

Data for Individual Health

Prepared for:

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

Contract No. JSR-14-TASK-007

Prepared by:

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

AHRQ Publication No. 15-0006-EF
November 2014



Support
provided by

Robert Wood Johnson Foundation



Data for Individual Health

Contact: Dan McMorow — dmcmorow@mitre.org

November 2014

JSR-14-TASK-007

Approved for public release; distribution unlimited

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

Contents

1 EXECUTIVE SUMMARY	1
1.1 Introduction	1
1.2 Study Charge	1
1.3 Summary	2
1.4 Relevance of the 2013 JASON Report on Health Information Technology	3
1.5 Data Associated with Health	4
1.6 Progress since the 2013 JASON Study	4
1.7 Strategies to Accelerate Progress towards a Learning Health System	5
1.8 The Need for a Nimble Regulatory Environment	6
1.9 Findings	7
1.10 Recommendations	8
2 INTRODUCTION	11
2.1 JASON Study Charge	13
2.2 JASON Study Process	15
2.3 Current State	16
2.4 Overview of Previous Work	20
2.5 JASON 2013 Report on Robust Health Data Infrastructure . .	22
3 DATA AND ANALYTICS	27
3.1 Genomics and Other “omics” Data	27
3.2 Personal Health Monitoring Devices	29
3.3 Moving Diagnostics to the Point of Care	31
3.3.1 Testing for biomarkers	32
3.3.2 Miniaturizing the diagnostics laboratory	33
3.4 Patient Reported Outcomes and Collaborative Networks . . .	34
3.5 Social Media Data	35
3.6 Data Analytics	36
3.6.1 Risk stratification example	37
3.6.2 Clinical decision support system example	39
3.6.3 Natural language processing example	41
3.6.4 Health and environment example	43
3.7 Findings and Recommendations for Data and Analytics	45

4	INTEROPERABILITY	47
4.1	Opening the EHR System APIs	47
4.2	Recent Progress on Standards	52
4.2.1	The clinical document architecture (CDA) standard . .	52
4.2.2	The proposed FHIR standard	54
4.3	REST and FHIR as the Foundation of a Future Health IT System	59
4.4	User Interfaces	61
5	OPPORTUNITIES AND CHALLENGES	69
5.1	Communities	69
5.2	Leveraging Non-Profit Organizations	71
5.3	The Continued Evolution of the Health Care Team	73
5.4	Health Care Teams and Soft Regulation of Health Education .	76
5.5	Privacy and Security Considerations	80
6	CONCLUSIONS	83

1 EXECUTIVE SUMMARY

1.1 Introduction

The promise of improving health care through the ready access and integration of data continues to draw significant national attention and federal investment. Information technology is rapidly expanding this data universe beyond traditional information associated with health care providers to embrace information in the larger spheres of health and wellness. This includes not only electronic health records (EHRs), but also personal health records (PHRs) and sources such as environmental data and social media data, some of which may be related only indirectly to the delivery of health care. To date, federal investments in health data infrastructure development, through mechanisms such as the Centers for Medicare and Medicaid Services (CMS) Medicare and Medicaid Electronic Health Record Incentive Program, have focused on the medical care of individuals. This report discusses how to expand this vision, with a focus on the health of individuals and the development of a Learning Health System.

1.2 Study Charge

JASON was asked to address how to bridge to a system focused on health of individuals rather than care of individuals. The questions posed to JASON include:

- How can EHRs highlight opportunities to engage individuals as they try

to achieve health, and to learn from failed efforts to improve treatments for individuals?

- How would an EHR of the future help a care team whose goal was health, rather than health care, and who, along with the individual, have access to all the data?
- How can data analytics be used to support high quality, patient centered care and offload the large requirements of processing?

The ultimate goal is to achieve an agile, national-scale “Learning Health System” for identifying and sharing effective practices of care.

1.3 Summary

Today, the delivery of health care moves in a linear fashion, proceeding from preventive medicine, to diagnosis, to treatment, and ultimately to outcomes. This process is informed by clinical research, but there is an inadequate feedback loop between health care outcomes and clinical research, reducing opportunities for further learning in this system. Additionally, population health research and community engagement are not adequately connected. A “Learning Health System” would connect the medical system with broader societal inputs, creating important links between health and wellness and health care. This concept highlights natural roles for EHRs and PHRs, but also points to a level of data access, integration, and scalability that goes well beyond the interoperability of EHR systems.

1.4 Relevance of the 2013 JASON Report on Health Information Technology

In its 2013 report to HHS/AHRQ [23], JASON described a set of principles for a health information technology ecosystem that would open entrepreneurial opportunities and accelerate progress. This ecosystem is to be underpinned by a proposed software architecture that could serve as an organizing framework for the development and implementation of a health data infrastructure. Such an infrastructure can also be leveraged to facilitate a Learning Health System. Here, architecture refers to a collection of components of a software system that interact in specified ways and across specified interfaces for the purpose of guaranteeing a specified functionality. The software architecture devised by JASON is not a monolithic system; rather it is a framework for the development and implementation of a health data infrastructure. This architecture and the associated infrastructure focus on the role of data and data exchange. To gain interoperability, JASON recommended in 2013 and recommends again in this report, the establishment of publicly available APIs to bridge from existing systems to a future software ecosystem that can ingest, protect, integrate, and share the knowledge gained from the vast stores of data. The 2013 report pointed to an inevitable shift from a small number of proprietary systems to a software ecosystem with a diversity of products and interacting applications or “apps” that would increasingly make it possible to achieve interoperability and, more importantly, a focus on the individual’s access to his or her data.

1.5 Data Associated with Health

In extending the ideas of the JASON-proposed architecture to the broader realm of health, it will become necessary to expand greatly the types of data that can be ingested and analyzed. In addition to the traditional data associated with health care (e.g., EHR data), it will also be necessary to assimilate data from PHRs. These include, for example, data from personal health devices, patient collaborative networks, social media, environmental and demographic data, and the burgeoning data streams that will soon become available through progress in genomics and other “omics.” Despite the profusion and complexity of new data sources associated with personal health, the architecture for a learning health system would look essentially the same as that proposed by JASON for EHRs except that the data layer must also encompass these highly diverse forms of personal health information. The requirement for interoperability through the adoption of open APIs becomes even more critical here; without this interoperability it will be extremely difficult to scale up today’s health information technology (IT) systems to assimilate and analyze these new data sources.

1.6 Progress since the 2013 JASON Study

There has been significant progress on some of the recommendations of the 2013 JASON report. There are now promising approaches to the problem of representation of atomic data, with the inclusion of metadata to establish data provenance. These approaches can be extended to the larger realm of health. There have also been important developments in the construction of reusable and distributed user interfaces, which are essential for interoperabil-

ity between health information systems. Another key development is the establishment of data interchange APIs for mobile health, making the installed base of 140 million smartphones a natural platform for collection, assimilation, and exchange of EHR and PHR data. However, there remains a critical need for open APIs for EHR systems to further open the entrepreneurial space. In JASON’s view, any API that is exposed to EHR customers should also be exposed to the general application development ecosystem. This, in turn, will enable increased partnering among health care providers and empower individuals to increase their meaningful participation in their own health and wellness.

1.7 Strategies to Accelerate Progress towards a Learning Health System

In addition to the development of interoperable IT systems described above, JASON suggests implementing several strategies to accelerate progress. First, it would be desirable to leverage the work of various nonprofit institutions that focus, for example, on specific medical conditions or on the overall health of certain population groups, such as the American Heart Association or AARP. These organizations could advocate for interchange of EHR and health data, as well as encourage the development of applications that further their mission and promote overall health.

Second, there is a need to improve data exchange among members of the individual’s health care team. This would have the benefit of facilitating the efficient use of the diverse sources of data associated with PHRs, genomics, etc., and might be achieved with the addition of a “care coordinator” to the team who can serve as an integrator of the data streams. In any case, in the

members of the health care team in this anticipated data-rich environment will require training to a higher level of quantitative literacy.

Lastly, developing reliable indices of the health of a community, based on both geographic and socioeconomic indicators, could lead to a better understanding of how best to integrate community support for health.

1.8 The Need for a Nimble Regulatory Environment

As the level of “wellness information” increases, both in amount and sophistication, a question arises as to when such information impinges on the practice of medicine. For example, the FDA has enjoined genetic testing companies from providing customers information connecting the existence of various single nucleotide polymorphisms (SNPs) in their genome to the potential occurrence of certain diseases because of concerns over demonstrated clinical accuracy of the diagnostic results. FDA argues that such SNP analysis, although analytically accurate, falls in the space of medical diagnosis rather than patient education. Such datastreams, provided they are accurately measured and reported, are relevant to medical diagnosis and prognosis, but it would also be desirable to allow consumers to continue to make use of such services, and, importantly, to provide mechanisms by which this information could be shared with the consumer’s health care team for appropriate medical interpretation. JASON believes that a new more nuanced approach in making such information available to consumers should be applied here and in other similar cases.

1.9 Findings

1. There have been numerous previous reports regarding health care information, all are in broad agreement regarding the need for greater interoperability and data integration.
2. There is an explosion of data from many and varied sources. Yet there is little understanding of how to parse, analyze, evaluate, merge, and present these data for individuals and for the health care team. The health data infrastructure currently does not have the capability to make the data accessible in usable form, including the associated meta-data and provenance.
3. There is a critical need for open Application Programming Interfaces (APIs) that effectively support an entrepreneurial ecosystem.
4. Pilot projects are emerging that encourage intra-community partnering to create healthier environments for citizens. However, metrics are lacking that provide a reliable index of community health.
5. Non-profit organizations, for example those devoted to finding cures for particular diseases, have strong patient and community support and trust. These attributes could be leveraged to support health and wellness.
6. Health care teams are growing in size and diversity of expertise. The effectiveness of such teams will critically depend on intra-team communications enabled by access to data and fluency in a common parlance of health informatics. A higher level of quantitative literacy will be required among team members to make informed health decisions from the growing body of health-relevant data and analytics. There are op-

portunities for new professional roles, integrating medicine, analytics, and social and behavioral sciences in support of health care consumers.

7. There is an inherent tension between the flow of health information to the individual and the need to protect individuals from misuses of that information. Individuals are already seeking such information and acting upon it. Hard lines currently exist between regulated and unregulated products and services. However, technology is evolving quickly in this arena. There are opportunities to update the regulatory process to benefit health and wellness, while also accelerating innovation of accurate, personalized health care.
8. The learning health system needs to be “closed loop” to ensure a continuous and transparent cycle of research, analysis, development, and adoption of improvements relevant to health and wellness and to the delivery of health care.

1.10 Recommendations

1. It is time to act on the broad agreement found in past reports regarding health information. HHS should take the lead in harmonizing the recommendations and promoting paths forward. HHS should establish a framework for measuring progress along these paths.
2. HHS should adopt standards and incentives to allow sharing of health data. HHS policies should require that metadata and provenance be associated with all data so that data quality and use can be evaluated.
3. HHS should adopt policies of support for proposed open API standards. These policies should make it advantageous for one or more leading

EHR vendors to be the first to propose such standards. HHS should recognize ecosystem-friendly EHRs as a public good. HHS should adopt reimbursement differentials, initially small, for institutions that adopt ecosystem-friendly EHRs.

4. HHS, in partnership with private foundations, should establish “race to the top” challenges for community involvement. These would seek to demonstrate, through the integration of community services and broader wellness services, a measurable increase in health and wellness, and a concomitant reduction in the number of encounters with the health care system.
5. Relevant non-profits should be encouraged to assess their goals with respect to health data streams, and to provide “stamps of approval” for applications (apps) and other consumer tools. This will speed adoption of these tools.
6. Accrediting bodies should require training of all health care team members to achieve required levels of numeracy and fluency in a common parlance of health informatics. Professional schools should develop continuing education and certification programs that cross-educate team members regarding the diverse informatics expertise required by the team.
7. FDA and other agencies should seek a nuanced approach to adjudicating the regulatory line. Products and services that now risk being construed as “practicing medicine” could be allowed to exist in a more nimble regulatory space that fosters rapid innovation and adaptation to new developments while mitigating concerns over demonstrated accuracy. For example, products and services in this space could be subject

to a requirement to simultaneously report information to the designated health care team, as well as to the consumer, as a risk mitigation strategy.

8. HHS should accelerate the development and adoption of a robust health data infrastructure based on the principles outlined in the ONC 10 Year Vision and other consensus studies.

2 INTRODUCTION

The promise of improving health care through the ready access and integration of data continues to draw significant national attention and federal investment. The universe of health-relevant data is rapidly expanding beyond traditional information associated with managed health care, to include new and emerging sources of information on health and wellness. To date, federal investments in health data infrastructure development, through mechanisms including the Medicare and Medicaid Electronic Health Record Incentive Program, have focused on the medical care of individuals. This report addresses how to expand this vision, with a focus on individual-centric health, culminating in the development of a Learning Health System that serves the present and future needs of the US population.

Two oft-reported observations provide motivation for moving the focus from health care to the broader health and wellness of the individual. First, **most health care doesn't occur in medical centers**. The table in Figure 2-1, adapted from [13] characterizes how a population of 1000 individuals interact with the health care system in an average month. It is clear that only a small fraction of individuals with health issues visit a physician, only 1% are hospitalized, and less than 0.1% are hospitalized in an academic medical research center. Some argue that current investments in health care are inconsistent with this reality.

The second trend is that **behavioral risk factors strongly impact health** and wellness. Table 2-2 summarizes causes of death in 1990 and in 2000 from [29, 30]. The entries in bold font are considered modifiable behavioral risk factors and represent 40% of the deaths in both years. These data suggest

Table 2-1: Estimated use of health care system for a population of 1000 individuals during an average month. Adapted from [13].

Reported State	Estimated %
Report discomfort, illness, or injury	80%
Consider seeking medical care	33%
Visit a physician's office	22%
Visit a complimentary or alternative medical provider	7%
Receive home health care	2%
Visit an emergency department	1%
Hospitalized	1%
Hospitalized in an academic medical center	<0.1%

that current health care research and investments do not adequately address the role played by social determinants in health and wellness.

Both of these trends highlight the need for broadening health care beyond traditional medical care as practiced by physicians. This new, broader health system should be designed to improve the health of individuals by linking traditional health care with new sources of relevant data, and by embracing the critical role of societal support in modifying behavior and making informed choices. This broadening can best be enabled through the development and implementation of a robust health data infrastructure, one which is

Table 2-2: Estimated causes of death in 1990 and 2000 (from [29, 30]).

Actual Cause	% of Total Deaths 1990	% of Total Deaths 2000
Tobacco	17%	18%
Poor diet and physical inactivity	14%	15%
Alcohol consumption	5%	4%
Microbial agents	4%	3%
Toxic agents	3%	2%
Motor vehicle	1%	2%
Firearms	2%	1%
Sexual behavior	1%	1%
Illicit drug use	<1%	1%
Total in Table	50% (1.1M)	47% (1.1M)

able to integrate and interact with data beyond the current electronic health records.

2.1 JASON Study Charge

Health and Human Services (HHS), through the Agency for Healthcare Research and Quality (AHRQ), requested this JASON study. AHRQ, an agency within HHS, promotes research on the quality, safety, accessibility and affordability of health care, with the goal of improving health care decision-making and the quality of health care for all Americans. HHS asked JASON to address the nationally significant challenge of health and wellness, recognizing health care as one key resource helping the public achieve and sustain health, but not necessarily the most important. HHS is interested in how best to

achieve and sustain individual health by using data in predictive, preventive, personalized, and participatory ways. This includes leveraging tools such as health IT (e.g. electronic health records), mobile devices and wireless sensors, web and social media, and genomics related to personalized medicine and tailored to behavioral interventions.

