A Robust Health Data Infrastructure

Prepared for:
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. JSR-13-700

Prepared by:
JASON
The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102-7508

AHRQ Publication No. 14-0041-EF
April 2014

A Robust Health Data Infrastructure

Contact: Dan McMorrow — dmcmorrow@mitre.org

November 2013

JSR-13-700

Approved for publication 4/09/2014. Distribution only by sponsor: Director, Health IT
Agency for Healthcare Research and Quality

JASON
The MITRE Corporation
7515 Colshire Drive
McLean, Virginia 22102-7508
(703) 983-6997

Contents

1 Executive Summary

1.1 Introduction

The promise of improving health care through the ready access and integration of health data has drawn significant national attention and federal investment. David Blumenthal (former National Coordinator for Health Information Technology) and Marilyn Tavenner (current Administrator for the Centers for Medicare & Medicaid Services, CMS) have characterized the situation well:

> "The widespread use of electronic health records in the United States is inevitable. EHRs will improve caregivers' decisions and patients' outcomes. Once patients experience the benefits of this technology, they will demand nothing less from their providers. Hundreds of thousands of physicians have already seen these benefits in their clinical practice.
>
> But inevitability does not mean easy transition. We have years of professional agreement and bipartisan consensus regarding the potential value of EHRs. Yet we have not moved significantly to extend the availability of EHRs from a few large institutions to the smaller clinics and practices where most Americans receive their health care." [1]

The two overarching goals of moving to the electronic exchange of health information are improved health care and lower health care costs. Whether either, or both, of these goals can be achieved remains to be seen, and the challenges are immense. Health care is one of the largest segments of the US economy, approaching 20% of GDP. Despite the obvious technological aspects of modern medicine, it is one of the last major segments of the economy to become widely accepting of digital information technology, for a variety of practical and cultural reasons. That said, the adoption of electronic records in medicine has been embraced, particularly by health care administrators in the private sector and by the leaders of agencies of the federal and state governments with responsibility for health care. Although the transition to electronic records now seems a foregone conclusion, it is beset by many challenges, and the form and speed of that transition is uncertain. Furthermore, there are questions about whether that transition will actually improve the quality of life, in either a medical or economic sense.

1.2 The Promise of a Robust Health Data Infrastructure

In principle, a combination of electronic health records (EHRs) and improved exchange of health information could serve a number of useful purposes. Frequently cited benefits include:

- Satisfy the growing demand of patients for flexible access to their own health information
- Offer faster, interoperable access to patient records by health care providers
- Reduce errors within individual records and across records
- Reduce redundant testing and diagnostic procedures
- Produce more complete health records and more accurate health data

- Promote better longitudinal tracking of patients and patient groups
- Promote improved standards of care and reduce the incidence of errors in clinical practice
- Provide research data of unprecedented power to inform clinical care, public health, and biomedical research
- Facilitate better communication among health care providers and patients
- Enable electronic detection of health care fraud
- Improve tracking of health care costs and benefits, thereby enhancing understanding of the economics of health care delivery.

Whether any of these benefits can be realized depends not only on the framework for health information technology and exchange, but also on the details of any such implementation. It is therefore vitally important to get those details right.

1.3 Evidence of Benefits

Evidence has been slow to accumulate that the widespread use of EHRs and health information exchanges (HIEs) actually improves the quality, safety, or efficiency of health care in the US. This lack of evidence is partly attributable to slower-than-anticipated adoption rates of computerized health information technology (HIT) systems, especially among small health care organizations and individual providers [2,3]. EHR adoption has been incentivized by HITECH (Health Information Technology for Economic and Clinical Health), a program that was part of the American Recovery and Reinvestment Act of 2009. That program has made more than $15.5 billion available (through July 2013) to hospitals and health care professionals based on their meeting certain EHR benchmarks for so-called "meaningful use." HITECH funds also were used to stand up hundreds of HIEs with the goal of mobilizing the information contained in EHRs. Collectively, these efforts are one of the largest investments in health care infrastructure ever made by the federal government.

The evidence for modest, but consistent, improvements in health care quality and safety is growing, especially over the last few years [4]. Evidence has recently emerged to indicate that EHRs can indeed reduce the costs of health care in the general community setting, and not just in an academic hospital and its affiliated practices or in a large-scale health care enterprise. A recent study of 180,000 outpatients and 800 clinicians in communities that had adopted EHRs from multiple vendors found that, over a multi-year period, the overall cost of outpatient care was reduced by 3.1% relative to the control group [5]. These and other encouraging findings suggest the potential for improved efficiency program-wide.

1.4 Facing the Major Challenges

A meaningful exchange of information, electronic or otherwise, can take place between two parties only when the data are expressed in a mutually comprehensible format and include the information that

both parties deem important. While these requirements are obvious, they have been major obstacles to the practical exchange of health care information.

With respect to data formats, the current lack of interoperability among the data resources for EHRs is a major impediment to the effective exchange of health information. These interoperability issues need to be solved going forward, or else the entire health data infrastructure will be crippled. One route to an interoperable solution is via the adoption of a common mark-up language for storing electronic health records, and this is already being undertaken by the HHS Office of the National Coordinator for Health IT (ONC) and other groups. However, simply moving to a common mark-up language will not suffice. It is equally necessary that there be published application program interfaces (APIs) that allow third-party programmers (and hence, users) to bridge from existing systems to a future software ecosystem that will be built on top of the stored data.

There is a natural tension between the private and public use of health-related data. Individual patient health data are sensitive and therefore must be carefully safeguarded, whereas population health data are a highly valuable, and largely untapped, resource for basic and clinical research. It is in the public interest to make such information available for scientific, medical, and economic purposes, thereby helping to realize the promise of a robust health data infrastructure. Any HIT system for health care must strive to balance these countervailing demands.

1.5 A New Software Architecture

The various implementations of data formats, protocols, interfaces, and other elements of a HIT system should conform to an agreed-upon specification. Nonetheless, the software architecture that supports these systems must be robust in the face of reasonable deviations from the specification. The term “architecture” is used in this report to refer to the collective components of a software system that interact in specified ways and across specified interfaces to ensure specified functionality. This is not to be confused with the term “enterprise architecture,” referring to the way a particular enterprise’s business processes are organized. In this report, “architecture” is always used in the former sense.

To stimulate discussion, JASON proposes in this report a possible software architecture for the exchange of health information. That architecture is based on the following core principles:

- Be agnostic as to the type, scale, platform, and storage location of the data
- Use public APIs and open standards, interfaces, and protocols
- Encrypt data at rest and in transit
- Separate key management from data management
- Include with the data the corresponding metadata, context, and provenance information
- Represent the data as atomic data with associated metadata
- Follow the “robustness principle”: be liberal in what you accept and conservative in what you send

- Provide a migration pathway from legacy EHR systems.

The architecture that JASON proposes allows for various specific implementations, including as possibilities integrated software suites that run on a single box, a cloud implementation, or a widely federated system of systems with shared responsibilities across different organizations. At the architecture's top layer are the applications that interface with the physical world. Stakeholders interact with the architecture through these applications. The bottom layers of the architecture are for physical and logical data storage, including accompanying metadata that provide information about what the data are and where they came from. All data are encrypted, both at rest and in transit. Between the encrypted data layers and the user interface applications layer are intermediate layers that search, index, process, and organize the data.

JASON believes that patient health privacy issues should be mapped onto well-defined architectural elements in the health data infrastructure. The software architecture that JASON proposes adopts the principle that the ultimate owner of a given health care record is the patient him/herself. Thus, the intermediate and top layers of the architecture can gain access to the stored data only through "Identification, Authorization, and Privacy Services." These services include cryptographic key and certificate management, which are handled in accordance with privacy choices made by the patient and the various health care providers. The proposed software architecture capitalizes upon best practices developed in the information technology community to protect electronic information by encrypting it at all times, and by separating key management from data management. Maximal flexibility is achieved to implement various security regimes by associating distinct user permissions with each atomic data element (e.g., blood pressure measurement, serum glucose level) and accompanying metadata. It is anticipated that different individuals will opt for different levels of assumed risk associated with sharing their personal data, in exchange for different perceived benefits that might result from that sharing.

The architecture incorporates a migration pathway from the current legacy software used to store and process EHRs to the future system of broad interoperability. This pathway could be provided by published APIs mandated through the CMS Stage 3 Meaningful Use program, which aims to provide incentives for improving health care outcomes through the adoption of EHRs

There would be opportunities to operate within the new software architecture even as it is starting to be implemented. The APIs provide portals to legacy HIT systems at four different levels within the architecture: medical records data, search and index functionality, semantic harmonization, and user interface applications. These interfaces would allow the architecture to be populated from the legacy systems until the time when all data and functionality are fully contained within the architecture. For example, search functions could pull data from the legacy systems and index those data so that they are more amenable to general queries. User interface applications could capture formatted screens from the legacy systems and reformat the information to better meet the needs of individual users. In this way, the interoperability of the new system would begin to take shape even before all of the data reside within the architecture.

1.6 Benefits of Rapid Adoption

Even in the early stages of adoption, a new software architecture for health information would offer potential benefits, including the opportunity to enhance both clinical care and biomedical research. Above all, it will begin to shift control from a small number of software vendors to a software ecosystem with a diversity of products and "apps," focused on the patient, and enabling health care providers to partner with patients in data sharing. The patient will have increasing control over his/her own data and will take responsibility for that information by reviewing the elements of the EHR, setting access permissions, and making his/her own contributions to the dataset. Increased patient engagement will foster improved patient education, health maintenance, and treatment compliance. Physicians and other health care providers will become discerning customers of a robust health data infrastructure, rather than slaves to a closed-box system. Patients and providers will gravitate toward user interface applications that provide the best functionality and convenience. Vendors will need to serve these consumers if they are to be successful in the HIT marketplace.

A new software architecture will make aggregated health care data available to all biomedical researchers, not just those who happen to work at a large academic center with strength in a particular specialty. The federated database will provide large effective sample sizes, both to support statistical significance and to identify statistical outliers. In the near term, the data will consist mostly of traditional EHRs, including information about medical history, physical examination, physicians' notes and orders, laboratory reports, and medical treatments. These data are already being supplemented by genomic data, expression data, data from embedded and wireless sensors, and population data gleaned from open sources, all of which will become more pervasive in the years ahead. Biomedical researchers in the US will be able to draw upon what amounts to an ongoing clinical trial with over 300 million potential enrollees who report their individual outcomes in relation to their individual medical history and treatment record.

Adoption of a new software architecture for the exchange of health information also is expected to have economic benefit, even in the short term. Through data mining and predictive analytics, methods analogous to those used in the financial services sector, it should be possible to reduce significantly the estimated $60–100 billion of annual health care fraud in the US. Even the partial recovery of fraudulent billing for duplicate claims, unbundled services, and services not rendered would more than cover the cost of implementing the architecture. The data also will contribute to improved understanding of the economics of health care delivery, both in the aggregate and for particular instantiations that either outperform or underperform the aggregate in achieving beneficial outcomes.

1.7 Findings

The following two findings are fundamental and mutually dependent, and the challenges that they identify must be overcome to enable further progress in developing a robust health data infrastructure.

1. The current lack of interoperability among data resources for EHRs is a major impediment to the unencumbered exchange of health information and the development of a robust health data

infrastructure. Interoperability issues can be resolved only by establishing a comprehensive, transparent, and overarching software architecture for health information. (Section 2.4)

2. The twin goals of improved health care and lowered health care costs will be realized only if health-related data can be explored and exploited in the public interest, for both clinical practice and biomedical research. That will require implementing technical solutions that both protect patient privacy and enable data integration across patients. (Section 2.4)

The findings listed below concern the migration pathway from present HIT systems to a unifying HIT software architecture that is agnostic as to the scale and location of the stored data, protective of the data as they move across components of the architecture, and tolerant of variation in the receipt of information.

3. The criteria for Stage 1 and Stage 2 Meaningful Use, while surpassing the 2013 goals set forth by HHS for EHR adoption, fall short of achieving meaningful use in any practical sense. At present, large-scale interoperability amounts to little more than replacing fax machines with the electronic delivery of page-formatted medical records. Most patients still cannot gain electronic access to their health information. Rational access to EHRs for clinical care and biomedical research does not exist outside the boundaries of individual organizations. (Section 3.2)

4. Although current efforts to define standards for EHRs and to certify HIT systems are useful, they lack a unifying software architecture to support broad interoperability. Interoperability is best achieved through the development of a comprehensive, open architecture. (Section 5.1)

5. Current approaches for structuring EHRs and achieving interoperability have largely failed to open up new opportunities for entrepreneurship and innovation that can lead to products and services that enhance health care provider workflow and strengthen the connection between the patient and the health care system, thus impeding progress toward improved health outcomes. (Section 5.1)

6. HHS has the opportunity to drive adoption and interoperability of electronic health records by defining successive stages of Meaningful Use criteria that move progressively from the current closed box systems to an open software architecture. (Section 5.2)

7. The biomedical research community will be a major consumer of data from an interoperable health data infrastructure. At present, access to health data is mostly limited to proprietary datasets of selected patients. Broad access to health data for research purposes is essential to realizing the long-term benefits of a robust health data infrastructure. (Section 6.2)

8. The data contained in EHRs will increase tremendously, both in volume and in the diversity of input sources. It will include genomic and other “omic” data, self-reported data from embedded and wireless sensors, and data gleaned from open sources. Some types of personal health data, especially when combined, will make it possible to decipher the identity of the individual, even when the data are stripped of explicit identifying information, thus raising challenges for maintaining patient privacy. (Section 6.3)

9. The US population is highly diverse, reflecting much of the diversity of the global population. Therefore, important research findings applicable to Americans are likely to come from shared access to international health data. Currently there is no coherent mechanism for accessing such data for research. (Section 6.4)

10. Electronic access to health data will make it easier to identify fraudulent activity, but at present there is little effort to do so using EHRs. (Section 6.5)

1.8 Recommendations

1. CMS should embrace Stage 3 Meaningful Use as an opportunity to break free from the *status quo* and embark upon the creation of a truly interoperable health data infrastructure. (Section 3.2)

2. An immediate goal, to be sought within 12 months (including time for consultation with stakeholders), should be for ONC to define an overarching software architecture for the health data infrastructure. (Section 5.1)

2.1. The architecture should provide a logical organization of functions that allow interoperability, protect patient privacy, and facilitate access for clinical care and biomedical research. JASON has provided an example of what such an architecture might look like.

2.2. The architecture should identify the small set of necessary interfaces between functions, recognizing that the purpose of a software architecture is to provide structure, while avoiding having "everything talking to everything."

2.3. The architecture should be defined, but not necessarily implemented, within the 12 month period. During that time, ONC should create (or redirect) appropriate committees to carry out, continuing beyond the 12 month horizon, the detailed development of requirements for the functions and interfaces that comprise the architecture.

3. To achieve the goal of improving health outcomes, Stage 3 Meaningful Use requirements should be defined such that they enable the creation of an entrepreneurial space across the entire health data enterprise. (Section 5.2)

3.1. EHR software vendors should be required to develop and publish APIs for medical records data, search and indexing, semantic harmonization and vocabulary translation, and user interface applications. In addition, they should be required to demonstrate that data from their EHRs can be exchanged through the use of these APIs and used in a meaningful way by third-party software developers.

3.2. The APIs should be certified through vetting by multiple third-party developers in regularly scheduled "code-a-thons."

3.3. Commercial system acquisition by the VA and DOD should adhere to the requirements for creating public APIs, publishing and vetting them, and demonstrating meaningful data exchange by third-party software developers.

4. The ONC should solicit input from the biomedical research community to ensure that the health data infrastructure meets the needs of researchers. This would be best accomplished by convening a meeting of representative researchers within the immediate (12 month) time frame for architecture definition. (Section 6.2)
5. The adopted software architecture must have the flexibility to accommodate new data types that will be generated by emerging technologies, the capacity to expand greatly in size, and the ability to balance the privacy implications of new data types with the societal benefits of biomedical research. (Section 6.3)
6. The ONC should exert leadership in facilitating international interoperability for health data sharing for research purposes. The genomics community is already engaged in such efforts for the sharing of sequence data, and the ONC should consider adopting a similar process. (Section 6.4)
7. Large-scale data mining techniques and predictive analytics should be employed to uncover signatures of fraud. A data enclave should be established to support the ongoing development and validation of fraud detection tools to maintain their effectiveness as fraud strategies evolve. (Section 6.5)

JASON is grateful to have had this opportunity to examine the challenging and important topic of enhancing the adoption and interoperability of electronic health records. The body of this report provides the details of an example software architecture that breaks the stranglehold of current stovepipe systems and facilitates migration to a software ecosystem, with a diversity of products and apps, that fosters innovation and entrepreneurship. JASON believes that now is time to define such an architecture, leveraging the opportunity to specify CMS Stage 3 Meaningful Use requirements to drive implementation. A fundamental precept of medicine is: “Above all, do no harm.” A software architecture that is broadly tolerant of different scales, input types, and sites for data storage and processing offers a sure pathway, and one that will be open to future innovation. Patients and health care providers will be in a position to choose which particular implementations within the architecture have the most utility for their needs.

2 Introduction

An electronic health record (EHR) is a longitudinal digital record of an individual's health information. There are many different motivations for moving to EHRs and the electronic exchange of health information, and these tend to vary depending on the person or agency within the overall health care system. However, the two overarching goals most commonly associated with the increased use of health information technology (HIT) are improved health care and lower health care costs. Whether either, or both, of these goals can be achieved remains to be seen, and the challenges are immense.

Health care is one of the largest segments of the US economy, approaching 20% of GDP. Despite the obvious technological aspects of modern medicine, it is one of the last major segments of the economy to become widely accepting of digital information technology, for a variety of practical and cultural reasons. That said, the adoption of electronic records in medicine has been embraced, particularly by health care administrators in the private sector and by the leaders of agencies of the federal and state governments with responsibility for health care. Although the transition to electronic records now seems a foregone conclusion, it is beset by many challenges, and the form and speed of that transition is uncertain. Furthermore, there are questions about whether that transition will actually improve the quality of life, in either a medical or economic sense.

The situation was well summarized in a 2010 Perspective in the *New England Journal of Medicine* by David Blumenthal (former National Coordinator for Health Information Technology) and Marilyn Tavenner (current Administrator for the Centers for Medicare & Medicaid Services, CMS) [1]:

> "The widespread use of electronic health records in the United States is inevitable. EHRs will improve caregivers' decisions and patients' outcomes. Once patients experience the benefits of this technology, they will demand nothing less from their providers. Hundreds of thousands of physicians have already seen these benefits in their clinical practice.
>
> But inevitability does not mean easy transition. We have years of professional agreement and bipartisan consensus regarding the potential value of EHRs. Yet we have not moved significantly to extend the availability of EHRs from a few large institutions to the smaller clinics and practices where most Americans receive their health care."

