Demystifying Healthcare Data Governance

As the age of analytics emerges in healthcare, health system executives find themselves increasingly challenged to define a data governance strategy that maximizes healthcare data’s value to the mission of their organizations. “Data is the new oil,” Andreas Weigend, former chief scientist at Amazon.com recently said. That statement might be a bit dramatic—data can’t heat homes or power cars—but there is no denying the dramatically growing importance of data to business success. This is especially true for those businesses based in knowledge management delivery, which is certainly the case with healthcare.

The rising hype of the data warehouse and analytics market has resulted in significant noise from vendors and consultants who promise to help health systems develop their data governance strategy. The purpose of this paper is to share my observations and lessons learned in data governance, as a CIO, analytics specialist, and vendor; and in so doing, remove the mystery and confusion about how and where to start the data governance journey in healthcare and provide a roadmap to evolve that journey incrementally over the lifespan of the organization.
AN ELUSIVE TARGET: THE PERFECT DATA GOVERNANCE ENVIRONMENT

It is difficult to find a perfect data governance environment. Over the course of my career, starting in the early 1980s when I was a young Air Force information systems officer, I’ve seen attempts at data governance swing between too much and too little, rarely finding equilibrium. Until the early 1990s, the term data governance wasn’t commonly used, and the body of knowledge around data governance was almost non-existent.

The chart depicts a small but significant sampling of projects I was involved in and the initial state of the data governance for each. While each project has a unique story of data governance, there are common themes and patterns that emerge as indicators of success and failure. The point of this paper is to expose those common patterns.

The Maintenance Management Information and Control System (MMICS, 1987) is an Air Force-based system that can be thought of as an EMR for aircraft. Continuing that analogy, Air Force crew chiefs function as the primary care physicians for their assigned aircraft. These crew chiefs are responsible for maintaining and optimizing the health of their aircraft while at the same time, consuming as few supplies and resources as possible. The development of MMICS was prompted by a 1983 GAO report criticizing the Air Force for its poor ability to accurately track and manage the health and maintenance costs for individual aircraft, as well as entire fleets of aircraft. It is easy to see the parallels between patient care and population health management.

In my career, MMICS still stands out as the only project that achieved a perfect balance of user interface, data collection, and analytic governance, from its origins. MMICS was designed from the back, forward. That is, the Air Force first decided what type of data was needed to be collected to effectively manage the health of its vast variety of aircraft and other assets. Then they designed the MMICS user interface to support the collection of that data, as well as the workflow efficiency of the crew chiefs using the system. The efficiency of user interface and data collection was a very important attribute.
because those crew chiefs had to operate in less-than-ideal conditions on active flight lines in scorching heat and subzero cold. The overriding design motives were analytic output and workflow efficiency.

It is not hard to imagine the benefits of designing an EMR with those same two design motives: analytics and clinician workflow efficiency.

The Sanders Philosophy of Data Governance

My philosophy of data governance is to be as lean as possible—govern to the least extent necessary in order to achieve the greatest common good. Healthcare organizations tend to govern too much too soon, which results in unnecessary constraints on data and wasted labor. These health systems govern data that, in practice, doesn't need any governance yet. Pairing the data governance function with overseeing the development and evolution of an enterprise data warehouse (EDW) gives the data governance committee something tangible to govern. A Late-Binding™ data engineering architecture and an EDW, combined with lean data governance, work very well together. Bind no data before its time, and govern no data before its time.

DATA GOVERNANCE CULTURES

Data governance tends to mirror societal governance in the three categories shown in the diagram: authoritarian, democratic, and tribal.

Authoritarian data governance cultures are typically associated with a centralized and closely managed EDW. These data warehouses are usually designed around a monolithic, early-binding data model, and the data governance culture tends to believe that most, if not all, analytic use cases are well known and persistently agreed upon. Access to the EDW is tightly controlled through a top-down bureaucratic approval process, and there is very little—if any—tolerance for unconstrained and unsupervised data exploration within the EDW. In healthcare, the dominant users of the data warehouse in an authoritarian governance culture are typically from finance. Occasionally, users are data aggregators in the organization responsible for submitting reports to
external agencies, but they are not data analysts who play an active role in process improvement or cost reduction.

The tribal form of data governance is characterized by the lack of a centralized data governance function as well as the lack of an EDW. Analytics and reporting functions are resourced from decentralized business and clinical units, resulting in significant resource inefficiencies and redundant, confusing reports. Without a centralized EDW, there is no ability to fuse data together in such a way that supports a better understanding of longitudinal care and population health. The value of data to the organization is much lower than it could be, if that data were used more synergistically.