The challenges noted by HHS are that:

- The role of the patient in the process has not been well-defined,
- Presentation, analysis and interpretation of large amounts of data is challenging,
- Incorporation of mobile technologies, personal monitoring devices and other self-reported information has not been standardized,
- Work force training is required to service an expanding patient population,
- Integration of all members of the Care Team is needed in managing a patient's care,
- Home health care and care provided in rural settings must be integrated into the data-informed health care system.

Specifically, JASON was asked to address how to bridge, on the national scale, to *a system focused on health of individuals rather than care of individuals*. The questions posed include:

- How can EHRs highlight opportunities to engage individuals as they try to achieve health, and to learn from failed efforts to improve treatments for individuals?

- How would an EHR of the future help a care team whose goal was health, not health care, and who, along with the individual, have access to all the data?
- How can data analytics be used to support high quality, patient centered care and offload the large requirements of mental processing?

The ultimate goal is to architect an agile, national-scale, *Learning Health System* for developing and sharing effective practices of achieving and maintaining health.

2.2 JASON Study Process

JASON was introduced to the topic through presentations by, and discussions with, the briefers listed in Table 2-3. These individuals represented patients, physicians, care givers, medical center leadership, entrepreneurs, electronic health record and IT vendors, and researchers. Most briefers attended the full set of presentations and participated in the accompanying discussions. Materials recommended by these individuals, together with a wide range of other publicly available materials, were reviewed and discussed by JASON. JASON gratefully acknowledges the efforts of Dr. P. Jon White (HHS AHRQ), Joy Keeler Tobin (CMS Alliance to Modernize Health) and Dr. Michael Painter (Robert Wood Johnson Foundation) who provided important help and guidance in coordinating the briefings.

Table 2-3: Briefers for the 2014 Study.

Briefer	Affiliation	Briefing title
Ruben Amarasingham	Parkland Medical Center	Electronic Healthcare Predictive Analytics (e-HPA)
Patti Brennan	University of Wisconsin	Patients, Caregivers and Providers Panel
John Brownstein	Harvard Medical School	Computational Epidemiology
Carol Cain	Kaiser Permanente	Health and healthcare delivery system
Shub DebGupta	WiserTogether	Technology Panel
Karen DeSalvo	HHS/ONC	ONC's Vision
Carl Dvorak	Epic	JASON Briefing
Kathy Ensor	Rice University	Environmental Sensors
Brian Fitzgerald	FDA	Technology and Software Architecture for High Performance Computing
Dave Gustafson	University of Wisconsin	Active Aging
John Halamka	Beth Israel Deaconess Medical Center	Connecting Payers Providers and Patients
Sridhar Iyengar	Misfit Wearables	Technology Panel
Ross Koppel	University of Pennsylvania	Socio-technical systems
David McCallie	Cerner	Data for Individual Health
Erin Moore	Cincinnati Children's Hospital	Patients, Caregivers, and Providers Panel
Mike O'Reilly	Apple	Apple HealthKit
Carlos Rodarte	PatientsLikeMe	Technology Panel
Danny Sands	Beth Israel Deaconess Medical Center	Patients, Caregivers and Providers Panel
Ida Sim	UC San Francisco	mHealth Data
Bill Stead	Vanderbilt University	Path to a Learning Health Care System
Paul Tang	Palo Alto Medical Foundation	Disrupting the Status Quo: Putting Healthy People First
Joy Tobin	MITRE	Health It Safety
Daniel Wattendorf	DARPA	DARPA Biotechnologies
Jenna Wiens	MIT	Leveraging the EMR

2.3 Current State

The Office of the National Coordinator (ONC), defines an electronic health record (EHR) as a digital version of a patient's (individual's) paper chart, noting that this is managed and curated by authorized medical providers [21].

ONC makes the distinction between an EHR and a personal health record (PHR), where a PHR is similar to an EHR, but the content is managed by the individual. These records may overlap in content and need not be identical. For example, the PHR may contain fitness and wellness information that is not part of the EHR and has not been used in the delivery of the individual's

Health Care Delivery

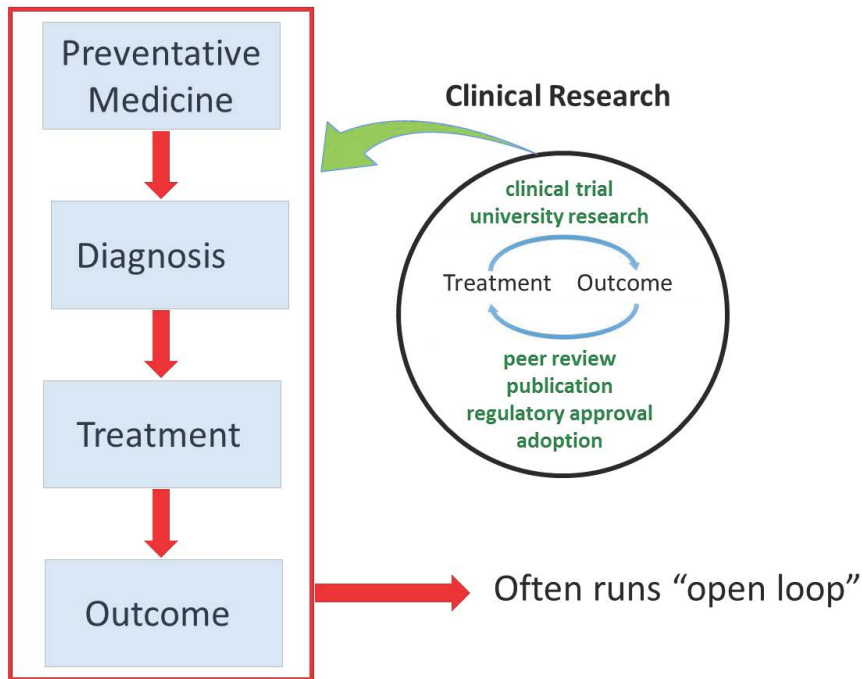


Figure 2-1: Current relationship between health care delivery and clinical research.

health care. In 2005, the Institute of Medicine (IOM) established the IOM Roundtable on Evidence Based Medicine [7] (now called the Roundtable on Value and Science-Driven Health Care). In their original charter and vision statement they introduce the concept of a Learning Healthcare System:

“We seek the development of a learning health care system that is designed to generate and apply the best evidence for the collaborative health care choices of each patient and provider; to drive the process of discovery as a natural outgrowth of patient care; and to ensure innovation, quality, safety, and value in health care.”

It is useful to assess where we are with respect to this vision. Today, the delivery of health care moves in a linear fashion, proceeding from preventive medicine, to diagnosis, to treatment, and ultimately to outcomes, as depicted in Figure 2-1. Clinical research has a well-established learning and discovery cycle between research treatment and outcome; this feeds into health care delivery. However, there is typically an open loop between health care outcomes and clinical research, resulting in a failure to effectively link research to treatment. An important step in achieving a Learning Healthcare System, as envisioned in Figure 2-1, is to close this loop and create a full learning cycle between clinical research and health care delivery. However, there is an important distinction between a Learning Healthcare System as envisioned in Figure 2-2 and the desired Learning Health System as envisioned by the ONC. As noted above, the change of focus from Healthcare to Health requires incorporation of an entirely new set of health and wellness inputs and associated research, as shown in Figure 2-2.

The set of activities captured in the second circle added in Figure 2-2 currently run largely parallel to the traditional health care delivery system. However, integration of these two components is critical to a Learning Health System. Figure 2.3 illustrates this vision. Here, there is a closed loop between health care and health and wellness, each of which is supported by a set of data and institutions. For example, EHRs exist mainly in the health care domain, whereas PHRs can include broader health and wellness information, at the individual's discretion. The institutions feeding into health care include clinics, pharmacies, hospitals, the National Institutes of Health (NIH), CMS, EHR vendors, Emergency Medical Services (EMS), insurers, and urgent care, whereas those feeding into health and wellness include community services, public health services, health clubs and gyms, social media, and

Health Care Delivery

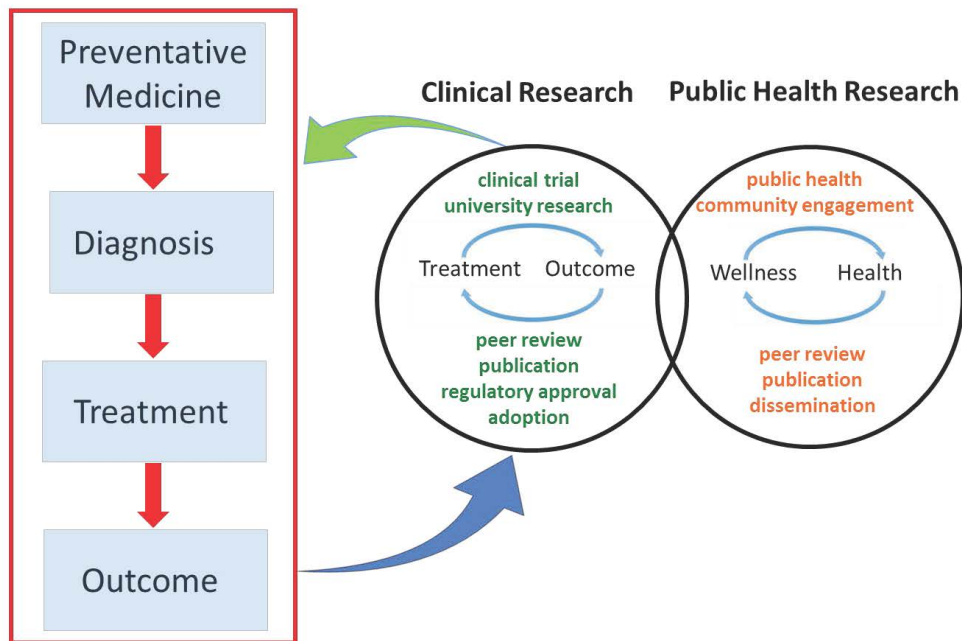


Figure 2-2: Desired relationships between health care delivery, clinical research, and public health research.

also health software. Such a system could support current and future clinical practice, delivery of health care, basic research, and public health. The IOM Roundtable has also evolved in its thinking. Its recent vision captures the concept of a continuously learning health system[8].

“Our vision is for the development of a continuously learning health system in which science, informatics, incentives, and culture are aligned for continuous improvement and innovation with best practices seamlessly embedded in the care process, patients and families active participants in all elements, and new knowledge captured as an integral byproduct of the care experience.”

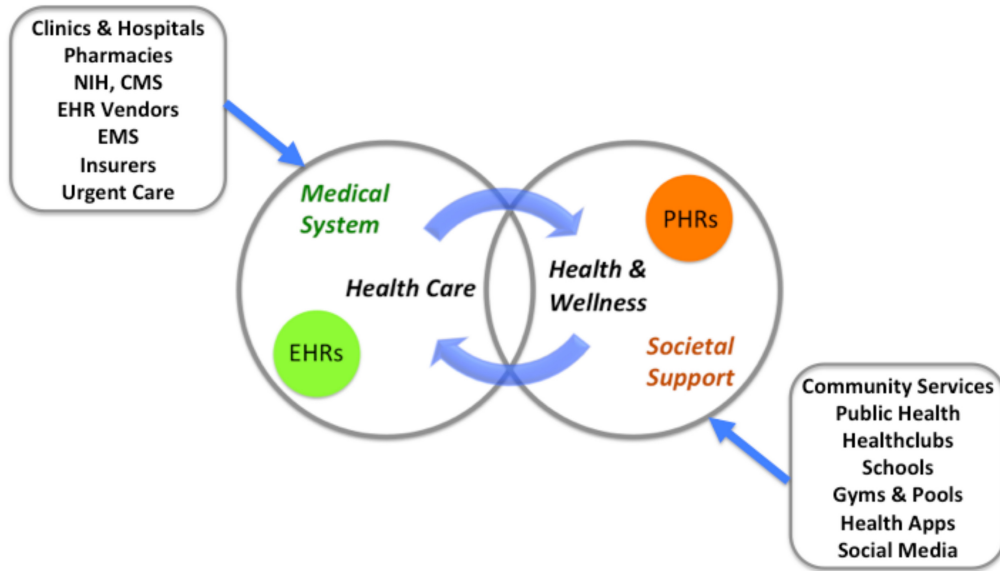


Figure 2-3: Integration of data required for a Learning Health System.

2.4 Overview of Previous Work

There have been numerous, excellent reports on health IT, EHRs, Learning Healthcare Systems, and measurement of health and health care. In addition, JASON has written reports on similar topics in other domains. Figure 2-4 provides a list of some of these reports.

There is **commonality in the conclusions across these reports**. Common recommendations include:

- Improve access to data,
- Create open interfaces,
- Define objectives and match analytics and data streams,

- National Research Council
 - Institute of Medicine: 15 workshops and 1 consensus panel on learning health systems (2007-2014) [7]
 - Committee on National Statistics: Accounting for Health and Healthcare: Approaches to Measuring the Sources and Costs of Their Improvement [6]
- President’s Council of Advisors on Science and Technology
 - Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward (2010) [34]
 - Better Health Care and Lower Costs: Accelerating Improvement Through Systems Engineering (2014) [33]
- JASON (including parallel topics in other domains)
 - Robust Health Data Infrastructure(2013) [23]
 - Data to Decision (not publicly released)
 - Data Analysis Challenges(2008) [22]

Figure 2-4: A non-exhaustive list of reports issued on health IT, EHRs, Learning Healthcare Systems and measurement of health and health care.

- Improve coordination and communication across agencies and with public/private sectors,
- Pose challenge problems,
- Develop workforce, build competencies.

This leads to a finding and recommendation for turning previous advice into action.

Finding: There have been numerous previous reports regarding health care information technology, and there is broad agreement regarding the need for greater interoperability and data integration.

Recommendation: It is time to act on these common recommendations found in past reports regarding health information. HHS should take the lead in harmonizing the recommendations and promoting paths forward. HHS should establish a framework for measuring progress along these paths.

2.5 JASON 2013 Report on Robust Health Data Infrastructure

It is worthwhile to revisit some of the key points from the 2013 JASON report [23]. That report concluded that interoperability is within reach, provided the health data infrastructure adopts a common architecture that adheres to the principles listed below. The report called for an architecture that

- Is agnostic as to data type, data scale, platform, and storage location,
- Uses published APIs, open standards and protocol,
- Includes metadata, context, and provenance of the data,
- Represents the data as atomic data with associated metadata,
- Provides a migration path for currently implemented software systems,
- Encrypts data at rest and in transit,

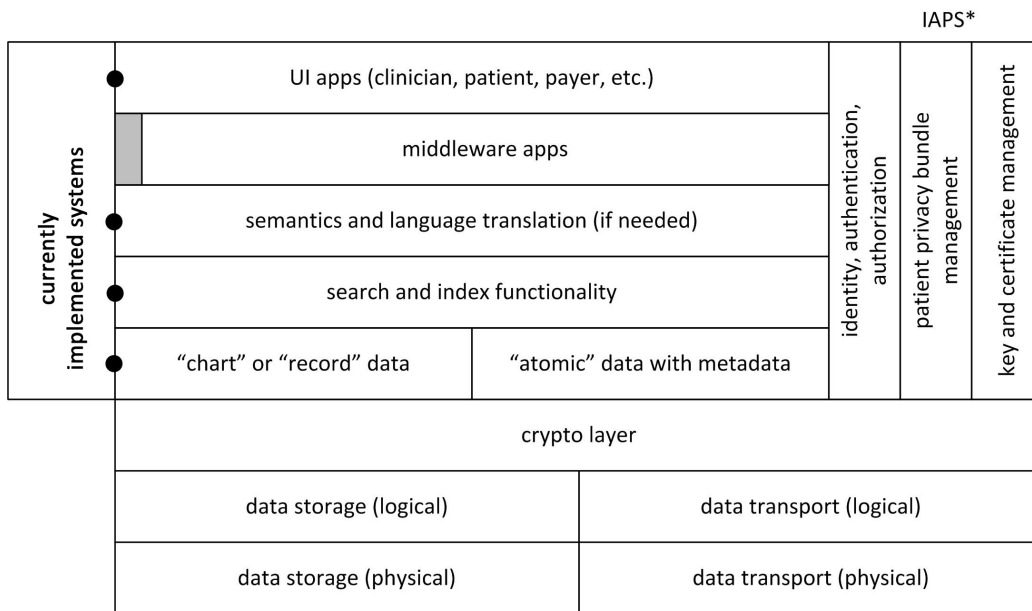


Figure 2-5: A diagram of the Health IT architecture proposed by JASON in 2013 [23].