2.1 The Promise of a Robust Health Data Infrastructure

The promise of improving health care through the ready access and integration of health data and records has been offered for nearly 150 years, beginning with Florence Nightingale in 1863 [6]:

> "In attempting to arrive at the truth, I have applied everywhere for information but scarcely an instance have I been able to obtain hospital records fit for any purpose of comparison. If they could be obtained they would enable us to decide many other questions besides the ones alluded to. They would show subscribers how their money was spent, what amount of good was really being done with it or whether their money was not doing mischief rather than good."

Yet more than a century after Florence Nightingale pointed out the obvious benefit of broad access to and integration of health records, health care still had not been empowered with such information. In 1967 a cry went out from the National Advisory Commission on Health Manpower to encourage the harnessing of technology to improve the efficiency and effectiveness of health care delivery [7]:

> "The Panel on the Impact of New Technologies was asked by the National Advisory Commission on Health Manpower to suggest specific areas of technological innovations that could improve the efficiency and effectiveness of health manpower. [I]t became apparent that the problem is as much one of bringing existing technology into working support of physicians as it is one of developing new technology; it appeared that much of the technology that is needed already exists. [Technology] will surely provide new and more excellent means for coping with unsolved problems."

Nearly thirty years later, in 1995, the Veteran's Administration (VA) took the Commission's recommendation to heart and made unprecedented advances in using existing technology to reform the delivery of health care to veterans, reducing costs and overcoming significant cultural barriers. However, this has been followed by more than 15 years of stalemate in developing the infrastructure and means for exchanging information between the VA system and the DOD health care systems for active military personnel.

Another example of progress only slowly won is attributed to Sidney Garfield, one of the co-founders of Kaiser Permanente, who said of health information technology in 1970 [8]: "Matching the superb technology of present-day medicine with an effective delivery system can raise U.S. medical care to a level unparalleled in the world." It was forty years later, in 2010, that Kaiser Permanente completed the implementation of its "HealthConnect" system across 533 medical offices and 37 hospitals. Even now, interoperability among separate geographic blocks of these facilities remains elusive.

In 2009 the Health Information Technology for Economic and Clinical Health (HITECH) Act became law as part of the America Reinvestment and Recovery Act. This was evidence of the US government recognizing the need for more aggressive progress on the challenge of exchange of health information. Through a series of current incentives and future disincentives, the HITECH Act aims to move the US health care system toward adoption of a national-scale information technology with broad interoperability to meet the promise of improved health and health care delivery at reduced cost. As distinct from some of the earlier proposals, this vision encourages the integration of clinical research with health care delivery and the monitoring of population health, enabling the current standard of practice to evolve toward more personalized medicine.

In principle, a combination of EHRs and improved exchange of health information could serve a number of useful purposes. Frequently cited benefits include:

- Satisfy the growing demand of patients for flexible access to their own health information
- Offer faster, interoperable access to patient records by health care providers
- Reduce errors within individual records and across records
- Reduce redundant testing and diagnostic procedures

- Produce more complete health records and more accurate health data
- Promote better longitudinal tracking of patients and patient groups
- Promote improved standards of care and reduce the incidence of errors in clinical practice
- Provide research data of unprecedented power to inform clinical care, public health, and biomedical research
- Facilitate better communication among health care providers and patients
- Enable electronic detection of health care fraud
- Improve tracking of health care costs and benefits, thereby enhancing understanding of the economics of health care delivery.

Whether any of these benefits can be realized depends not only on the framework for health information technology and exchange, but also on the details of any such implementation. It is therefore vitally important to get those details right.

2.2 JASON Study Charge

Health and Human Services (HHS), through the Office of the National Coordinator for Health IT (ONC) and the Agency for Healthcare Research and Quality (AHRQ), requested this JASON study. ONC, reporting directly to the Secretary of HHS, has been legislatively mandated to be the principal federal entity responsible for the coordination and implementation of nationwide efforts for the electronic exchange of health information. AHRQ, an agency within HHS, promotes research on the quality, safety, efficiency, and effectiveness of health care, with the goal of improving health care decision-making and the quality of health for all Americans.

HHS asked JASON to address the nationally significant challenge of developing comprehensive clinical datasets, collected in real world environments and accessible in real time, to support clinical research and to address public health concerns. These datasets could be used to guide clinical research, enhance medical decision-making, and respond quickly to public health challenges. The specific challenge is to derive relevant information from the population as a whole in a way that is timely, cost-effective, and responsive to new research directions that evolve from health trend observations. These datasets also could enable comparative effectiveness research, resulting in more accurate and individualized medical decision-making. There might be further benefits to public health resulting from comprehensive syndromic surveillance and adverse event monitoring that would be coupled to the rapid assessment, dissemination, and utilization of health data.

Specifically, JASON was asked to address the following questions.

- How can complex data handling techniques and Internet-based technologies be applied to health care to promote the development of real-time integrated datasets at a scale seen in other industries?

- How can the various users of health data in the clinical research and public health communities be presented with tailored and highly specific data views in near real time based on routinely collected health data?
- As health data grows from megabits to gigabits per individual, what fine-grained analytics should be made available to patients and health care providers to guide health care decisions?
- What fundamental data management capabilities are needed to support potential future requirements in an open-ended manner?
- What are the national security consequences of not addressing comprehensive health data opportunities in clinical research and public health?

2.3 JASON Study Process

JASON was introduced to the topic through briefings by various experts, listed in Table 1. Materials recommended by these individuals, together with a wide range of other publically available materials, were reviewed and discussed by JASON.

Table 1. JASON Study Briefers

Name	Organization
David Altshuler	The Broad Institute
Marc Armstrong	University of Iowa
Murray Campbell	IBM
Christine Cassel	National Quality Forum
Deborah Estrin	Cornell University
Kenneth Kizer	UC Davis
John Mattison	Kaiser Permanente
Craig Mundie	Microsoft
Sean Nolan	Microsoft

Name	Organization
Ryan Panchadsaram	The White House
Kevin Patrick	UC San Diego
Bharat Rao	Deloitte
Dan Roden	Vanderbilt University
Ben Sawyer	Digitalmill
Ted Shortliffe	Arizona State University
Carla Smith	HIMSS
Michael Snyder	Stanford University
Robert Sorrentino	IBM

Note: Joy Keeler Tobin (MITRE) and Michael Painter (RWJF) played principal roles in coordinating the briefings.

Most briefers attended the full set of presentations and participated in the accompanying discussions. This was a candid exchange that led to the emergence of the following themes:

- EHRs and health information exchanges (HIEs) are currently woefully inadequate in what they provide to health care professionals
- Data collection interrupts workflow and needs to be made less intrusive
- The level of interoperability set forth through the CMS Meaningful Use criteria, as a result of the HITECH Act, is too low to drive meaningful progress

- Innovation in health care appears to be frozen by a deluge of overly ambitious, insufficiently practical, and often conflicting advice
- Vendor proprietary issues are a barrier to interoperability and innovation
- There are a plethora of standards and national deployment organizations, but none that might be regarded as a consensus for adoption
- A HIE infrastructure that will support research has not yet been identified, and current efforts toward this infrastructure may have the opposite effect
- EHRs should not be things that one buys, but rather things that evolve through cultural change aided by technology
- It is desirable to have a continuous rather than episodic personal health record
- At present, HIEs are largely seen as replacements for fax machines.

One participant eloquently paraphrased W. Edwards Deming: "If you invest in automating bad things, you just make bad things happen faster." Clearly it is necessary to devise a better path forward.

2.4 Key Findings

The above discussion leads JASON to the following two key findings. These findings are fundamental and mutually dependent, and the challenges that they identify must be overcome to enable further progress in developing a robust health data infrastructure.

- The current lack of interoperability among data resources for EHRs is a major impediment to the unencumbered exchange of health information and the development of a robust health data infrastructure. Interoperability issues can be resolved only by establishing a comprehensive, transparent, and overarching software architecture for health information.
- The twin goals of improved health care and lowered health care costs will be realized only if health-related data can be explored and exploited in the public interest, for both clinical practice and biomedical research. That will require implementing technical solutions that both protect patient privacy and enable data integration across patients.

The remainder of this report will develop specific recommendations to address these key findings and will present additional findings and recommendations. The complete list of findings and recommendations appears in the Executive Summary in sections 1.7 and 1.8, respectively. Chapter 3 provides some background on how the development of EHRs and the exchange of health information are currently conducted. Chapter 4 lays the foundation for the development of a HIT software architecture. Chapter 5 provides an example of such an architecture. Chapter 6 points out important research directions that would benefit from wide access to EHRs, and Chapter 7 offers some concluding remarks.

3 Health Information Exchange Today

There is a growing consensus in the biomedical community, especially at the administrative level, that the appropriate use of EHRs and HIEs could lead to improved health outcomes overall, and help to lower health care costs in the long term. The movement towards EHRs (and their exchange via HIEs and other mechanisms) enjoys support from many funding and regulatory agencies, health care providers, health entrepreneurs, and other stakeholders within the biomedical community. Limited but seemingly successful implementations of EHRs and HIEs already exist in selected cases, for example, at the VA, Kaiser Permanente, Vanderbilt University Medical School, and in a few other countries.

Evidence that the widespread use of EHRs and HIEs actually improves the quality, safety, or efficiency of health care in the US has been slow to accumulate. This lack of evidence is partly attributable to slower-than-anticipated adoption rates of computerized HIT systems, especially among small health care organizations and individual providers [2,3]. EHR adoption has been incentivized by HITECH (Health Information Technology for Economic and Clinical Health), a program that was part of the American Recovery and Reinvestment Act of 2009. That program has made more than $15.5 billion available (through July 2013) to hospitals and health care professionals based on their meeting certain EHR benchmarks for so-called “meaningful use.” This is one of the largest investments in health care infrastructure ever made by the federal government.

The evidence for modest, but consistent, improvements in health care quality and safety is growing, especially over the last few years [4]. Evidence has recently emerged to indicate that EHRs can indeed reduce the costs of health care in the general community setting, and not just in an academic hospital and its affiliated practices or in a large-scale health care enterprise. A recent study of 180,000 outpatients and 800 clinicians in communities that had adopted EHRs from multiple vendors found that, over a multi-year period, the overall cost of outpatient care was reduced by 3.1% relative to the control group [5]. These and other encouraging findings suggest the potential for improved efficiency program-wide.

3.1 The Stakes are High and Complex

Even with the emerging evidence of benefit, implementing the useful exchange of EHRs across the entirety of the US presents an enormous challenge. That challenge is made more difficult by the multifactorial nature of the problem. Serious issues need to be faced at every level of implementation. These issues do not fit neatly into a traditional classification scheme of technical, organizational, sociocultural, and regulatory concerns. Rather, most of the issues and challenges cut across several of those areas. The following list of problems gives a window into the complexity of implementing a comprehensive and robust national heath data infrastructure.

1. *The federation problem.* There are at least 50 separate agencies charged with health care responsibilities in the federal government. There are also the 50 states, each running its own health care and public health systems. There are tens of thousands of private-sector health care providers

and enterprises, both large and small, delivering medical services. HHS might be regarded as the closest approximation to a central authority for health care issues in the US government, but its jurisdiction is limited.

2. *The turf problem.* Given the federation problem, different entities have assumed overlapping jurisdiction for various aspects of health care delivery. Different federal agencies have authority over different aspects of EHR implementation and exchange, but there does not appear to be any single interagency group charged with coordinating and harmonizing these efforts. That said, two Federal Advisory Committees have been created to assist with coordination: the Health IT Policy Committee and the Health IT Standards Committee [9]. These advisory groups report to ONC, which has a broad mandate for the coordination of HIEs. However, there has been lack of movement by these groups since the release of the 2011 report by the President's Council of Advisors on Science and Technology (PCAST) on health information technology [10].

3. *The scalability problem.* A federated national database of all health-related information would be enormous (on the order of exabytes). As the number of patient records grows, several "big data" issues loom large, for example, pertaining to access, collation, storage, transport, maintenance, and security of the data. The issues surrounding the management of vast amounts of data are being addressed in many other research communities, and great strides are being made. Fortunately, many of those solutions also could be applied to health-related data. The key to addressing the big data issues surrounding improved health care and lower health care costs is the need for interoperability among the data sources. This simply does not exist in current health care practice and is impaired by the proprietary manner in which the data are curated.

4. *The user interface problem.* User issues are serious and threaten to scuttle the entire enterprise. Data entry for EHRs is almost universally acknowledged to be cumbersome, offering no perceived improvement over traditional handwritten charts. This is not merely an issue of training, but also a serious workflow issue. The user interface is not up to modern standards and there is little incentive to improve this situation. As a result of most current EHRs being part of a closed, vertically integrated system, there is limited interoperability. The lack of competitive pressure gives little opportunity for innovation that could result in better user interfaces.

5. *The interdisciplinary problem.* This problem is related to, but goes beyond, the user interface problem. The EHR/HIE arena needs more players with interdisciplinary skills, particularly individuals with training in both computer science and biomedicine. Such people do exist, but they are too few in number at present. It will take some time for a new specialty discipline of biomedical information technology to emerge and provide a substantial cohort of skilled workers. This issue is a common theme across the many reports on modern health care from the US National Academy of Sciences [11].

6. *The front-loaded cost problem.* The development and implementation costs for EHRs and HIEs are acknowledged to be substantial. Who will pay for these costs? Up-front expenses are anticipated to be more than offset by long-term health care savings. However, such savings may be realized

elsewhere in the health care economic system and won't necessarily provide a good return on investment to the entities directly responsible for implementing EHRs and HIEs. It has been estimated that the implementation costs can exceed $32,000 per physician [12].

7. *The payer problem.* This problem is related to the front-loaded cost problem, but goes beyond it. The ultimate payer for health care in the US is the individual consumer, but payment is carried out through many intermediaries and indirect mechanisms, including via government taxation and private health insurers. There is no clear answer as to which groups should bear the implementation and maintenance costs for EHRs and HIEs. Some of this responsibility has been incentivized by Meaningful Use funding provided through HITECH, but that seems insufficient and not sustainable to drive the entire payment system.

8. *The business model problem.* There are several new business models for implementing limited aspects of EHRs and HIEs, including Microsoft HealthVault [13] and BlueButton+ [14]. Importantly, these all rely on external ways of collecting and entering data into EHRs, and so only address part of the problem. None of these models have yet been established to be economically viable. The decision in 2013 by Google to terminate Google Health [15] is an indicator that success is not guaranteed. If any self-sustaining, private sector alternative to government-funded EHRs and HIEs is to arise from the free market and exist stably, it must establish and prove a viable business model. For such a model to survive, it needs an innovation boost. That will require breaking down the barriers of closed proprietary systems that are highly limited in their interoperability. It will require publishing interfaces and protocols, built around an agreed-upon software architecture that will provide the basis on which new business models can develop and evolve.

9. *The exchange concept problem.* Widely different concepts currently exist among health care professionals regarding what a HIE can and should do. These concepts differ as to what types of data the HIE should manage, as well as who should have access to the data and for what purposes.

10. *The data security problem.* Individual health data are almost universally considered to be sensitive and in need of protection. Beyond privacy concerns, which are considerable, the safeguarding of individual health care information is vital to protect against fraud, identity theft, and other types of criminal abuse. If protection and security are not part of the systems that are developed, people will not trust the technology and will not participate in it. Conversely, population health data are considered to be extremely valuable assets for clinical practice and biomedical research, and therefore very much in the public interest to be exploited. The Health Insurance Portability and Accountability Act of 1996 (HIPAA) [16] enacted complex privacy provisions for health care records, with stiff penalties for violation, enforced by HHS. Some regulatory aspects of HIPAA are controversial, and some others are viewed as an impediment to biomedical research. Regardless, there are many difficult security issues associated with keeping individual data secure while moving forward with population studies in the public interest.

11. *The data integrity problem.* There are many issues pertaining to the accuracy and consistency of the data, including the need for highly reliable disambiguation of an individual's health information. The

absence of a unique identifier for an individual can lead to record misidentification, and thereby to the accidental release of information to the wrong individual (a HIPAA violation) or the conflating of two persons' medical records in the context of either medical treatment or research. Especially with the redaction of records for privacy protection and the sharing of partial datasets, the potential for misidentification is large.

12. *The access and curation problem.* This problem concerns not only practical issues of maintenance and secure storage of data records, but also the reconciling of errors or discrepancies, safeguarding of access, and consolidation of data. How, for example, might individuals correct errors that they discover in their own records? Will they be permitted to elect proxies to check their individual records and make the necessary corrections? Will individuals or their proxies be allowed to delete information or to change the access permissions? Could such deletions or adjustments be made retrospectively?
13. *The consent problem.* This problem is closely related to the access and curation problem. If an individual is allowed to set the access permissions for information stored in his/her own EHR, how will such permissions actually get set? Will there be general categories of access permissions to which a user would subscribe? Would the default trust levels be set in advance so that most types of research use would be allowed by default (opt out), or would research use have to be specifically enabled (opt in)? Would the user be expected to sign an electronic indemnification against re-identification?
14. *The intellectual property problem.* It is becoming increasingly accepted that each person owns his/her individual medical data. However, there has been controversy about health-related products, tests, and inventions that arise out of these data. Would individuals be eligible for any form of compensation for their individual data if these contributed to the economic success of a business enterprise? Is there an implicit contract when an individual gives up a measure of personal privacy and releases his/her medical information? What do they receive in return, and should this return be made more tangible? Would it include any direct form of compensation or would profits from the information be restricted to the commercial developer?
15. *The legal liability problem.* Many physicians may be reluctant to embrace EHRs because of malpractice concerns. They may believe that they are better protected against malpractice lawsuits by the handwritten chart system. Furthermore, HIPAA has raised many new issues about data handling. There are also international legal issues about sharing health information. Many unresolved legal concerns surround legal liability in the event of medical errors that are byproducts of health analysis software or EHR data encoding.

Addressing all of these problems is well beyond a reasonable scope for this report. Instead the focus is on the technical issues that are within JASON's range of expertise, including items 3, 4, and 9–13.

3.2 HITECH and Meaningful Use

The HITECH Act, enacted in 2009, aimed to accelerate the development and adoption of EHRs and to synergize the creation of a robust, sustainable, HIT infrastructure in the US. HITECH included commitment to significant financial incentives and investment ($27 billion over 10 years) by the US government to health care providers and to the states for demonstratively improving health care delivery, quality, and outcomes. The targeted health care providers include physicians, nurse practitioners, certified nurse midwives, dentists, physician assistants, acute care hospitals, and children's hospitals. With the exception of the hospitals and pediatricians, eligible health care providers must have a 30% minimum volume of Medicaid patients.