The balanced, democratic governance model is characterized by a centralized data warehouse, and these data warehouses typically follow a Late-Binding™ data engineering architecture. This type of governance culture offers the ability to produce meaningful, actionable analytical insights across the continuum of care. A centralized data governance body exists and is staffed by executive leadership from across the organization (such as the COO, CFO, CIO, CMIO, CMO, and CNO). This group establishes overall principles of data management and analytics, then entrusts adherence to those principles to the data analysts that access the EDW. There is a cultural commitment to participate in data governance according to shared values, policies, and procedures. Significant emphasis is placed upon the protection of sensitive patient information, but with that exception, there is a high degree of tolerance on the part of the data governance committee for unconstrained data access and exploration. Access to the EDW is encouraged, as is cultural data literacy and data-driven decision making.

Too Little Data Governance: What Does It Look Like?

Too little data governance is marked by analytic inefficiency and uncoordinated data analysis resources. Analytic redundancy and expenses are high, in both human labor and technology. Multiple, redundant resources result in inconsistent analytic results when attempting to answer the same question. Poor data quality runs rampant. When there is a data quality problem, there is no formal body for appeal, and no process exists for fixing data quality problems. Thus, data quality problems tend to languish. Finally, there is an inability to respond to analytic use cases and requirements. Without a data governance function in place, when new analytic requirements emerge, there is no appropriate method for discussing and supporting the new requirements. The net business impact of these characteristics is lengthy and inaccurate decision cycles.
Too Much Data Governance: What Does It Look Like?

As expected in any culture characterized by highly centralized and authoritarian decision-making, too much data governance is also very inefficient. One of the vital sign indicators of a burdensome and overbearing data governance culture is an unhappy collection of data analysts and their internal customers. Approving and loading new data content into the EDW takes too long. Making changes to data models to support analytic use cases takes too long. Getting access to data takes too long. Resolving data quality problems takes too long. Developing new reports takes too long. In sum, everything takes too long. Interestingly, the consequences of too much data governance are the same as too little data governance—decision-making cycles are longer than they should be, the decisions are less accurate, and the mean time to improvement for the organization is longer than it should be.

THE TRIPLE AIM OF DATA GOVERNANCE

Borrowing from the familiar IHI Triple Aim initiative for healthcare improvement, the Triple Aim of Data Governance is: (1) Ensuring data quality; (2) Building data literacy; and (3) Maximizing data exploitation for the organization’s benefit.

Ensuring data quality is the first step in a data governance mission. In this context, data quality is defined by the completeness of the data times its validity (Data Quality = Completeness x Validity). That is, collecting all the needed data for a particular analytic use case and ensuring that the data is valid. The data governance committee and function must have strategies to support and improve data quality—ensuring completeness and validity of the data to support analytics.

The second aim is building data literacy throughout the organization, and the data governance committee should champion this initiative. It makes no sense to build a library in an illiterate community. Similarly, it makes no sense to invest in the technology and data content of an EDW in an organization that suffers from a lack of data literacy. The data governance committee must sponsor training, education, and hiring practices that build the data literacy of the organization.

Finally, the third aim is data exploitation—maximizing the value of data to the organization, creating a data-driven culture that lowers costs, improves quality, and reduces risk. It’s not enough to support data quality and data literacy. Those attributes alone will not serve the betterment of the organization. That data and those skills must be put to good use by creating a culture that constantly seeks self-improvement through the spotlight of data.
Mindset, Skillset, and Toolset

Another useful three-part paradigm to guide the data governance committee is: mindset, skillset, and toolset, in that order of importance. The data governance committee must play an active executive leadership role in the development of a data-driven mindset throughout the organization. This is an important first initiative for the data governance committee—simply communicating from the executive level that the organization is, from this point forward, becoming a data-driven culture, constantly searching for ways to reduce their mean time to improvement. The next step is the development of the skillset among the employees to support this data-driven mindset. Finally, the data governance committee is the most logical choice for executive sponsorship of the toolset, such as an EDW, necessary to support the analytics journey.

THE DATA GOVERNANCE LAYERS

As seen in the diagram, there are multiple layers in the data governance process, flowing down from the executive and board leadership, then the data governance committee, data stewards, data architects and programmers, the database and systems administrators, and finally the technical data access control system that surrounds the data warehouse and the analytics platform.