- Separates key management from data management,
- **Follows the principle that the patient participates in the management of his or her data.**

Those principles highlighted in bold are revisited in this report:

JASON also provided an exemplar architecture that follows those principles. In this context, architecture refers to a collection of components of a software system that interact in specified ways and across specified interfaces for the purpose of guaranteeing a specified functionality. The JASON example architecture, diagrammed in Figure 2-5, resembles a protocol stack and provides an organizing concept for developing the infrastructure, opening entrepreneurial spaces throughout. Such a principle has been used in evolving

the Internet as we know it today. This architecture and the resulting infrastructure is all about data and data flow. This includes the ability to ingest all of the data that will inform the Learning Health System; the functionality to keep the data safe; the information (metadata, provenance, and context) required to find, understand and integrate the data; as well as the ability to communicate the knowledge gained. The remainder of this report will develop these concepts as they relate to health as opposed to health care.

It is important to clarify how JASON envisions that the architecture be applied. The idea is to use the architecture as a guide for the separation of concerns. In the design of any EHR system the various functions associated with the boxes in the architecture diagram, Figure 2-5, are handled through specific interfaces between those boxes. The actual implementation of the functionality need not be part of the application being designed. In the integrated setting of a Learning Health System this approach becomes especially important. The overall vision is shown in Figure 2-6. Regardless of where information is being gathered, if the relevant application is designed with the desired design principles, the overall benefits as described in the 2013 JASON report [23] are preserved.

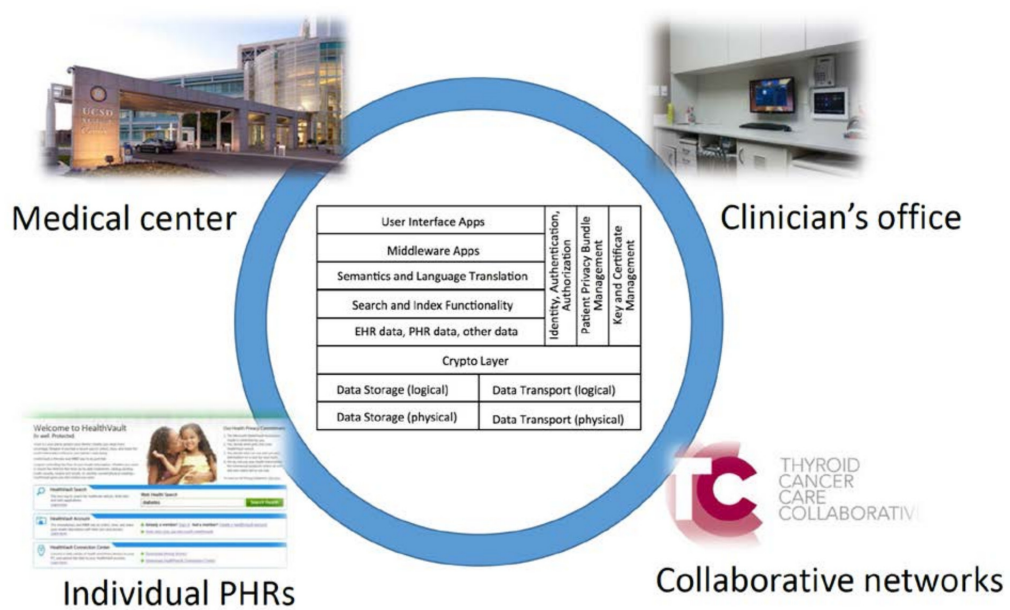


Figure 2-6: The architecture proposed by JASON is meant to tie together diverse information resources and so is also applicable to a learning health system.

3 DATA AND ANALYTICS

The Learning Health System will need to link the medical system (delivery of health care) to societal support and inputs. This expands the types and roles of data and analytics beyond current EHRs. The JASON 2013 report [23] considered various data types that will need to be supported in a robust data infrastructure, in both EHRs and PHRs, although the term PHR was not used in that study.

3.1 Genomics and Other “omics” Data

The cycle between health and wellness and health care will be informed through the capture, use, and combination of genotypic, environmental, and phenotypic data. This will allow the tailoring of health care to the individual. Advances in the collection and use of genomics and other “omics” data continues to grow. These “omics” technologies seek to identify and quantify the molecules, large and small, in cells of the human body, and in microbes that inhabit it. Phenotypic information is being gathered through surveys and self-reporting on the part of individuals. This includes, for example, the reporting and sharing of family history, treatment experiences, and social consequences associated with a chronic disease within a collaborative digital network. In addition, there is wide-spread collection of health-relevant information from wearable devices and smartphone-based apps. The combination of all of these data streams constitutes, arguably, a more sensitive indicator of human health-related phenotypes than has ever existed before. There is tremendous potential for health benefits to be derived from leveraging all of this data in a detailed read-out of the phenotype that can be used to assess,

for example, susceptibility of an individual to diseases and sensitivity to medical treatments. The challenges lie in the fact that much of the phenotype and environmental data vary widely in quality and utility. It will be essential to capture such information for a future health data infrastructure. Equally essential is the requirement to associate this new data with metadata that provide provenance. Genomics and other omics were covered in detail in the JASON 2013 report [23] and that discussion will not be repeated here. Rather, other emerging data types and analytics will be highlighted.

There is, however, one issue common to omics and many of these new data types that is worth considering at the onset. For these data, the technology involved in making the measurement is often highly advanced and the accuracy and reliability of the measurement can be very high. However, the interpretation of the resulting data is, in most cases, still in its infancy.

Genome sequence analysis is a good example of this dichotomy. Genotype can be determined at a limited number of variable loci in the genome, by analysis of single-nucleotide polymorphisms (SNPs), across the protein-coding regions of the genome, by exome sequencing, or of the entire genome. Each of the technologies required for these determinations is mature (or at least “adolescent” in the case of full genome sequencing), and can generate highly accurate and reproducible data. However, there is little high-confidence, actionable information that can be gleaned from such data because our understanding of the genotype – phenotype relationship is so poor. This is largely a basic research question, but it is one that is fueled by the existing human genetic variation, and that will benefit strongly from closing the loop between basic and clinical research as discussed earlier. JASON considers that most of the data types discussed below suffer from a similar



Figure 3-1: Personal health monitoring devices.

problem of limited interpretability with respect to health care, but that this will change rapidly for the better.

This issue lies at the heart of concerns over the value of this new type of data. Indeed, these concerns have led recently to action by the FDA to enjoin the 23andMe genetic information service [1] from reporting on possible connections between genetic information and disease. We discuss this in more detail and offer some additional recommendations in Section 5.4. Nevertheless, JASON feels it is important to have these new data sources available as part of a learning health system. Omics data, for example, are not static. As more is learned about the genotype-phenotype relationship can be assimilated into a dynamic and evolving health IT system with future benefits for health care and health.

3.2 Personal Health Monitoring Devices

A growing number of companies now produce personal monitoring devices to measure various aspects of health and wellness ranging from step counters to blood oxygen saturation sensors to sleep monitors. Examples are shown in

Figure 3-1. These devices can allow continuous measurement and autonomous communication of the data, possibly directly to an individual's PHR.

Whereas these devices may have a promising future for informing health assessment and treatments, there are several practices that inhibit or limit such use. Data from these devices are not yet adequate to draw detailed comparisons between individuals or against a population, except for a limited number of specific activities, such as sleeping. The devices are able to provide coarse-grained activity monitoring, such as assessing whether an individual is sedentary, sleeping, or active, and are able to make relative comparisons of a single individual at different times.

There is insufficient openness of data formats and algorithms for these devices, preventing interoperability and innovation in synthesis of individual health data. Although many of today's activity monitors include some open protocols, the data are usually locked in data structures that make it difficult for individuals to directly use the data. For example, service agreements have significant restrictions on how individuals may use what is in fact their own health data. While it is understandable that vendors wish to protect their investments and limit their liability, restrictions may stifle long-term opportunities for the application of these devices to improve individual health. Such industry practices make it clear that service providers consider themselves the owners of the health data, and that individuals are merely given limited access (that can be revoked) to interpretations of their health data.

While standards such as the IEEE Personal Health Data Standards (ISO/IEEE 11073) do exist, the accuracy of the devices appears to be based on mostly proprietary algorithms and calibration processes. As a result, de-

vices from different vendors measuring the same health or fitness activity will provide significantly different and thus incomparable data, e.g., numbers for steps, distance, and calorie counts. In fact, even the same device used in a slightly different way (e.g. attached to one’s hip as opposed to one’s wrist) will produce different results. Whether this is a problem or not depends on the intended use of the device. If one is interested in relative improvement in number of steps or pace (for example as part of recuperation from an orthopedic procedure), then accuracy is not really an issue. But if the intended use does require some accuracy (for example calories expended) then these issues will need to be addressed.

There is an opportunity for personal health monitoring devices to play a larger role in future health, wellness, and health care. To truly enable patients to improve their health and wellness with better knowledge from such devices, the industry should establish meaningful statements of uncertainty for both fitness measurement and fitness calculations so that data are comparable and interpretable. Metrics and standards should be independently reproducible from raw sensor data. Any device falling short of this end-to-end requirement lacks a fully scientific basis as reproducibility and verification are central to science, health, and the practice of medicine.

3.3 Moving Diagnostics to the Point of Care

An important trend for monitoring health is marked by recent advances in technology that allow performance of sophisticated medical diagnostics at the point of care, as opposed to a hospital or diagnostic center. New instruments are being developed that exploit advances in microfluidics, photonics, and microelectronics. They are typically smaller, lighter and less expensive than

conventional instruments for health care, and they can be used to monitor the condition of people before they seek care, as well as for bedside testing of patients seeking treatment. Miniaturization often reduces the size of the sample of blood or bodily fluids required to carry out a test, and the fact that the instrument lies next to the patient often permits test results to arrive more quickly than they would from a hospital laboratory.

As for personal monitoring devices, measurements can be taken at frequent intervals and the results can be autonomously communicated. As distinct from the situation described above with personal monitoring devices, the data quality and potential for use coming from these devices is being carefully scientifically justified. Since these devices are designed to be used in medical diagnosis, FDA certification is typically required. In this section two examples of these new approaches to point-of-care medical testing will be highlighted.

3.3.1 Testing for biomarkers

JASON was briefed by Dr. Dan Wattendorf, currently attached to DARPA, on the promise of a new generation of diagnostics that combine sensitivity to a wide variety of biomarkers that are related to the expression of various proteins. Such markers might be expressed, for example, prior to or at the onset of disease. Traditionally, testing for a limited set of such markers has to be performed in a diagnostic laboratory setting and the assays are limited in the number of proteins that could be detected. Technologies have been developed to collect tissue samples with minimal pain to the patient and the



Figure 3-2: Right: a benchtop NMR system. Left: an equivalent hand held unit.

samples do not need to be obtained in a clinical setting. This allows the performance of “diagnostics on demand” [38].

Such diagnostics will be performed using a forthcoming suite of “in-vivo” nanosensors currently under development. These can be ultra-small scaffolds inserted directly into the body. An example is the use of fluorescent nanospheres that are functionalized to detect biomarkers of interest. The sensor output is read either directly from a thin patch or via a hand-held wand. Such future developments will facilitate health monitoring on a much larger scale with far more frequent updating of medical information.

3.3.2 Miniaturizing the diagnostics laboratory

The miniaturization trend outlined above is also relevant to the instrumentation used to measure diagnostic data. A conventional Bruker Minispec NMR relaxometer system is shown on the right in Figure 3-2. It is a fairly large benchtop unit that must be operated in a hospital clinic, due to its size and weight. By developing a custom integrated circuit [24], it is possible to place all of the required electronics onto a single chip, and to reduce the size of

the NMR relaxometer to a hand-held unit, shown on the right of the Figure. Because a permanent magnet's field stays constant as its size is reduced, an ice-cube sized magnet produces sufficient field. The hand-held unit operates with smaller blood samples, but shows a factor $150\times$ higher spin-mass sensitivity compared with the benchtop unit [24].

An additional aspect of this miniaturization trend is that the new devices are able to upload their information wirelessly. Indeed, the example above shows a sophisticated measuring device, but the ability to transmit diagnostic information is now a common feature of many diagnostic instruments. It can be anticipated that the level of detail and frequency of transmission will both increase over time as these instruments see greater use. The potential benefit is a more responsive learning health system that facilitates the closed loop interactions discussed earlier. But it will also be necessary to be able to assimilate this data in such a way so that it can be meaningfully exchanged.

3.4 Patient Reported Outcomes and Collaborative Networks

Collaborative networks are “patient powered” networks in the sense that individuals share their data and experiences about their health and wellness. For example, the Collaborative Chronic Care Network project (C3N) [32] provides a web-based platform for patients or caregivers, clinicians, and researchers to share information about their disease(s). Participants interact and provide their own patient-reported data through a collection of apps.

The integrated web pages provide a wealth of information, from the synthesis of patient-provided data and experiences to links to new research and information.

Another patient-powered network is PatientsLikeMe [27]. This for-profit network encourages participants to “donate their data” for the common good. There is a lengthy agreement and privacy policy that explains that the data provided will be shared with both members of the network and others not in the network. It is not disease-specific, rather it encourages everyone to join and share their health experiences.

The data from these networks could be gathered and mined to learn about health and wellness. Some of this is being done within the specific sites. Standards for how such data should be, or will be, used in the delivery of health care are still evolving.

3.5 Social Media Data

The abundance of data available through the internet and social media is already having an impact on health and wellness. This creates the potential for improved health awareness. For example, HealthMap [14] reports disease outbreaks through monitoring a large collection of information sources, ranging from the World Health Organization to crowd-sourced surveillance tools and apps. Such a report for the San Diego area is shown in Figure 3-3. For comparison, HealthMap reports nearly 10,000 food-borne illnesses in the US over the course of a year, an important adjunct to the data reported by CDC which typically does not report as many incidents. As we see in many areas of data driven discovery based on “big data” there is also cause for concern.



Figure 3-3: Healthmap data for the San Diego area.

For example, the comparison of HealthMap reports to CDC reports above is given without regard to the scientific validity and appropriateness of the data and the methods used to combine the information.

3.6 Data Analytics

The ability to access and integrate data associated with health through informatics and data science will accelerate learning in a Learning Health System. The integration needs to be able to draw on information from the medical system, as well as incorporating broader societal and environmental inputs. There are natural roles for EHRs and PHRs, but this also points to a level of data access, integration, and scalability that must go beyond the interoperability of EHR systems. To motivate the value of gaining such access this Section will highlight a few examples that demonstrate the learning processes. In each example there is a close partnership between data scientists and health care researchers and professionals.