The incentive program is based on health care providers meeting the CMS Meaningful Use criteria, that is, meaningfully using federally certified EHR technology. There are three stages to the incentive program, and failure to participate will result in financial disincentives in the future. Stage 1 will be completed in 2013, Stage 2 has been defined and will commence in 2014, and Stage 3 has not yet been defined. Table 2 summarizes the Stage 1 objectives. Stage 2 requires increased level of implementation for most of the Stage 1 criteria; eliminates a few objectives, such as providing patients timely electronic access to their health information; and adds a requirement of secure electronic messaging to at least 5% of the patients. CMS maintains a complete description of Meaningful Use and current information on the incentive program [17].

The following finding and recommendation summarizes JASON's assessment of the current situation regarding the adoption of EHRs.

Finding

- The criteria for Stage 1 and Stage 2 Meaningful Use, while surpassing the 2013 goals set forth by HHS for EHR adoption, fall short of achieving meaningful use in any practical sense. At present, large-scale interoperability amounts to little more than replacing fax machines with the electronic delivery of page-formatted medical records. Most patients still cannot gain electronic access to their health information. Rational access to EHRs for clinical care and biomedical research does not exist outside the boundaries of individual organizations.

Recommendation

- CMS should embrace Stage 3 Meaningful Use as an opportunity to break free from the *status quo* and embark upon the creation of a truly interoperable health data infrastructure.

Table 2. Summary of Core Objectives for Stage 1 Meaningful Use

Objective	Measure
Record patient demographics (sex, race, ethnicity, date of birth, preferred language, and in the case of hospitals, date and preliminary cause of death)	Over 50% of patients' demographic data recorded as structured data
Record vital signs and chart changes (height, weight, blood pressure, body-mass index, growth charts for children)	Over 50% of patients 2 years of age or older have height, weight, and blood pressure recorded as structured data
Maintain up-to-date problem list of current and active diagnoses	Over 80% of patients have at least one entry recorded as structured data
Maintain active medication list	Over 80% of patients have at least one entry recorded as structured data
Maintain active medication allergy list	Over 80% of patients have at least one entry recorded as structured data
Record smoking status for patients 13 years of age or older	Over 50% of patients 13 years of age or older have smoking status recorded as structured data
For individual professionals, provide patients with clinical summaries for each office visit; for hospitals, provide an electronic copy of hospital discharge instructions on request	Clinical summaries provided to patients for >50% of all office visits within 3 business days; >50% of all patients discharged from the inpatient or emergency department of an eligible hospital or critical access hospital and who request an electronic copy of their discharge instructions are provided with it
On request, provide patients with an electronic copy of their health information (including diagnostic-test results, problem list, medication lists, medication allergies, and for hospitals, discharge summary and procedures)	Over 50% of requesting patients receive electronic copy within 3 business days
Generate and transmit permissible prescriptions electronically (does not apply to hospitals)	Over 40% are transmitted electronically using certified EHR technology
Computer provider order entry (CPOE) for medication orders	Over 30% of patients with at least one medication in their medication list have at least one medication ordered through CPOE
Implement drug–drug and drug–allergy interaction checks	Functionality is enabled for these checks for the entire reporting period
Implement capability to electronically exchange key clinical information among providers and patient-authorized entities	Perform at least one test of EHR's capacity to electronically exchange information
Implement one clinical decision support rule and ability to track compliance with the rule	One clinical decision support rule implemented
Implement systems to protect privacy and security of patient data in the EHR	Conduct or review a security risk analysis, implement security updates as necessary, and correct identified security deficiencies
Report clinical quality measures to CMS or states	For 2011, provide aggregate numerator and denominator through attestation; for 2012, electronically submit measures

Reproduced from Blumenthal and Tanner [1].

3.3 Health Information Exchange (HIE) versus the Exchange of Health Information

In addition to providing financial incentives for meaningful use, the HITECH Act provides financing for grants through the ONC for the creation of statewide HIEs. These are organizations that oversee and loosely govern the exchange of electronic health information. There are now more than 200 of these organizations, 22 in California alone. These organizations act as consolidation points for health records, with the goal of reducing the high costs of information flow by using electronic means. HIEs may be run by state agencies or privately, and are subject to numerous state and federal regulations, some which are still being defined.

A sustainable business model for HIEs remains to be established [18,19]. A recent report [20] found that 74% of the HIEs surveyed have issues of financial viability, reporting that the development of a sustainable business model is a moderate or substantial barrier. The various HIEs use disparate and largely incompatible technical approaches (e.g., query models, push models, end-to-end interrogation). Only 30% of hospitals currently receive data from HIEs, and only 10% of ambulatory practices do so.

In contrast, the exchange of health information refers more generally to the mobilization of health care information in electronic form, including the contents of EHRs, within or across organizations. This exchange may, or may not, be carried out through a HIE. The remainder of this report will focus on the exchange of health information in the general sense, rather than HIEs *per se*.

A meaningful exchange of information, electronic or otherwise, can take place between two parties only when the data are expressed in a mutually comprehensible format and include the information that both parties deem important. While these requirements are obvious, they have been major obstacles to the practical exchange of health information.

With respect to data formats, the current lack of interoperability among the data resources for EHRs is a major impediment to the effective exchange of health information. These interoperability issues need to be solved going forward, or else the entire health data infrastructure will be crippled. One route to an interoperable solution is via the adoption of a common mark-up language for storing electronic health records, and this is already being undertaken by ONC and other groups. However, simply moving to a common mark-up language will not suffice. It is equally necessary that there be published application program interfaces (APIs) that allow third-party programmers (and hence, users) to bridge from existing systems to a future software ecosystem that will be built on top of the stored data. The issues of data formats and interoperability will be discussed in more detail in the next two chapters.

With respect to the value of the information, there is a natural tension between the private and public use of health-related data. Individual patient health data are sensitive and therefore must be carefully safeguarded, whereas population health data are a highly valuable, and largely untapped, resource for clinical practice and basic research. It is in the public interest to make such information available for scientific, medical, and economic purposes, thereby helping to realize the promise of a robust health data infrastructure. Any HIT system for health care must attempt to balance these countervailing demands. The research value of EHRs and the consequences associated with implementing data redaction as an approach to safeguarding privacy will be discussed in Chapter 6.

3.4 Lessons Learned from Abroad

There is a tendency to look abroad and believe that solutions to the nation's most difficult HIT problems may be found there. This "grass is greener" mentality ignores the difficulties that have occurred with the development of other health data infrastructures, and limits the benefit of lessons that can be learned from abroad. It is important to look at the failures as well as the successes, and to take into account the different social and political realities when seeking guidance from the experience of other countries.

This section highlights international lessons learned that have informed JASON's thinking. It is not intended to provide a complete international perspective on HIT, which has been the subject of many papers and monographs [21.22]. Rather, the focus is on the lessons most relevant to this study. The core lessons include the following.

- Do not underestimate the importance or challenge of security and privacy.
- What works at a provider or regional level will not necessarily scale well to a national level.
- The ability to carry out search and indexing of EHRs holds substantial promise for improving public health outcomes and should be a core component of the software architecture for HIT.

3.4.1 Sweden

The Swedish Ministry of Health and Social Affairs sets the agenda and provides principles and guidelines for the Swedish health care system. The actual implementation of the health care system is through 21 county councils, which have the power of taxation and operate independently. As a result, there are several different electronic health information systems in Sweden, represented by five different vendors. The 21 independent health care systems have coevolved over decades. The services they provide are comparable, but the systems differ significantly across regions. For example, although Sweden has had EHRs for several decades, the records could not be exchanged across regions until recently, due to differing capabilities.

In 2008 Sweden launched an initiative for the development of a national EHR that would provide electronic access to health records for patients, health care professionals, and health care facilities in the nation. A key issue to moving forward with this initiative was to address privacy and security in a transparent manner based on unique identifiers. Sweden implemented a national authentication service and a common e-prescription service, which has the highest utilization level of any country. It is encouraging that in just five years since the initiative was launched, Sweden has created a system with 100% electronic health records and nearly 90% electronic prescriptions.

In late 2010 Sweden launched a national on-line medical records system for access by patients and health care professionals. In just a few years, this has enabled patients and their health care providers to access health records at home and while traveling in other parts of Sweden. Sweden also has a common

Picture Archiving Communications System (PACS) for radiology. In order to exchange health information between health organizations, a patient must first give consent. Five major vendors supply EHR technology and three major vendors provide PACS technology in Sweden.

The situation in the US is significantly more complex than in Sweden. In the US there is no unique identifier for each patient, and the US system of both private insurers and public insurance (Medicare and Medicaid) provides many more payment models and incentives compared to the situation in Sweden. However, one can draw lessons regarding HIT expectations from the Swedish experience. First, it takes years, if not decades, to grow a functioning health data infrastructure from the seeds of EHRs, but once there is sufficient buy-in and cooperation, then progress can be rapid. One of the largest challenges is coping with legacy systems and work processes. A second lesson is not to underestimate the need for early design of security and privacy measures, beginning at the concept phase of a HIT system. Sweden saw late-stage delays because of a lack of effective security controls. It will be much less expensive and more effective in the long run to build upon deeply integrated security and privacy measures from the start.

3.4.2 United Kingdom

The National Health Service (NHS) in the UK is often used as the exemplar of a single-payer health care system. Operational since 1948, it enjoys widespread support among the population, although in terms of health outcomes it is ranked 15th in Europe and 18th worldwide. The NHS is a confederation of four national health systems (England, Scotland, Wales, and Northern Ireland). All residents of the UK are entitled to health care under these four systems. Despite the unifying role of the government, there is significant variation in the quality of care across regions, including with regard to the ability to exchange electronic health information.

There are several lessons to learn from the successes and failures of the NHS and the exchange of health information in the UK. First, one should not underestimate the threat posed by insiders regarding the risk of health information disclosure. Second, one should be skeptical of claims of security when no mechanism exists for independent validation and verification.

The NHS has suffered several embarrassing security and privacy violations that stem from its lack of capability to control access to databases. In 2009 a staff member compromised the medical records of then Prime Minister Gordon Brown via insider access [23,24]. This happened despite assurances from the NHS that data were protected using the “highest standards of security.” Unfortunately, such vacuous claims of security are all too common. Security only comes through careful engineering and constant vigilance and refinement. The trusted computing base of the NHS is too large to prevent authorized individuals from inappropriately disclosing health information. As has been pointed out [25], a national system “holding 50,000,000 records is too big a target, will be cumbersome, fragile, and unsafe, and failures to properly protect privacy will have real costs in safety and access — particularly for the most vulnerable or at risk sections of societies.” While an audit-based access control may suffice

within a single clinic or hospital, the same processes will fail to provide adequate privacy protection at a national level.

A second key lesson is to be skeptical of vendor claims about security and privacy without proof that can be independently validated and verified in an irrefutable manner. One should be especially dubious of claims that encryption "solves" a security problem because poorly used encryption is virtually indistinguishable from properly used encryption. Both poor and proper use of encryption employ the "highest standards of security." Many poorly engineered systems employ encryption, but were designed by people with little or no data security expertise, including smart meters, automobiles, and SCADA systems. These systems are deployed and fill important roles, but are highly vulnerable to attack due to poorly designed security. The security of EHRs cannot be left to the vendors of HIT systems, and must instead be part of an encompassing, robust health data infrastructure.

3.4.3 Taiwan

Taiwan implemented a national HIE that is tightly coupled with smartcards carried by individual patients. These smartcards serve to ferry compact EHRs, with remarkable success. They also serve to authenticate the individual to deter fraud, with mixed success. Taiwan has approximately 500 hospitals and 20,000 clinics [26]. Only about half of the hospitals participate in the Taiwanese HIE, a system that has been in place since 2009. Under strict privacy controls, cleared personnel can access medical records from a central database to identify urgent public health trends, such as the SARS outbreak.

One lesson to draw is that the US health data infrastructure should not eliminate the possibility of smartcards or their equivalent, patient-controlled cloud storage, or some other future technology from being used in conjunction with traditional storage for medical data. A second lesson is the value of a rapid search and indexing capability to support public health.

3.5 Veterans Administration and Department of Defense

The VA and DOD operate two of the largest health care systems in the world. They have independently developed EHR systems and for more than 15 years they have made failed attempts to achieve some level of interoperability between their systems. The two systems are described briefly here, especially with regard to their current state of interoperability.

3.5.1 VA VistA System

The VA VistA system is widely regarded as one of the best electronic medical information systems in existence. This HIT transformation of the VA health care system is often cited as the largest and most successful health care turnaround in US history [27]. Whereas the development of VistA started before 1985 (initially called the Decentralized Hospital Computer Program, DHCP), the actual transformation began in 1995, and by 1999 the VA was able to:

- Treat 24% (>700,000) more patients per year
- Reduce staffing by 12% (25,867) of full-time employees
- Implement universal primary care for veterans and their families
- Close 55% (28,986) of acute care hospital beds
- Improve access and reduce waiting times by opening 302 new community clinics
- Implement a National Formulary, which improved evidence-based drug utilization and reduced the total cost of pharmaceuticals by $650 million per year
- Reduce bed days of care per 1,000 patients by 68%
- Reduce in-patient admissions by 350,000 per year.

No appropriations were designated for this transformation; essentially all of the changes were implemented by redirecting savings. Today the VA system has ~8.3 million enrollees, ~235,000 employees, an operating budget of $49 billion per year, and manages ~1,400 sites of care. More than 60% of the physicians who are trained in the US rotate through the VA system, gaining valuable experience with a successful EHR system. The VA is rightfully proud of what it has accomplished.

VistA was developed in-house around an interesting licensing model, often called open source, but different from what is normally meant by the term in the commercial world. The source code is in the public domain and is available through a FOIA (Freedom of Information Act) request that allows interested parties to have access to the code base. However, there does not appear to be any mechanism for new innovations created outside the VA to be merged back into the VistA system.

The VistA system is composed of approximately 160 modules (or applications), built around a commercial implementation of MUMPS (a programming language incorporating database primitives). VistA has been ported to other MUMPS implementations, which is a testament to the quality of the software engineering. The most visible application is the Computerized Patient Record System (CPRS), which is a client-server based application that provides a consistent graphical user interface for many clinical functions. VistA was used as the basis for the DOD's Composite Health Care Systems (CHCS), but the two have diverged. VistA has continued to evolve into a more modern system while the DOD approach, described below, has become an amalgamation of several largely text-based systems to meet its various EHR needs.

3.5.2 DOD CHCS System

The current DOD electronic medical records system, the Armed Forces Health Longitudinal Technology Application (AHLTA), is an agglomeration of several legacy systems. For example, DOD still relies on CHCS, which was developed more than 25 years ago, for pharmacy, radiology, and order management. All appointments are still booked using CHCS, except for walk-ins and telephone consultations, which are now booked in AHLTA. DOD also uses the Clinical Information System (based on

Essentris, a commercial product), which has been customized to support inpatient treatment at military medical facilities.

AHLTA is the clinical document engine used by physicians for orders, notes, and other documentation. It is also used as the basis for medical coding. It uses CHCS and its various modules to store this information via the Comprehensive Ambulatory Patient Encounter Record (CAPER) interface. CHCS was developed under contract by SAIC in 1988 for $1.02 billion. It uses the original code base of the VistA system. As a result, it too is module-based and built around a MUMPS engine. Despite their common code legacy, the two systems are not compatible. While VistA has been modernized, CHCS remains a text-based system that requires the computer to emulate a DEC VT320, a terminal that has not been manufactured for almost 20 years.

3.5.3 VA and DOD Interoperability

In 2008 the VA and DOD were directed through the National Defense Authorization Act to jointly develop some interoperable EHR capability. This failed and in 2009 the two entities created the Virtual Lifetime Electronic Record initiative (VLER) [28]. Conceptually, VLER looks and sounds promising, but has yet to be fully implemented. In 2010, and again in 2011, joint initiatives to achieve some level of EHR interoperability between the VA and DOD were initiated. These efforts were to continue through 2017. However, after this long and expensive attempt to meet the Congressional mandate of interoperability, the VA and DOD recently announced that they would no longer pursue a single integrated electronic medical information system [29–31]. The GAO cited cost and management problems, and indeed these were formidable. The cost rose from an already astounding $4 billion to an estimated $12 billion before the project was abandoned. The VA stated that it would cost $16 billion to replace its current VistA system.

In May 2013 Secretary Hagel, following a 30-day review, decided that the DOD would seek a commercial solution [32]. DOD has pledged that the commercial solution it adopts will be compatible with and be able to exchange medical information with VistA, as well as other major commercial electronic medical information systems. It may turn out, and is perhaps likely, that the system the DOD adopts will be a commercial derivative of VistA. Whereas seeking a commercial solution seems like a sensible forward path, there are hazards that DOD should be aware of. Chief among these is treating this as a fixed procurement process and not taking into account the rapid evolution of technology and the changes that a robust health data infrastructure will bring to clinical practice and the delivery of health care. Both the VA and DOD should take this opportunity to agree upon a set of interfaces and data formats that will allow them to exchange information freely. Chapter 6 will address this point more specifically and provide a recommendation for a new path forward for the VA and DOD.

4 Underlying Concepts for a HIT Software Architecture

The term "software architecture" can mean many things. For the purposes of this report, a software architecture defines a set of interfaces and interactions among the major components of a software system that ensures specified functionality. Said another way, the architecture decomposes a complex problem into smaller, more manageable sub-problems that interact only in specified ways and across specified interfaces. "Architecture" is used here in a way similar to its usage in defining the Open Systems Interconnect (OSI) protocol stacks for network communication. A hierarchy of layers is identified, with different layers responsible for the various functions. For OSI, the lowest layers deal with actual physical transport of the bits across a communications channel, while the higher layers are used for management of traffic and scheduling of packet transmissions. In this way there is a clear separation of concerns and various tasks can be delegated to the appropriate layer without knowing details about how that layer implements a given task.

The HIT software architecture presented here is not a design for any particular EHR system, nor for a national HIT system, nor for anything that falls in-between. It should not be confused with "enterprise architecture," which refers to the way a particular enterprise's business processes are organized. The principles of a HIT software architecture, its posited functionalities, should be identifiable in every system that conforms to that architecture at any relevant scale. If such systems share a common architecture, even in an abstract way, then the task of making them interoperable is vastly simplified.

4.1 Principles for a HIT Software Architecture

The following principles have guided JASON in its articulation of a unifying HIT software architecture for the exchange of health information.