Combined effectively, these layers of data governance result in happy data analysts. At first glance, these layers might appear bureaucratic, but if implemented properly, the layers efficiently complement one another.

From the high-level guidance and aspirations of the executive and board layer down to the low-level, embedded technology of the information system supporting the EDW, every layer plays an important role in data governance. Below is a description of the typical functions and behaviors that occur in each layer.
Data Governance Role: Executive and Board Leadership

The executive and board leadership sets strategic goals for analytics, which in turn initiates an implementation strategy by the data governance committee. For example, leadership of a large ACO may decide that their healthcare organization needs a longitudinal analytic view across the ACO of patient treatments and costs, as well as all similar patients in the population served.

Data Governance Role: Data Governance Committee

The data governance committee’s job is to take that aspirational goal and translate it into analytic skillsets and toolsets. The committee evaluates options to achieve that goal. In this example, for the continuum-of-care-analysis and population health, a data governance committee would likely conclude the need for an EDW. They would also recognize that, in this data warehouse, the organization will need to consolidate all clinical and financial data associated with the beneficiaries in the ACO, as well as implement a master patient identifier and common coded terminologies to tie it all together.

In addition to the data warehouse toolset described above, the data governance committee is the proper executive sponsor for the development of the skillsets necessary to achieve the aspirations of the organization. Those skillsets include a data analysis team and the IT skills to manage the data warehouse. At this point, the committee would also start defining the general principles for determining who will have access to which data in the EDW, and the processes for approving and auditing that access. A critically important role in the processes for reviewing and approving access to the data content of the data warehouse is the data stewards.

Data Governance Role: Data Stewards

The data steward is an emerging and very important role in healthcare. Data steward candidates are those employees who are at the frontlines of data collection in the organization and understand how data is collected in their area of responsibility, how the information systems that support their areas collect this data, and any shortfalls or quality problems in that data. They are the equivalents of subject area experts in a library. Healthcare data stewards can include nurses, physicians, registration clerks, registry managers, billing and coding staff, cost accounting staff, and researchers. They typically have been underappreciated in their role as the originators and stewards of data in the organization.

In the implementation of an analytic use case, the data governance committee will rely heavily upon the data stewards whose data areas will be included in the use case. For example, it is common for the manager of the hospital registrars to serve as the data steward for the data collected in the process of registering
and scheduling a patient for a clinical encounter. This includes critical data such as the patient's name, gender, insurance coverage, and—especially—the master patient identifier. It is also very common for nurse managers to function as data stewards for the data collected in their areas of responsibility. Chief medical informatics officers (CMIOs) also commonly serve as data stewards for the data collected by physicians in the organization's EMR. All of these data stewards will participate in the development of an analytic use case to ensure that the data targeted for use in support of the analysis is, in fact, appropriate for the analytic use case. The data stewards will also serve as data content experts and advisors to data architects and data analysts, in general.

Data Governance Role: Data Architects and Programmers

With the consultation of the data stewards, data architects and programmers will translate the aspirations of the executive team and data governance committee into the technical implementation of the analytic use case in the EDW. They will extract data from the source information systems such as registration, EMR, revenue cycle, and cost accounting; ensure that that data is properly modeled and stored in the EDW; bind and organize that data to support the specific analytic use case; and determine the best way to expose and share that data with the executive team, using tools for data visualization and manipulation, such as Excel, Cognos, QlikView, SAS, etc.

Data Governance Role: Database and System Administrators

Database and system administrators implement and configure the auditing and access control systems of the EDW in a manner that reflects the high-level principles for data access and security as outlined by the executive team and data governance committee. The database and systems administrators must also work closely with the data architects and programmers to ensure that data content and visualization applications are integrated with the operating systems and the database management systems supporting the data warehouse. Upon approval by the data stewards, the data architects and programmers will provide the systems and database administrators with the names and roles of the people in the organization approved for access to the contents of the data warehouse.

Data Governance Role: Data Access Control System

The data access control system is embedded in the operating system and database management system supporting the EDW. The system provides the means for associating the user names of the people logging into the EDW with their authorized levels of access to the data content and visualization applications. The access control system also provides an audit trail for tracking who accessed what data and when.
Data Governance Role: Data Analyst

If functioning properly, all six layers in the data governance ecosystem work together in complementary fashion to provide a supportive environment for data analysts. It is worth noting, in this context, a data analyst can be anyone from a very technical data engineer working in a business or clinical unit to a member of the board interacting with an online organizational scorecard. In a fully mature, data literate culture, everyone is a data analyst in the scope of their role in the organization.