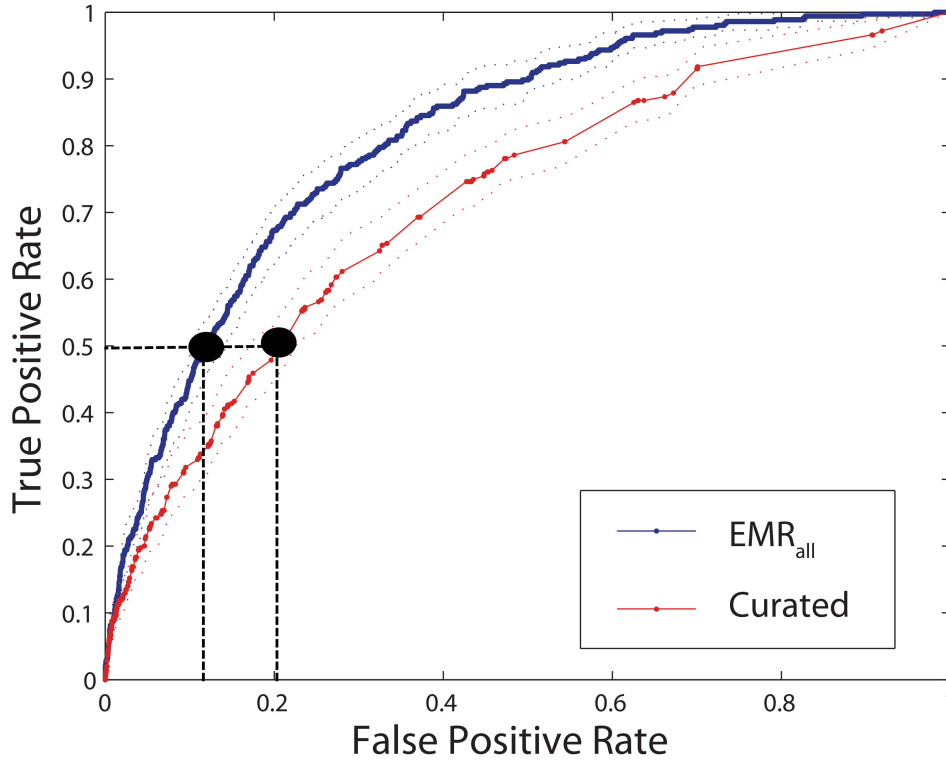


Figure 3-4: A ROC curve developed for the assessment of risk of bacterial infection with *C. diff.* via the use of machine learning on patient EHRs (labeled EMR in the figure).

3.6.1 Risk stratification example

In this section we discuss the use of large scale data assimilation and machine learning as a way of building data-driven models useful for predicting bad healthcare outcomes. It is estimated that, for example, medical errors or in hospitals account for 98,000 deaths a year [5]. Broader access and integration of EHR data within a medical facility, or across the facilities associated with medical providers, could lead to better awareness about the causes of these errors and improved patient safety and patient treatment.

Dr. Jenna Wiens briefed JASON on the development of a hospital-specific risk stratification procedure for estimating the probability a patient will test positive for *Clostridium difficile* (*C. diff.*)[39]. *C. diff.* is a bacterium responsible for infectious diarrhea and commonly occurs in hospital settings. It is typically resistant to many antibiotics and for some patients can be fatal, making it a serious risk factor for hospitalized patients.

Wien's method is based on supervised learning models, with the goal of predicting the risk for infection within the first 24 hours after patient admission. Traditional approaches are based on a small number of known clinical risk factors, such as age, admission source, recent hospitalization, and previous CDI. The new approach leverages all of the structured patient data within the hospitals EHR system, including lab results, medications, and procedures.

The specific application compared an expert-driven curated risk model with 14 features to a data-driven model with 10,859 variables from the structured contents of the EHRs. There was a significant gain in predictive power, as measured by the area under the receiver operating curves. The method could be easily automated. Some of the results are shown in Figure 3-4. The Figure compares two ROC curves showing the percentage of correct inferences of elevated risk for *C. diff.* from the method of assimilating medical records vs. the use of traditional curated risk factors. It is seen that the machine learning model has a higher true positive rate and lower false positive rate than the traditional approach. One open issue is that while the new approach method is more accurate at predicting a patients risk for *C. diff.*, it does not provide an easy way to learn new causative factors. This particular modeling approach is not designed to identify causative factors.

Nevertheless, used properly it can save lives by identifying high risk patients with greater accuracy.

The data-driven approach captures more than clinically relevant impacts, such as work and information flows, and so is most appropriately used in the larger overall system of health care delivery to optimize its use. In addition, the transferability of such models across hospitals and facilities must account for institutional differences in underlying populations served and in practices and workflows. The current lack of interoperability between EHR systems makes it a very slow process to obtain and assimilate the needed data.

3.6.2 Clinical decision support system example

The previous example highlighted the possibility of building predicative analytics to support the delivery of health care. Amarasingham et al. (2010) [3] take this one step further and demonstrate the **use of a systems engineering approach to reduce the probability of readmission or death for patients with cardiac care.** The process is depicted in Figure 3-5. The model is used to prioritize which patients gain rapid access to cardiologists post-discharge. The analytics system leverages data from the EHR system, as well as additional information collected regarding social determinants, such as income, occupation, address changes, homelessness, language preference, transportation assistance, and social support. This first study resulted in significant reductions in re-admissions. This process has now becoming part of Parkland Intelligent e-Coordination Evaluation System (PIECES) [2].

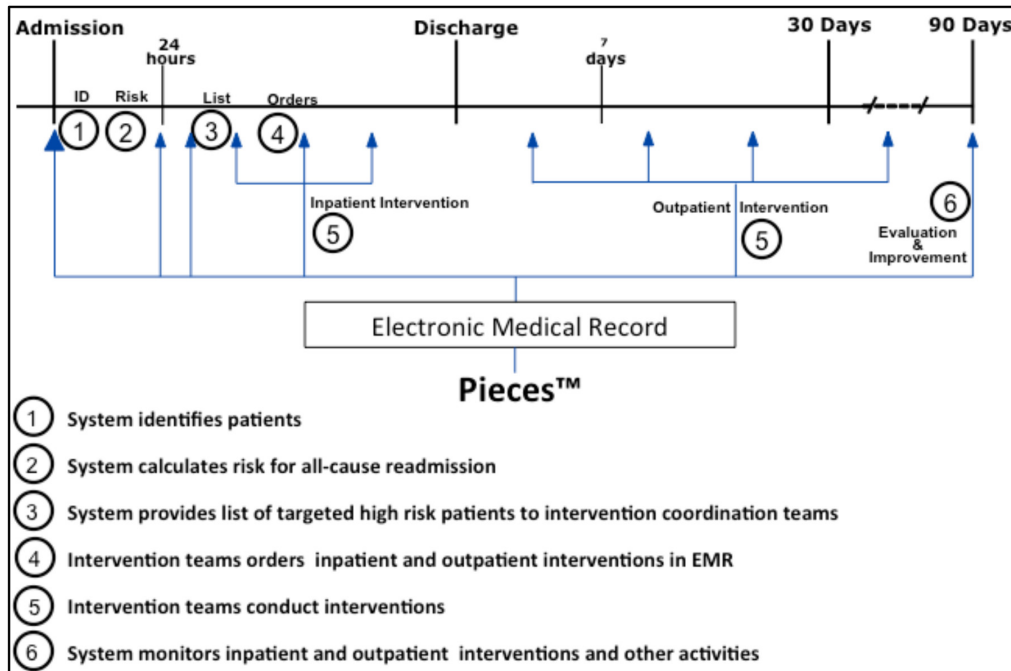


Figure 3-5: The PIECES system as used throughout the time-span of the interaction of a patient with cardiac disease. Evaluations of risk and prognosis are performed at early stages (admission) and also after discharge to minimize the possibility of readmission due to incomplete follow-up.

A recent PCAST report [33] focuses on the under-utilization of systems engineering approaches in health care. The example here demonstrates the power and effectiveness of such approaches in the development of clinical support systems. However, as PCAST points out, this is not enough. The full “system” must connect the medical system to broader societal inputs. The PIECES vision is to expand into a system that has strong connections into other community services, thus providing more guidance to the health care team. This example clearly points out again the need for data access, integration, and scalability that goes well beyond the interoperability of EHR systems. But broad-scale interoperability is an important first step.

3.6.3 Natural language processing example

The advent of modern machine learning approaches has made it possible to develop natural language processing as a powerful tool to ingest and organize the largely text-based information sources associated with patient records. In the past, most natural language processing systems attempted to deal with natural language through the use of complex conceptual ontologies. The idea was to encode language constructs into computer-usable form. This leads to large sets of rules which are then applied to attempt to understand and also to communicate in a given language. This proved to be quite unwieldy and the results were often unsatisfactory.

More recently, statistical machine learning approaches have been applied with favorable results. Such models of language make decisions based on soft probabilistic analyses that attach weights to the various features of the input. Such systems tend to be far more robust in the face of new or unfamiliar input and can often translate through errors. There are several advantages to these approaches that make them very suitable for taking large volumes of medical information including, for example, the inscrutable (to the layman) notes made by doctors as they take patient histories. The use of machine learning allows for focus on the most commonly used cases in medical text. Such systems can be tailored even to specialties where the vocabularies are more restricted and can capture the dominant scenarios more quickly. Because these approaches use statistical methods they are much more tolerant of errors. We will show an example of this below. Finally, as more data are included in the training set, the system is able to perform more accurate inference.

For example, we were briefed by Dr. Ruben Amarasingham of UT Southwestern Medical System on the PIECES [2] suite of medical analytics tools. One component is the use of natural language processing to identify risk factors for patients. The following example of a note written in a chart was provided:

“55 yo m c h/o dm, cri. Now with adib rvr, chfexac, and rle
cellulitis going to 10W, tele”

This is written in the commonly used shorthand used by medical professionals in writing patient notes. A translation using natural language processing by the PIECES system is:

“55 year old male Caucasian with a history of diabetes mellitus and chronic renal insufficiency now with atrial fibrillation, rapid ventricular rate, congestive heart failure exacerbation and right leg cellulitis going to 10W telemetry unit.”

There are several notable features about the translation. First, and most obviously, it can translate quickly the shorthand into text that can be read and then further scanned by others. Second, the doctor who either wrote or dictated the note mistyped the common abbreviation for atrial fibrillation (which is afib) as adib (d and f are next to each other on standard keyboards). But because the system is based on statistical modeling of the language the correct meaning could be inferred from the context. Third, the last part of the note refers to clinical workflow and not to any medical condition. For this particular hospital 10W refers to a location on the 10th floor and tele indicates that telemetry will be required for the patient. The system is also

Table 3-6: An example of the use of Natural Language Processing to identify workflow from doctor’s notes.

Diagnosis present	Acute	Body part	Location	Severe
Heart failure	Yes			
Atrial fibrillation	Yes			Yes
Cellulitis	Yes	Leg	Right	
Diabetes				
Renal insufficiency				

then able to categorize the concepts so that they can be recorded in a more systematic way in the hospital’s information system. For this patient the following (now machine readable and parsable) description was generated as shown in Table 3-6.

3.6.4 Health and environment example

Access to EHR and PHR data provides opportunity to directly link other information, such as air quality, to health events. Dr. Kathy Ensor briefed JASON on a study that integrates integrate point-in-time and point-in-space (location) air quality data with emergency medical services (EMS) data to identify a pathophysiological link between air pollution and acute health events: out-of-hospital cardiac arrest (OHCA) [10]. The integration of the multiple data sources is shown in Figure 3-7.

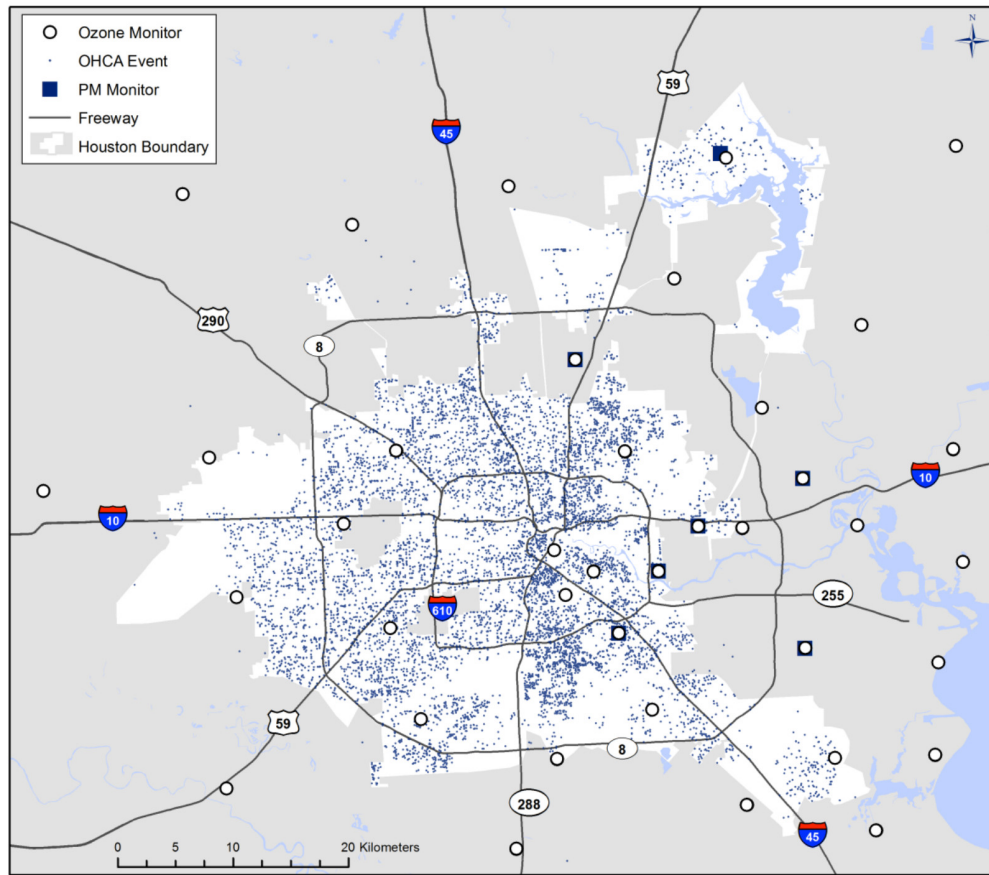


Figure 3-7: Environmental sensors for detection of air pollution in the Houston area and locations of out-of-hospital cardiac arrest during the period 2004-2011.

Approximately 270,000 deaths a year in the US are attributed to OHCA, representing 90% of the 300,000 individuals that experience an OHCA. Historically, air pollution has not been directly linked due to the aggregation of air pollution exposure over 12 or 24 hours. This study was able to disaggregate the air quality data and align pollution levels at 15 minute increments (point-in-time) to the locations in Houston (point-in-space) of the OHCA occurrences over an eight year period, 2004-2011 and 11,754 cases. The key statistically significant finding is that for each 20 parts per billion by volume

(ppbv) increase in ozone over the previous 1-3 hours leading up to the OHCA event is associated with a 4.4% increased risk of cardiac arrest. The study also found differences associated with various social determinants, such as race and gender. Finally, the study demonstrated that air quality differed geographically across Houston. This has led to new practices by Houston EMS, such as deploying extra resources in areas across Houston expected to have higher ozone levels. This work is now being extended to the study of asthma and the interplay of multiple pollutants [37].

This example shows how community health can be impacted by local environmental conditions and how community services can take advantage of such information. The success of these studies was due to the collaboration between the statisticians, physicians, and EMS and access to the relevant health and environmental data. A robust health data infrastructure able to ingest pertinent point-in-time and point-in-space data, such as air quality, noise, water quality, etc., will play a principal role in understanding health and wellness at the community level.

3.7 Findings and Recommendations for Data and Analytics

Throughout this Section the need for a robust data infrastructure with the ability to ingest “all” the data, keep it safe, understand it to build useful health informatics, and to communicate the knowledge gained from the data has been highlighted. Data opportunities are emerging faster than existing infrastructure(s) and current state of interoperability can handle. This leads to the following finding and recommendation.

Finding: There is an explosion of data from many and varied sources. Yet, there is little understanding of how to parse, analyze, evaluate, merge, and present these data for individuals and for the health care team. The health data infrastructure currently does not have the capability to make the data accessible in usable form, including the associated metadata and provenance.