1. The architecture must be agnostic as to type, scale, platform, and storage location of the data.

For the health data infrastructure problem, the architecture must be agnostic with regard to scale and to the actual locations of the stored data to allow various specific implementations, including as possibilities integrated software suites that run on a single box, a cloud implementation, or a widely federated system of systems with shared responsibilities across different organizations. In short, the architecture must be flexible enough to incorporate any particular technology, but specific enough to ensure adherence to system-wide principles for the exchange of health information.

2. The architecture must be based on open standards and published application program interfaces (APIs) and protocols.

Standards, APIs, and protocols all aim to achieve the same basic outcome, which is to enable the seamless interaction among components. To achieve interoperability for EHRs and to open the entrepreneurial space for software development, all of these elements must be made public. People frequently encounter standards in their daily lives. For example, an E26 light bulb has a standard base diameter and conductor position to allow mating with a compatible socket, which also has standard properties. One uses different names for this same basic concept at different levels. The term

"standard" typically is used for basic components, especially those that have a physical instantiation or interaction. Standards usually are established through a formal process and are endorsed by a standards organization, such as the IEEE, ISO, or ANSI. There are hundreds of such organizations, most of which are centered on a particular industry.

In contrast, an API is seldom a standard and is usually dictated by a vendor, although there are some APIs that are highly standardized (such as POSIX for operating system compatibility). A well-known API is WinAPI, which provides the published programming interface to the Microsoft Windows operating system. In the past there also were unpublished Windows interfaces known only to Microsoft and not formally supported. Such private APIs are not uncommon in the industry, but exploiting them is a path to incompatibility and therefore should be avoided. An API is a software concept, and in the most basic sense it is the set of procedure calls and persistent variables that a module presents as a way for other modules to communicate with that module. In the case of the HIT software architecture, the APIs will need to be negotiated by the stakeholders and codified through an open process.

At a higher level of abstraction, it is useful to think in terms of a protocol. A protocol is similar in concept to an API, but governs the interaction amongst independently acting entities. For example, computer programs that interact across networks must follow protocols in order to communicate. Protocols dictate the form, content, timing, and order of messages that can be exchanged among the cooperating entities. Protocols always have associated APIs that implement the protocol exchange.

3. Data must be encrypted at rest and in transit.

For the electronic exchange of health information to be widely adopted, patients and health care professionals must have trust in the system and in the security of the data. Encryption is a mechanism to achieve basic security properties such as confidentiality. The software architecture should minimize inadvertent exposure of health data by keeping all health data encrypted, both at rest on storage systems and in transit across networks. While encryption will not solve all security problems, it helps to minimize data breaches and is not costly to implement.

4. Key management must be separated from data management.

The architecture should consolidate policy-related mechanisms into a well-defined access control system that, at a minimum, provides identity management, user authentication, and user authorization. While encryption ensures the confidentiality of data, an authority is needed to control distribution of the cryptographic keys required to access the data. The architecture should include a module for managing cryptographic keys to access encrypted data and certificates to authenticate public keys, respectively. The public keys enable users to establish trustworthy network links across the HIT infrastructure. The access control system never sees the underlying data, separating policy management from data management and the mitigation of some kinds of insider threats.

Health care providers tend to use *post hoc* audit-based access control rather than preemptive per-object-based access control. That is, any user of the system can access all health records within the system, but will suffer consequences if found to exceed his/her intended authority during a subsequent

security audit. The exchange of health information at the national level is too large to manage using an audit-based model. Instead, the software architecture must provide an integrated mechanism for cryptographic key management that is separate from management of the data.

5. Data must be accompanied by relevant metadata and provenance information.

Provenance refers to the chain of custody of the data, from its inception and through its entire history of access, transmission, or modification. Having knowledge of what the data are and where they came from gives context when interpreting the data for either clinical care or biomedical research. One of the lessons learned from past standardization of medical imaging is that metadata and provenance are important to capture along with the primary data. The DICOM (Digital Imaging and Communications in Medicine) standard for radiological imaging, established in the early 1980s, performs well with exchanging pixels. However, the early designers of that standard did not anticipate the specificity and tremendous quantity of data that would ensue. As a result, DICOM fumbled with regard to metadata, such as procedural and patient-specific information. Users of DICOM systems still have difficulty today distinguishing whether a particular image came from an echocardiogram or coronary ultrasound because of the lack of metadata.

6. EHRs should be represented as a collection of atomic data items and associated metadata.

A software architecture that can represent the individual elements in an EHR (e.g., blood pressure measurement, serum glucose level), together with the associated metadata, will provide maximum flexibility in data handling and security. The atomic data elements can be reassembled in various ways for either clinical or research purposes, and distinct user permissions can be associated with each data element.

7. The robustness principle should be followed: be liberal in what you accept and conservative in what you send.

Because the software architecture will contain many interacting components, developed by multiple vendors and used in various contexts, interoperability is a key concern. As a guide to how broad interoperability can be achieved, one should look to the Internet communications protocols. In developing robust communication among the various layers, Internet protocols observe Postel's Law, first articulated in RFC 760 [33]: "An implementation should be conservative in its sending behavior, and liberal in its receiving behavior." This approach has served Internet communications extremely well, and was restated as the "robustness principle" in RFC 1122 [34], which defines the lower layers of the Internet protocols.

For the HIT software architecture to follow the principle of robustness means that implementations of the data formats, APIs, protocols, and other elements of the system should make every effort to adhere to the agreed-upon specification. At the same time, developers of these implementations should realize that the developers of other components of the system might deviate from the specification, by either mistake or design. By being tolerant of these imperfections, small deviations of little consequence

do not bring the system to a halt. For example, if a vendor adds additional fields to the metadata of a laboratory test result, that should not cause the data to be rejected.

Another lesson from the computer networking community regarding interoperability is the value of conducting a "connect-a-thon" or "bake-off" to test whether disparate components can operate together seamlessly to form the network. These hands-on events bring together the vendors, who must demonstrate that their particular component can connect and interoperate correctly with those provided by other vendors. Current EHR systems do not interoperate at all, and in many cases are unable to even exchange data between hospitals running the same system from the same vendor. Moving forward will require HIT vendors to demonstrate that they can meaningfully exchange information in a seamless fashion.

8. A migration pathway must be provided for legacy EHR systems.

Today's EHR systems are already legacy systems, many of which are built on the MUMPS database technology first developed in the 1960s [35]. Unfortunately, these systems are likely to dominate the HIT landscape for years to come. The development of a unifying HIT software architecture needs to move aggressively forward in light of this reality. There must be an opportunity for the legacy systems to operate within a new and evolving software architecture. This can be accomplished through public APIs that provide portals between the legacy systems and the modern architecture. These APIs would allow the new architecture to be populated from the legacy systems until the time when all data and functionality are fully contained within systems embodying the new architecture. For example, search functions could pull data from the legacy systems and index those data so that they are more amenable to general queries. User interface applications could capture formatted screen shots from the legacy systems and reformat the information to better meet the needs of individual users. In this way, the interoperability of the new system would begin to take shape even before all of the data reside within the architecture.

4.2 Focusing on the Patient

The software architecture that JASON proposes adopts the principle that the ultimate owner of a given health care record is the patient him/herself. Before discussing the practical aspects of this position, it should be noted that focusing on the patient draws upon a higher principle to which all health care professionals subscribe: it is not merely to "do no harm" [36], but to do one's best for the patient. Despite cynicism about the US health care system, the 878,000 licensed physicians [37], 2.8 million registered nurses [38], and nearly 10 million other medical professionals are focused primarily on caring for their patients. When faced with decisions about how to implement systems for exchanging health information, one should ask: "What is best for the patient?" The answer usually provides clarity to help cut through the debate about these matters.

The patient knows who he/she is, which facilitates accurate identification and disambiguation of conflicting identifiers. The patient also knows whom he/she trusts, which forms the basis for granting

authorization to access health information. The Health Insurance Portability and Accountability Act of 1996 (HIPAA) includes a Privacy Rule that prevents the disclosure of protected health information without written authorization from the patient, other than to facilitate treatment, payment, or health care operations [16]. Reasonable effort must be made to minimize such disclosures and the patient must be notified of any use of his/her health information. These protections can be met through electronic authorization services, that apply to both the data and accompanying metadata.

The patient is aware of his/her own medical condition. The standard doctor visit includes a ritualized account of the patient's chief complaint (reason for the visit), history of present illness, medications, medical history (e.g., allergies, immunizations, major illnesses, surgeries), review of systems (checklist of positive and negative symptoms), social history, and family history. Typically this ritual is repeated many times, generating redundant and sometimes conflicting information that is dutifully logged in separate physical or electronic records. If instead this information were to travel virtually with the patient, continually updated and cross-checked, the patient's awareness of his/her medical condition would be enhanced.

Patients who are minors or are physically or mentally incapacitated may be unable to fulfill the responsibility of managing their electronic health information. There are legal provisions that allow a parent or legal guardian to act on behalf of these patients in accessing their health information. The parent or guardian has the authority and the duty to act in the patient's best interests. There also are circumstances where other entities require access to portions of the patient's health data, for example, for public health reporting of cancers and communicable diseases, in meeting law enforcement requirements, and as part of the implied consent that occurs when a patient accepts treatment or hospitalization. Thus a health data infrastructure that is focused on the patient still allows other parties to gain access to the data as necessary.

A likely benefit of empowering the patient as the ultimate owner of his/her electronic health information is that it places increased responsibility on the patient for health maintenance. The patient will have responsibility for becoming educated and staying informed about his/her condition and for making good lifestyle choices. He/she will play an active role in data gathering through web-based reporting, wireless sensors, and other electronic communications. He/she will be expected to engage in preventative care and seek early intervention for adverse conditions. Finally, the patient will have responsibility for complying with medical treatment, including prescribed medications, physical therapy, and follow-up care. Historically the patient was a passive recipient of medical treatment, although this situation has changed somewhat over the past few decades with increasing emphasis on informed consumerism. Now there is the opportunity for patients to become more thoroughly engaged, and enlightened, about their own medical care as EHRs transform to PHRs (personal health records) [39]. Health care professionals, in support of what is best for their patients, will want their patients to become active partners in achieving good health.

4.3 Winning Trust

Data security and data integrity are paramount for any HIT system. Few things are considered more personal and private to an individual than those relating to his/her current state of health or personal history of physical or mental illness. HIPAA set national standards for the security of electronic personal health information, and this led to the specification of privacy, security, and breach notification rules aimed to protect this information. Emerging technologies, including DNA sequence analysis, are expected to make it increasingly possible to predict medical status based on a person's biomedical data, making such information even more sensitive. Widespread public concern about possible abuses of DNA sequence information led to the passage in 2008 of the Genetic Information Nondiscrimination Act. In addition, a major provision of the 2010 Patient Protection and Affordable Healthcare Act was the requirement that insurance policies be issued without respect to pre-existing medical conditions that might otherwise be used as a basis for discrimination. The call for this legislation underscores the need to protect the privacy and security of personal data within a system for exchanging health information.

It will also be necessary for the procedures for protecting health data to be understood and appreciated by the public. If the security of their health data is not perceived to be robust, then patients will not trust the technology and will not agree to participate in it. Two problems that must be addressed are the need for a flexible electronic health data infrastructure that can accommodate various security requirements, and the need to educate the public about the privacy protections afforded by the HIT system. Patient education about security issues is beyond the scope of this report, but JASON is mindful of the challenge and the sensitivities involved. Here the focus is on how a HIT software architecture can be defined to facilitate information exchange while implementing a flexible security model that can be adapted to changing constraints, including evolving technologies, health regulations, needs for access, and public attitudes about privacy.

In the move to EHRs, traditional notions of patient privacy, at least as applied to medical records, will require some adjustment. The transition is made easier by noting that the IT concept of user security for data files has some parallels to traditional practices for handling confidential medical records (Table 3). Traditionally, patient information has been stored as hardcopy in medical charts. Patient privacy is implemented by storing medical charts in a secure location, restricting their access to authorized medical personnel, and not permitting duplication or sharing without further authorization. Security for IT systems is implemented by safeguarding physical access to computers that store the data and by restricting electronic access to those with appropriate user privileges. Additional security for electronic information may be provided by data encryption. It will be important for any computer-based security system for health data to embody some of the same functionality as traditional security practices for medical records. However, computer-based systems have the advantage of greater flexibility in authorizing access and controlling the disposition of records.

Table 3. Correspondences Between Patient Privacy and User Security

Patient Privacy (health care usage) *Information in your medical records is:*	**User Security (IT usage)** *Information in your electronic records is:*
Physically safeguarded	Physically safeguarded
Not revealed without express permission(s)	Not decipherable without decryption key(s)
Only changed by health care professionals	Not modified without user privileges
Not duplicated without permission(s)	Not copied without user privileges
Not shared without permission(s)	Not transmitted without user privileges

4.4 Fine-grained Permission Model

The information in traditional medical charts is recorded page-wise, placed into folders, and stored in filing areas. Access to charts is most often to one folder at a time or to selected pages from a folder that have been reproduced and sent remotely. Backup copies of medical charts may or may not exist. By contrast, the information contained in EHRs is found within multiple computer files, and commonly as independently-accessible data within those files, which are likely to be stored in multiple locations. Maintaining backup copies of electronic records is the norm. Computerized records can be a collection of atomic data and associated metadata and can be structured so that each individual piece of information in a given record carries its own set of privileges. Thus, the structure and versatility of electronic records admits, in principle, to a finer granularity in setting authorizations for user access compared to traditional charts.

The implementation of such a fine-grained permission model offers maximal flexibility in controlling access to the data in different parts of an EHR, but without imposing any particular hierarchy or fixed set of rules upon such access. This is a desirable feature, given the changing landscape of regulations and public expectations for health care access. Issues surrounding what user privileges are granted, how these privileges become inherited when information is added to a record, and who gains access to the data under what conditions, could all be handled separately, irrespective of the EHR file structure. A fine-grained permission model would be agnostic about the system of permission controls that is imposed upon it, and it would be compatible with many different types of such systems.

4.5 Patient Privacy Bundles

Implementing fine-grained permissions by associating distinct user permissions with each atomic data element and accompanying metadata in a health record allows for separate privileges to control the visibility, encryption status, read-write access, copying, deletion, linkage, and transmission of the data elements. Because there are more privileges than data elements in this fine-grained system, it becomes impractical to set these manually or individually, except in special cases. Instead, the privileges can be designed to be adjusted automatically based on information supplied by a patient privacy bundle. JASON defines a patient privacy bundle as a predetermined set of default permission and inheritance settings for the atomic data elements that comprise an EHR, set according to some

predefined security policy. The election of a given privacy bundle and the corresponding choice of policy governing data access would reside with the patient, to be chosen in consultation with his/her health advisors and health care providers. This security model encapsulates the notion that the patient ultimately owns his/her own data. It is anticipated that different patients would opt for different levels of assumed risk associated with sharing their personal data, in return for different perceived benefits that may accrue from that sharing, both for themselves and for society.

Patient privacy bundles can be flexible because they enable the patient to share information selectively. They represent, in effect, the personal security policy for an individual. For example, patient privacy bundles may be set up to restrict access to certain types of information to designated individuals or groups only (e.g., mental health records, family history, history of drug abuse) while making other types of information more generally available to medical personnel (e.g., known allergies, vaccination records, surgical history). In addition, patient privacy bundles represent a practical way to reveal selectively to insurers or US government agencies only those data that may be required by law under the current regulatory regime, such as reportable diseases and conditions. Patient privacy bundles also afford the opportunity for an individual to elect to make some, or all, of his/her health data available for biomedical research. In this regard, JASON strongly favors the creation and implementation of patient privacy bundles whereby safeguarded data sharing for research purposes is permitted by default (an opt-out system). Genomic research programs presently being carried out at major research hospitals with opt-out policies for sharing DNA sequence information (e.g., Vanderbilt University Medical Center) have achieved impressively high levels of patient participation [40].

Where will the patient privacy bundles come from? In practice, few patients would ever deal with setting permissions for their own health data at the level of individual data elements. Instead, after consultation with their health advisors and health care providers, they would elect to take a pre-packaged patient privacy bundle designed and recommended to them by trusted parties. Examples of such trusted parties include health care providers, health insurance providers, governmental advisory bodies, medical advisory bodies, patient advocacy groups, and consumer advocacy groups — whomever the patient, in his/her sole discretion, elects to trust as advisor.

In keeping with the principle that the patient owns his/her data, JASON envisages that patient privacy bundles would enable health IT users to set the majority of access permissions to their own EHRs. Nevertheless, any valid set of permissions will clearly need to comply with applicable federal and state regulations regarding access to certain types of personal health information. For example, access would be required by state public health systems and the National Notifiable Diseases Surveillance System to information about reportable communicable diseases, foodborne and waterborne disease outbreaks, pesticide-related illness and injury, cancer incidence, and lead exposure. Thus certain privacy settings could be overridden, based on legal authority, by an authorized agency. In addition, for minor children and others who require a legal guardian, the obligation for setting patient privacy bundles would fall to the guardian or other responsible individual.

The extent to which patients will be willing to share their personal health information under a future health data infrastructure remains to be seen. The situation is similar to other issues regarding sharing

of information that people must confront, given increasing reliance on the Internet, electronic transactions, and electronic devices. It is hoped that information sharing for the purpose of biomedical research will be widely encouraged, for example, through the promulgation of patient privacy bundles that supply such permission on an opt-out basis. Conversely, failure to educate the public about the benefits of sharing their EHRs for research purposes or failure to control the misuse of shared information would have negative consequences for research.

5 The JASON HIT Software Architecture

Based on the underlying concepts presented in Chapter 4, and as a stimulus to further discussion, JASON proposes a unifying software architecture for the exchange of health information. A possible migration pathway is presented from the current legacy software used to store and process EHRs to a future system of broad interoperability. This pathway could be provided through the use of published APIs mandated through the CMS Stage 3 Meaningful Use program, which aims to provide incentives for improving health care outcomes through the adoption of EHRs.

5.1 Overview of a Proposed Architecture

A diagram of the JASON HIT software architecture is shown in Figure 1. The meaning of the various boxes and interfaces between the boxes is described below.