Managing the Roles of a Data Warehouse

Role management in a data warehouse is a worthy of a separate blog or white paper, but in summary, constrain the number of roles in an EDW as much as possible. The director of the EDW should be the data steward for EDW roles; role management should not fall to the systems or database administrators. Experience shows that if the roles are not carefully managed and constrained, the number of roles grows exponentially, making management of them very difficult. These difficulties manifest themselves in security risks and inappropriately limiting access to data—that is, some analysts will accidentally receive access to data that they are not authorized to view and some analysts will be restricted from data that they need and are authorized to view. From a data analyst perspective (not a programmer or data architect working at the system level, which will require more complex read/write access to data content areas) it is best to use “late binding” philosophy, starting with only two access roles and expanding the number of roles from there only when thoroughly justified by a use case. Those two roles are: (1) Personal Health Information Access (PHI); and (2) De-Identified Data Access—that is, all data that is not patient identifiable. Over time, additional roles will likely be required to accommodate other use cases and data content, such as employee identifiable data, physician identifiable data, and extremely sensitive patient data (such as behavioral health data, HIV status, drug and alcohol addiction, and others). State regulations and laws typically define these sensitive data areas.

WHO IS ON THE DATA GOVERNANCE COMMITTEE?

The data governance committee is an executive-level committee and should include the chief analytics and/or chief data officer, chief information officer, chief medical officer, chief medical informatics officer, chief nursing officer, chief financial officer, and if applicable, chief clinical research officer. In today's healthcare industry, these people and roles represent the vast majority of data in the environment, as both consumers of data analytics and producers of data content. In rare cases, data governance committees in healthcare also include patient representatives; this will become more common in the future.
The chief analytics/chief data officer is a new role in healthcare that will become progressively important as the focus on accountable care grows and healthcare organizations realize the value of data in improving the quality of care they deliver and the role that data can play in reducing waste and improving financial margins. People in this position will be business and data generalists, capable of understanding and supporting the analytic needs of all C-levels computers using Excel and Access. White-space data management tools replace the need for these desktop spreadsheets and databases by providing an easy-to-use data entry tool that is tightly coupled with the EDW. The white-space data entered through this tool is then naturally integrated with other analytically valuable data in the data warehouse. It is also more secure than Excel spreadsheets or Access databases.

**Metadata Repository**

Finally, one of the most important tools to data governance is a metadata repository. The metadata repository serves as the “Yellow Pages” for the EDW, providing data analysts and members of the data governance committee with a tool for browsing the various types of data in the EDW and seeing the attributes of that data. This includes information such as how far back in history the data has been collected, the number of records in the data set, any known data quality problems, and the data stewards who can be contacted for more information. A metadata repository is critical to the democratization and full utilization of the data in an EDW.

Misguided metadata strategies place too much emphasis on the objective computable metadata that can be collected automatically and too little emphasis on the more important subjective metadata that can only be provided by human beings who have been collecting and managing data for a number of years and understand its nuances in ways that a computerized tool cannot. The best metadata repositories contain information that is a combination of human-generated content and computer-generated content in a 50/50 split. The human generated content should be collected and curated in a wiki-style contribution model. The computer-aided metadata should be collected through the database management system and ETL (extract, transformation, and loading) tools.

**HEALTHCARE ANALYTICS ADOPTION MODEL**

Modeled after the HIMSS Analytics EMR Adoption Model, the Healthcare Analytics Adoption Model provides a framework for evaluating an organization’s adoption of analytics. It also provides a roadmap for developing analytics strategies, both for vendors and for internal use by healthcare delivery organizations.
The progressive development of an analytics strategy in healthcare starts tribal-like at Level 0, where little or no data governance exists in the organization, all the way up to Level 8, at which the data governance committee is operating in a very deliberate and formal manner, driving the strategic acquisition and exploitation of data in the organization (e.g., genomics data initiatives).

**Progression of Data Governance in the Model**

Several patterns emerge as organizations progress through the analytics adoption model. First, data content expands with each progressive level as new sources of data are added to the EDW. Second, the timeliness of data refresh and data-driven decision-making increases, as the organization becomes more agile and comfortable operating as a data-driven culture. The data literacy of the organization increases progressively as well, just as might be expected of students in the progressive completion of a college curriculum. Finally, the complexity of the analytics employed by the organization increases, and the mission of the data governance committee expands from the governance of data to the governance of algorithms and rules about binding their data.