Recommendation: HHS should adopt standards and incentives to allow sharing of health data. HHS policies should require that metadata and provenance be associated with all data so that data quality and use can be evaluated.

4 INTEROPERABILITY

Interoperability remains a challenge for achieving a robust data infrastructure that would enable the Learning Health System. Some progress has been made since the JASON 2013 [23] report on health data interchange standards and also on user interfaces. This progress will be highlighted here. In contrast, relatively little progress has been made on the creation of application programming interfaces (APIs) for commercial EHR systems. As discussed in Section 4.1, this lack of progress serves to significantly impede reaching an acceptable level of interoperability to support the integration of EHRs and PHRs for the Learning Health System.

4.1 Opening the EHR System APIs

In its 2013 report [23], JASON described a health information technology ecosystem in which a rich set of applications, including views designed for both individuals (patients) and clinicians, could be built by entrepreneurial third parties on top of the platform of stored EHR data, and, in the context of this report, PHR data as well. JASON commented that this inevitable shift from a small number of proprietary systems to a software ecosystem with a diversity of products and applications interacting would increasingly make it possible to achieve an individual-centric focus. This, in turn, would enable increased partnering between health care providers and empowered individuals with increasing responsibility for their own wellness. Such engagement could help foster improved patient education, health maintenance, and treatment compliance. Physicians and other health care providers would become more discerning customers of a robust health data infrastructure, in-

stead of being trapped in closed-box systems. Individuals 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 health IT marketplace.

Revisiting this issue in this report, JASON observes that, a year later, there has been movement in the right direction. Future progress along the lines described seems more inevitable than ever. Market leaders, including Epic Systems, whose Epic EHR has previously been considered among the most closed systems, have announced some steps towards creating application programmer interfaces (APIs) that would allow applications developers, medical device manufacturers, and others to interoperate directly with their proprietary products. However, the initiatives of which JASON is aware are far from complete solutions. For example, they may support the export of data only as complete documents (akin to exporting whole EHRs); or they may not expose all of the patient's data; or have sufficient versatility for consuming incoming data from innovative new sources. A particularly troubling issue, in JASON's view, arises when APIs that are open in a technical sense are accompanied by contractual limitations that discourage their entrepreneurial use. For example, there may be restrictions allowing their use only internally by existing customers. As a general rule, any API exposed to a customer should also be exposed to the entrepreneurial space. While there is no way to enforce such a requirement, JASON believes it is in the long term interests of EHR software vendors to adopt such an approach. As discussed further below, there is a trend away from monolithic software architectures and so those vendors adopting a more open approach for their core software may improve their market positions by encouraging entrepreneurial software development in a way similar to the approach being adopted by

Apple through its recently announced HealthKit platform [20].

Information technology companies create public APIs that expose proprietary interfaces and data structures when it is in their own economic interests to do so. It is worth thinking about how this has happened in other industries, and what steps the federal government might take to incentivize incumbent EHR vendors to make such decisions in the space of health care.

APIs are often exposed in the form of standards. JASON applauds progress that has been made in health information technology standards in the last year, such as FHIR [16] which is discussed in further detail below. However, progress seems to be most rapid in the area of taxonomic standards or whole document exchange standards, with much less visible progress towards standards for APIs. JASON continues to believe that taxonomies and document exchange standards alone are not sufficient to enable an entrepreneurial ecosystem of EHR products to develop as rapidly as it otherwise might. Standard and/or open APIs are a critical need.

In other sectors of IT, APIs move from proprietary to open when a company perceives the advantage of maintaining its centrality, rather than see a migration away from its core products that, over time, would put it into a less competitive position. As a strategy, this is distinct from, and complementary to, strategic positioning for market share or overall market size. It is a kind of insurance against future disruptive changes that can drive companies with no strategy to address this issue into irrelevancy. The classic successful example of this strategy, though not specifically in the area of standards, is Apple's business model of combining closed hardware with open software APIs. Apple's operating systems function only on the hardware platforms that it manufactures with open programming APIs that encourage an entire

developer ecosystem (both “cottage” industries and large companies) that provide a large part of the Apple customers’ user experience.

More specific examples of companies that have successfully transitioned proprietary APIs into open standards, as part of a strategy to enhance their centrality in the market include:

- ATA – This ANSI Standard was originally developed by Western Digital. It is now widely used for storage connections, including compact flash.
- Ethernet 802.3 – This was developed at Xerox PARC, then commercialized by 3COM. Only after it became a standard did it displace proprietary token ring and token bus protocols.
- FAT file system, ISO 9293:1994 – This was originally developed by Microsoft. It is now a widely used standard disk format. This is a mixed example, because Microsoft has patents on some small aspects of the standard, and extracts a small royalty from firms that use it.

In health IT, it seems possible that an incumbent vendor with significant market share, or a joint venture among smaller vendors, could propose as a standard a set of detailed specifications for an API. The federal government should encourage such a proposal, and it should be prepared to give support to its development in the consensus (i.e., non-governmental) open standards process. A declared policy of this type could encourage market leaders to consider the advantages of being the prime mover of such a standard, versus the disadvantages of impeding the process and of later having to play catch-up.

A more radical, but perhaps workable, proposal would be that the federal government recognize that health data in EHRs that support robust APIs are a more valuable public good than health data in siloed EHRs. It would make sense, then, to include a measure of this in CMS's Hospital Value-Based Purchasing Program, and/or as a part of any similar program of differential reimbursements to hospitals or providers. For example, EHRs with the most robust APIs might qualify the adopting hospitals for 0.25% additional reimbursements, while hospitals with the poorest APIs might lose 0.25%. Although this seems similar to the meaningful use strategy of rewarding for the adoption of EHRs, or EHRs supporting information exchange, the goal here is actually somewhat different: it is to recognize the long-term public good of facilitating the development of an entrepreneurial ecosystem of companies in the health IT sector, and that such an ecosystem will be, in the long run, a way of improving quality of care and reducing costs. Over time, as the ecosystem of applications develops, it may be possible to measure the added value of this kind of interoperability and to adjust the reimbursement differential accordingly.

This discussion leads the following finding and recommendation:

Finding: There is a critical need for open Application Programming Interfaces (APIs) that effectively support an entrepreneurial ecosystem.

Recommendation: HHS should adopt policies of support for proposed open API standards. These policies should make it advantageous for one or more leading EHR vendors to be the first to propose such standards. HHS should recognize ecosystem-friendly EHRs as a public good. HHS should adopt reimbursement differentials, initially small, for institutions that adopt ecosystem-friendly EHRs.

4.2 Recent Progress on Standards

A key layer in the proposed JASON health data architecture is the data layer which calls for the decomposition of data into atomic components that possess metadata and provenance information. It was argued that this would allow much better aggregation and semantic processing of the data so that the relevant pieces could be re-aggregated into forms that would be meaningful to various users of the data.

4.2.1 The clinical document architecture (CDA) standard

Until recently, the main proposed approach taken for interchange of clinical documents was the **HL7 Clinical Document Architecture (CDA)** [17] developed by the Health Level Seven (HL7) standards organization [19]. This is an XML-based markup standard that is intended to specify the encoding, structure and semantics of clinical documents that were to be exchanged by various medical service providers. As a document format, the standard possesses desirable characteristics. For example, one can infer the context of a CDA document. More importantly, using an XML aware web browser one can read the contents of a CDA document. The electronic format was developed for what is called the **Continuity of Care Document** [18] which is to be exchanged among health providers for a specific patient. Many EHR systems now support this document architecture.

In Figure 4-1, various levels of the CDA are shown with the most sophisticated usage labeled at CDA level 3. Any clinical document can be embedded inside the “body” and need only be identified by a header. It is true that there are specifications for headers and a plan for standardizing as-

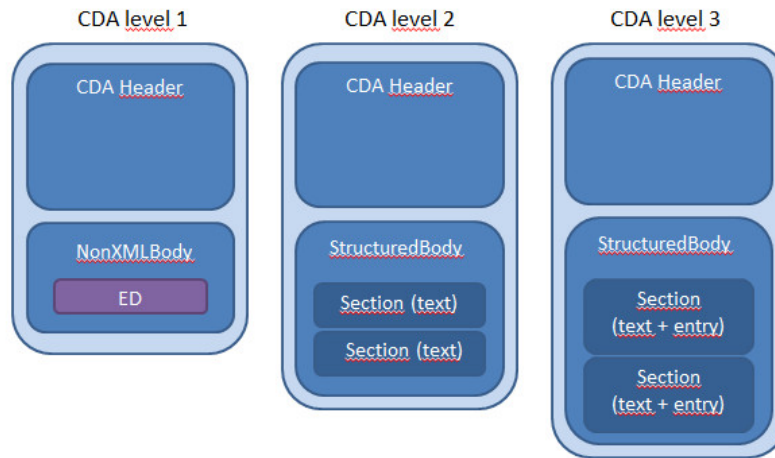


Figure 4-1: Graphical representation of the Clinical Document Architecture (CDA).

pects of the various body entries. But they are quite broad. For example, at level 1, any external document could be embedded with a descriptive header. This is essentially the analog of faxing a doctor's chart to another medical provider. At Level 2 of the CDA, there is a requirement to structure the document much like a doctor's note. There are a set of templates that are meant to document the patient's history and physical condition. These contain a subjective section to report the patient's description of their condition, a vital signs section, an assessment section and a plan section. Level 3 is meant to add additional indicators like billing codes or even further representations of findings, symptoms and diagnostic data. This is a good example of an open standard that can be further evolved.

The problem with this approach is that the CDA is really only a container for the information. In principle, the use of XML will allow disaggregation of the atomic data, but the parsing would be left to the particular

application and each provider of the information would have to publish details of their particular XML schema. Because it is in some sense such an open-ended standard, supporting it is made quite difficult.

4.2.2 The proposed FHIR standard

The recent introduction of FHIR[16] by HL7[19] is in JASON's view a significant improvement over CDA. FHIR attempts to standardize the exchange of information through a set of modular components that are called Resources. Resources have standardized names and provide basic pieces of information but can be extended to fulfill specialized requirements. Examples of Resource names are CarePlan, FamilyHistory, Medication, Patient, etc. Resources are grouped by functions. For example the Administrative Resource grouping contains essential information about the patient, the treating practitioner, workflow associated with that patient's care and so forth.

Each Resource contains a human readable summary, a set of entries for agreed-upon standard data like medical record number, name, gender, etc. and the ability to tie this data to local workflow. An example of a Patient Resource is shown in Figure 4-2.

If this Patient resource is requested it is then possible to parse the various fields simply because the standard is clear about the meanings of the various entries. At the very least, it is possible to extract a human readable summary which then reports in readable form the patient's name and medical record number so that if another EHR system is being used the patient can then be located within that system in an interoperable fashion.

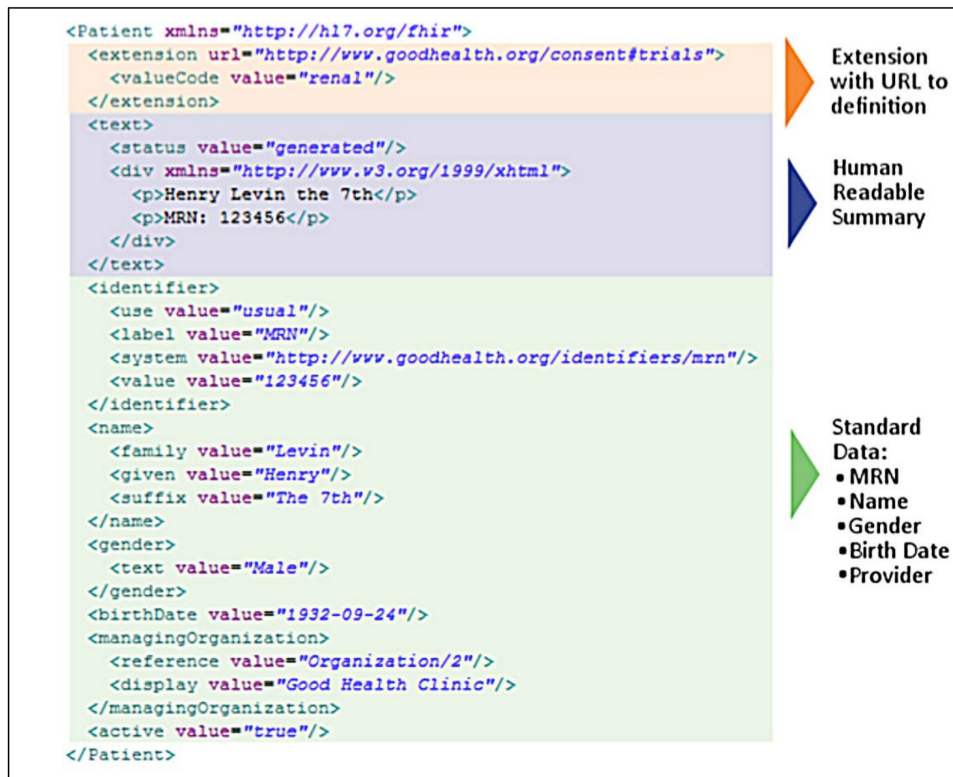


Figure 4-2: An example of a Patient Resource in FHIR.

The main focus of FHIR is to use the notion of web application interoperability standards so that the EHR (as well as PHR) development process is relatively rapid. This is accomplished first by making consistent use of web standards, such as those listed below:

XML: Extensible mark-up language standard that defines a set of rules for encoding documents in a format that is text based and so is human readable but can also be parsed by machine. There are a wide variety of tools to process XML based documents.

JSON: JavaScript Object Notation - a light weight data interchange format based on a subset of the JavaScript language. JSON uses only two data structures - name-value pairs and ordered lists. The objects encapsulate these structures, and because all languages support these basic structures, JSON is easy to parse.

HTTP: Hypertext Transfer Protocol HTTP is the communication protocol used in the World Wide Web. A key element is the use of uniform resource locators (URLs) to identify data elements. FHIR makes heavy use of this idea to standardize queries for FHIR Resources.

ATOM: a web standard for creating web feeds as well as or creating and updating web resources. ATOM is an alternative and updated version of the RSS (Rich Site Summary) used to develop information feeds that can then be syndicated so that subscribers can get updated versions of the posted information as it changes.

Oauth: an open standard for authorization. It provides secure delegated access to the resources of a web server on behalf of a resource owner. A third party can be authorized to access the information once properly authenticated and without sharing their credentials. This is done through the generation of access tokens.

Support of web standards, while desirable, is not the main reason FHIR could become a useful standard for HIT. Of greater importance is the support for representational state transfer (REST). REST is not a web standard but an approach to building web-based applications. It is an architectural style that consists of components and connectors between those components. “RESTful” applications are built using a specific set of design principles:

- The interaction is of the client-server type which implies that there is a separation of concerns about the storage of data and the interface to that data. The clients do not store the data; it remains on the server. Servers, on the other hand, have no responsibility for user interfaces, this is handled by the client.
- There is no client “context” ever stored on the server . The communication is essentially stateless between requests. When a client wants to change its state (e.g. update some information it needs) it sends a request to the server and is said to be in transition while all such requests are outstanding. In practice servers do try to create some temporary state by sending the queries to databases so they can be referenced later but there is no permanence guaranteed.
- For efficiency, a client can cache a request. If the server tells the client the information can be cached.
- A client cannot actually tell if it is talking to some specific server; the server can delegate client requests to other servers and the client cannot tell the difference. This makes it possible to avoid overloading a single server.
- Servers can send clients code to extend the client functionality. This is done by sending Java applets or JavaScript. This part of REST called “code on demand” is optional and may not be available for security reasons.
- Applications present a uniform interface. This is the most important aspect of a “RESTful” application. There are four guiding principles for a RESTful interface:

1. Resources are identified via a uniform approach usually using a Uniform Resource Identifier (URI). For example, FHIR does this via a URL. This simplifies things considerably as there are standard naming conventions and so locating a resource is performed through an HTTP request. The server need not be organized according to the web-based naming system. Internally there may be a database or other internal representation for the data but the server has a mechanism for translating the information into HTML, XML, JSON etc.
2. Clients can modify the resource on the server (assuming they have the appropriate permissions) through the web-based resource. The resource therefore is like a handle into the data. Even though the client does not use the server's native information formats it can still access, modify, or, if allowed, delete the data.
3. The messages interchanged between client and server are self-descriptive. For example various types of data (e.g. images, text etc.) have associated with them an internet media (MIME) type so the proper parser can handle the message
4. Transitions of the data are performed using hyperlinks. For example, HTTP GETs, PUTs etc. can be used to manipulate the data once the URI is in hand and appropriate authorization is provided.

