The case for a robust data analytics platform

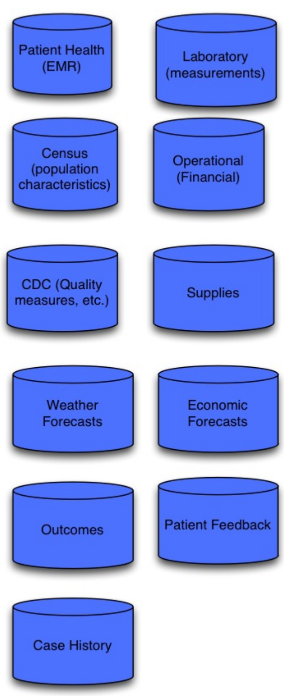
Persistent White Paper

10/8/2016. V0.8

Abstract: Shifting provider remuneration models, from pay-for-service to pay-for-outcome, are changing the way in which information systems are used within healthcare. Providers are assuming risk in exchange for being able to capture cost-saving efficiencies as increased revenue. Accurately measuring that risk, quality of services rendered, and patient and population health outcomes requires curation of data from a wide variety of sources and making that data available to sophisticated analytical, clinical, administrative, and patient applications.

Efficiency, risk, and the case for healthcare data analytics

Of the nearly $3 trillion dollars spent annually on healthcare in the United States it is estimated that as much as $1 trillion is simply wasted due to inefficiencies in the delivery of health care.



Under existing fee-for-service (FFS) remuneration models, when providers address those inefficiencies it often results in a lowering of revenue to the provider[1]. New provider remuneration models, such as capitation and bundled payments, are designed to correct this and allow providers to capture the value of

In a clinical trial, intubation rates fell from 78% to 18%. The children stayed in the nursery, not the far more expensive newborn ICU. With the simpler, less invasive care, the hospital's total operating costs for these children fell by $544,000 a year. But fee-for-service insurance payments dropped by $873,000, causing a $329,000 dip in the hospital's operating income. The hospital also had to bear the costs of developing and implementing the change. Moreover, when Intermountain decided to deploy the new methods across all its hospitals—clearly the right thing to do for the children—that $329,000 turned into more than $5 million in annual losses.

Brent C. James and Gregory P. Poulsen, "The Case for Capitation," *Harvard Business Review* (July-August 2016)

---

[1] Brent C. James and Gregory P. Poulsen, "The Case for Capitation," *Harvard Business Review* (July-August 2016)

increased efficiencies while simultaneously improving the quality of health care at both the population and individual patient level.

"… reimbursement models are changing; meaningful use and pay for performance are emerging as critical new factors in today's healthcare environment. Although profit is not and should not be a primary motivator, it is vitally important for healthcare organizations to acquire the available tools, infrastructure, and techniques to leverage big data effectively or else risk losing potentially millions of dollars in revenue and profits"

Wullianallur Raghupathi and Viju Raghupath, "Big data analytics in healthcare: promise and potential." *Health Information Science and Systems* (2014 2:3)

In order to do so, however, requires that providers have access to a vast array of data and the sources of that data in order to support decision-making. At the same time, providers must now record a wider set of operational and clinical data in order to support the accurate measurement of and reporting on outcomes.

Bringing this all together requires a *data-centric* approach to information processing that de-siloes individual data sets and allows access to that de-siloed data by a wide variety of applications.

The Triple Aim

- Improve the individual experience of care
- Improve the health of populations
- Reduce the per-capita cost of care for populations

Those applications enable the transformation of the raw data into "information, knowledge, and wisdom"[2]. For example, emerging patient relationship management systems enable capturing individual patient histories, greatly reducing issues of institutional amnesia and duplicated procedures on every individual patient encounter.

Other sources of that information, knowledge, and wisdom include

---

[2] John D. Halamka, "Early Experiences With Big Data At An Academic Medical Center,: *Health Affairs*, 33, no.7 (2014):1132-1138 .

clinical decision support systems, "big-data" quality measurement systems, and patient-generated data. Application infrastructure and platforms must robustly support not only data ingestion from a wide variety of sources, but also ***data accumulation*** from a wide variety of sources (such as the just-mentioned patient-generated data) into deep and wide Hadoop data lakes, relational databases, and graph databases.