There are six distinct phases that the data governance committee will pass through, as the organization progresses through the Analytics Adoption Model.

**Six Progressive Phases of Data Governance**

Phase 1 for the data governance committee is relatively simple, but critically important, as it sets the stage and foundation for all other phases and progression. In this phase, the executives on the data governance committee are setting the tone for becoming a data-driven organization. In staff meetings, in emails, in their decision-making behavior, in their priorities of projects, they constantly reinforce the importance of using data to make better decisions, faster. They communicate to their staff that all employees who play a part in data collection and origination are responsible for data quality—i.e., the collection of valid and complete data that they capture in the course of their duties.
Phase 2 involves knocking down technical and cultural barriers that stand in the way of access to data. Typically, the most visible sign of this phase is the establishment of an EDW. While the data warehouse is being built or procured, the data governance committee can work in parallel on the establishment of the principles, policies, and procedures for gaining access to the EDW and maximizing its value to the organization. Key to deriving that value is the establishment of a data stewardship program.

In Phase 3, the data governance committee identifies data stewards who can function as domain experts in specific data content areas of the organization. Data stewards are typically from the director or manager level of the organization, and have leadership responsibility for staff who originate, collect, and enter data into the organization’s primary information systems such as registration, the EMR, laboratory information system, and revenue cycle system. In the early lifecycle of a healthcare EDW, approximately eight to 10 data stewards will play a critical role, and thus should be identified early.

Shown in the chart is an excerpt from a data steward policy written for a large academic medical center. It identifies the key source systems and data stewards required in the initial development of the organization’s data warehouse. Over time, as the EDW matures and the data content grows, additional source data information systems and data stewards will be assigned. It is not unusual for a mature data warehouse environment, one that has been in operation for more than five years, to be associated with up to 40 different source information systems and data stewards. For this particular academic medical center, its EDW became operational in 2006. As of October 2014, there are over 100 different source data systems feeding their EDW.

The previous phases established cultural tone for data; defined the processes and technology to enable greater access to data; and established data

<table>
<thead>
<tr>
<th>Source Data Information System</th>
<th>Responsible Data Steward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Identifiable Data/PHI</td>
<td>Patient Identifiable Data/PHI</td>
</tr>
<tr>
<td>Patient Registration</td>
<td>Director of Ambulatory Operations</td>
</tr>
<tr>
<td>Patient Scheduling</td>
<td>Director of Ambulatory Operations</td>
</tr>
<tr>
<td>Electronic Medical Record - Physicians</td>
<td>Chief Medical Information Officer</td>
</tr>
<tr>
<td>Electronic Medical Record - Nurses</td>
<td>Chief Nursing Informatics Officer</td>
</tr>
<tr>
<td>General Ledger</td>
<td>CFO</td>
</tr>
<tr>
<td>Accounts Payable</td>
<td>Director of Finance</td>
</tr>
<tr>
<td>Payroll</td>
<td>Director of Payroll</td>
</tr>
<tr>
<td>Billing and Accounts Receivable</td>
<td>Director of Business Operations</td>
</tr>
<tr>
<td>Ancillary Departmental Systems</td>
<td>Affiliated Department Administrator</td>
</tr>
</tbody>
</table>
stewards that could support data quality issues and serve as data domain experts in their areas. Phase 4 shifts to data quality. All of the cultural themes, people, and technological tools have been established to address the inevitable data quality concerns that will arise as a consequence a greater access and visibility to data.

In Phase 5, the data governance committee can shift its attention to the full exploitation and utilization of the data and the data management processes that surround it. The organization's board and executive teams will now have access to data and processes that were not available previously. Developing internal management dashboards will be less labor-intensive and more accurate. External reporting to government and professional organizations will also improve and become more consistent and less error-prone. Clinical and financial process improvement initiatives will begin to thrive. Contract negotiations with payers will become more data-driven and transparent. Opportunities to identify wasteful care processes and supply chain management issues will emerge. Research grants will be easier to submit and complete, and high-quality researchers will be easier to recruit, as they are attracted to the availability of high-quality data.

