A *Cyber-Social Systems* Approach to the Engineering of Ultra-Large-Scale National Health Information Systems

Kevin Sullivan
University of Virginia
Institute of Medicine, August 18, 2011 (revised August 22, 2011)

Introduction: An Ultra-Large-Scale Systems Perspective on National Health Information Systems

In previous discussions here at the IOM, a group that I have been working with, including Mary Shaw, William Knaus, and Richard Marks, has argued that work in the area of ultra-large-scale systems, or ULS systems, can help us to better understand how to define, design, deploy, operate and evolve a national health cyber-infrastructure system. The concept of ULS systems was first presented in a report that was published by the Carnegie Mellon Software Engineering Institute, based on a study conducted by a team of computer scientists from around the country. Mary Shaw and I were members of that team.

ULS systems are complex, information-intensive, large-scale amalgams of people, organizations, technology, policies, objectives, incentives, finances, laws, and so forth, that provide services that are essential for meeting major societal needs, in such areas as defense, healthcare and energy.

ULS systems, in turn, are generally supported by digital computing infrastructure systems and systems of systems. Decades of painful experience show that attempts to develop or even to significantly improve such systems are often at, or even beyond, our abilities as systems and software engineers and domain experts. Many such efforts have come to grief, often at the cost of billions of dollars. The Army's Future Combat Systems (FCS) project, for example, was meant to integrate a large number and a diverse range of Army assets. That project, among many similar, floundered and was ultimately cancelled.

The ULS Systems report emphasized several points about such projects. First, they are at high risk of failure due to systems complexity. Second, an important source of risk is that traditional engineering approaches are based on assumptions that do not hold in ULS systems: e.g., that design is centrally controlled. A national health information network would instead integrate many autonomously evolving systems. Third, the ULS systems of the future, and the kinds of digital information technology systems that they require, will be even more demanding than today's systems. Fourth, we need to significantly rethink our approaches to software and systems engineering in such complex environments.

The national healthcare system is clearly an ultra-large-scale system in the sense stated here. More to the point, building a national-scale cyber-infrastructure for healthcare information, such as the National Health Information Network (NHIN) as authorized by the HITECH Act (part of the American Recovery and Reinvestment Act of 2009), must be understood as project in this class of cyber-infrastructure projects for ULS systems. Such a project demands not just a rigorous, modern software and systems engineering effort, but an approach at the cutting edge of our understanding of information processing systems and their development and deployment in complex socio-technical environments.

The purpose of the ULS systems study and report was to gain a better understanding of the essential characteristics of such systems in order to better understand how to approach them from a software and systems engineering perspective as well as from a domain (e.g., healthcare or energy) perspective.

A key part of the report enumerates major characteristics of ULS systems that tend to distinguish them from more ordinary domains. Your briefing book lists a range of such ULS Systems characteristics. For example, they tend to include large numbers of software components, sensors, computers, networks, and other such devices spread across organizational boundaries.

A Refined Perspective: Cyber-Social Systems

In this talk, I will focus on one particular characteristic of ULS systems. From it I will derive new insights into requirements for a national health information system, and from these requirements I will argue for the need for some new thinking about systems design. I will then briefly describe a small model of a national scale system that my students and I have been building in our lab to gain some early insights in these areas. I conclude with observations on engineering of the NHIN and some recommendations.

The characteristic of ULS systems that I want emphasize today is that, in ULS systems, people are not just users of information processing technology systems, but are actually parts of the system. The ULS Systems report called this the *erosion of the people/system boundary.* What does this really mean? The ULS systems report was not entirely clear on this matter. In this talk, I will propose that in these systems, people, networks of people, and organizations *are,* in a very fundamental sense, information processing elements in the systems. The system as a whole, and in particular people and social networks of people within the system, *compute.* Computing is not limited to the digital technology components. Indeed, the people in the system perform the most difficult and demanding computations.