An example of how the methods associated with HTTP can be used in a RESTful web service API is shown in Figure 4-3. **Note that REST isn't really a standard but it relies on well-accepted standards.** This makes it relatively easy to program the relevant applications.

Resource	GET	PUT	POST	DELETE
Collection URI	List the members in the collection	Replace the entire collection	Create a new entry in the collection	Delete the entire collection
Element in a collection URI	Retrieve the element (translate into appropriate MIMEtype)	Replace the element	Typically not used	Delete the member

Figure 4-3: Uniform command list for a RESTful architecture.

4.3 REST and FHIR as the Foundation of a Future Health IT System

We argue here that the introduction of a proposed standard like FHIR [16] can facilitate not only the interoperability of EHRs but also, with some additional work, can make it possible for other components of an overall health data infrastructure system to interoperate. In the 2013 JASON report on health IT [23], an idealized architecture was laid out that embodied many of the desiderata of an EHR system. But it was also noted in that report that it was possible to evolve to such an architecture in stages. For example, if it is agreed that the main objective of interoperability is to query health information systems via an external interface and then process the results via a set of user interface applications, then one can consider a simple architecture of the sort shown in Figure 4-4.

In this picture, diverse EHR systems allow for interrogation (as well as information input) via web interfaces. It is easy to see that FHIR can serve as the basis of such a system. Because the resources carry agreed-upon atomic data, it is possible for any EHR system to add a RESTful web interface and with proper authentication facilitate the interchange of information in a form

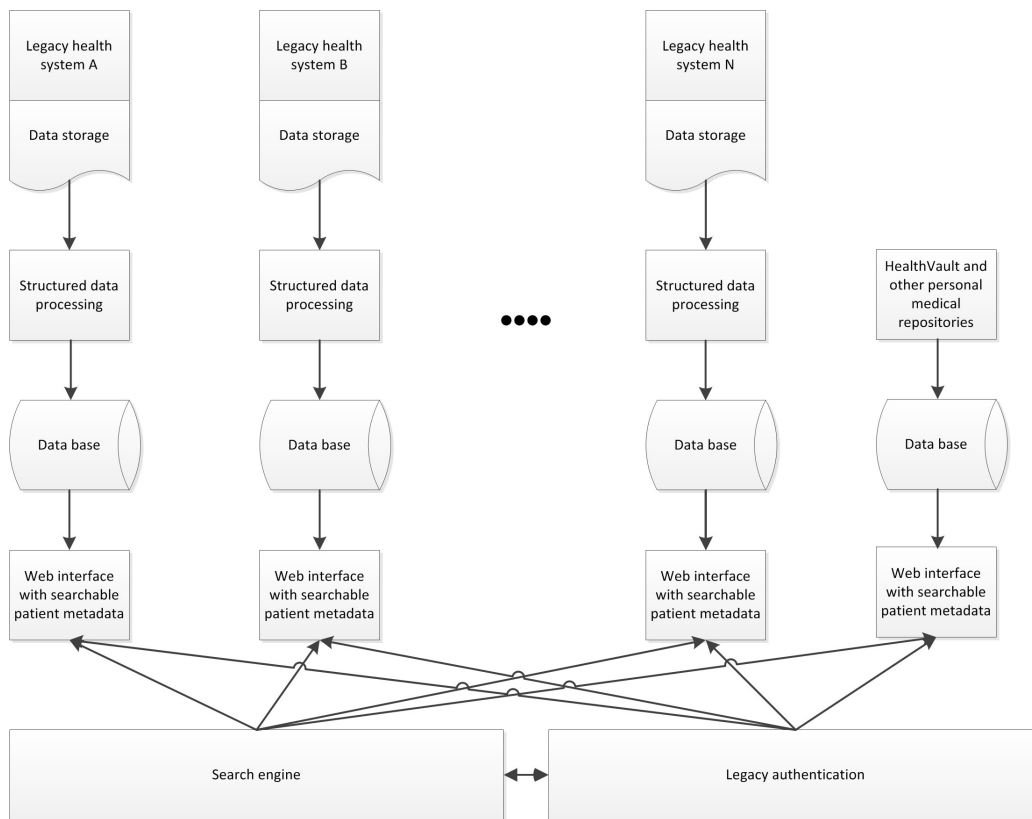


Figure 4-4: A proposed HIT system built on web interfaces.

that can then be processed for future use. Indeed, there is no reason this type of interoperability need be limited to EHR systems. With a FHIR query layer built using RESTful principles, any PHR system can also so interoperate. The addition of such layers to, for example, patient collaborative networks or medical social media systems, can make it possible to gather the external data required for a broader assessment of health.

Another appealing feature of this approach is that no health IT system needs to go “off line” in order to build the interface. All that is required is a web server front end and a database back-end to translate from the proprietary formats of commercial EHR systems to a form that can then be packaged as FHIR Resources. Objections have been raised that such a system

cannot cope with the complexities of what a modern EHR system must do, such as complex transaction processing. This is partly a result of the fact that EHR vendors encourage vertical integration of their systems for reasons of efficiency (and of course for economic reasons). But **the objective here is not to replace all the functionality of an EHR system which necessarily is a complex enterprise-wide system but to provide a much less time-sensitive layer to facilitate interchange of information** and support better informed decisions regarding health.

To be sure, the FHIR system is now only in a demonstration phase, but with a concerted effort to standardize it and perhaps extend some of the Resource types, the goal of interoperability can be largely met for EHRs as well as PHRs and other data sources.

4.4 User Interfaces

Once an appropriate approach to data interchange (such as FHIR [16] discussed above) is available it is possible then to utilize modern web application technologies to create “apps” that can display relevant health information in meaningful ways. Indeed, today, many cell phone users do not make use of just one application to access various information sources. There is typically a wide variety and each user can adopt (or reject) an application (or make use of multiple applications) depending on how they want the relevant information displayed.

In this picture, the user interface becomes a distinct component of the health data infrastructure and the health record is no longer a monolithic file stored in an EHR database system, but a collection of diverse information re-

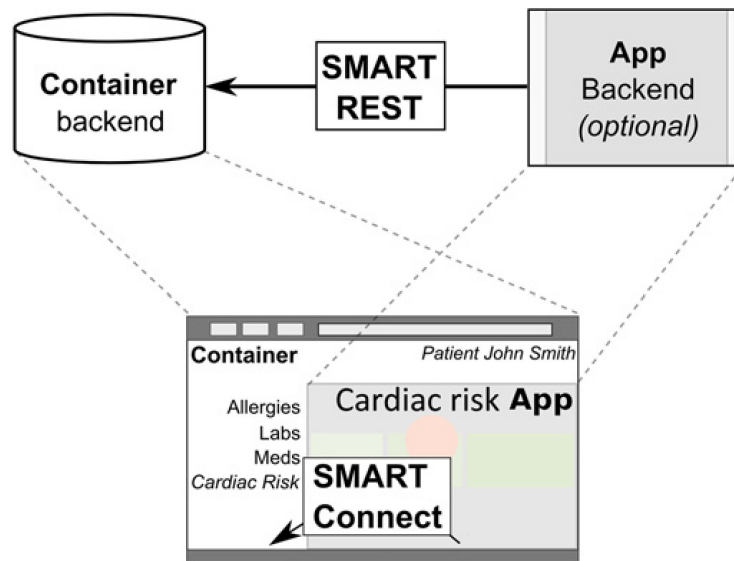


Figure 4-5: The use of RESTful approaches in SMART Applications [25].

sources. A health care provider would want certain pieces of information (e.g. patient lipid levels etc.) displayed prominently, but may not be interested in other pieces of information. The same information would be displayed in a different way for a consumer as part of their PHR. Alternatively, a social welfare agency may want to aggregate other pieces of information (perhaps suitably anonymized) as part of its efforts to assess community health.

An example of this approach is provided by the Smart Medical Applications Reusable Technologies (SMART) Platforms project [25]. SMART takes the point of view that the data container need not be connected to the user interface via the health IT system. The overall philosophy is illustrated in Figure 4-5.

Because the application is now distinct from the EHR, it is possible to make it portable through the use of the new syntactic features introduced in HTML5 which allow for standardized embedding of multimedia components.

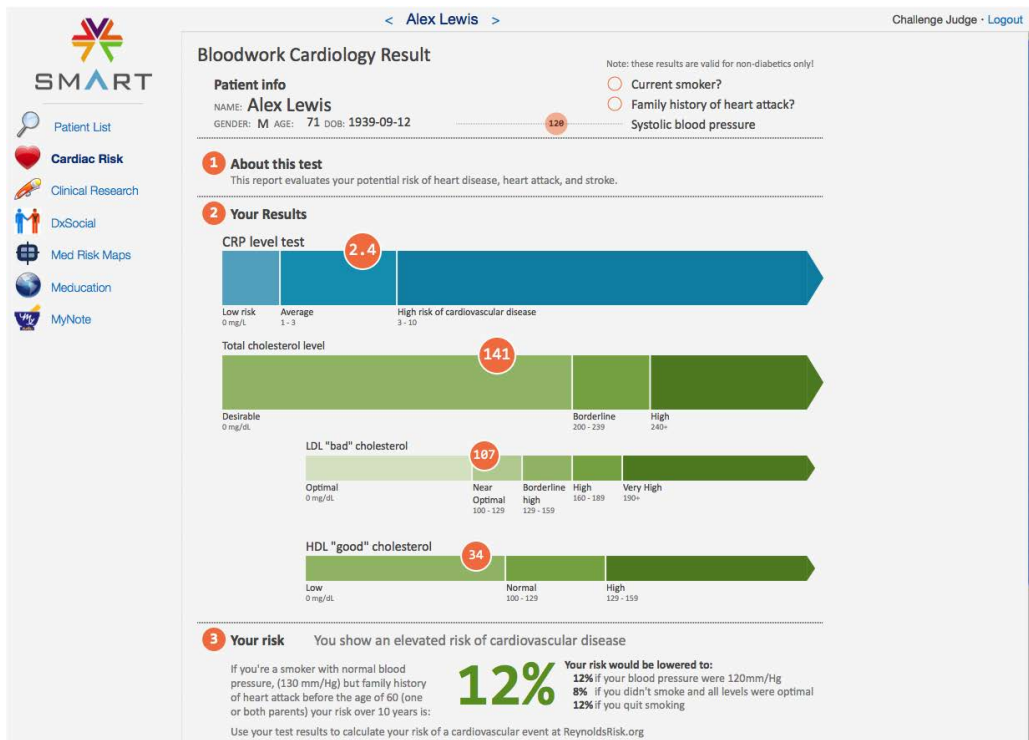


Figure 4-6: SMART Cardiac Risk app running standalone [25].

This allows one to view the output of the application without having to resort to the use of proprietary plugins or APIs. As part of the SMART project, its developers designed the applications so that they could be directly utilized in a number of open frameworks designed for integration of medical and biological data.

An example of this is shown in Figures 4-6, 4-7, and 4-8. These Figures show the same SMART Cardiac Risk application running in three environments: a simple SMART application reference container, the Indivo Personally Controlled Health Record (PCHR) framework developed at Boston Children's Hospital and the Informatics for Integrating Biology and the Bedside (i2b2) analytic framework. In each case, the same cardiac risk application which interrogates an EHR using a RESTful set of queries, is embedded in

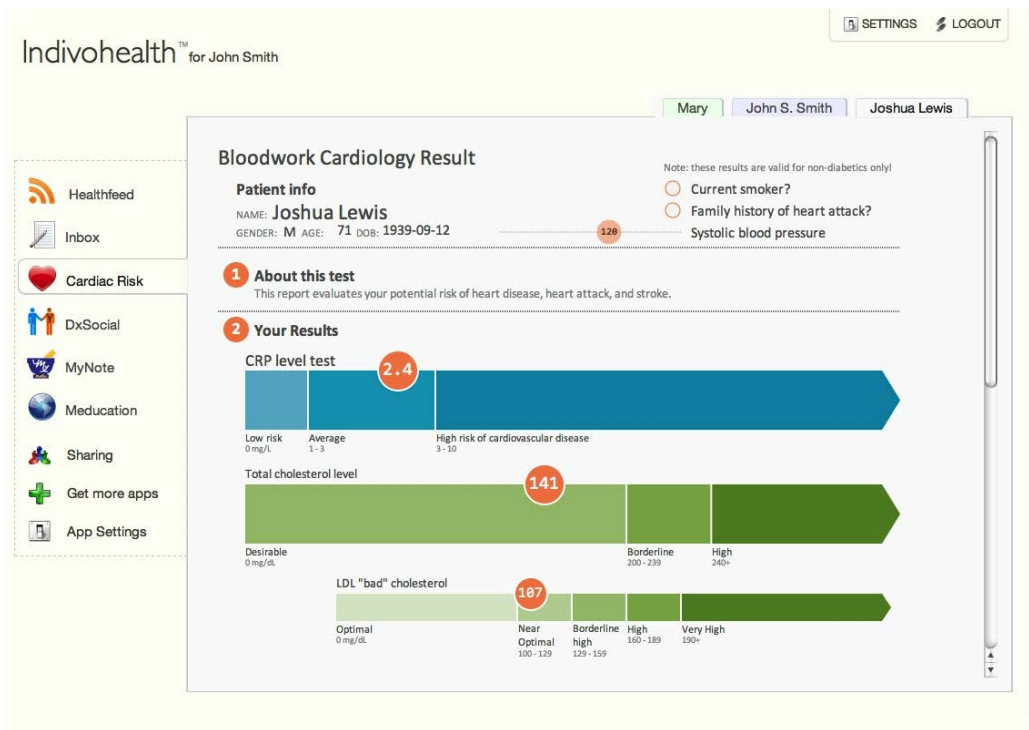


Figure 4-7: SMART Cardiac Risk app running as part of the Indivo framework [25].

a particular framework for analysis of medical data. The original SMART applications used a query approach designed by the SMART developers, but with the advent of FHIR, it was a simple matter to convert the applications so as to make use of the proposed FHIR standards. The new version is called SMART on FHIR.