Finally, in Phase 6 of the data governance committee’s evolution, members will be operating at a very strategic level with the deliberate development of roadmaps to acquire the data necessary to achieve accountable care and population health in the truest sense. In the future of healthcare, mergers, acquisitions, and partnerships will be characterized as much by the acquisition of critical data as by people and facilities. The geographic shift from delivering care in hospitals and clinics, to people’s homes and workplaces, will also move healthcare toward becoming a digital knowledge delivery industry. Data governance committees will play a critical role in the procurement of data and information systems that add content to the data warehouse and facilitate new forms of healthcare. The committee members will play a part in the acquisition of activity-based cost accounting systems, genomic information systems, patient-reported outcomes systems, 24 x 7 biometric monitoring systems

**Six Phases of Data Governance**

You need to move through these phases in no more than two years

**Phase 6: Acquisition of Data**

**Phase 5: Utilization of Data**

**Phase 4: Quality of Data**

**Phase 3: Stewardship of Data**

**Phase 2: Access to Data**

**Phase 1: Cultural Tone of “Data Driven”**
for the home, and patient social interaction systems—all motivated by the value of data that they contribute to the analytics of the organization.

WHAT IS THE DATA GOVERNANCE COMMITTEE GOVERNING?

Shown in the graphic are 16 different categories of data that require some form of governance. There are undoubtedly more in today’s healthcare environment and more are coming. But at the present time, these are the categories of data that most organizations should be concerned about acquiring, managing, and integrating into the EDW. Organizations should address the first six of these data types immediately. These six core data sources are: billing data, lab data, imaging data, inpatient EMR data, outpatient EMR data, and claims data.

The next six, data types 7 through 12, should be on every organization’s strategic data acquisition roadmap for acquisition within the next one to two years. These include health information exchange (HIE) data, bedside monitoring data, external pharmacy data, familial data, and home monitoring data.

Over the next two to four years, the data governance committee should sponsor the acquisition of patient-reported outcomes data, long-term care facility data, genomics data, and 24 x 7 biometric monitoring data.

A significant amount of money and time have been invested in our country’s acquisition of EMRs and HIEs. Many healthcare executives exhibit a sense of achievement, and in some cases, fatigue, as a consequence of these acquisitions and projects. However, EMRs and HIEs are just the beginning of healthcare’s digital transformation. Now is not the time to celebrate false summits or plan for a break in activity; the real climb and summits are still ahead.
**MASTER DATA MANAGEMENT**

*Master data management* is comprised of processes, governance, policies, standards, and tools that consistently define and manage the critical data of an organization to provide a single point of reference. Tangibly, that means codes such as ICDs, CPTs, SNOMED, LOINC, RxNorm, facility and department codes, master patient identifiers, cost accounting codes, and master clinician identifiers are standardized and sequestered in a data warehouse where they can be referenced and utilized.

There are no organizations in U.S. healthcare that have a perfect adherence to local, regional, national, or international master reference standards throughout the organization. Master data management is a long-term journey in healthcare, with no end. Adherence to master reference data can seem overwhelming, and it will become so if not approached carefully, in small and manageable objectives. Establishment of master patient identifiers and master physician identifiers are the two most critical forms of master data management. Without those two identifiers, healthcare analytics is left to very broad, population-based use cases, not specific clinical outcomes transformation. The transition to ICD-10 and annual updates to CPT codes are other examples of master data management. Mapping an organization’s SNOMED, LOINC, and RxNorm data for submission to federal agencies such as CMS and NLM is also a form of master data management.

The data governance committee’s role in master data management is, once again, at the executive level. Rarely will the committee members engage in the details of master data management, unless it requires resolution of a priority or dispute. The data governance committee will endorse and support the concept of a master data management function and a master reference area to support that function, in the EDW, but they will leave the details of implementation up to a subcommittee and the data stewards who are most affected by master data. CIOs, chief data officers, and chief analytics officers can play a very hands-on role in the development and implementation of a master data management program.

While standard codes and vocabularies are critical to a master data management strategy, they also might be the easiest to manage. The more complicated aspect of master data management is related to the standardization and management of the algorithms and rules that bind data together.
DATA BINDING AND DATA GOVERNANCE

Data binding goes hand-in-hand with good data governance. For example, two sets of data composed of the numbers 115 and 60, by themselves, are meaningless pieces of data. They are simply numbers. The numbers themselves do not mean anything until they are “bound” through analytic software, to a vocabulary and the clinical or business rules about that data. In the example, as seen in the diagram below, the values 115 and 60 are bound to the vocabulary “systolic” and “diastolic” and “blood pressure.” With that vocabulary binding, there is meaningful context for the numbers; now the context must be further bound to another form of context that represents knowledge and understanding of the data in a clinical sense—Are those blood pressure numbers low, normal, or high? More specifically to chronic disease management, does the patient with whom these numbers are associated exhibit signs of hypertension? Furthermore, the definition of “hypertensive patient” requires yet another rule for data binding, such as “three concurrent readings of high over a period of one month.”