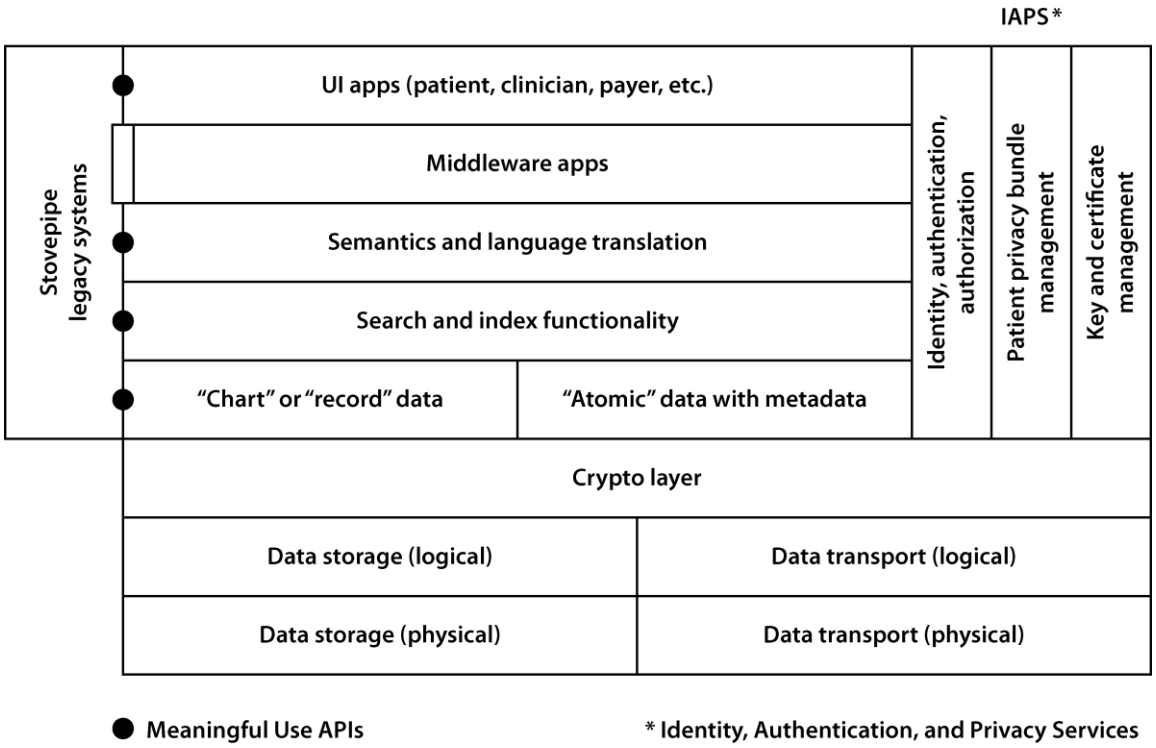

Figure 1. JASON's proposed software architecture for the exchange of health information.

The top of the architecture, labeled "UI apps" (user interface applications), contains all of the applications that interface with the physical world. In the JASON architecture, a clinician's tablet display interacts with the HIT system through UI apps. The patient's interface to his/her personal health record is through an app, perhaps running on a mobile device. If the results of a diagnostic test are entered automatically into a health record, the software that does so is an UI app. Payers also interface with the HIT system through UI apps. All stakeholders in the system interact with the architecture through applications in the UI apps layer.

Next consider the bottom three layers of the architecture, which define how patient data and metadata are actually stored or transported between physical locations. A key design specification in the JASON architecture is that all such data are encrypted, both in storage (at rest) and in transport (in motion). It should be emphasized that data are not decrypted and then re-encrypted for the purposes of transport and relocation. Data at rest or in motion can be meaningfully accessed only through a "crypto layer," and only when the crypto layer has been given, for one-time use in the current transaction, the appropriate decryption keys. Put differently, data that reside below the crypto layer appear as just so many random bits; they can be stored or moved, but (with strong cryptographic guarantees) they have no interpretable meaning so long as the integrity of the files containing the data is preserved.

The adoption of these cryptographic principles as part of the architecture is completely agnostic as to where the data are stored or how they are actually transported. The security remains in place whether the data are in (or transported between) the cloud, or on unguarded machines in the offices of solo practitioners, or anything in-between. A distinction is made between "logical" and "physical" stored data and transported data as abstractions that make it simpler to upgrade storage and transport mechanisms to new physical technologies as they become available in the future. The logical layers can remain the same even as the physical instantiations change.

It remains necessary to discuss the four layers between the UI apps layer and the crypto layer, but before doing so, it is useful to discuss the three vertical "pipes" of the architecture, shown at the right in Figure 1. JASON terms these pipes Identification, Authorization, and Privacy Services, or IAPS. These IAPS are similar to what PCAST termed Data Element Access Services [10], except that the usage here is more neutral as to whether they are implemented across a network, in a single box, or anything in-between; and more neutral as to the types of data that they service. The three IAPS are: (i) identity, authentication, and authorization (IAA); (ii) patient privacy bundle management (PPBM); and (iii) key and certificate management (KCM). These pipes capture the patient-centric nature of the architecture and its capacity to implement fine grained permissions.

The functionality of the IAPS enables the top five layers (from the UI apps layer down) to process actual patient data. The IAPS do this by mediating an interaction with the crypto layer and, when the IAPS are satisfied that all policy requirements are met, providing to other layers the crypto keys that will unlock pieces of data for immediate use in appropriate ways. These principles are illustrated in the first six steps of the example patient query depicted in Figure 2.

The reason that IAA and PPBM are made separate functions in the JASON architecture is to allow (though not require) their functions to be performed by different entities. The entity that patients choose to trust to validate and enforce their patient privacy bundles might not, for example, be the same entity that checks physician and hospital credentials. Similarly, the entity that actually holds the crypto keys (and, for example, generates one-time session keys from longer-term cryptovariables) may be different from both of the first two. Note that the IAPS themselves never access patient data, but only judge the validity of requests and pass appropriate keys.

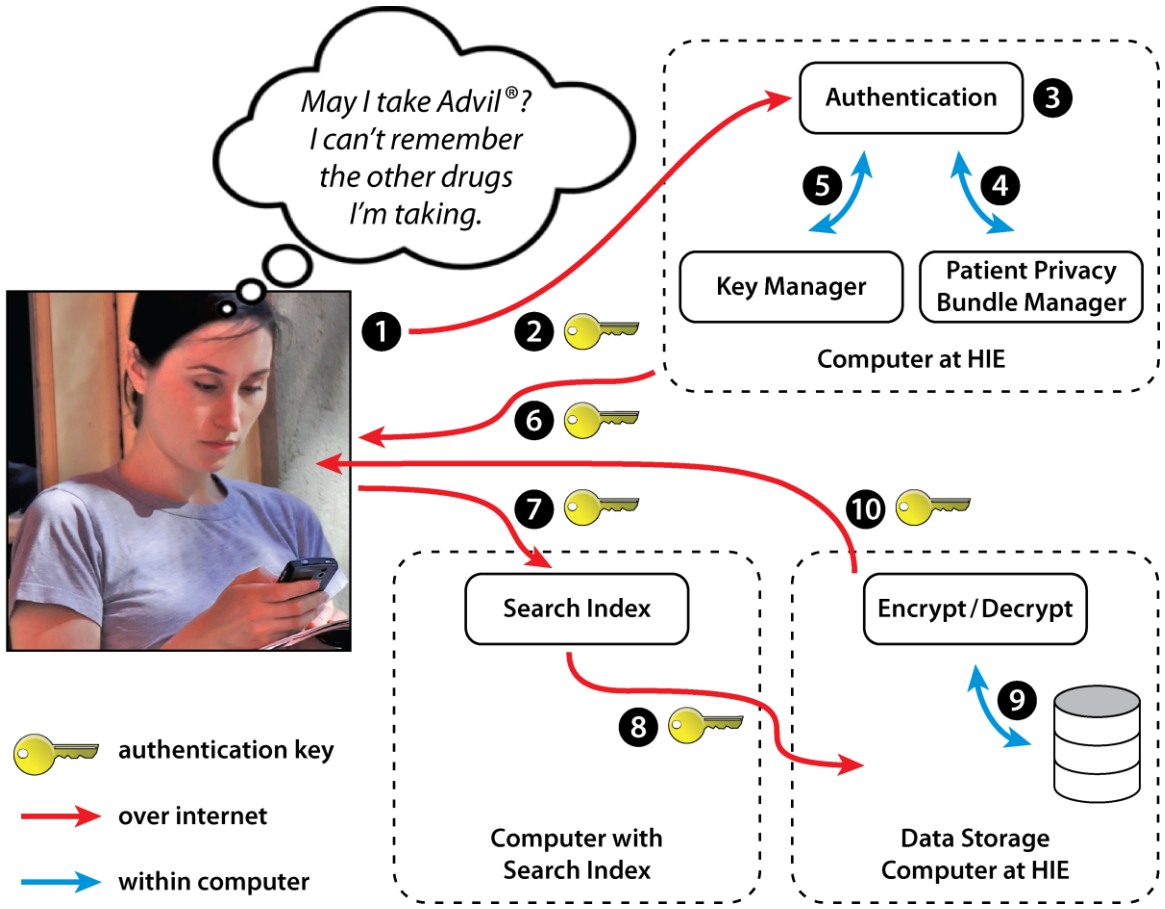

Figure 2. Example of a patient query within the proposed architecture. The patient wonders if Advil® will interact adversely with any of the other medications she is taking. (1) She enters her question into the user interface of an application (UI app) on her smartphone. (2) Her phone establishes a secure wireless connection to her health information exchange. (3) The Authentication Server at the exchange verifies, by key exchange, that she is who she says she is. (4) The Patient Privacy Bundle Manager checks that she is allowed to access the record her smartphone app has requested. (5) The Key Manager generates a unique key for this transaction. (6) The key is transmitted back to her smartphone for use in the remainder of the transaction. (7) Her smartphone sends the key and the request to the Search Computer. (8) The Search Index validates the key, locates the record, and sends a request to the computer storing the record. (9) The record is accessed via the encrypt/decrypt layer. (10) The encrypted information and key are returned to the UI app on her smartphone, which displays the answer. Assuming normal Internet connectivity, this transaction would be completed in less than one second.

Now consider how the request goes down through the remaining layers, and how the data come back up to the user. This is illustrated in the last four steps in Figure 2. The layer immediately below the UI apps layer is where middleware applications reside. For example, if an MRI image is requested, the middleware might be computation-intensive software for slicing the three-dimensional MRI data in different ways, or for de-blurring or segmenting it. The reason that the JASON architecture has a separate middleware layer is to allow for separate market niches, one for developing UI apps, which are focused on the user interaction, and another for back-end processing, which may best be developed and

sold by different entities and run on different machines. Separating these layers in the architecture allows for the development of standards that can connect multiple UIs to multiple back-ends for a given application.

The middleware layer is also where there would be hosting software that carries out data aggregation for research purposes. In this case, the IAPS will have permitted the release of crypto keys that unlock only the data that can be aggregated. The IAPS also are responsible for the imposition of whatever further policy constraints are in force that are associated with the patient privacy bundles.

The next layer down is for semantic harmonization and language translation. Whenever data elements are stored (which occurs below the crypto layer), they are accompanied by metadata about their semantics, including the source of nomenclature, standard terminology, and controlled vocabulary. When data elements come back up from storage, they pass through the semantics layer. Applications above the semantics layer may declare to the semantic layer that they can only accept certain semantic standards. So, at a minimum, the semantics layer reports to the application whether the requested data are available and are in the correct form. However, the application might say that it is willing to accept translated semantics, if available. For this possibility, the semantics layer may contain programs that provide these translation services. Just as for the other layers, the intent of the architecture is to open the opportunity for innovative products specific to this layer.

Particular to the semantics layer is an important design principle of the JASON architecture: the architecture should be agnostic as to whether the future is one of strong semantic harmonization, or is one in which multiple semantics flourish with software available to mediate among them. Harmonization would be beneficial, but its rate of progress should not become the limiting factor for progress in other areas. The semantics layer in the architecture takes up the slack on this issue.

Below the semantics layer, but above the layer where decrypted data first appear, is a layer for search and index functionalities. This is where search software for locating a particular patient's records resides, as well as software for locating sources of data for authorized research. As with the layers above and the layer below, the search and index layer is subject to strong cryptographic control by the IAPS. That is, identity, authentication, authorization, and patient privacy policies must all be satisfied before search queries or index functions are allowed. To define a layer like this is not to say that one knows exactly how to implement such functionalities, which is a task at the intersection of technology and policy issues. Rather, the architecture simply specifies a place for such solutions in the stack: it is below applications and semantics and above (because it must search and index data) the layer where unencrypted data are present.

It is important to emphasize that the IAPS are not beholden to a single, centralized computer. While the architecture is silent on the specifics of its implementation, there are numerous examples of highly scalable naming and authentication systems, including Public Key Infrastructure (PKI) and the Domain Name Service (DNS). As a distributed architecture, it is likely that components of the IAPS will exist from the smallest implementation (e.g., an app running on a smartphone) to a hierarchical system of servers that allows interoperation among all users of the health information system.

Although one could make a different policy choice, the preference would be to require that search and index programs store all their nonvolatile data and metadata with the same data storage requirements as patient data, that is, always encrypted by the crypto layer and subject to control by the IAPS. This will add some encryption and transaction overhead to the search and index functions, but the benefit of having indices being given the same protection as patient data is likely to be large. Modern cryptography is very efficient and does not impose an appreciable burden on modern platforms.

Just above the crypto layer is the layer where unencrypted data come out of data storage or transport. Data here are required by policy (and possibly with engineering safeguards) to be only held in volatile memory, that is, used by applications in only a single session or clinical encounter and then destroyed or with modifications sent back down through the crypto layer. Here the data layer has been divided into two, allowing for both "chart" or "record" data (whose format resembles current EHRs) and "atomic" data with accompanying metadata.

The architecture is agnostic as to the actual location or locations where the data are stored. It seems unlikely that there will be a central repository, but it would be natural to envision a virtual repository in the cloud. It is also possible that the data will be stored in a federation made up from the existing repositories, including individual medical practices, hospitals, and the data centers that support medical records systems. Given the appropriate set of interfaces, it should not matter where the data are physically stored and, other than a central repository, all of the options can coexist and interoperate seamlessly.

It remains to discuss the "large vertical region" at the left side of the JASON architecture. This is where legacy EHR systems reside. Because the legacy systems will persist for years or decades, the JASON architecture does not specify any internal constraints on these systems, which (logically and in Figure 1) extend from stored chart or record data up to the user interfaces. What the JASON architecture does require, presumably as a part of future Meaningful Use requirements, is that legacy systems, within a small number of years, present useable and standard API interfaces to four different levels of the architecture, thus allowing the development of innovative, independently implemented products that interface to the legacy systems.

The four required APIs (from the bottom up in Figure 1) are: (i) an interface for obtaining chart or record data for a specified patient; (ii) an interface for obtaining data that can be used by external search and index programs; (iii) an interface for obtaining metadata about what semantic standards are used by the legacy system; and (iv) an interface for driving the front end of the legacy system (e.g., mouse clicks or command lines) so that better or more convenient user interfaces can be built on top of the legacy system.

The above discussion regarding a proposed HIT software architecture leads to the following findings and recommendations.

Findings

- Although current efforts to define standards for EHRs and to certify HIT systems are useful, they lack a unifying software architecture to support broad interoperability. Interoperability is best achieved through the development of a comprehensive, open architecture.
- Current approaches for structuring EHRs and achieving interoperability have largely failed to open up new opportunities for entrepreneurship and innovation that can lead to products and services that enhance health care provider workflow and strengthen the connection between the patient and the health care system, thus impeding progress toward improved health outcomes.

Recommendations

- An immediate goal, to be sought within 12 months (including time for consultation with stakeholders), should be for ONC to define an overarching software architecture for the health data infrastructure.
    - The architecture should provide a logical organization of functions that allow interoperability, protect patient privacy, and facilitate access for clinical care and biomedical research. JASON has provided an example of what such an architecture might look like.
    - The architecture should identify the small set of necessary interfaces between functions, recognizing that the purpose of a software architecture is to provide structure, while avoiding having "everything talking to everything."
    - The architecture should be defined, but not necessarily implemented, within the 12 month period. During that time, ONC should create (or redirect) appropriate committees to carry out, continuing beyond the 12 month horizon, the detailed development of requirements for the functions and interfaces that comprise the architecture.

5.2 Initial Approaches to Health Data Exchange within the JASON Architecture

The JASON software architecture presented in Figure 1 provides a potential solution to the most challenging issues in developing a modern system for the exchange of health care information. It builds in patient privacy from the outset through its adoption of the principle that data are born encrypted and remain encrypted throughout their existence. Strong authentication mechanisms are in place from the beginning and clear interfaces are denoted to existing legacy systems. A major benefit of the architecture is that it allows and encourages the storage of information as searchable atomic data with accompanying metadata. This provides, for example, information on provenance. More importantly, the metadata are an inseparable component of the data rather than a feature to be bolted on later. Another major benefit of the architecture is that it will facilitate aggregation of the data, properly safeguarded, so that large-scale studies of the efficacy of patient care can be performed.

In this section, incremental approaches are discussed that are consistent with the JASON architecture but do not implement all the layers from the outset. The point here is that there are ways

to begin using the architecture even without implementation of all the layers, and the resulting software can still offer some (but not all) of the benefits. This is a result of the requirement that the APIs used to implement software within the architecture are public, so that one can begin to design software with partial functionality and then iteratively improve that software over time.

These ideas are put forth in response to the tenet that the overarching goals of health information exchange are improved health care and lower health care costs, and recognizing that health care professionals are focused primarily on caring for their patients. The theme of this section is that priority should be given to services that make it possible for health care providers to rapidly access patient clinical information in the near term. The use of patient data for research purposes would become a priority only after significant exchange of health records for clinical purposes is established. If one takes this view, then the application developers' view of the JASON architectural diagram might look as shown in Figure 3.

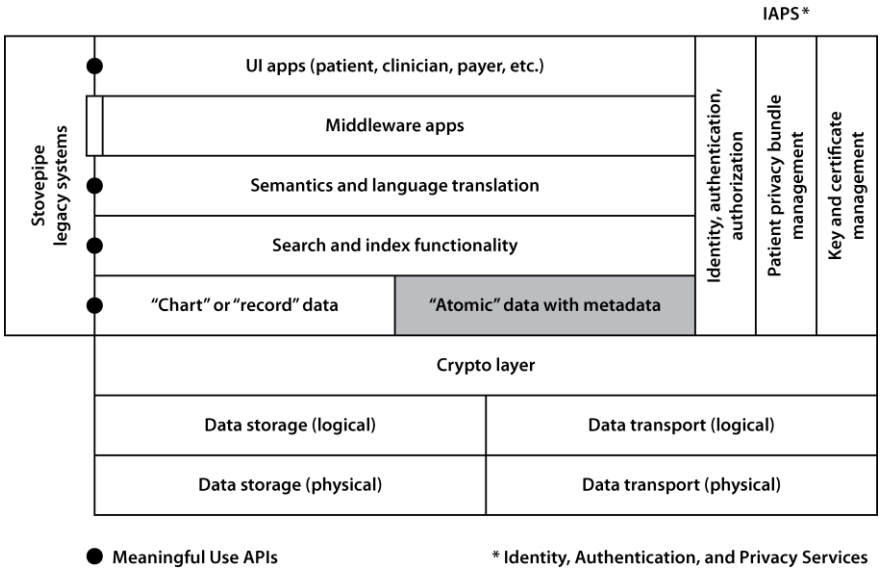

Figure 3. The proposed JASON architecture, but without the capability for processing chart or record data into atomic data. This makes it more difficult to mine the data for research purposes because the data are not disaggregated into atomic data with accompanying metadata.