The overall architecture of SMART is shown in Figure 4-9. The container of the SMART data is an EHR, PHR or other information resource which communicates using FHIR. Thus, a user never needs to directly interact with the native data store of the EHR or, for that matter, does not have to limit themselves to one data store. The user interface is created by assembling a set of component SMART applications from a gallery of such “apps”. This makes it possible to design custom interfaces depending on the

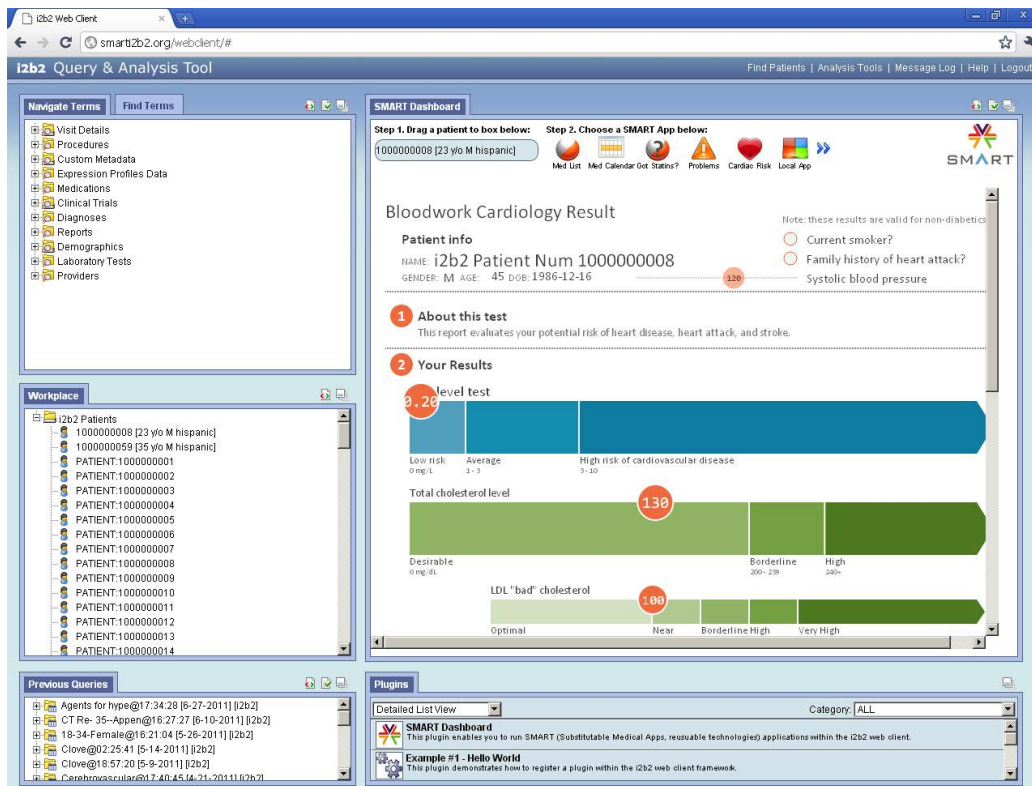


Figure 4-8: SMART Cardiac Risk app running as part of the i2b2 framework [25].

requirements of the individual requiring the information. Doctors will want to assemble the information differently from the consumer. In addition, because of the high level of potential customization, the view of the data may differ from clinician to clinician or from hospital to hospital depending on the components of greatest importance or the relevant workflow of the clinic.

An example of this is feature is shown in Figure 4-10 where two SMART apps are used simultaneously. The first provides information on diabetes risk. The second is a query tool for the personal genetic information web site 23andme [1]. By bringing together diverse information resources clinicians can better inform patients of their health status. This is an example in which generic information from a personal health resource is merged with EHR

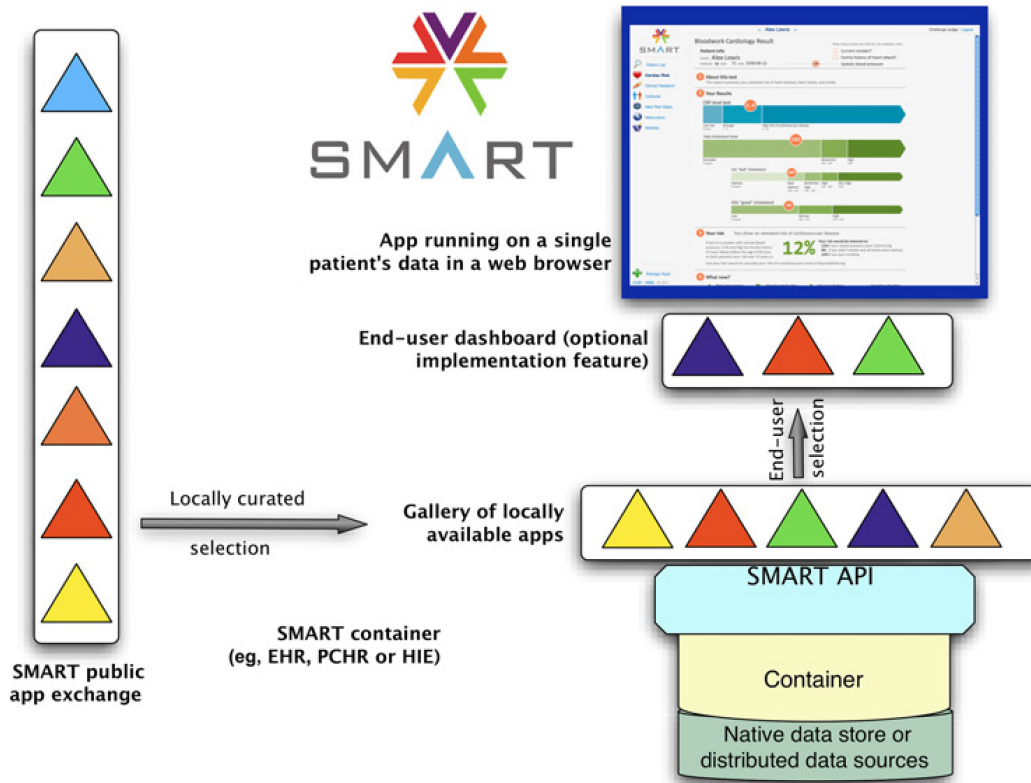


Figure 4-9: SMART application architecture.

information. We discuss further the role of such personal health information services in Section 5.4.

Perhaps most notably, using API's published by the Cerner Corporation for their Millenium EHR system, it has been possible to implement the SMART API on top of the existing EHR system at Boston Children's Hospital [25]. Microsoft has also produced a SMART-enabled version of the HealthVault PHR [25]. The picture that emerges is very much in keeping with the concepts outlined in the 2013 JASON architecture [23]. These concepts are directly extended to the concepts outlined in this report on data infrastructures for health.

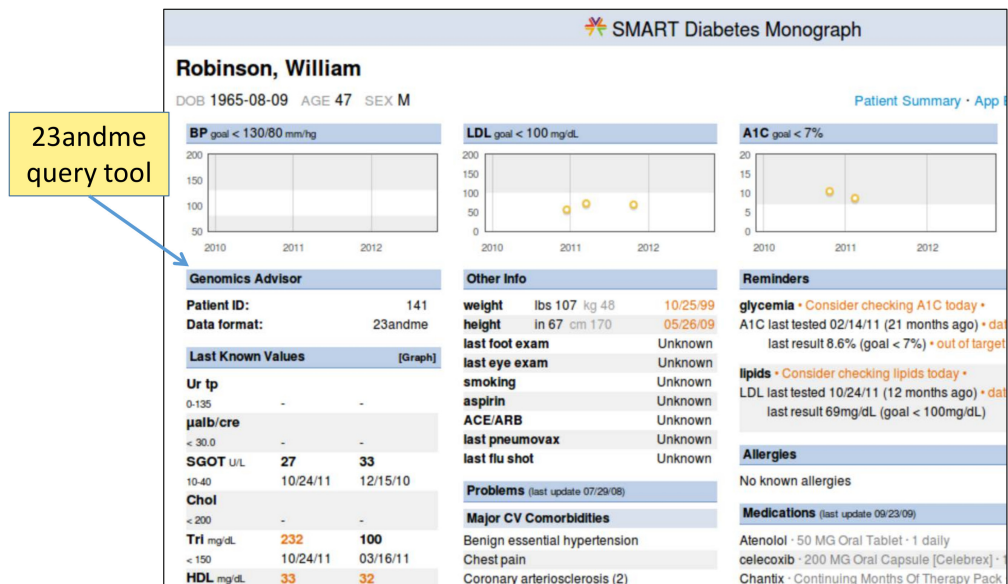


Figure 4-10: Combining EHR data with 23and me genetics data using the SMART platform [25].

5 OPPORTUNITIES AND CHALLENGES

The US Department of Health and Human Services (HHS) has released a 10 year agenda for realizing an interoperable health information technology infrastructure able to support a Learning Health System [35]. This calls for:

- Fully interoperable EHR systems in three years,
- The emergence of a Learning Healthcare System that will realize improvement in health care quality and lower costs in six years,
- The national health IT infrastructure supporting the Learning Health System in 10 years.

This plan is achievable provided the recommendations from the JASON 2013 report [23] are adopted, many of which have been reinforced in this report. This chapter discusses some additional strategies to help ensure this 10 year vision stays on track. These include accelerating the role of community engagement; leveraging existing organizations focused on chronic conditions, including aging; recognizing and enhancing the expanding health care team; and revisiting the role of regulation for health and wellness products and services.

5.1 Communities

One important source of good health for an individual is the support provided by a caring and public-spirited community. This could be a community defined by a specific geographical location, or defined within a region based on

socio-economic conditions, i.e., vulnerable populations. A healthy community can be defined by the quality of its education, housing, and environment; its availability of jobs and income; its access to and quality of health care; and its adoption of healthy behaviors, including diet and exercise. Indices, such as those developed in the joint Robert Wood Johnson Foundation [12] and University of Wisconsin County Health Rankings and Roadmaps Program [36], exist as tools to compare health and wellness across communities. These are based on state and national level surveys and are weight-adjusted to produce community-level estimates. The rankings provide a good start for measuring community health. However, more effort is needed to develop additional science-based indices of community health that leverage the entire health data infrastructure, and not just state and national data collections, of the Learning Health System. This would include integrating information from the medical system to societal support within and across communities.

PCAST (2014) [33] discusses the benefit of re-engineering health care delivery to bring together community partners with community health care highlighting several emerging examples that show some promise. PCAST points out that the assessments of community health needs being mandated by the Affordable Care Act, or by other Federal programs, are not being leveraged to build capacity and momentum in these partnerships. In addition, the information technology infrastructure to support such partnering, such as the capture and sharing of information between community services and health care providers, is lacking.

Foundations and other private organizations have begun supporting pilot efforts for communities to define health and wellness goals and measure progress. These come in the forms of grand challenges and prizes and span

both geographic communities and sub-populations within these communities. The Robert Wood Johnson Foundation “Culture of Health Prizes” [12] program, initiated in 2013, is gaining traction with more than 250 communities competing for the prizes. Another example is The Way to Wellville, a privately funded contest [15]. In this contest, five communities with populations under 100,000 will be selected to compete over five years on the improvement of five community health metrics that are yet to be defined. Forty-two communities applied and the down selection is currently in-progress. The results of this experiment and of the lessons learned with the Culture of Health Prizes may help guide other communities to healthier futures, as well as serve as a basis for larger-scale programs.

This discussion leads to the following finding and recommendation:

Finding: Pilot projects are emerging that encourage intracommunity partnering to create healthier environments for citizens. However, metrics are lacking that provide a reliable index of community health.

Recommendation: HHS, in partnership with private foundations, should establish “race to the top” challenges for community involvement. These would seek to demonstrate, through the integration of community services and broader wellness services, a measurable increase in health and wellness, and a concomitant reduction in the number of encounters with the health care system.

5.2 Leveraging Non-Profit Organizations

Organizations focused on various chronic conditions, such as the American Heart Association, American Lung Association, or the Cystic Fibrosis Foun-

dation, or on the overall health of various population groups, such as the AARP, are very well informed within their mission spaces. They typically have registries of information and deep familiarity with ongoing research, treatment, and practice. Their expertise should be continuously leveraged to enable learning in the Learning Health System.

These organizations could, for example, advocate for the use of specific PHR data, facilitating access to, and integration of EHRs, PHRs, and other data for education, thus accelerating the advancement of cures and treatments. They could encourage the development of applications, and even endorse products and services related to their mission spaces.

Perhaps one of the most important contributions these non-profit organizations could make is in enabling the health data infrastructure to support learning by helping to make the data available for basic research, public health, and clinical practice. This means helping individuals manage their data by suggesting, maintaining and fine-tuning privacy requirements. The JASON 2013 report [23] introduced the concept of fine-grained permissions and patient privacy bundles. JASON defined a patient privacy bundle as “a predetermined set of default permission and inheritance settings for the atomic data elements.” It could be a daunting task for the individual to decide how to associate user permissions with each of their atomic data elements in their EHR or PHR. These non-profit, trusted, organizations could help by recommending (advocating for) sets of prepackaged privacy bundles.

This leads to the following finding and recommendation.

Finding: Non-profit organizations, for example those devoted to finding cures for particular diseases, have strong patient and community sup-

port and trust. These attributes could be leveraged to support health and wellness.

Recommendation: Relevant non-profits should be encouraged to assess their goals with respect to data streams, and to provide “stamps of approval” for “apps” and other consumer tools. This will speed adoption of these tools.

5.3 The Continued Evolution of the Health Care Team

There is a rapid evolution occurring in the composition of, and relationship among, the individuals that deliver health care and advise on health and wellness. Mitchell et al. [28] point out “health care has not always been recognized as a team sport”. However, the complexity of today’s health care; the rapidity of technological advances; the heterogeneity of patient (individual) types, necessary tasks and settings; have given rise to a team-based approach involving physicians, nurses, physician assistants, pharmacists, therapists, and technicians, as well as administrators, dietitians, social workers, fitness trainers, family members and others. There are, in fact, a wide variety of health care teams, and while such teams have been in operation for decades, research on health-based care and the optimal organization and effective assessment of these teams is still nascent.

Further demands on this still-evolving system are anticipated. It has become important to consider the changes implied by teams dedicated to health rather than just to health care. These are teams that more fully integrate the active participation of the individuals themselves, as well as their families, and teams that can best leverage and incorporate the multitude of direct and

inferred data and informatics made available directly through ever-growing digital and mobile sources. The health care team of the future must operate on a variety of time scales and also must be able to accommodate and ingest a continuum of information at short time intervals, rather than the episodic updates (e.g. physician’s appointments) that have been more typical of the earlier team-based health care.

There are many analogies in effective operation of the health care team of the future and the optimal performance of the robust health data infrastructure discussed in the JASON 2013 report [23]. Both the architecture and the health care teams are comprised of diverse functionalities with multiple interfaces that must be well integrated. Trust and security must be built into the system and team, and a clear communications protocol must be established and enforced. Communication does not only imply a protocol for information transfer across interfaces or between members of the health care team, but also assumes a means of translating information (data) that relate to a common series of health outcomes, but which are acquired, initially described or even stored in different formats.

Given the present-day increase in the number of sensors and information sources that may prove important in providing signatures for health, there will be **ever-greater challenges in storing, validating, processing and interpreting massive amounts of data.** Given the increasing number of relevant indicators of health that are available, and given that much of that data is expressed in numbers, it is **important to incorporate a high degree of “numeracy” into the health care team.** “Numeracy” denotes having appreciation for and understanding of the information presented in those numbers [9]. Numeracy is an important precursor to being able to make critical evaluations of

the data for decision-making. This points to the importance of augmenting the education of physicians and all health care team members so they will achieve greater proficiency in computer science, statistics and data analysis.

The advantage of a team-based approach to health is the ability for the team members to contribute special expertise, providing the highest quality outcomes. With the diversity of the evolving health care teams, it is likely team members do not know each other, and possibly do not even know of each other. This makes the data exchange even more important as it will provide the foundation for communication across the team, making the team a team. As new technologies and information sources are incorporated into integrated health record, it will be important for the team to have representation of expertise in those technologies and their interpretations. This opens opportunities for the inclusion of new important expertise and skills that could manifest itself in the form of a “care coordinator” or “data coach.”

The following finding and recommendation provides some guidance on how to approach the future of the evolving health care teams.

Finding: Health care teams are growing in size and diversity of expertise.

The effectiveness of such teams will critically depend on intra-team communications enabled by access to data and fluency in a common parlance of health informatics. A higher level of quantitative literacy will be required among team members to make informed health decisions from the growing body of health-relevant data and analytics. There are opportunities for new professional roles, integrating medicine, analytics, and social and behavioral sciences in support of health care consumers.