Knowing when to bind data, and how tightly, to vocabularies and rules is critical to analytic success and agility. Data governance committees in healthcare will play an increasingly important role in facilitating the definition of these clinical rules that bind to data. As an industry, healthcare is amazingly immature and nonstandard in these definitions. There is significant variability across organizations, and even within organizations, in the data definition of common disease states such as hypertension and diabetes. Data governance committees will encourage, stimulate, and identify these standard data binding rules by catalyzing comprehensive and persistent agreement about those rules, within their organizations and within the industry. When comprehensive and persistent agreement is achieved, the data governance committee should act quickly to establish that rule or rules as a standard and ensure that the analytic software programming of the data binding follows that standard.
VOCABULARY: WHERE TO START?

Data governance committees are frequently overwhelmed by the notion of master data management. The easiest place to start the master data management journey is in vocabulary management. In today's data ecosystem, only 19 vocabulary elements constitute 80 percent of the analytic use cases in healthcare. The data governance committee and data stewards should focus their master data management efforts on these 19 vocabulary areas first.

Master patient identifiers and master provider identifiers are among the most important. In the absence of those two identifiers, almost all downstream analytic use cases will suffer or be impossible to achieve. These vocabulary data elements also represent a significant overlap of data elements among the primary source information systems in healthcare today. Therefore, resolving inconsistencies among these vocabularies will greatly facilitate the ability to integrate and analyze this data in the EDW. It is important to note that the organization does not necessarily have to impose vocabulary standards on the primary source systems, initially. That can occur more slowly and over a longer period of time because it involves significant changes to those primary source systems, most of which will be constrained by commercial vendors. The best alternative initially, is to utilize tools and methods within the EDW to map and enforce these vocabulary standards.

Where to Start, Clinically?

Identifying where to start clinically is considerably more complicated than where to start from a vocabulary perspective. The seven criteria for choosing a clinical data opportunity are the following:

1. Comprehensive agreement about the data definition of the disease state. An organization cannot improve what it cannot measure, and it cannot measure what it cannot agree upon measuring.

2. Persistent agreement about the data definition of the disease state. It is not enough to agree on a comprehensive and widespread basis about the data definition of the disease state. The organization must also agree persistently. The agreement cannot be volatile.

Vocabulary: Where Do We Start?

In today’s environment, about 20 data elements represent 80-90% of analytic use cases. This will grow over time, but right now, it’s fairly simple.

- Charge code
- CPT code
- Date & Time
- DRG code
- Drug code
- Employee ID
- Employer ID
- Encounter ID
- Gender
- ICD diagnosis code
- ICD procedure code
- Department ID
- Facility ID
- Lab code
- Patient type
- Patient/member ID
- Payer/carrier ID
- Postal code
- Provider ID

Source data vocabulary Z (e.g., EMR)

Source data vocabulary Y (e.g., Claims)

Source data vocabulary X (e.g., Rx)
The availability of data in the organization's data ecosystem to support the analysis of the disease state and the care processes around it. It makes no sense to target the improvement of a disease state or care process if the organization does not have the data to analyze the problem.

Significantly meaningful patient volumes associated with the disease state. Unfortunately and realistically, an organization cannot afford to spend scarce resources on small volumes of patients unless those small volumes represent a significant overall cost to society, such as ALS.

Significantly meaningful costs associated with treating the disease state. Fortunately, there is a general correlation between the severity of the disease or condition and the cost for treating it. Focusing data management and process improvement activities on high-cost diseases and conditions is common sense.

Significant cultural interest in addressing improvements in either quality of care or cost of care, or both, for patients in the disease state. It makes no sense to undertake a complex data management activity and process improvement project if there is no cultural inspiration to do so.

The ability to clinically intervene and show improvement for patients in that disease state. Once again, it makes no sense to undertake a complex project if there is little or no means technically, geographically, economically or otherwise to intervene for the purposes of improving the quality of care or cost of care associated with the disease or condition.

My empirical observations over the past 17 years indicate that the following areas of healthcare delivery seem to consistently meet the criteria listed above.