The only difference here is that, in the initial stages, one postpones the requirement to store the information as atomic data elements with accompanying metadata that can be searched and reconstituted to facilitate analyses of patient care or public health trends. To symbolize this, there is a shaded region in Figure 3 to indicate that the APIs have not yet been used for atomic data and metadata handling. Initial attempts at HIE systems could make use of this approach to speed up implementation of services that can connect diverse legacy systems to a unifying architecture. An application developer would still proceed according to the structure of the architecture, but would not be using the APIs for atomic data and metadata handling because there may not yet be an implementation of these services. When such an implementation becomes available the capability can be added.

It should also be appreciated that adoption of even the above reduced implementation may prove difficult, given the scale and diversity of existing HIT systems. Because these systems are vertically integrated, the use of pervasive encryption as a fundamental attribute may require significant time for

development and implementation. In this case, one might be forced to accept some security risk and consider an implementation that delivers many of the requirements but does not use innate encryption, as shown in Figure 4.

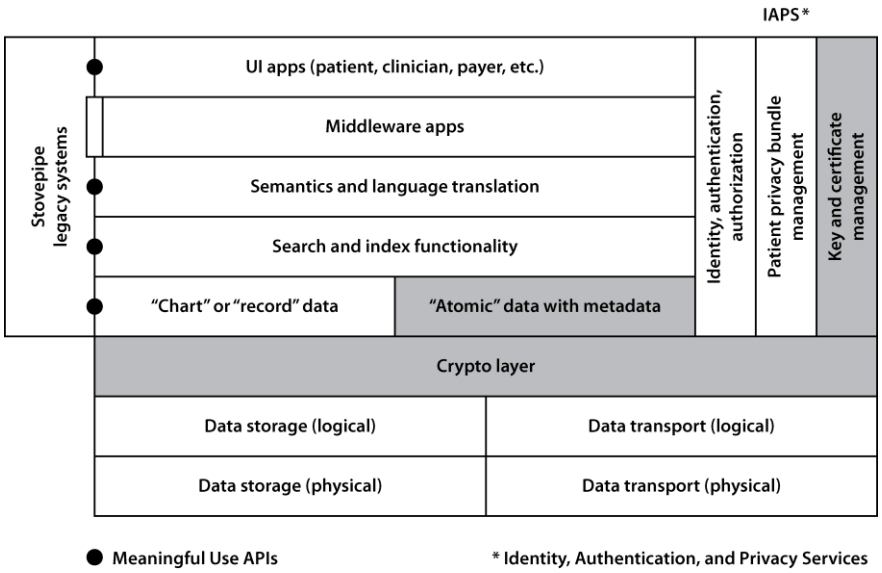

Figure 4. The proposed JASON architecture, but without the capability for processing chart or record data into atomic data, and without intrinsic encryption being implemented in the target application.

In this case, the data storage could be realized using existing data warehouses owned and operated by vendors of current EHR software. Encryption could still be used in storing data on an individual vendor's system (if it isn't already) and could also be used for data in transit, but because there is no intrinsic encryption architecture, there would be more risk here that patient data could be improperly released. Note that it is still necessary to retain mechanisms for authentication, search, and indexing. This reduced implementation may provide only modest savings with regard to the development effort required, but may be useful in promoting greater interoperability among data resources for EHRs.

As a final backward step in this progression, one could argue that the essential components of a system for exchanging health information are the publication of patient data, the ability to search those data, and the ability to access the data once authorization is granted. This is much like what one does with web applications for corporate IT today. Data are stored in a variety of formats, but it is possible to understand what is available if those stores have a web interface. Highly sophisticated search engines are able to crawl and index large data stores and often pinpoint the information one needs in response to queries. An organization today can make all of its internal data available to authorized viewers within the organization by creating an intranet and then using web technology to index the data with an internally deployed search appliance. With proper authentication measures, selected data can be shared with other corporate partners or with consumers. Such an architecture is shown in Figure 5.

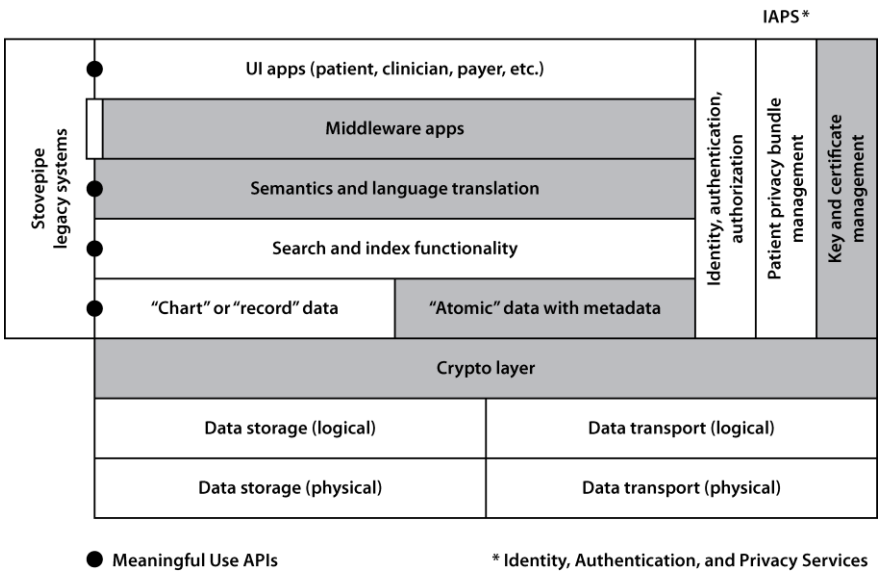

Figure 5. The proposed JASON architecture, but without the capability for processing chart or record data into atomic data, and without intrinsic encryption, semantic translation, and middleware applications being implemented in the target application.

In this approach, indexing, search, and authentication are being used at the least common denominator of a federated health information system. Each provider processes its own data, perhaps using the HL7 standards or something even simpler. The provider then publishes metadata associated with the patient so that it can be indexed by a web search engine. Such searches will only succeed if one is properly authenticated and the patient's permissions have been provided for data access. This could be done via a typical certificate authority and a public key infrastructure. Again, all operations would be performed though the use of the public APIs of the JASON architecture, once these are established.

There are significant drawbacks to this highly reduced approach. For example, if a patient wishes to have their records examined by another health care provider, it will be important to make sure that it is the right patient, which is made more difficult by the lack of interoperability within a common architecture. Once the patient's records are located from a remote provider they must be entered into the local system. This most likely will require some human intervention, and it will be essential to update the record at the remote provider once the patient is treated. Much is lost in terms of the ability to monitor overall health trends and efficacy of treatment modalities because the data cannot easily be disaggregated for research purposes. In addition, privacy can be at risk if information is improperly released because it is not always encrypted.

As collective experience with this type of simplified system grows, health care providers and health information system vendors should be pushed to take steps towards implementing missing functionality, such as intrinsic encryption and metadata storage. When the APIs are finalized, the vendors can then develop and provide products that:

- Convert patient data into atomic elements annotated with metadata
- Store the data using intrinsic encryption in a uniform manner
- Provide uniform mechanisms for accessing the encrypted data

- Publish the metadata so that they can be indexed and searched
- Provide middleware that performs semantic analysis of the data
- Provide the applications for diagnosis, treatment, payment, and other functions.

The basic approach outlined in this section may be achievable in the near term so that the Stage 3 Meaningful Use criteria could be imposed to drive improvements in exchange of health data among diverse providers. This could in principle be done in steps as outlined above, wherein the providers of legacy systems develop back ends to their established software products so that, at least for the near future, the look and feel of the legacy applications is preserved. This helps preserve the original and often significant investments made by health care providers in these legacy systems. However, once the APIs as outlined in the JASON architecture are finalized and published, both existing and new vendors could bring forward the requisite products to move beyond the legacy systems.

One danger in putting forward this reduced approach is that vendors might envision that the modification of their legacy systems to enable the interchange of health information is the end state of EHR Meaningful Use goals. However, this must be viewed as only a waypoint, which leads to the following additional finding and recommendations.

Finding

- HHS has the opportunity to drive adoption and interoperability of electronic health records by defining successive stages of Meaningful Use criteria that move progressively from the current closed box systems to an open software architecture.

Recommendations

- To achieve the goal of improving health outcomes, Stage 3 Meaningful Use requirements should be defined such that they enable the creation of an entrepreneurial space across the entire health data enterprise.
    - EHR software vendors should be required to develop and publish APIs for medical records data, search and indexing, semantic harmonization and vocabulary translation, and user interface applications. In addition, they should be required to demonstrate that data from their EHRs can be exchanged through the use of these APIs and used in a meaningful way by third-party software developers.
    - The APIs should be certified through vetting by multiple third-party developers in regularly scheduled “code-a-thons.”
    - Commercial system acquisition by the VA and DOD should adhere to the requirements for creating public APIs, publishing and vetting them, and demonstrating meaningful data exchange by third-party software developers.

5.3 Relationship to Other ONC Efforts

The ONC is currently sponsoring several efforts to develop interoperability and facilitate the exchange of health information. Two of the main activities, CONNECT [41] and DIRECT [42], are described here.

CONNECT is a software platform designed for the exchange of EHRs using the Nationwide Health Information Network (NwHIN) standards. The software is open source and any health care provider can use it to establish mechanisms for exchanging health information. It is meant to be the software glue that binds that organization to a larger HIE. It can also be used to tie an HIE to a larger regional network. CONNECT provides several services that can be used to enable exchange, including the following.

1. *Core services gateway.* These services are meant to run as part of existing Java-based web application servers, such as WebSphere or GlassFish, and can be used to query the provider to locate patient records and to retrieve documents once they are located. The gateway services also provide mechanisms to authenticate various parties and to ensure that patient privacy choices are respected.
2. *Enterprise service components.* These components can be used as part of the core services gateway if the provider has not already implemented them. The components provided include:
    - Master patient index
    - Document registry
    - Authorization policy engine
    - Preferences manager
    - Audit log.
3. *Universal client framework.* Using the gateway services and enterprise service components, the client framework enables the provider to create a system meant to sit at the edges of a HIE network.

CONNECT was started in 2006 through the joint efforts of 20 government agencies. At present, it is in trial use by the VA, DOD, CMS, and Social Security Administration. So far, 69 medical software vendors have written applications using the provided open source tools and a series of “connect-a-thons” have been held to assess interoperability and to identify barriers. The goal is eventually to transition the ownership of the open source tools from ONC to a private sector entity, but to retain the open source policies.

Interoperability in this approach is achieved through the exchange of “direct messages.” Each direct message is received by the web application server of a given provider and processed. The core gateway services then are used to query the back end of a proprietary system. The vendor of the proprietary system must write a software “adapter” that processes the messages through the vendor’s proprietary system and replies to the querying system, again via the core services gateway. The adapter could be developed using the service components provided from CONNECT or perhaps developed independently by a vendor.

This approach has some overlap with the ideas associated with the JASON software architecture. For example, CONNECT utilizes the notion of interfaces while allowing the vendors to implement functionality. But there are also important differences. There is no clear notion of how CONNECT would be scaled up to a national HIE because the system does not allow one to find patients across the various health care providers. One must know where the patient has received services in the past in order to make the connection. There also is no clear picture of how the information is packaged. This may be justified by the need to be agnostic with regard to data formats, which could include everything from an organized XML document to a scanned PDF of the patient's chart. Finally, because each software vendor is managing the patient's privacy settings in their own way, there is no clear approach to enforcement of patient privacy settings across multiple health care providers.

DIRECT is a formal specification for secure message exchange of health information between providers. It is essentially a collection of existing standards for secure e-mail transport using well-understood protocols, such as SMTP, S/MIME, and authentication via X.509 certificates. For example, if a provider wishes to send a message to a physician at Sunny Family Practice they might use a standard e-mail addressed to <johndoe@direct.sunnyfamilypractices.example.org>. The fact that this is the DIRECT address will ensure that proper checking of the message takes place and that security practices are observed. A large number of EHR providers have agreed to adopt this standard, and it certainly has benefit for secure exchange of patient information. However, if this is the highest level of adoption of health information exchange the vendors will agree to, then adoption of more meaningful information exchange appears to be very challenging. The efforts described above share some similarity with the more comprehensive architecture described this report, but much more could be done in the interest of the patient if a more systematic and patient-centric architecture is adopted and used as a template for future HIE development.

6 Research and the Health Data Infrastructure

Much of the promise of a robust health data infrastructure depends on benefits to be derived from research making use of that infrastructure. That research includes clinical research aimed at discovering new treatments and improving existing ones, and basic research aimed at revealing the underlying mechanisms of human disease. The nature of both types of research is changing as more, and more diverse, patient information is becoming available. The current system for accessing this information for research purposes suffers from the same problems as for health care in general. The HIT software architecture proposed in this report is consistent with the requirements for the use of health data for research. In addition to the national research imperative, there is a strong interest in being to able to access international health data for research. Thus it will be important to develop agreements that address at the international level many of the same issues that are being addressed at the national level.

6.1 Relationship Between Research and Health Data

Basic biomedical research and clinical research have overlapping goals, which are to understand the molecular mechanisms of human disease and to identify effective treatments for disease. In many cases this can be reduced conceptually to understanding the relationships among the genotype, environment, and phenotype of individuals. "Genotype" refers to the genome sequence of the individual, or to some meaningful subset of that sequence such as the sequence variants (alleles) at known sites of polymorphism in the human genome. "Phenotype" refers to the observable traits of that individual, for example, their serum cholesterol level or their response to a particular therapeutic intervention. "Environment" refers to everything other than genotype that affects the phenotype.

The ultimate goal of personalized medicine is to use combined genotypic, environmental, and phenotypic data to tailor treatment to the individual. An example is the breast cancer drug trastuzumab (Herceptin®), a monoclonal antibody directed against human epidermal growth factor receptor (HER2). Trastuzumab is only effective in tumors in which HER2 is overexpressed (approximately 25% of breast cancers [43]), usually as a result of amplification of the corresponding *ERBB2* gene [44]. HER2 overexpression is assayed in biopsy samples by immunological methods or by testing for *ERBB2* amplification by *in situ* hybridization. Identification of patients likely to be responsive to trastuzumab is critical because the drug is expensive and treatment is associated with potential cardiotoxicity [45]. Another example is sensitivity to the anticoagulant warfarin (Coumadin®), which is influenced by genotype at both the cytochrome P450 2C9 (CYP2C9) and vitamin K epoxide reductase complex 1 (VKORC1) loci [46]. Warfarin is the drug most commonly implicated in emergency hospitalizations due to adverse drug events [47]. Prescription of warfarin currently does not require genetic testing, and it is still common to determine empirically the effective dose for each patient. However, this is likely to change as more data become available relating genotype to drug response.

The trastuzumab and warfarin examples illustrate how information of the sort that might be in an EHR can directly influence treatment choices. These examples also illustrate how research into the basic mechanisms of disease and therapeutic response increasingly makes use of genotypic and phenotypic

data to develop and test hypotheses. Because of the ethical and practical limitations on performing patient research, most such studies rely strongly on statistical power to come to meaningful conclusions. In the case of warfarin sensitivity, the strength of association between particular cytochrome P450 alleles and drug response was such that it could be conclusively demonstrated with only a few hundred test and control patients. However, the trend is towards much larger studies, driven by the desire to identify risk variants that have small but important effects, and that occur at a lower frequency in the population. A recent landmark study on genetic factors affecting the risk of cardiovascular disease assayed 200,000 individuals and identified 16 new risk factors [48]. Thus, there are two research-critical areas that would immediately benefit from a robust health data infrastructure that takes into account the needs of researchers:

- *Statistical power of studies:* removing the ownership and interoperability barriers to sharing data would facilitate clinical and basic research by increasing the size of datasets
- *Identification of rare variants:* access to a larger, more detailed datasets will aid in the discovery and characterization of informative rare variants, enhancing the understanding of disease mechanisms.

6.2 Data Types in the EHRs of the Future

In assessing the potential for a robust health data infrastructure to support research, it is useful to consider the nature of EHRs today and how this is likely to change as new technologies are developed and implemented. The information in a typical EHR includes a combination of patient medical data and administrative information about the patient and his/her health care provider. The medical data might include demographic information, patient medical history and family medical history, current and past medications, history of allergic reactions, vaccination records, laboratory test results, imaging results, and so on. The administrative information might include billing information, insurance provider, dates and nature of previous visits to the health care provider, referrals received, and so on. The challenge posed by consolidating this disparate information in an interoperable EHR is large and is the main focus of this study. The discussion below will emphasize how the technologies used in medical testing and diagnosis are changing and how this will impact the EHRs of the future and the associated HIT infrastructure.

There is a growing trend towards capturing large quantities of data associated with particular aspects of patient phenotype, analyzing those data, and reporting relevant information back to the patient. These come under the general heading of “omics” technologies, a designation derived from genomics, the first of such data types. A brief description of the currently most important types of “omics” information is provided below.

1. *Genome sequence.* The haploid human genome contains 3×10^9 base pairs of DNA. Humans are diploid, so each person has two copies of their genome, one maternal and one paternal. These two copies differ by approximately 0.1%, so it is necessary to sequence the DNA sufficiently deeply to

capture all of the genetic variation of an individual in comparison to the reference human genome sequence. The current standard for individual genomes is to sequence to approximately 30-fold coverage, or approximately 10^{11} bases of sequence data. In the case of cancer, for which it is important to know the genotype of the tumor in comparison to that of normal tissue, a similar level of sequencing might be applied to a tumor sample, and this could include a sample of both the primary tumor and its metastases. Although these data can be compressed by denoting only the difference with respect to the reference human genome sequence, there is clearly a rapidly growing need to incorporate vast amounts of genome sequence information into individual EHRs.