Issues

Before getting into business specifics and solutions around healthcare data it would be prudent to outline the real challenges facing even the most technically sophisticated organization. Healthcare data has a unique set of attributes, such as data hygiene, data provenance, and data assurance, all of which will have to be comprehensively addressed in any effective platform.

1. There's a lot of data

The adoption of electronic record systems within healthcare is leading to an explosion in the amount of raw data available. In 2011 it was estimated that over 150 exabytes of data were being used in the healthcare industry. Kaiser Permanente is estimated to have between 26.5 and 44 petabytes of rich patient health data stored in its EHR systems[3].

It is also not enough to simply pool data into a single data lake. Any system that effectively supports provider efficiency and quality improvements must contain mechanisms to curate, normalize, and link the data from the disparate sources together. Another necessary function is the creation of metadata and schemas that tie data elements from different sources together and give semantic meaning to the data. Some of that metadata generation can be automated. For example extracting structured data from within a

[3] Institute for Health Technology Transformation, *Transforming Health Care Through Big Data.* 2013.

continuity of care document (CCD) and then relating those structured values where they appear in other data sets.

It must also support interoperability with the other systems, such as EMRs, from which the data is pulled and to which data will need to be pushed. For example, EMRs are likely to remain the systems of record for patient health information. Changes to the health information of a given patient must be pushed to the appropriate system of record (EMR) as needed.

> The typical medical chart is actually stored in various fragments in many different locations and systems. Imagine your entire medical record as a jigsaw puzzle in which the pieces are scattered and stored in different locations and in different types of boxes, each of which is hard to open. Your cardiologist has her record of you but not the record from your endocrinologist or from the urgent care doctor you saw six months ago for bronchitis, for example. No wonder people feel as if they are repeating themselves every time they visit a medical facility. There is no master record of your care.
>
> Mark Scott, "The Data is In: 3 Ways Analytics Will Improve Healthcare," http://dataconomy.com/the-data-is-in-3-ways-analytics-will-improve-healthcare/ (March 2016)

The location of patient health data is another significant issue that any effective platform must confront and resolve.   For example, a given patient may have health data in their primary care physician's EMR, any number of specialists' EMRs, as well as hospital EMR systems. Applications that deliver a full Patient360 view must be able to transparently draw on a patient's health information irrespective of where it is located.

And they must be able to do so securely. Both the HIPAA and the HITECH acts convey extreme caution when interacting with and creating protected health information (PHI). Penalties for disclosing PHI inappropriately are becoming severe. Healthcare data must be protected by well-architected systems that are regularly audited for faults and vulnerabilities.

2. Structured Data: Data cleanliness and normalization, specificity, and ambiguity

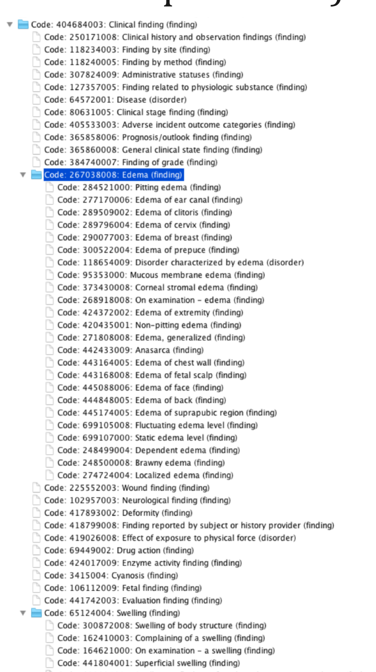
A large health care system may have several hospitals and other facilities contained within it. Each facility probably implemented its own electronic health record system and even when different facilities use systems from the same EHR vendor, location specific customizations can mean that when one facility stores laboratory results one way and another way. When bringing data together from each facility into a useful platform, care must be taken with variations in coding, units, and even medical shorthand between facilities if meaningful correlations are to be found in the data.