Should I be tested for elevated cancer risk? If I am at elevated risk, should I take medicine? If so, which one? What is the risk of breast cancer for this patient? What procedure should I perform on this patient now? How shall I coordinate with my colleagues in the care of this cancer patient? These are decisions that made by way of what we can think of as *human computations*, generally made by people working in social networks: a patient and her physician, a patient and her family, a physician and her professional colleagues, etc.

The slightly jarring position that I will take in this talk is that people, social networks of people are, in a literal sense, *computers*. People and social networks compute; and complex computations move across these social networks, e.g., during transitions in care. People obtain information from many sources; they process it systematically, albeit in ways that are often beyond our ability to formalize or automate today; they make vitally important decisions and act accordingly, producing new information and taking actions that have real consequences in their physical, economic and social environments.

People and social networks are unique and remarkable computers capable of processing information in ways that no digital computers can match. They are the most profoundly important and in many ways the most capable computers in the system. They are not *digital* computers, and they certainly are not machines, but they are computers. They are essential to and largely driven by the overall information processing dynamics of the larger system. Indeed, the entire healthcare system, in this view, is engaged in an enormous set of non-terminating, distributed, decentralized, largely event-and-condition driven computations.

The role of digital technology, in this view, is to support the computing needs of people, social networks of people, and organizations. Digital technologies play essential, enabling, even transformative roles, but ultimately remain as components that handle only selected parts of the information processes that drive such systems, whose role is to support the information processing tasks that people, organizations and social networks perform. From a computational perspective, then, the principal subject of study and of design needs to be the computational behavior of the whole system: of what I will call the *cyber-social system.*

A Cyber-Social Systems Approach to Cyber-Infrastructure Requirements and Architecture

How does this strange point of view help? The key idea is that it can yield important insights into requirements--and from requirements, architectures--for digital cyber-infrastructure systems. This view leads the designers of a digital infrastructure to ask such questions as this: Whose *human* computing needs are to be supported? What range of computation do they need to perform? What information do they need for their computations? How are their individual computations composed into processes? How do computational behaviors move across social networks as healthcare processes evolve? How well are these processes engineered: for performance, safety, portability, and so forth? What properties must digital technology systems have to support the “computing” needs of people, social networks, and organizations, to improve their performance and the human and social outcomes of the overall system?

The focus shifts from information processing by machines to information processing by people, teams and organizations, and *then* on to the possibilities for improving human information processing through the use of digital computation and communication systems. As digital technology continues to advance at a remarkable pace, revolutionary possibilities emerge to automate and enhancing human computing, but in the end, people will remain as the ultimate computers: receiving complex information, evaluating alternatives, communicating with others, using services to aid in decision making, making decisions and taking actions within complex, evolving environments, and producing new information that flows back into the ongoing computational behavior of the cyber-social system.

As computational actors in complex systems, people, networks of people, and organizations need both formal and informal decision rules and procedures ("human software"); they need to be called upon to perform required actions at the right times; they must be presented with analyses to be performed, decisions to be made, actions to be taken, and information relevant to these needs; they must know what information to produce and must be given effective means to communicate it; and they often need information services to help them with their information processing and communications tasks.

Key questions for the designer of a new or improved cyber-infrastructure for a ULS system are what *parts* of the overall computation of the cyber-social system can best be handled by digital technology, and how can digital technology systems best improve, transform, and support the human information processing needs of *all* of the principal actors in the system, including, notably, patients and families?

Concrete Examples and Derived Requirements

Example 1: Should I, as a patient, undergo a thorough evaluation of my cancer risk? An important input in my computing of an answer to this question might be the results of an early, partial risk assessment based on a few known factors, with others remaining unknown. To get this assessment, I might decide to engage a third-party cancer risk assessment service. This service will need specific, selected clinical data on from my medical records, which are held by several institutions, as well as certain non-clinical data such as illnesses and causes of death among my relatives. (I thank William Knaus for this example.)

From this simple scenario we can infer basic requirements for a national health information system: (1) patients are among the *principal* "computers" that a national health information system must support; (2) a national system must enable patients (and other principals) to *query* for selected information from medical records that are scattered across institutional boundaries; (3) the system must allow principals to route responses to such queries onto patient-selected digital channels to reach patient-selected destinations, such as third-party services; (4) the system must *require* that healthcare providers or their proxies respond to (and a fortiori be able to respond to) such queries in a timely manner.