2. *Transcriptome.* The transcriptome is a quantitative description of the types and amounts of messenger RNA molecules transcribed from the genomic DNA. Most cells in the body have the same genome sequence, but differential expression of that genome allows cells to become differentiated. Differential expression also defines disease states; for example, breast cancers can be divided into subtypes based on gene expression patterns. The transcriptome can be assessed by microarray analysis or, increasingly, by "RNAseq," in which DNA copies of the messenger RNAs are sequenced with high coverage. The amount of information generated in a transcriptomics experiment is typically similar to that of a genome sequence, although because every cell type is different and there are many possible variables of cell state, there is the potential for much larger datasets.

3. *Epigenome.* The epigenome is a description of the modification states of the genomic DNA and the RNA and proteins that are physically associated with the DNA in the form of chromatin. These modifications are part of the basis for the differential expression of the genome that is manifested in the transcriptome. The epigenome is assessed by a variety of methods that allow for spatial resolution of particular modifications in the genome (e.g., "ChIP-Seq" for measuring modifications of histone proteins, bisulfite sequencing for determining sites of methylation along the DNA, and DNase I hypersensitivity analysis for assessing chromatin structure). Efforts are currently underway to establish reference epigenome information for all genes in all tissue types.

4. *Proteome.* The proteome is a description of the types and amounts of proteins expressed from the genome; it is the protein analog of the transcriptome. The proteome is determined in part by the transcriptome from which it is derived, but also by the many subsequent processes that affect proteins, including their translation, transport, post-translational modification, and degradation. The proteome is usually assessed by mass spectrometry. Both the sensitivity of detection and the methods for determining the amount of each protein detected by mass spectrometry are improving rapidly.

5. *Microbiome.* The human body contains approximately 10 times more microbial cells than human cells by cell number (although only about 1% by mass). The microbiome is the complete description of this microbial population, including commensal and symbiotic organisms as well as pathogens. The microbiome of an individual is a unique signature, changing with time and environment, and likely responsible for some elements of phenotype. Because many of the microorganisms living in and on humans cannot be cultured, the microbiome is usually assessed by deep sequencing of the genomic DNA of microbiome organisms. There is growing evidence that several pathogenic

conditions are due to aberrant states of the microbiome, some of which can be corrected by altering or replacing an individual's microbiome.

6. *Immunome.* The immunome is a description of the state of the immune system of an individual, focusing on the diversity of immune responses based on past exposures. In a narrower sense, such information has long been a part of health records. For example, the Mantoux (or PPD) skin test, and its predecessor the tuberculin tine test, assess the immune response to *Mycobacterium tuberculosis* antigens as a measure of previous exposure to this pathogen. High-throughput methods now allow testing for reactivity to thousands of antigens at once, in combination with deep sequencing to characterize the genome rearrangements that occur in each immune cell and define its reactivity.

JASON's assessment is that each of these "omics" approaches to characterizing individuals will be important elements of health care in the future, and therefore that a robust health data infrastructure must be able to incorporate the corresponding data and to facilitate sharing the data for research purposes. It is important to be aware that these types of data, alone or in combination, can provide a sufficient level of detail to uniquely identify a person in the world population. This should be obvious based on the Combined DNA Index System (CODIS) used in criminal forensics, which relies on 13 polymorphic loci in the human genome, which is a minute subset of the entire genome. Combining genome information with other types of metadata and large publicly available datasets, it has been possible to re-identify individuals in genetic studies using de-identified data. Failsafe de-identification is already impossible, and it is likely to become more difficult to avoid re-identification in the future given the increasing specificity of medical data. Therefore, other means of mitigating privacy risks must be pursued, with the goal of protecting patients from discrimination, and fostering confidence in the research enterprise. This issue is discussed in detail in section 6.3, below.

There is a growing trend on the part of researchers to seek phenotypic information through surveys and on the part of individuals to gather personal health data relevant in digital form. Both are likely to be important contributors to EHRs and fall in the category of self-reported data. For self-reported surveys, there have been several recent examples in which relatively simple questions regarding phenotype and family medical history were combined with genomic information in genome-wide association studies to identify loci involved in human traits [49–51]. There is some evidence that anonymous self-reporting is more accurate than information gathered in consultation with a physician, presumably because some patients are uncomfortable revealing personal information directly to another individual.

In a further step away from involvement of trained medical personnel in data collection, there is a growing market for wearable fitness trackers and smartphone-based apps that generate health-related data continuously or on demand. The self-reported data from surveys and personal monitors will vary widely in quality and utility for research, but will be an increasingly important source of phenotype information. As for "omics" data, the health data architecture should be tolerant of inputs from these self-reported sources, and allow for indicators of uncertainly in the data based on the means of its acquisition.

It is useful to consider how these new types of data might be used for research in the context of a broader system for sharing health data. There are several instructive examples already underway at the NIH, VA, Wellcome Trust, and Kaiser Permanente. However, a program at Vanderbilt University Medical Center is perhaps the most advanced with respect to biomedical research. The Vanderbilt BioVU program is a large program that establishes a biobank of DNA samples for nearly all patients entering the Vanderbilt hospitals and clinics, and matches the banked samples with EHRs that have been de-identified by standard methods [40]. This is an opt-out program, meaning that patient genetic material is acquired and added to the biobank unless the patient specifically indicates a desire not to participate after having been given a brief description of the potential research uses of those samples. Remarkably, only about 3% of patients opt-out [40].

BioVU began collecting samples in 2007 and now has 167,250 samples banked. Access to BioVU is limited to Vanderbilt researchers, and the Vanderbilt genomics group has developed new analytical tools to make use of this treasure trove of data. An example is their use of combined EHR and genomic data to carry out a phenome-wide association study, analogous to the more common genome-wide association study [52]. The principle, which is broadly applicable to any similar large dataset, is to use indicators of phenotype to identify traits that are associated with a particular genetic variant. To standardize phenotype descriptors, the Vanderbilt researchers used the International Classification of Disease (ICD9) billing codes in EHRs as indicators of the conditions individuals might have. They then sought statistically significant association of those phenotypes with a particular genotype. This approach has successfully identified several connections between diseases and single-nucleotide genetic variations [53].

This discussion about health data for research leads to the following finding and recommendation.

Finding

- The biomedical research community will be a major consumer of data from an interoperable health data infrastructure. At present, access to health data is mostly limited to proprietary datasets of selected patients. Broad access to health data for research purposes is essential to realizing the long-term benefits of a robust health data infrastructure.

Recommendation

- The ONC should solicit input from the biomedical research community to ensure that the health data infrastructure meets the needs of researchers. This would be best accomplished by convening a meeting of representative researchers within the immediate (12 month) time frame for architecture definition.

6.3 Data Access

There is a natural tension in the research community between the desire for open access to health data, with its advantages for increasing the power and relevance of studies, and the desire to receive individual credit for scientific advances by restricting access to a smaller set of collaborating researchers.

JASON notes that a similar tension existed at the beginning of large-scale sequencing of the human genome and the genomes of various model organisms. In those cases, open access to complete genome sequence information threatened the old model of research in which a researcher would identify and characterize a single gene based on some property of interest, using the relatively arduous technology of the time. Based on this perceived threat, many researchers opposed the large and expensive sequencing projects. However, once it was appreciated that open access to genome sequences could lead to a new and much more productive model for research in which many genes and their interrelationships could be characterized, the sequencing effort was embraced by the research community. In hindsight, it is now clear that open access to sequence information revolutionized biomedical research and created rich entrepreneurial opportunities. JASON believes that a revolution of similar scale will be driven by open access to electronic health data, and that the research community will fully embrace this once a plausible architecture for access exists and the potential benefits for research become apparent.

There is a second tension for data access regarding the balance between patient privacy and the potential societal benefit of access to patient data. In an effort to protect patient privacy, a general practice in biomedical informatics has been to "de-identify" the data before sharing them with researchers for analysis. De-identification (de-ID) is the process whereby strong identifiers are selectively removed from the individual records in the dataset, for example, the patient's name, address, and social security number. Alternatively, data may be obfuscated by replacing strong identifiers with more generic, or operationally ambiguous, information. However, as many in the biomedical research community have come to realize, true de-ID cannot be achieved by simply redacting strong identifying information. A person is often uniquely represented in his/her health records by a constellation of data that are specific to that individual. Patient records carry a sufficient amount of weakly identifying information that de-ID can almost never be guaranteed. Re-identification (re-ID) has been shown to succeed on other types of data as well, and has spawned an entire industry of consumer tracking and web-based data analytics.

In the biomedical arena, partial and complete re-ID has recently been demonstrated using the published DNA sequences of nominally anonymous individuals who participated in the Personal Genome Project. Their published DNA sequences were cross-correlated with information obtained online by querying recreational genealogy databases, employing a custom algorithm [54]. The available genomic information was first used to associate an unknown individual with his or her family tree, then additional sequence information was used to pinpoint the individual.

The prospects for successful re-ID are expected to grow in the future as much greater amounts of information-rich data are included in EHRs. As whole genome sequencing becomes routine, each person will be uniquely represented in his/her health data. Transcriptome, epigenome, proteome, microbiome, and immunome data will provide additional potentially uniquely identifying information. Health data may also be accompanied by geolocation information, such as data generated by smartphone apps. Soon it will no longer be feasible to carry out meaningful de-ID without redacting much of the medical information itself.

JASON finds that de-ID is not a viable approach for ensuring patient privacy going forward. Furthermore, de-ID is highly undesirable for clinical and basic research because it impedes discovery in two significant ways. First, there is the problem of misidentification (mis-ID), whereby an individual's health record, in the absence of sufficiently strong identifiers, is mistakenly duplicated, merged, or conflated with another person's health record. This corrupts many types of analyses by destroying the statistical independence of the data. The second major problem for research is data loss. Data that are redacted or obfuscated in the de-ID process may lose information that is useful for discovery. For example, an individual's home address may be very useful in epidemiological research, such as in tracking communicable diseases or pinpointing the source of toxic exposures.

A straightforward way to prevent mis-ID problems is to associate a unique identifier (UID) with each EHR. There is currently substantial opposition to implementing a national system for assigning UIDs for health care. However, alternatives to a true UID system are routinely implemented. All major health care providers assign an ID that is unique to the patient, but is only operable within their own system. Smaller health care providers often use the patient's social security numbers or a number supplied by the patient's health insurance provider. Problems can arise with mis-ID within a health care organization, and are much more common when trying to combine data from different organizations. Ironically, algorithms for re-ID may be useful in disambiguating these inconsistencies.

With regard to problems of data loss for research, it would be preferable to avoid any redaction or obfuscation of health care data and provide some other means to protect patient privacy and ensure data security. JASON recommends moving towards a security model where large datasets may be collected for research purposes complete with identifying information, with the caveat that any patient identifying information would subsequently be removed from the results of the research analyses. The results would be de-identified analogous to the way results from clinical trials are de-identified; the clinician is fully aware of the patient's identity during the course of the trial, but the results of the trial contain no uniquely identifying information.

The HIPAA Privacy Rule protects patient health information in the clinical setting. In the research setting it would be necessary to establish real or virtual data enclaves ("walled gardens") where the unanonymized data can be analyzed in a protected setting. Successful examples of such data enclaves already exist for other types of sensitive data [55,56]. Because there can be no guarantees against re-ID, and because individual health data are highly sensitive, it must be recognized that any attempt to balance the private and public uses of medical data necessarily involves compromise. It would be preferable to move towards a system like HIPAA, which penalizes by law any abuses of data, rather than attempting to perpetuate the fallacy that the data can truly be anonymized.

Data access models that feature a fine granularity of permissions, such as the patient privacy bundles for atomic data and metadata within the JASON architecture, offer maximal flexibility in allowing the patient to specify the balance between privacy and perceived benefit for research. Fine-grained permissions will facilitate whatever security policy is eventually adopted for the exchange of health information, but without rendering the records unsuitable for research purposes. These

permissions make it possible to adapt access to all or part of an individual EHR easily and without modification of its contents. This discussion leads to the following finding and recommendation.

Finding

- The data contained in EHRs will increase tremendously, both in volume and in the diversity of input sources. It will include genomic and other "omic" data, self-reported data from embedded and wireless sensors, and data gleaned from open sources. Some types of personal health data, especially when combined, will make it possible to decipher the identity of the individual, even when the data are stripped of explicit identifying information, thus raising challenges for maintaining patient privacy.

Recommendation

- The adopted software architecture must have the flexibility to accommodate new data types that will be generated by emerging technologies, the capacity to expand greatly in size, and the ability to balance the privacy implications of new data types with the societal benefits of biomedical research.

6.4 International Nature of Research

The US population is less than 5% of the world population, but is very diverse in terms of representation of racial and ethnic groups. Aside from Native Americans, all US residents are relatively recent immigrants when compared to populations in Africa and Europe. Thus the genetic makeup of an individual in the US is as likely to closely resemble that of an individual outside the US as one inside. This argues that the US has advantage in carrying out biomedical research of global import because it can draw on the genetic diversity of its people. It also argues that it ultimately will be beneficial to achieve international interoperability for the sharing of health data. The ability to access international health data will aid the US research community in the ways described above for the benefits of increased numbers, and will aid international health care by making use of what is learned from the US population.

There are serious challenges to developing international interoperability that mirror those for a US national system, but are amplified by differences in national priorities, privacy expectations, legal systems, medical and insurance infrastructure, and more. The genomics research community has similar concerns, although on a smaller scale, and provides an example for how to proceed. Researchers at institutions in 40 countries have initiated a global alliance with the common goal of enabling the secure sharing of genomic data. This alliance is non-governmental and not-for-profit. In the letter of intent for this alliance, its members have outlined a set of core principles, which are largely applicable to the broader sharing of health data [57]:

- *Respect* data sharing and privacy preferences of participants
- *Transparency* of governance and operations
- *Accountability* for best practices in technology, ethics, and outreach

- *Inclusivity* based on partnering and building trust among stakeholders
- *Collaboration* based on shared information to advance human health
- *Innovation* in developing an ecosystem that accelerates progress
- *Agility* by acting swiftly to benefit those suffering with disease.

Agreement on these core principles provides a framework for developing international protocols for the exchange of biomedical data, as described in detail in the position paper from this group [58].

These considerations lead to the following finding and recommendation.

Finding

- The US population is highly diverse, reflecting much of the diversity of the global population. Therefore, important research findings applicable to Americans are likely to come from shared access to international health data. Currently there is no coherent mechanism for accessing such data for research.

Recommendation

- The ONC should exert leadership in facilitating international interoperability for health data sharing for research purposes. The genomics community is already engaged in such efforts for the sharing of sequence data, and the ONC should consider adopting a similar process.

6.5 Electronic Health Records and Health Care Fraud

It may seem out of place to include a section on health care fraud in a chapter promoting research. However, the same data access issues and health data analysis requirements that pertain to research underpin the ability to carry out fraud detection. In both case, broader access to health data and new health data analytic methods are needed.

The FBI estimates that the total annual loss due to health care fraud exceeds $80 billion [59]. In addition to this loss, the Government Accounting Office reported a federal government allocation of at least $0.6 billion to investigate and prosecute alleged health care fraud cases in 2011 [60]. In their 2012 annual report on health care fraud, the Health Care Fraud and Abuse Control Program noted that the federal government won or negotiated approximately $3 billion that year in fraud judgments and settlements [61]. Thus only a small fraction of the money lost to health care fraud is being recovered by existing fraud discovery and fraud recovery methods. This creates a substantial opportunity for exploiting EHR data to reduce costs from health care fraud.

The sources of fraudulent activities in health care are various and include patients (or fake patients), health care providers [62], health care equipment manufacturers, pharmaceutical companies [63], organized crime, terrorist groups, and foreign governments [64]. Therefore, the design of EHR systems should take into consideration what data could be collected to defend against or respond to fraud. Regardless, large collections of health care data likely could be analyzed in innovative ways to reveal

normal patterns of patient care and corresponding expenditures. Patterns outside these norms could receive special attention and require validation for payment by human investigators, as is common today for private medical insurance reimbursement.

Paradoxically, initial launches of local and regional EHR systems have generally been met with increases in health care costs, rather than the decreases one might expect if fraudulent activity were more transparent [64,65]. Two emergent phenomena can explain some of these unexpected increases. Electronic records can more easily be "cloned," whereby a patient is charged for services not received because the provider found it easy to copy-and-paste a record from a different patient. Also, it is easier for providers to exaggerate the level of care or the severity of a disease by "upcoding," which may involve simply clicking a box to trigger a higher billing charge. A combination of clearly articulated expectations for appropriate billing, training on the appropriate use of EHRs, and the use of EHR data analytics will be required to reduce these sources of unnecessary health care costs.

Beyond the more insidious problems noted above, certain abuses of the health care system should be readily identified by electronic analysis of EHRs. However, little effort is being made today to use EHRs to identify and reduce even the simplest of health care fraud tactics. In 2012 HHS announced a public-private partnership to prevent health care fraud [66]. This initiative includes the long-range goal to "use sophisticated technology and analytics on industry-wide health care data to predict and detect health care fraud schemes." In opposition to this intended goal, the Office of Inspector General recently terminated 400 of its fraud detection staff, which corresponds to about one fourth of its employees in this area [67]. This reduction in staff is likely to reduce actions taken by CMS in response to their existing predictive analytics software that is designed to spot patterns of fraud.

Clear indications of fraud should be easy to identify and simple actions can be taken to eliminate their sources. For example, delivery of disease-specific health care products to patients who have not been diagnosed with the corresponding disease could be uncovered by matching claims to diagnoses in EHRs. A major target for fraud detection could be the fraudulent prescription of schedule II drugs (medically-useful drugs of abuse such as opiates, sedatives, and stimulants). Inappropriate prescriptions, usually made with abnormally high frequency by a small number of doctors, can be identified by the fact that no prior medical exam has been completed for the "patient." Also, schedule II drugs cannot be refilled without a new prescription (with exceptions for patients in long-term care facilities or hospice care), and therefore associated claims should not be accepted. The CDC reports that schedule II drugs have replaced illicit drugs as the leading cause of death by drug overdose [68]. Therefore, reducing schedule II drug fraud will have a positive effect on both health care finances and the major societal problem of drug abuse.

An indirect mechanism of health care fraud is stealing EHR data that can be monetized. As health care records become electronic, there is greater risk that large amounts of data can be stolen during a single operation. Unlike conventional credit card or ID theft, a person's health history cannot be cancelled or restored to secrecy if publically disclosed, making the problem distinct from other types of data security challenges. Also, given that some diseases are hereditary, the disclosure of one person's medical records might allow someone to realize financial gain by predicting the disease risk of the

victim's relatives. It is possible that cyber criminals may be more likely to target someone for identity theft if their immutable personal history is already known. A complete medical record for an individual could be used to submit fraudulent medical bills, in this instance using a fake patient with a real and verifiable medical history.