- CAUTI
- CLABSI
- Pregnancy management, elective induction
- Discharge medication adherence for MI/CHF
- Prophylactic pre-surgical antibiotics
- Materials management, supply chain
- Glucose management in the ICU
- Knee and hip replacement
- Gastroenterology patient management
- Spine surgery patient management
- Heart failure and ischemic patient management
It is worth noting that the conditions and disease states listed above are heavily weighted towards inpatient, acute care processes. This is probably due to the relative difficulty in intervening in chronic condition diseases because so much of what affects those interventions lie outside the organization’s boundaries and scope of influence; those conditions are significantly affected by lifestyle and other socio-economic factors. Whereas, in an inpatient setting, the organization, in theory, has nearly complete control over the environment that surrounds the patient. Also, the data in a chronic-condition, outpatient-focused environment is very sparse in comparison to the data compiled during an inpatient encounter. Accurate and meaningful analytics depend on deep and voluminous data.

The diagram depicts this relationship between the degree of inpatient-centered care; and the availability of data and the opportunity for influence and intervention on a patient’s disease or condition.

DATA GOVERNANCE: IN CONCLUSION

Healthcare is on the brink of becoming a truly digital, knowledge delivery industry. The geography of care is shifting, making digital connections with patients more and more a requirement instead of an option. Five-star hotel-hospitals that encourage a long length of stay are, largely, a thing of the past. These hospitals are already seeing double-digit decreases in admission rates in those areas served by ACOs. The economics of healthcare has reached a tipping point. The industry can no longer operate with such high waste; various and numerous studies reveal waste levels of at least 30 percent to as high as 60 percent. The integration and analysis of data will play a critical role in helping healthcare organizations maintain a strong financial balance sheet while also improving the quality of care and health for their patients.

The Triple Aim of Data Governance is found in the constant and simultaneous management of data quality, data literacy, and data exploitation. Data governance committees need to avoid the most common failure modes: wandering, technical overkill, political infighting, and bureaucratic red tape.
Healthcare organizations that are undergoing analytics adoption will also go through six phases of data governance:

1. Establishing the tone for becoming a data-driven organization
2. Providing access to data
3. Establishing data stewards
4. Establishing a data quality program
5. Exploiting data for the benefit of the organization
6. The strategic acquisition of data to benefit the organization

As U.S. healthcare moves into its next stage of evolution, the organizations that will survive and thrive will be those who most effectively acquire, analyze, and utilize their data to its fullest extent. Such is the mission of data governance.

RESOURCES

http://www.healthcatalyst.com/5-reasons-healthcare-data-is-difficult-to-measure
http://weigend.com/files/speaking/Weigend_WorldMarketingForum_MEX_2013.06.27.pdf
http://www.healthcatalyst.com/population-health/
http://www.healthcatalyst.com/late-binding-data-warehouse-platform
http://www.healthcatalyst.com/late-binding-data-warehouse-explained/
http://www.ihi.org/Engage/Initiatives/TripleAim/Pages/default.aspx
http://www.healthcatalyst.com/aco/
http://www.healthcatalyst.com/healthcare-analytics-adoption-model/

Dale Sanders, Senior Vice President

Prior to his work in the healthcare industry, Dale Sanders worked for 14 years in the military, national intelligence and manufacturing sectors, specializing in analytics and decision support. In addition to his role at Health Catalyst, Dale served as the senior technology advisor and CIO for the National Health System in the Cayman Islands. Previously, he was CIO of Northwestern University Medical Center and regional director of Medical Informatics at Intermountain, where he served in a number of capacities, including chief architect of Intermountain’s enterprise data warehouse. He is a founder of the Healthcare Data Warehousing Association. He holds Bachelor of Science degrees in Chemistry and Biology from Fort Lewis College and is a graduate of the U.S. Air Force Information Systems Engineering Program.
ABOUT HEALTH CATALYST

Health Catalyst is a mission-driven data warehousing and analytics company that helps healthcare organizations of all sizes perform the clinical, financial, and operational reporting and analysis needed for population health and accountable care. Our proven enterprise data warehouse (EDW) and analytics platform helps improve quality, add efficiency and lower costs in support of more than 30 million patients for organizations ranging from the largest US health system to forward-thinking physician practices. Faster and more agile than data warehouses from other industries, the Health Catalyst Late-Binding™ EDW has been heralded by KLAS as a "newer and more effective way to approach EDW."

For more information, visit www.healthcatalyst.com, and follow us on Twitter, LinkedIn, Google+ and Facebook.