The required legal and regulatory regime is already partially in place in the form of HIPAA regulations, which allow for queries and require timely responses. One big problem of course is that the rules do not demand the construction and use of a national cyber-infrastructure standard for such functions. As a result, executing queries and processing the results is tedious, time-consuming and error-prone to the point that all but the most determined patients are unwilling to use the current query-response system.

Example 2: I don't need to be tested for cancer risk today, but I want to be alerted when emerging conditions dictate that I should be tested. Again I might engage a third party service to decide when to notify me. Now, however, the service requires ongoing updates on changes in my conditions, including clinically ascertained conditions. The trigger could be that my weight has increased or I have received a new diagnosis of a risk-related condition. The requirement for a national system implied by this scenario is that patients must be able not only to *query* providers for information already in the patient's medical records, but *subscribe* to notifications of selected ongoing updates to records--and again to direct that update notification be sent to particular channels connected to patient-specified destinations.

A national health cyber-infrastructure that met such requirements would also satisfy other needs that some have articulated. For example, it would be trivial for patients or their proxies to use the system to create life-long, continuously updated personal health records, e.g., by subscribing to any changes and directing that change notifications be sent to a channel that patients connect to their PHR providers.

Example 3: Should a public health agency declare an epidemic condition? Now we recognize institutions as principals. To this end, the agency could subscribe to all healthcare providers for changes in medical records indicating reportable diseases. There are many details, of course, but the essential structure is pretty clear. The computation that the agency must perform (determine whether to raise an epidemic alert flag) requires ongoing updates of changes in medical data from across the healthcare system.

here what is needed in a national system differs mainly in the details of queries/subscriptions: they are not on a per-patient, but on a per-condition, basis. The agency would direct that notifications be sent on a channel it controls to deliver data into its systems. Once again, the legal/regulatory framework that is needed to drive the required data flows appear to be already partly in place, in the form of established reporting requirements. What is missing once again, however, is a standardized cyber-infrastructure to support the computational needs of the health agencies in an efficient manner.

At first glance, the requirements articulated here are not that different from those embedded in current National Health Information Network projects of the U.S. Department of Health and Human Services or recommended by the President's Council of Advisors on Science and Technology on Health Information Technology (PCAST). There are however real differences. Recognizing patient/citizens as principals whose computing needs must be supported leads to a requirement that a national cyber-infrastructure enable patient-citizens to set up authorized queries and subscriptions and to direct resulting data flows to channels connected to destinations of their choice. No current major architecture, to my knowledge, provides this function. Identifying the full set of principals and considering their computing needs leads to fundamental requirements not met by architectures currently under consideration and development.

From Requirements to System Architecture

In my laboratory at the University of Virginia, my students and I are developing an experimental systems architecture for, and implementation of, a system that shows how such a requirement for patient and other principal control over clinical data flows, could be met in a scalable, secure, reliable manner. The system provides for principal-controlled publish-subscribe *dissemination* of queries and subscriptions, across institutional boundaries, and the direction of responses onto principal-controlled channels for distribution to ultimate destinations.

The model system simulates patients visiting multiple hospitals, hospitals producing encounter and episode records as a result, hospitals listening for and maintaining pools of queries/subscriptions for such data, and replies being sent on principal-controlled channels, on which they flow, under principal control, to destinations such as personal health records, clinical data repositories for analysis, or third-party services. It also simulates several principals' use of the system, including hospitals subscribing to records of patients from other hospitals, a public health agency receiving ongoing notifications of new incidences of reportable diseases, clinical data repositories accumulated data for research, and other repositories indexing data in the style suggested by the 2010 report of the President's Council of Advisors on Science and Technology (for Health Information Technology). The current system is based on a REST architectural style, emphasizing the need to support evolving standards for data types, and minimizing dependencies on technology standards that are not likely to survive for the long term.