"Admittedly, the [health information exchange]() process necessary for getting that patient data isn't easy, Halamka says. Even when data's in hand, analytics can be complicated; what one electronic health record (EHR) system calls "high blood pressure" a second may call "elevated blood pressure" and a third "hypertension." To combat this, Beth Israel is encoding physician notes using the [SNOMED CT standard](). In addition to the benefit of standardization, using SNOMED CT makes data more searchable, which aids the research query process."

Brian Eastwood, "6 Big Data Analytics Use Cases for Healthcare IT," *CIO Magazine*, April 2013

Structured data, such as ICD-10 codes, also present issues. For instance, the great specificity of the ICD-10 taxonomy can be an issue when aggregating diagnostic codes into higher classifications, such as for display in a patient portal.

Structured data can be both too specific (ICD-10 is a particular example of this) and introduce issues of ambiguity at the same time! For example, under SNOMED code 404684003 (Clinical Finding) we find code 267038008 (Edema) as well as code 65124004 (Swelling) at the same "is-a" level within the SNOMED taxonomy.   Edema is largely understood as either a form of swelling or synonymous with the term swelling within the medical community. It is possible that the same patient, visiting two separate facilities, could have the same diagnosis coded in two entirely different ways and most forms of automatic data discovery would not be aware that the two findings were in fact just one finding.



3. Unstructured data: Narratives, processing, and concepts

The vast majority of qualitative data in healthcare exists in unstructured forms (upwards of 95%). Clinician's notes, care plans, imaging data, etc. are all examples of unstructured data. Extracting useful information from unstructured data requires machine learning applications and sophisticated parsing mechanisms.

Issues with patient narratives

EHR systems store a lot of patient health information but they do not capture patient socio-economic profiles. Patient socio-economic information is crucial to accurate predictive analytics about future patient health conditions. It's also a crucial component in the quality delivery of patient care through enhanced ability empathize with the patient at an individual level. Modern healthcare-specific customer (patient) relationship management (PRM) systems are highly effective tools with which to capture each patient's individual "story" for qualitative use.

Text Analytics in Healthcare

Text analytics in healthcare involves the following practice areas (per Miner et al., 2012):

- Information Extraction (IE)
- Natural Language Processing (NLP)
- Concept extraction (CE)

Free-form (unstructured) clinical notes, patient social histories, etc. must first be extracted (IE) from the relevant system-of-record (EMR, PRM, social media, etc.) then subject to natural language processing in preparation for concept extraction. This is particularly important (and difficult) within the healthcare industry as synonyms such as “kidney” and “renal” abound in the data, as well as the use of different levels of specificity within taxonomies to refer to the same root issues. Finally as with all disciplines and specialties, healthcare abounds in its prevalence of the use of abbreviations (and inconsistently).

Other issues common to most forms of textual analysis also abound, such as machine understanding of homonyms. Does “left” refer to location within the body, or the fact that a surgical instrument remained inside a patient post-operation? Is “MD” an abbreviation for “Medical Doctor” or “Mental Disorder?”

Structured data is the easiest type of data to capture and categorize in a database. Accounting data, an example of structured data, includes numbers with a specific value in a particular column. Structured data in healthcare would be a lab value or patient demographic data that is entered from a dropdown box or radio button. Structured data is consistent and resides in pre-defined fields within the record.

Unstructured data is unorganized, may have irregularities or be ambiguous, and is typically “text-heavy.” A prime example of unstructured data in Health IT is a paragraph about the history of present illness. It is hard to condense patient complaints or physician assessments into a series of checkboxes and radio buttons and yet there is great value in analyzing patient information without having to manage free text. It is helpful to know if there is pertinent information about a cardiac procedure in the physician note even if it is not listed in the structured data of the problem list.

It is estimated that 95% of the worlds’ data is unstructured, so healthcare is not unusual. Today e-mails, text messages, Word documents, videos, and pictures are all unstructured data. The document or file has structure, but the content or text within it is unstructured. Solutions are being developed to parse unstructured data into patterns that improve the value and usability of the data. Natural Language Processing tools and other data-mining tools create opportunity to glean structure from free text.

