHSs will never be complete and will always continue to evolve. LHSs represent a socio-technical, as opposed to almost purely technical, enterprise, far more complex than sending a human to the moon and safely back. There are deep open scientific, engineering, and design questions, invoking a broad range of sciences, that must be addressed to achieve highly functional LHSs [22]. By sharp contrast, when President Kennedy announced in May of 1961 the goal of putting a person on the moon by the end of the decade, he did so knowing that the Saturn booster rocket, representing the most fundamental engineering challenge, had been built and would be test-flown four months later [67].

This conception of LHSs as ultra-large scale, socio-technical systems is inconsistent with a belief that any one entity, any one approach, or any one design can be imposed upon the existing health ecosystem to create an LHS. We offer instead an evolutionary vision of the journey toward LHSs that will invoke
different modes of thinking than have been customarily applied to health and informatics problems. These new modes of thinking will be multi-disciplinary [68] and will invoke important concepts such as co-production [69, 70]; “chaotic” organizations [71]; ultra-large-scale systems [72, 73]; co-operation [74, 75]; evolution at whole-industry scale; social sensing [76]; learning over highly distributed, locally autonomous nodes [77]; and methods for developing ultra-high-assurance software-intensive systems, such as those employing emerging techniques of proof engineering [78, 79, 80].

Progress toward LHSs will be an incremental journey toward an ecosystem that exists only in fragmented elements at the present time. A vision for how this evolution may occur includes three channels of development: 1) continuing development of the existing elements, each of which, in its own way, comprises a component of LHS at scale, 2) integrating the elements into more complete and scalable working systems, and 3) developing an international-scale program of basic and applied multi-disciplinary, scientific, engineering, and design research to inform all of the above.

1. Continuing development of the existing elements. Existing LHS elements include organizations that have achieved LHS-like capabilities and deployed many of the LHS infrastructure services depicted in Figure 3. In the U.S., individual organizations such as Intermountain Healthcare [81] Johns Hopkins [60], and the Mayo Clinic [82], have developed programs following the LHS model; and many more, by virtue of endorsement [18] of the LHS Core Values, have signaled future steps in that direction. At a higher level of scale, organizational networks exhibiting limited LHS properties are rapidly appearing. These include many of the PCORI Clinical Data Research Networks [83], the NIH Collaboratory [84], disease-focused initiatives such as CancerLinQ [30], and the Collaborative Quality Initiatives (CQIs) in the State of Michigan [85]. European initiatives such as TRANSFoRm [39] and EHRA4CR [38] complement this list.

2. Integration and scaling. A key necessary step toward scalable and high-functioning LHSs will require the adoption over time of common infrastructural elements and overarching cyber-social system architectures across levels of scale. As this process occurs, it will enable entities at one level of scale to function virtually at higher levels of scale, in much the same way that individual organizations participating in PCORI Clinical Data Research Networks, through their use of a common data model and deployment of the PopMedNet data aggregation software, and inherent the ability to conduct clinical research at national scale [86]. While local implementations will necessarily and desirably be unique to given institutions, they will increasingly reflect and benefit from shared architectural characteristics: support for shared APIs and processes, conformance to consensus design rules [87], standardized formats, and interpretations for data and procedures [88]. This in turn will require agreements on minimal constraints that all participants agree to be bound by so that the system can work as a whole even while substantial autonomy to act and evolve is left to the individual components. The processes of integration and scaling will require effective mechanisms of governance, which themselves will evolve with experience. In particular, the world will need to learn how LHS policy making and governance functions should be distributed across private sector and governmental entities. Experience with the U.S HITECH program [89], the European epSOS program [90], the U.K. National Information Governance Board [91], and other similarly-focused initiatives, may provide useful experience to guide the way to effective governance.

3. Expanding research. As noted previously, the challenge of achieving high-functioning LHSs raises a broad array of open research questions [92, 93]. Major advances in deep machine learning, artificial intelligence, parallel processing, connectivity, “big data”, and other areas of computing must be complemented by equally profound advances in behavioral, social, and organizational sciences embracing ethics, policy and governance, legal and regulatory mechanisms. Particularly in light of the Core Values of the LHS described earlier, ethical dimensions of LHSs are of particular importance and have been addressed in several recent publications [94, 95]. Knowledge derived from many disciplines, from computer science and engineering, to the social, behavioral, and economic sciences, to architecture and the humanities, will be necessary for LHSs to realize their positive disruptive potential. Research initiatives that join the public and private sectors, such as The Farr Institute of Health Informatics Research of the U.K. [96] may provide important models for stimulating and supporting the necessary level of interdisciplinary research.

In conclusion, the authors recognize that by conventional modes of thought, LHSs likely appear quixotic and the products of magical thinking. This essay has sought to bring LHSs into the realm of the achievable by describing mechanisms through which belief in the LHS concept, adoption of a multi-stakeholder evolutionary approach, careful attention to work already done, dedicated trial-and-error implementation, application of new ways of thinking, and rigorous research could combine to generate sustained progress toward increasingly functional LHSs in pursuit of improved global health. We indeed need to learn how to learn, and this will not in any way be easy, but momentum is increasing. To the extent that the international informatics community has been searching for an ultra-grand challenge to address—a challenge that invokes every aspect of the field, integrates with all other success-critical disciplines, and unites all members in a global enterprise—it has found one in the Learning Health System.

References


5. Alper BS, Hand JA, Elliott SG, Kinkade S. How


33. Health Affairs. Using big data to transform care; 2014 Jul 1; Volume 33, Issue 7. http://content.healthaffairs.org/content/33/7/309


49. Senge PM. The fifth discipline: The art and practice