Key elements of this architecture include the following: (1) A scalable, redundant, secure, reliable, open source, publish-subscribe messaging system that is used both to distribute queries and subscriptions and to accept and to forward responses on principal-designated channels. (2) Principal (e.g., patient) control over these channels and the ability to query/subscribe to information to which a principal is entitled by law, rule or regulation. (3) The ability to support a wide range of principal players, including individual patients, hospital and other clinical data producers, and large clinical data repositories (e.g., for research or PHRs). (4) Assumptions about a legal/policy framework that requires provider institutions to accept and honor queries and subscriptions, with real-time timeliness requirements, security, and so forth; (5) use of highly scalable data repository technologies; (6) use of modern authentication and encryption technologies (in development); (7) use of widely employed web server and web service technologies.

From the view that the principals in the ULS healthcare system have to compute, and work together to compute, and that the principals must include patient-citizens, we reached crisp statements of some of the fundamental requirements for national system. This is treatment of requirements is, of course, not remotely adequate. It is meant only to illustrate how explicit consideration of the "computing" needs of principals can help to inform the development of properly engineered requirements, which are in turn one of the most essential inputs to system architectural design. Today, by contrast, we have numerous architectural proposals on the table, *none* of which are well rooted in a proper requirements analysis.

Conclusion and Recommendations

Many large IT systems efforts fail. Many large Federal IT efforts have failed, often incurring losses in the tens of billions of dollars (and perhaps even greater opportunity costs over the long run). We *know* that such failures are all too often due to inadequate treatment of requirements, validation of architectures in terms of such requirements, and the use of iterative, incremental, adaptive approaches to building systems that work at scale with limited function in the short run, but that are validated as providing a practical and affordable path to satisfying all major requirements in the long run. The absence of such basic software systems engineering practices creates high, unnecessary and indeed unacceptable risks of project and operational system failure: even for ordinary enterprise systems. Ordinary engineering practices will not suffice for a national cyber-infrastructure for the ULS healthcare system. Cutting-edge thinking is require. To have essentially *no* serious engineering team or process for the development of a National Health Information Network is astonishing and urgently requires attention at the highest levels.

Recommendation #1: State the Problem. My first recommendation is that the IOM should find a way to communicate to the relevant high-level policy makers that current efforts to develop a National Health Information Network, and thus to achieve benefits at a national scale that require health data liquidity, are at an unacceptably high risk of failure -- and are indeed likely to fail -- because they are proceeding without the involvement of substantial engineering talent or a cutting-edge engineering process. Such an effort requires extensive systems analysis; the production and maintenance of critical engineering artifacts (requirements, architectures *traceable to requirements,* and more); the iterative, incremental and adaptive development of working systems at national scale within an architectural framework and process that provides the flexibility and a realistic path to the satisfaction of long-term objectives; and rigorous and ongoing evaluation of the cyber-infrastructure development process. Policy and standards are critical, but it will take a serious and visionary computational engineering effort to produce a system analogous to the interstate highway system.

Recommendation #2: Call for the Problem to be Solved. My second recommendation is that the IOM should find a way to communicate to policy makers that this shortcoming should be corrected urgently.

Recommendation #3: Call for a Solution Commensurate with the Challenge. My final recommendation is that engineering expertise brought to bear on a national health information network should include experts in such areas as ultra-large-scale systems, cyber-social systems, and collaborative and virtual systems of systems, data-intensive computing and large-scale cyber-infrastructures, as well as expertise in medicine, health informatics, human factors, standards, policy and economics.

We are at a crucial moment in the history of our country. If we succeed, we can improve the lives of countless Americans for decades into the future, and lay foundations for reining in unsustainable trends in healthcare costs. However, we are not yet on the right track. A few key changes in the configuration of current efforts would go a very long way to putting us on a better track.

Acknowledgement: I would like to thank my colleagues, William A. Knaus (University of Virginia and Connecticut), Richard Marks (Patient Command), and Mary Shaw (Carnegie Mellon University) for many productive conversations leading to these ideas, and for their comments on earlier drafts of this paper.