“The Human Condition in Structured and Unstructured Data,” Acumen Physician Solutions http://acumenmd.com/blog/human-condition-structured-unstructured-data/

The business case for a data-driven digital transformation of healthcare

Doing what we do now better: Three Categories of Waste

Drags on Productivity (5% of waste in healthcare)

From an efficiency and risk-management angle, there are three types of waste. Production-level waste, case-level waste, and population-level waste[4]. The first category involves what are traditionally referred to as costs of goods sold, or "units of care" in healthcare. Pricing negotiations, inventory storage, production process streamlining (e.g. cost of producing X-rays), are all examples of waste falling into this category. It represents approximately 5% of total health care waste.

Unnecessary Redundancy (50% of waste in healthcare)

The second category involves misuse of resources during patient encounters. Redundant testing, supportive medications prescribed when the complications could have been avoided, end-of-life care to people who have asked not to receive it, etc. are examples of this type of waste, which comprises 50% of all healthcare waste.

Shifting from traditional fee for service (FFS) payments to person focused payments (in which all or much of a person's overall care or care for related conditions is encompassed within a single payment) is a particularly promising approach to creating and sustaining delivery systems that value quality, cost effectiveness, and patient engagement. Such payments can include accountability for the quality of care at the population level, rather than for the volume of particular services. Population-based payments give providers more flexibility to coordinate and manage care for individuals and populations. In combination with substantially reduced incentives to increase volume and increased incentives to provide services that are currently undervalued in FFS, there is a consensus that this flexibility will expedite innovations in care delivery, particularly for individuals with chronic, complex, or costly illnesses.

Health IT Playbook. Office of the National Coordinator for Health Information Technology
https://www.healthit.gov/playbook/

Lack of proactive prevention (45+% of waste in healthcare)

---

[4] James and Poulsen, "The Case for Capitation," *Harvard Business Review* (July-August 2016)

The third category involves unnecessary or preventable cases. This includes elective surgeries that patients would, with better information, have forgone. Expensive medical care (e.g. emergency room visits) where the emergency encounter might have been avoided with better medication conformance. In fact, any time a more expensive form of care is administered because a less expensive form of care was not earlier administered. This category constitutes about 45% of total waste. However its effects are compoundable as this category drives the first two – unnecessary care still consumes the "units of care" from the first category as well as offers the same misuse of resources from the second.

> Researchers at the Johns Hopkins School of Medicine discovered they could use data from Google Flu Trends to predict sudden increases in flu-related emergency room visits at least a week before warnings from the CDC… the analysis of Twitter updates was as accurate as (and two weeks ahead of) official reports at tracking the spread of cholera in Haiti after the January 2010 earthquake .
>
> Institute for Health Technology Transformation, *Transforming Health Care Through Big Data* (2013)

Payment-for-service revenue models do not allow providers to capture savings when waste is reduced. Payment-for-outcome revenue models, including capitation models, will allow providers to realize the savings on their own balance sheets from reduction in levels of waste. Effective application of information technology is critical to capturing those savings at the provider level.

Doing New Things: Two Categories of Opportunity – measuring outcomes and improving patient/provider satisfaction

In addition to opportunities arising from waste reduction, changes in the industry are opening new provider revenue opportunities as well. Reductions in waste will improve the quality of care (outcomes). Outcome measurement is a relatively new concept in the industry and in order to capture the financial rewards associated with better outcomes, providers will need to develop new and sophisticated ways of measuring treatment efficacy.

In addition to reporting and measuring (both forms of data analytics) new opportunities are arising in the area of patient relationship management (PRM, sometimes called CRM). Increased patient record portability and competition between health systems has moved patient satisfaction to a new level of importance.

Increasingly health systems, hospitals, and individual practices are implementing PRM systems to help capture new patients, make sure existing patients are not lost due to dissatisfaction, and to improve provider's ability to emphasize with the patient. Not only does this improve the patient experience but it has shown to dramatically increase job satisfaction within the provider community.