Only modest innovation is needed to identify some of the examples of health care fraud noted above. However, the development of more sophisticated data analytics would allow investigators to uncover more obscure patterns of fraud. Moreover, as fraud detection strategies improve, innovative fraudsters will be forced to find new ways to exploit the health care system for financial gain. Therefore, JASON recommends that collections of de-identified EHR data be made available for researchers to develop strategies and algorithms to uncover subtle patterns indicative of fraud, and to adapt these algorithms to the changing tactics of fraud perpetrators.

Finding

- Electronic access to health data will make it easier to identify fraudulent activity, but at present there is little effort to do so using EHRs.

Recommendation

- Large-scale data mining techniques and predictive analytics should be employed to uncover signatures of fraud. A data enclave should be established to support the ongoing development and validation of fraud detection tools to maintain their effectiveness as fraud strategies evolve.

6.6 Modeling and Simulation

This chapter concludes with a few words about modeling and simulation in the context of EHRs and the electronic exchange of health information. Modeling and simulation has been proposed as a valuable technology to explore models and scale-up of the health data infrastructure [7,69]. However, efforts to date have tended to focus on modeling health care delivery of individual entrerprises [70–72], rather than that of a regional or national interoperable system. The approaches use multi-level simulations, frequently combined with agent-based models to optimize medical outcomes. Large amounts of data are collected on the various components of a health care enterprise, and rules describing the behavior of patients, health care providers, and insurers are collected to inform the simulations. In addition, information regarding disease models may be incorporated so that the outcomes of various treatment strategies can be considered.

These types of analyses can have great value in guiding how providers manage various medical conditions, and are a good example of the potential utility of HIT for improving health care. However, these models are only as good as the data that informs them. It is therefore all the more urgent that high quality, searchable medical data be made available and that a broadly interoperable software architecture to handle such data be put in place.

7 Concluding Remarks

This report has expressed disappointment in current US progress towards the creation of a robust health data infrastructure, while praising ONC and HHS for their persistence in trying to tackle one of the most vexing problems of today's society. JASON believes that the two overarching goals, improved health care and lower health care costs, can be achieved by moving to EHRs and the comprehensive electronic exchange of health information. JASON has provided a path toward realizing the promise of a robust health data infrastructure through the development of a unifying HIT software architecture that adheres to the following core principles, all embodying a focus on the patient:

- Be agnostic as to the type, scale, platform, and storage location of the data
- Use public APIs and open standards, interfaces, and protocols
- Encrypt data at rest and in transit
- Separate key management from data management
- Include with the data the corresponding metadata, context, and provenance information
- Represent the data as atomic data with associated metadata
- Follow the robustness principle: be liberal in what you accept and conservative in what you send
- Provide a migration pathway from legacy EHR systems.

The path forward must also include full access to health data for clinical care, public health, and biomedical research. This will require building the public's trust. Patients must believe that their privacy protection can be balanced with the perceived social benefit of access to their data. In addition, the path forward must open an entrepreneurial space for innovation in the development of tools and approaches for the delivery of health care.

This report will close with a few comments about improving the nation's health and its relevance to national security. Improving the health of Americans is a worthy goal in its own right, and from a national security perspective it is important to have an accurate assessment of the current health and potential health vulnerabilities of the population. A robust health data infrastructure would help to provide that barometer, and the National Academy of Engineering has put this forward as one of its grand challenges for the 21st Century [73]. The health of the nation also has implications in terms of military readiness and effectiveness [74], although this discussion focuses more specifically on implications related to the exchange of health information.

The security of the nation is tied to its economic vitality and ability to withstand economic disruption. An integral part of economic growth and stability is continued innovation, which requires a healthy and productive population to explore new opportunities for innovation. The US is the world leader in biomedical research and technology, as well as computer science and information technology. The nexus of these two areas promises to be fertile ground for business and economic development, and that development will be fueled by health data, especially if the data can be harnessed in a coordinated manner. As discussed in section 6.4, the US has a special advantage compared to other

more ethnically homogenous countries: it is a genetic melting pot that can be a crucible for discoveries related to personalized medicine and the genetic basis of disease. New business opportunities exist in these areas, and the vendors of legacy EHR systems, once mandated to open their systems to broad interoperability, will be especially well poised to move toward a business model of innovation rather than entrenchment.

In a time of crisis, having a strong health data infrastructure will help to keep the people and responsible parties informed about current health conditions and available treatment options. A crisis may necessitate temporary changes in health care services, which can be better managed with better information. Hurricane Katrina provided an example of what can go right and wrong with medical care in an emergency. The VA health system was able to transfer electronic records and refer patients to hospitals and clinics outside the New Orleans area even before the evacuation began. Of 40,000 VA patients in the affected area, approximately 20,000 accessed their health records in the immediate aftermath of the disaster [75]. Those outside the VA system did not fare so well. Quoting Brown *et al.* from the *American Journal of Public Health* [75]:"The nation will probably never have complete data on how many evacuees were unable to continue their medical regimens without interruption and what consequences resulted."

Following a natural disaster, act of terrorism, or military attack, first responders will be able to operate more efficiently if they have rapid access to the names, locations, and medical history of persons who may need immediate or specialized attention. At present, first responders operate mostly in reactive, rather than proactive, mode when entering an area of devastation. Patients are triaged and their medical information is collected on the fly. It would be more effective to know in advance, for example, how many people will require dialysis within the next 72 hours, who might require vaccination against tetanus, who can and cannot tolerate first-line antibiotics, and so on. An efficient system for the exchange of health information also would help support the CDC in identifying a natural disease outbreak or biological attack by tracking occurrences of exposure-related symptoms across the nation. In the case of a communicable agent, rapid access to information will enable faster decision making and response, thereby helping to mitigate the medical and societal consequences of the contagion.

Improved integration and analysis of health information could help prevent violent attacks by lone individuals. The perpetrator of suicide, domestic violence, or mass killing is often revealed after the fact to have had a long and troubled involvement with mental health professionals and other health care providers. If these dots can be connected in advance, there is the possibility that some tragedies could be avoided. Of course, this treads on complex legal and ethical issues related to the presumption of innocence and the right to privacy of medical information. However, the duty to warn (so-called "Tarasoff laws") requires a clinician to report a patient's behaviors that may pose imminent danger to him/herself or others [76]. Even apart from specific cases, a statistical study using historical data could ask what indicators are most predictive of violent behavior and with what time lag before the violent event occurs.

Finally, the security of the nation relies its cadre of first responders, armed services personnel, and health care providers, all of who themselves must be in good health and fit for duty. Lifestyle illnesses,

such as diabetes, obesity, and addictive disorders, reduce the pool of citizens who are able to serve in these capacities and reduce the overall readiness of those who do serve. Electronic health care monitoring coupled with EHRs is already lessening the incidence of obesity, especially in children [77]. Wider application of these methods should help to improve the nation's health. Maintaining a healthy population will improve the security of the nation and the lives of its people. Ultimately this is the promise of a robust health data infrastructure.

JASON is grateful to have had this opportunity to examine the challenging and important topic of enhancing the adoption and interoperability of EHRs. This report has advanced many general principles, as well as a specific example an HIT software architecture to facilitate migration to a software ecosystem, with a diversity of products and apps, that fosters innovation and entrepreneurship. JASON believes that now is time to define such an architecture, leveraging the opportunity to specify CMS Stage 3 Meaningful Use requirements to drive implementation. A fundamental precept of medicine is: "Above all, do no harm." A software architecture that is broadly tolerant of different scales, input types, and sites for data storage and processing offers a sure pathway, and one that will be open to future innovation. Patients and health care providers will be in a position to choose which particular implementations within the architecture have the most utility for their needs.

8 References

1. Blumenthal D, Tavenner M (2010) *N. Engl. J. Med.* **363:**501–504.
2. Ford EW, Menachemi N, Phillips, MT (2006) *J. Am. Med. Inform. Assoc.* **13:**106–112.
3. Ford EW, Menachemi N., Peterson LY, Huerta, TR (2009) *J. Am. Med. Inform. Assoc.* **16:**274–281.
4. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D (2011) *Health Aff. (Millwood)* **30:**464–471.
5. Adler-Milstein J, Salzberg C, Franz C, Orav EJ, Newhouse JP, Bates DW (2013) *Ann. Intern. Med.* **159:**97–104.
6. Nightingale F (1863) *Notes on Hospitals*, 3rd Edition (Longman, Green, Longman, Roberts & Green, London), p. 175.
7. National Advisory Commission on Health Manpower (1967) *Report of the Panel on the Impact of New Technology, Volume 1* (U.S. Government Printing Office, Washington DC), pp. 165–179.
8. Garfield S (1970) *Sci. Am.* **222(4):**15–23.
9. http://www.healthit.gov/facas.
10. President's Council of Advisors on Science and Technology (2010) *Report to the President Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forwar*d (Executive Office of the President, Office of Science and Technology Policy, Washington DC), available at http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf.
11. Grossmann C, Goolsby WA, Olsen L, McGinnis JM (2011) *Engineering a Learning Healthcare System: A Look at the Future* (The National Academies Press, Washington DC), available at http://www.nap.edu/openbook.php?record_id=12213.
12. Fleming NS, Culler SD, McCorkle R, Becker ER, Ballard DJ (2011) *Health Aff. (Millwood)* **30:**481–489.
13. https://www.healthvault.com/us/en.
14. http://bluebuttonplus.org.
15. Brown A, Weihl, B (2011) *Google Official Blog*, July 15, 2011, available at http://googleblog. blogspot.com/2011/06/update-on-google-health-and-google.html
16. Health Insurance Portability and Accountability Act of 1976 (HIPPA), 42 USC §§ 300gg *et seq.*, Public Law 104-191.
17. http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html.
18. Murphy K (2012) *EHR Intelligence*, November 19, 2012, available at http://ehrintelligence.com/ 2012/11/19/ehealth-initiative-survey-shows-challenge-of-sustaining-hies.
19. Adler-Milstein J, Bates DW, Jha AK (2013) *Health Aff. (Milwood)* **32:**1486–1492.

20. Mathematica Policy Research, Harvard School of Public Health, and Robert Wood Johnson Foundation (2013) *Health Information Technology in the United States: Better Information Systems for Better Care* (Robert Wood Johnson Foundation, Princeton NJ).

21. Menachemi N, Singh S (eds.) (2012) *Health Information Technology in the International Context, Advances in Health Care Management, Volume 12* (Emerald Group Publishing, Boston MA).

22. Sinha PK, Sunder G, Bendale P, Mantri M, Dande A (2013) *Electronic Health Record: Standards, Coding Systems, Frameworks, and Infrastructures* (IEEE Press, Piscataway NJ).

23. http://www.theregister.co.uk/2009/03/02/nhs_database_breach.

24. http://www.bbc.co.uk/news/uk-18386968.

25. http://news.techeye.net/security/uks-anonymous-health-records-are-wide-open.

26. Reinhardt UE (2012) *New York Times Blog*, July 27, 2012, available at http://economix.blogs.nytimes.com/2012/07/27/taiwans-progress-on-health-care.

27. Edmondson AC, Golden BR, Young GJ (2007, revised 2008) *Harvard Business School Case 68-061* (Harvard Business School Publishing, Boston MA).

28. http://www.va.gov/vler.

29. U.S. Government Accountability Office (2011) *VA and DOD Health Care: First Federal Health Care Center Established, but Implementation Concerns Need to Be Addressed* (GAO-11-570, Washington DC).

30. U.S. Government Accountability Office (2011) *Electronic Health Records: DOD and VA Should Remove Barriers and Improve Efforts to Meet Their Common System Needs* (GAO-11-265, Washington DC).

31. U.S. Government Accountability Office (2011) *Electronic Health Records: Long History of Management Challenges Raises Concerns about VA's and DOD's New Approach to Sharing Health Information* (GAO-13-413T, Washington DC).

32. Hagel C (2013) Memo from U.S. Secretary of Defense to U.S. Undersecretary of Defense for Acquisition, Technology and Logistics and Acting Undersecretary of Defense for Personnel and Readiness, May 21, 2013.

33. http://tools.ietf.org/html/rfc760.

34. http://www.rfc-editor.org/rfc/rfc1122.txt.

35. Mandl KD, Kohane IS (2012) *N. Engl. J. Med.* **366:**2240–2242.

36. Smith CM (2005) *J. Clin. Pharm.* **45:**371–377.

37. Young A, Chaudhr HJ Thomas JV, Dugan M (2013) *J. Med. Regul.* **99:**11–24.

38. U.S. Department of Health and Human Resources (2013) *The U.S. Nursing Workforce: Trends in Supply and Education* (Health Resources and Services Administration, National Center for Health Workforce Analysis, Bureau of Health Professions, Washington DC), available at http://bhpr.hrsa.gov/healthworkforce/reports/nursingworkforce/index.html.

39. Steinbrook R (2008) *N. Engl. J. Med.* **358:**1653–1656.

40. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balser JR, Masys DR (2008) *Nature Clin. Pharmacol. Ther.* **84:**362–369.

41. http://www.healthit.gov/policy-researchers-implementers/connect-gateway-nationwide-health-information-network.

42. http://www.healthit.gov/policy-researchers-implementers/direct-project.

43. Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, Levin WJ, Stuart SG, Udove J, Ullrich A, Press MF (1989) *Science* **244:**707–712.

44. Hudziak RM, Lewis GD, Winget M, Fendly BM, Shepard HM, Ullrich A (1989) *Mol. Cell. Biol.* **9:**1165–1172.

45. Telli ML, Hunt SA, Carlson RW, Guardino AE (2007) *J. Clin. Oncol.* **25:**3525–3533.

46. Flockhart DA, O'Kane D, Williams MS, Watson MS, Flockhart DA, Gage B, Gandolfi R, King R, Lyon E, Nussbaum R, O'Kane D, Schulman K, Veenstra D, Williams MS, and Watson MS (2008) *Genet. Med.* **10:**139–150.

47. Budnitz DS, Lovegrove MC, Shehab N, Richards CL (2011) *N. Engl. J. Med.* **365:**2002–2012.

48. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, Hwang SJ, *et al.* (2011) *Nature* **478:**103–109.

49. Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, Francke U, Mountain JL, Goldman SM, Tanner CM, Langston JW, Wojcicki A, Eriksson N (2011) *PLoS Genet.* **7:**e1002141.

50. Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, Saxonov S, Avey L, Wojcicki A, Pe'er I, Mountain J (2010) *PLoS Genet.* **6:**e1000993.

51. Eriksson N, Tung JY, Kiefer AK, Hinds DA, Francke U, Mountain JL, Do CB (2012) *PLoS One* **7:**e34442.

52. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, Wang D, Masys DR, Roden DM, Crawford DC (2010) *Bioinformatics* **26:**1205–1210.

53. Denny JC, Crawford DC, Ritchie MD, Bielinski SJ, Basford MA, Bradford Y, Chai HS, Bastarache L, Zuvich R, Peissig P, *et al.* (2011) *Am. J. Human Genet.* **89:**529–542.

54. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y (2013) *Science* **339:**321–324.

55. http://www.dataenclave.org/index.php/home/welcome.

56. http://www.census.gov/ces/rdcresearch.

57. http://www.broadinstitute.org/news/globalalliance.

58. http://www.broadinstitute.org/files/news/pdfs/GAWhitePaperJune3.pdf.

59. http://www.fbi.gov/about-us/investigate/white_collar/health-care-fraud.

60. U.S. Government Accountability Office (2012) *Health Care Fraud: Types of Providers Involved in Medicare Cases, and CMS Efforts to Reduce Fraud* (GAO-13-213T, Washington DC).

61. U.S. Office of the Inspector General (2012) *Department of Health and Human Services and Department of Justice Health Care Fraud and Abuse Control Program Annual Report for Fiscal Year 2012* (Washington DC), available at http://oig.hhs.gov/publications/docs/hcfac/hcfacreport2012.pdf.

62. Abelson R, Creswell J, Palmer G (2012) *The New York Times*, September 23, 2012, p. A4, available at http://www.nytimes.com/2012/09/22/business/medicare-billing-rises-at-hospitals-with-electronic-records.html?pagewanted=all.

63. G. Harris G (2009) *The New York Times*, September 2, 2009, p. B4, available at http://www.nytimes.com/2009/09/03/business/03health.html.

64. Weaver J (2013) *Miami Herald*, April 25, 2013, available at http://www.miamiherald.com/2013/04/25/3364035/florida-man-behind-medicare-money.html.

65. Sebelius K, Holder EH (2012) U.S. Department of Health & Human Services letter to American Hospital Association, Association of Academic Health Care Centers, National Association of Public Hospitals and Health Systems, Federation of American Hospitals, and Association of American Medical Colleges, September 24, 2012.

66. Sebelius K, Holder EH (2012) U.S. Department of Health & Human Services Press Release, July 26, 2012, available at http://www.hhs.gov/news/press/2012pres/07/20120726a.html.

67. Bresnick J (2013) *EHR Intelligence*, July 1, 2013, available at http://ehrintelligence.com/2013/07/01/cms-cuts-400-oig-fraud-detection-staff-despite-expansions.

68. Paulozzi LJ (2011) *Morbidity and Mortality Weekly Report* **60(1):**60–61.

69. Valdez RS, Ramly E, Brennan PF (2010) *Industrial and Systems Engineering and Health Care: Critical Areas of Research - Final Report* (Agency for Healthcare Research and Quality, Rockville MD).

70. Park H, Clear T, Rouse WB, Basole RC, Braunstein ML, Brigham KL, Cunningham L (2012) *Service Science* **4:**253–268.

71. Bigus, JP, Chen-Ritzo C-H, Sorrentino R (2011) In *Proceedings of the 2011 Winter Simulation Conference* (IEEE, Piscataway NJ), pp. 1103–1116.

72. Bigus JP, Chen-Ritzo C-H, Hermiz K, Tessauro G, Sorrentino R (2012) In *Proceedings of the 2012 Winter Simulation Conference* (IEEE, Piscataway NJ), pp. 1–12.

73. http://www.engineeringchallenges.org/cms/8996/8938.aspx.

74. http://www.missionreadiness.org/wp-content/uploads/MR_Too_Fat_to_Fight-11.pdf.

75. Brown S, Fischetti LF, Graham G, Bates J, Lancaster AE, McDaniel D, Gillon J, Darbe M, Kolodner RM (2007) *Am. J. Public Health* **97:**S136–141.

76. Herbert PB, Young KA (2002) *J. Am. Acad. Psychiatry Law* **30:**275–281.

77. Coleman KJ, Hsii AC, Koebnick C, Alpern AF, Bley B, Yousef M, Shih EM, Trimble-Cox KJ, Smith N, Porter AH, Woods SD (2012) *J. Pediatrics* **160:**918–922.