Recommendation: Accrediting bodies should require training of all health care team members to achieve required levels of numeracy and fluency in a common parlance of health informatics. Professional schools should develop continuing education and certification programs that cross-educate team members regarding the diverse expertise required by the team.

5.4 Health Care Teams and Soft Regulation of Health Education

In a world of individual-centric health care delivery there is an intrinsic tension between the free flow of health and wellness-related information to the individual and the need for government to protect individuals from medical harm. On the one hand, enabled by the Internet, the individual is able to learn vastly more about health, diet, disease prevention, specific symptoms, diagnostic procedures, prognosis, alternative medicine, and so forth, than was ever before imaginable in history. Moreover, content providers on the web are often protected from censorship or regulation (at least in the US) by their First Amendment rights. The result is a web that is full of good information that can promote wellness, and also full of bad information that can lead an individual to make misguided, and sometimes dangerous, decisions about their own health.

The rocky relationship between the FDA and the company 23andMe perfectly illustrates the dilemma. 23andMe's business model is to offer personal genetic information to consumers in an understandable form. The company intermediates the consumer's access to a standardized single nucleotide polymorphism (SNP) genotyping test performed by an external laboratory

(currently using a standard DNA microarray, but, foreseeably in a decade, a full sequencing). It then provides web pages that match an individual's unusual SNPs to published papers in the peer-reviewed scientific literature that claim statistically significant relations between the individual's SNP and a health-related phenotype (that is, disease or condition propensity). 23andMe rates, by a system of one to four stars, the strength of the scientific evidence for each SNP, based largely on the number of publications in the literature, and perhaps in principle also on some measure of their quality. It summarizes each of an individual's unusual SNPs by comparing the individual's normal odds of acquiring a disease or condition to his or her apparent odds, according to its summarization of the published scientific evidence.

In one view, 23andMe is offering a purely educational service to consumers. The fact that it is able to educate consumers not just about the genetics of the population at large, but about their own specific genetic profile, only increases the usefulness of the educational information provided. As consumer's take greater responsibility for their own wellness, this is one kind of information that they evidently need. In another view, roughly the position taken by the FDA, it is precisely this specificity to the individual where 23andMe crosses the line from offering consumer education to offering a medical diagnostic test, regulated by the FDA with respect to its "safety and effectiveness", the latter including both analytical validity, the ability of a test to detect or measure the analyte it is intended to detect or measure, and clinical validity, its ability to accurately diagnose or predict the risk of a particular clinical outcome. In the extreme form of this view, each of the roughly 500,000 SNPs is a separate diagnostic test needing individual FDA approval. There is no history of the FDA approving diagnostic tests whose scientific base is as thin as one or two refereed scientific publications

(which may be different enough in experimental design that neither exactly replicates the other’s findings). By a standard like this, the vast bulk of the information that 23andMe offers consumers may not be approvable by the FDA.

Discussions between FDA and 23andMe are ongoing, and it is outside of this report’s scope to assess this particular case. The larger point, however, is that this kind of tension is intrinsic to empowering consumers with the knowledge and tools needed for them to assume more responsibility for their own wellness. JASON believes that, instead of trying to adjudicate a line between the totally regulated (e.g., by FDA) and the totally unregulated (e.g., First Amendment rights of publication), policy makers should look for new and more nuanced approaches.

This line can get confused and blurred. For example, FDA has proposed a new risk-based framework for health IT [11]. This framework includes IT functionality associated with health management which refers to clinical software and clinical decision support systems, and not to the products and services being discussed here. The new framework is concerned with software that is used intentionally to support health care diagnosis and decisions versus the products and services, like 23andMe, that are educational and being construed as “practicing medicine.”

There may be a path forward to manage this line. In the desired individual-centric world envisioned here, individuals will have standing health care teams (see Section 5.3), defined by the flow of their health-related information, consisting of provider institutions, individual clinicians, community resources, and applications that all appropriately share their health-related data. Rather than getting in the middle between the individuals and their

freely chosen sources of health and wellness information (for genomic information, a product and service like 23andMe), government’s protective role should be to ensure that the information flows in a way that enables the consumer to interpret and understand it accurately. For example, as a condition of its releasing interpretative personal genomic information to its customers, a product and service like 23andMe might require that the customer identify a provider interface within his or her health care team, to which the information would also flow. This information flow might result in something as simple as a clickable link on the consumer’s EHR that leads to a “for professionals” summary of the information that was also sent directly to the consumer. Or, with more advanced machine learning, it could trigger alerts within the individual’s health care team that could then be followed up.

This kind of approach could be applied more generally. For example, it might lead to better, safer, and more effective web-based methods by which an individual can get useful medical advice. Today, one observes (at least anecdotally) that more specific or unusual medical web queries tend to yield less reputable medical advice, often in the form of dubious patient-to-patient postings. This may be because reputable medical sites such as Mayo Clinic [4], Web M.D.[26], maintain a greater distance from information that could be construed as offering medical diagnosis, as distinct from patient education. If reputable institutions were free to offer web services much closer to diagnosis, but only on condition that the consumer’s input and advice provided to the consumer were automatically shared with the consumer’s health care team, the accuracy of health advice flowing to consumers might be significantly improved, and conditions requiring clinical intervention might be

flagged sooner. The following findings and recommendations provide an opportunity to create a space for nimble regulation of products and services at risk of being misinterpreted as diagnostic tools and practicing medicine.

Finding: There is an inherent tension between the flow of health information to the individual and the need to protect individuals from misuses of that information. Individuals are already seeking such information and acting upon it.

Finding: At present, hard lines exist between regulated and unregulated products and services. However, technology is evolving quickly in this arena. There are opportunities to update the regulatory process to benefit health and wellness while also accelerating innovation of accurate, personalized health care.

Recommendation: FDA and other agencies should seek a nuanced approach to adjudicating the regulatory line. Products and services that now risk being construed as “practicing medicine” could be allowed to exist in a more nimble regulatory space that fosters rapid innovation and adaptation to new developments while mitigating concerns over demonstrated accuracy. For example, products and services in this space could be subject to a requirement to simultaneously report information to the designated health care team, as well as to the consumer, as a risk mitigation strategy.

5.5 Privacy and Security Considerations

As discussed in the 2013 JASON report [23], privacy and security considerations for all types of information stored in a future health IT system are

paramount, with breaches having potentially serious consequences. While encryption was discussed in the JASON 2013 report [23], additional issues are best acknowledged as important and addressed from the start by design, rather than as afterthoughts.

The use of encryption is not a panacea. It is the case that some sensitive information will inevitably leak. For example sensitive health information must at some point be decrypted so it can be used to make decisions or be fused with other data sources which is seen in this report as potentially a benefit for more informed health decisions. There will be the danger that it will be possible to aggregate diverse pieces of decrypted leaked information for possible use in crime or fraud. As in all such complex information systems, it has been amply demonstrated that even the most sophisticated security systems can be defeated by means of social engineering. It is inevitable that this will happen here too.

The consequences of either errors or a breach can, of course, be very serious. Starting with a patient, malicious action may lead to an alteration of a patient's record that can have life-or-death consequences. A breach of a health care provider's records can have serious health or financial consequences. A breach at the State or Federal level can facilitate fraud on a massive scale.

As illustrated by the examples above, it cannot be taken for granted that security systems and processes exist today at a sufficient level of maturity and reliability to address the problem fully. It can be appreciated that the adoption of encryption measures can help protect vital, personal, and important information but does not, in and of itself, address all issues that must be faced. Indeed, such approaches can make violation of privacy more

difficult but cannot solve the problem completely. Thus continued thought must be given to privacy policies for any future collection of information systems geared towards individual health.

In particular, the 2013 JASON report [23] proposed the notion of privacy bundles. These are a set of permissions and authorizations chosen by an individual based on their privacy preferences. An implementation of privacy bundles specified either individually or via an acknowledged and trusted advocacy organization will be essential, especially if we are to enlarge information technology to the sphere of health as opposed to health care. Future systems should be designed so as to insure that forensic information can be readily obtained should a security violation occur. In this way attribution can be quickly achieved and appropriate legal remedies can then be pursued for the violation of a consumer's stated privacy bundle for the sensitive information that was exposed.

6 CONCLUSIONS

This report has addressed the possibility of developing a health data infrastructure capable of bridging from a health IT technology vision focused on the care of individuals to one focused on the health of individuals. Success would allow the emergence of a Learning Health System able to leverage electronic health records and an entirely new set of health and wellness inputs, from personal health records to environmental data. Such a system would also leverage basic research, clinical practice, public health, and community involvement.

Health and health care are linked, each being affected by the other. External influences also exist in this system, such as the delivery of health care (medical system) and societal support. It was noted that there are missing feedback loops in this system, between health care outcomes and clinical research and between clinical research and public health research. A robust data infrastructure that can enable a “Learning Health System,” requires the ability to ingest all the data, keep the data safe, understand it, integrate the data, and communicate the knowledge gained. This concept requires scalability that goes well beyond the interoperability of EHR systems.

The JASON 2013 report [23] laid the foundation for development of a robust health data infrastructure, focused on the delivery of health care. That foundation can also be used as a model for the broader infrastructure that could enable a Learning Health System. However, several serious impediments exist. First, there remains a critical need for open APIs for EHR systems to further open the entrepreneurial space. Second, FDA and other agencies should seek a nuanced approach to adjudicating the regulatory crite-

ria for products and services that are at risk of being construed as “practicing medicine.” If such products and services could be allowed to exist in a more nimble regulatory space, that could foster rapid innovation and adaptation to new developments while mitigating concerns over demonstrated accuracy.

Several new strategies were highlighted that could help in the acceleration of progress. These included leveraging the work of various nonprofit institutions that focus, for example, on specific medical conditions; developing quantitative capability across the evolving health care team; recognizing that data exchange is critical to making this team a team; and finding opportunities to integrate community support for health, such as the development of reliable geographic and socioeconomic indices.

The report concludes with the following finding and recommendation:

Finding: The learning health system needs to be “closed loop” to ensure a continuous and transparent cycle of research, analysis, development, and adoption of improvements relevant to health and wellness and to the delivery of health care.

Recommendation: HHS should accelerate the development and adoption of a robust health data infrastructure based on the principles outlined in the ONC 10 Year Vision and other consensus studies.

References

- [1] 23andme. <http://www.23andme.com> , 2014.
- [2] R Amarasingham. <http://www.pccipieces.org/pieces-software/>.
- [3] R Amarasingham, BJ Moore, YP Tabak, MH Drazner, CA Clark, S Zhang, WG Reed, Y Swanson, TS Ma, and EA Halm. *Medical Care*, 48(11):981–988, 2010.
- [4] Mayo Clinic. <http://www.mayoclinic.org> , 2014.
- [5] National Research Council. To err is human: Building a safer health system. <http://www.iom.edu/Reports/1999/To-Err-is-Human-Building-A-Safer-Health-System.aspx> , 1999.
- [6] National Research Council. Accounting for health and health care: Approaches to measuring the sources and costs of their improvement. http://www.nap.edu/catalog.php?record_id=12938 , 2010.
- [7] National Research Council. The learning health system and its innovation collaboratives (IOM roundtable on value and science-driven health care). <http://www.iom.edu/Activities/Quality/~media/Files/Activity%20Files/Quality/VSRT/Core%20Documents/ForEDistrib.pdf>, 2011.
- [8] National Research Council. Roundtable charter (IOM roundtable on value and science-driven health care). <http://www.iom.edu/~media/Files/Activity%20Files/Quality/VSRT/Core%20Documents/CharterandVision.pdf>, 2013.

- [9] National Research Council. Health literacy and numeracy workshop summary (IOM roundtable on health literacy). <http://www.iom.edu/Reports/2014/Health-Literacy-and-Numeracy.aspx>, 2014.
- [10] KB Ensor, LH Raun, and D Persse. *Circulation*, 127(11):1192–1199, 2013.
- [11] Food and Drug Administration. Fdasia health it report proposed strategy and recommendations for risk-based framework. available at <http://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHReports/UCM391521.pdf>, 2014.
- [12] Robert Wood Johnson Foundation. <http://www.rwjf.org/en/about-rwjf/newsroom/features-and-articles/culture-of-health-prize.html>, 2014.
- [13] LA Green, GE Jr Fryer, BP Yawn, D Lanir, and SM Dovey. *N Engl J Med*, 344(6):2021–2025, 2001.
- [14] Healthmap. <http://healthmap.org/en>, 2014.
- [15] Hiccup. <http://www.hiccup.co/contest/> and <http://www.hiccup.co/news/>, 2014.
- [16] HL7. <http://www.hl7.org/implement/standards/fhir/>, 2014.
- [17] HL7. HL7 CDA standard. http://www.hl7.org/implement/standards/product_brief.cfm?product_id=7, 2014.
- [18] HL7. HL7 continuity of care document. http://www.hl7.org/implement/standards/product_brief.cfm?product_id=6, 2014.
- [19] HL7. HL7 web site. <http://www.hl7.org/>, 2014.

- [20] Apple Inc. <http://developer.apple.com/healthkit>, 2014.
- [21] Health IT. <http://www.healthit.gov/providers-professionals/faqs/what-are-differences-between-electronic-medical-records-electronic>
- [22] JASON. Data analysis challenges. <http://fas.org/irp/agency/dod/jason/data.pdf>, 2008.
- [23] JASON. A robust health data infrastructure. <http://healthit.ahrq.gov/RobustHealthDataInfrastructureReport>, 2013.
- [24] H Lee, E Sun, D Ham, and R Weissleder. *Nature Medicine*, 14(8):869, 2008.
- [25] K D Mandl. et al., The SMART platform: early experience enabling substitutable applications for electronic health records. *J Am Med Inform Assoc*, 2012.
- [26] Web MD. <http://www.webmd.com> , 2014.
- [27] Patients Like Me. <http://www.patientslikeme.com>, 2014.
- [28] P Mitchell, M Wynia, R Golden, B McNellis, S , Okun, CE Webb, V Rohrbach, and I Von Kohorn. <https://www.nationalahecc.org/pdfs/VSRT-Team-Based-Care-Principles-values.pdf>, 2012.
- [29] AH Mokdad, JS Marks, DF Stroup, and JL Gerberding. *JAMA*, 291(10):1238–1245, 2004.
- [30] AH Mokdad, JS Marks, DF Stroup, and JL Gerberding. *JAMA*, 293(30):293–298, 2005.

- [31] MD Naylor, KD Coburn, and Kurtzman ET. Inter-professional team-based primary care for chronically ill adults. In *ABIM Foundation meeting to Advance Team-Based Care for the Chronically ill in Ambulatory settings*, Phila. PA, 2010.
- [32] Collaborative Care Network. <http://c3nproject.org/>, 2014.
- [33] Presidents Council of Advisors on Science and Technology. Report to the President: Better Health Care and Lower Costs: Accelerating improvement through systems engineering. http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_systems_engineering_in_healthcare_-_may_2014.pdf , 2010.
- [34] Presidents Council of Advisors on Science and Technology. Report to the President realizing the full potential of health information technology to improve healthcare for Americans: The path forward. <http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf>, 2010.
- [35] Office of the National Coordinator for Health Information Technology. Connecting health and care for the nation; a 10-year vision to achieve an interoperable health it infrastructure. <http://healthit.gov/sites/default/files/ONC10yearInteroperabilityConceptPaper.pdf> , 2014.
- [36] County Health Rankings. <http://www.countyhealthrankings.org/about-project/rankings-background>, 2014
- [37] LH Raun, KB Ensor, and D Persse. *Environmental Health*, 13:58, 2014.
- [38] D Wattendorf. DARPA biotechnologies. Briefing to JASON, 2014.

- [39] J Wiens, WN Campbell, ES Franklin, JV Guttag, and E Horvitz. *Open Forum Infections Diseases*, 1(2):1–9, 2014.