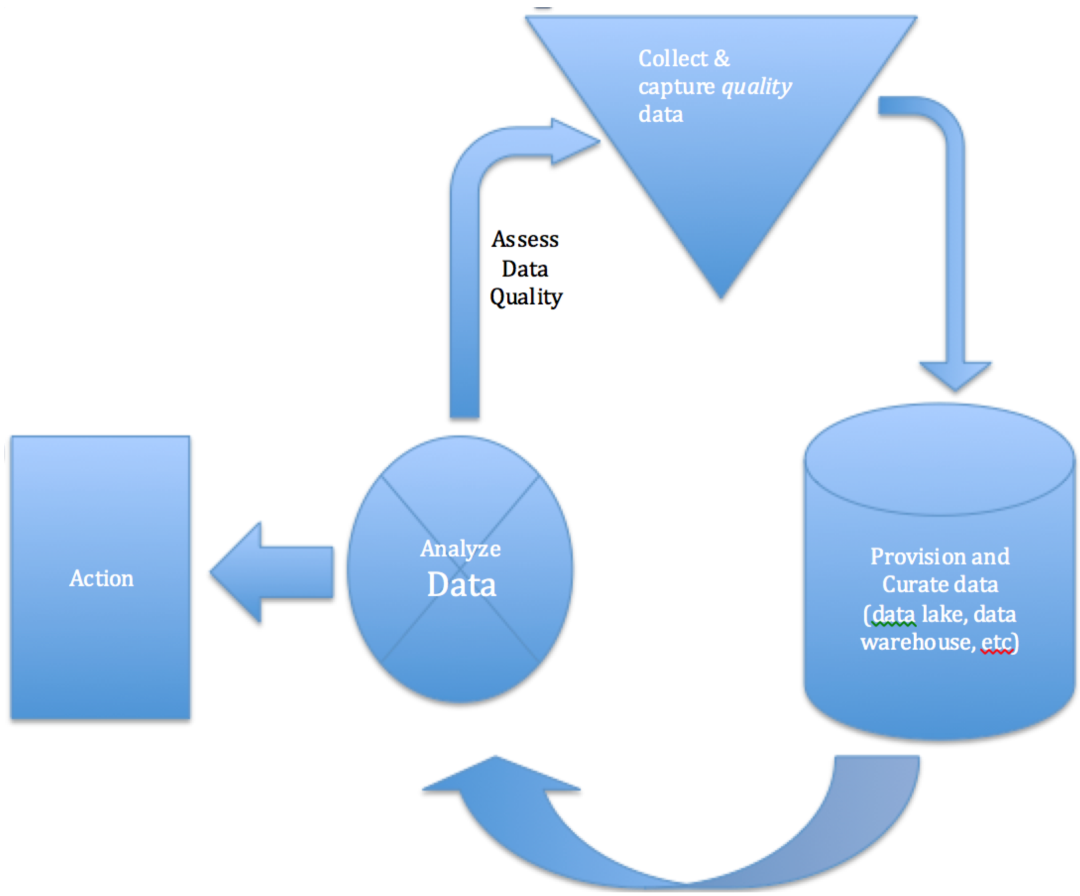
Data to Action

As we saw above, the healthcare industry is currently generating immense amounts of *data.* Whether from on-premise medical equipment, such as EKGs, or from "in the wild" equipment like personal Fitbits, home blood pressure machines, etc. the amount of raw data being collected is staggering and will continue to grow at an incredible rate.

But data alone is not useful. First the raw data must be turned into *information.* A patient may have had dozens of EKG readings, that's data, but when a meaningful drop of EKG voltages across several recent readings occurs, that's *information.* Relating that information to other information about the patient, such as a recent diagnosis of fatigue, turns the information into *knowledge.*

When that knowledge is turned into *wisdom*, such as understanding that the patient's heart may be subject to foreign body intervention such as in hemochromatosis, then the wisdom can be used to develop an effective plan of action. In this case, an iron-restricted diet for the patient will likely greatly improve outcomes while reducing waste.

Data to information to knowledge to wisdom to action. We have the data, the challenge is developing and deploying systems that result in action.



Opportunity Areas

At a high level, the opportunities are:

1. Risk Management
    a. Nowcasting
    b. Forecasting
2. Productivity Enhancement & Waste reduction
    a. Clinical Decision support/Expert Systems
    b. Earlier identification of at-risk individuals and populations
3. Care enhancement (improved patient experience)
    a. Clinical support (improved patient record access, access to social, family, etc. histories)
    b. Personalization of service
    c. Care coordination / Continuity of care
4. New business opportunity identification and enablement
    a. Remote Care & Telehealth
    b. Patient Relationship Management
5. Analytics
    a. Reporting
    b. Health forecasting
    c. Quality measurement
    d. Risk stratification

Applications

Mapping. Using GIS & spatial analysis to identify health care populations

Risk Stratification through identification and classifications of comorbidity factors

Hierarchical Condition Categories (Per Medicare Advantage). 70 Categories and expected health expenditures

Adjusted Clinical Groups (ACG). Johns Hopkins. Used to predict hospital utilization

Elder Risk Assessment (ERA)

Chronic Comorbidity Count (CCC) From Agency for Healthcare Research and Quality (AHRQ).

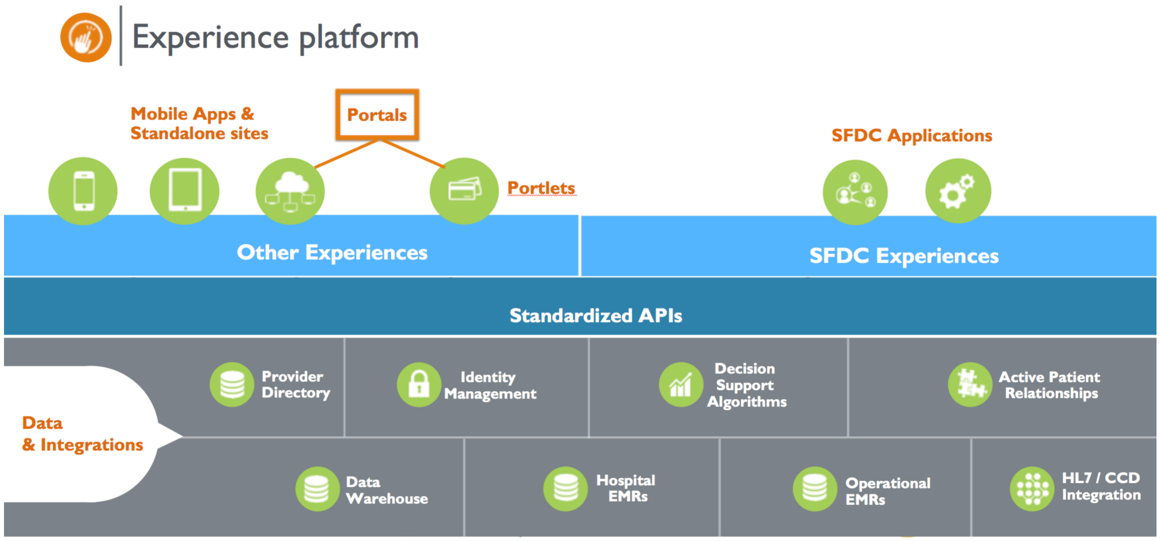
Minnesota Tiering. Five tiers: 0 (low) to 4 (Complex: 10+ conditions from extended diagnostic groups)

Charlson Cormobidity Measure. Risk of one-year mortality based on presence or absence of 17 comorbidity definitions.

Predictive Analytics

Using social media (Facebook, Twitter, etc.) to predict individual behavior, mining text to make analysis for suicidal behavior, medication conformance, activity, mood, etc. (Durkheim Project, DARPA & Dartmouth University & Attivio)

The Persistent Platform



Meaningful Use initiatives spurred the adoption of electronic health record systems (EHRs) throughout the provider space. Imaging data is no longer produced on film but rather generated and stored digitally. Accumulating health care data in digital form is no longer difficult. Turning that data into information, knowledge, wisdom and then action requires an application framework that provides applications with the services they need.

Persistent builds solutions in healthcare atop a sophisticated platform that provides applications with a set of standardized APIs giving access to a broad range of data from systems of record and systems of reference. The platform supports a number of standardized services, such as security and access control, messaging, etc. as well as micro services such as provider-patient relationships.

A number of healthcare-specific services are also provided as well as provisions for third-party plugins. Healthcare services, whether internal to the platform or provided by third party applications, support a broad number of use cases.

Connectivity & Curation

Middleware, Standard Services, and Add-Ons

Deployment and Scalability