

5 Reasons Healthcare Data Is Unique and Difficult to Measure

By Dan LeSueur

Healthcare data tends to reside in multiple places. From different source systems, like EMRs or HR software, to different departments, like radiology or pharmacy. The data comes from all over the organization.

Those of us who work with data tend to think in very structured, linear terms. We like B to follow A and C to follow B, not just some of the time, but all the time. **Healthcare data** isn't that way. It's both diverse and complex making linear analysis useless.

There are several characteristics of healthcare data that make it unique. Here are five, in particular:

1. Much of the data is in multiple places.

Healthcare data tends to reside in multiple places. From different source systems, like EMRs or HR software, to different departments, like radiology or pharmacy. The data comes from all over the organization. Aggregating this data into a single, central system, such as an enterprise data warehouse (EDW), makes this data accessible and actionable.

Healthcare data also occurs in different formats (e.g., text, numeric, paper, digital, pictures, videos, multimedia, etc.). Radiology uses images, old medical records exist in paper format, and today's EMRs can hold hundreds of rows of textual and numerical data.

Sometimes the same data exists in different systems and in different formats. Such is the case with claims data versus clinical data. A patient's broken arm looks like an image in the medical record, but appears as ICD-9 code 813.8 in the claims data.

And it looks like the future holds even more sources of data, like patient-generated tracking from devices like fitness monitors and blood pressure sensors.

Electronic medical record software has provided a platform for consistent data capture, but the reality is data capture is anything but consistent. For years, documenting clinical facts and findings on paper has trained an industry to capture data in whatever way is most convenient for the care provider with little regard for how this data could eventually be aggregated and analyzed.

2. The data is structured and unstructured.

Electronic medical record software has provided a platform for consistent data capture, but the reality is data capture is anything but consistent. For years, documenting clinical facts and findings on paper has trained an industry to capture data in whatever way is most convenient for the care provider with little regard for how this data could eventually be aggregated and analyzed. EMRs attempt to standardize the data capture process, but care providers are reluctant to adopt a one-size-fits-all approach to documentation. Thus, unstructured data capture is often allowed to appease the frustrated EMR users and avoid hindering the care delivery process. As a result, much of the data captured in this manner is difficult to aggregate and analyze in any consistent manner. As EMR products improve, as users become trained to standard workflows, and as care providers become more accustomed to entering data in structured fields as designed, we will have more and better data for analytics.

An example of the above phenomenon is found in a recent initiative to reduce unnecessary C-sections at a large health system in the Northwest. The first task for the team was to understand how the indications for C-section were documented in the EMR. It turned out that there were only two options to choose from: 1) fetal indication and 2) maternal indication. Because these were the only two options, delivering clinicians would often choose to document the true indication for C-section in a free text form, while others did not document it at all. Well, this was not conducive to understanding the root cause of unnecessary C-sections. So, the team worked with an analyst to modify the list of available options in the EMR so that more detail could be added. After making this slight modification to the data capture process, the team gained tremendous insight, and identified opportunities to standardize care delivery and reduce unnecessary C-sections.

3. Inconsistent/variable definitions; Evidence-based practice and new research is coming out every day.

Oftentimes, healthcare data can have inconsistent or variable definitions. For example, one group of clinicians may define a cohort of asthmatic patients differently than another group of clinicians. Ask two clinicians what criteria are necessary to identify someone as a diabetic and you may get three different answers. There may just not be a level of consensus about a particular treatment or cohort definition.

“A different approach is needed that can handle the multiple sources, the structured and unstructured data, the inconsistency, the variability, and the complexity within an ever-changing regulatory environment.”

Also, even when there is consensus, the consenting experts are constantly discovering newly agreed-upon knowledge. As we learn more about how the body works, our understanding continues to change of what is important, what to measure, how and when to measure it, and the goals to target. For example, this year most clinicians agree that a diabetes diagnosis is an Hg A1c value above 7, but next year it's possible the agreement will be something different.

There are best practices established in the industry, but there's always ongoing discussion in the way those things are defined. Which means you're trying to create order out of chaos and hit a target that's not only moving, but seems to be moving in a way you can't predict.

4. The data is complex.

Claims data has been around for years and thus it has been standardized and scrubbed. But this type of data is incomplete. Clinical data from sources like EMRs give a more complete picture of the patient's story.

While developing standard processes that improve quality is one of the goals in healthcare, the number of data variables involved makes it far more challenging. You're not working with a finite number of identical parts to create identical outcomes. Instead, you're looking at an amalgam of individual systems that are so complex we don't even begin to profess we understand how they work together (that is to say, the human body). Managing the data related to each of those systems (which is often being captured in disparate applications), and turning it into something usable across a population, requires a far more sophisticated set of tools than is needed for other industries like manufacturing.

5. Changing Regulatory Requirements.

Regulatory and reporting requirements also continue to increase and evolve. CMS needs quality reports around measures like readmissions, and healthcare reform means more transparent quality and pricing information for the public. The shift to value-based purchasing models will only add to the reporting burden for healthcare organizations.

“

Healthcare is not like those industries where business rules and definitions are fixed for long periods of time. The volatility of healthcare data means a rule set today may not be a best practice tomorrow.

”

Healthcare Data Will Only Get More Complex

Healthcare data will not get simpler in the future. If anything, this list will grow. Healthcare faces unique challenges and with that comes unique data challenges.

Because healthcare data is so uniquely complex, it's clear that traditional approaches to managing data will not work in healthcare. A different approach is needed that can handle the multiple sources, the structured and unstructured data, the inconsistency, the variability, and the complexity within an ever-changing regulatory environment. The solution for this unpredictable change and complexity is an agile approach, tuned for healthcare. As with a professional athlete, the ability to change directions on a dime when the environment around you is in constant flux is a valuable attribute to have. If I start out from point A in direct route to point B and the location of point B suddenly changes or an obstacle arises, I certainly wouldn't want to have to retrace my steps back to point A, redefine my coordinates, and set off on the new course. Rather, I need to take one step at a time, reevaluate, and pivot in flight when necessary.

Agility Compensates for Complexity and Uncertainty

Those are the core issues with healthcare data, and they are very real. Understanding that, and the fact that some of those issues will never change, the question becomes how you work within those limitations to deliver better information to those who need it.

The generally accepted method of aggregating data from disparate source systems so it can be analyzed is to create an enterprise data warehouse (EDW). It is a method common across many industries. Just as a physical warehouse is used to store all sorts of goods in bulk until they're needed, an EDW houses data from across the enterprise in a single place.

Yet how you aggregate that data can have a huge impact on your ability to gain maximum value from it. The early-binding methods that are prevalent in manufacturing, retail, and financial services don't work very well in healthcare, because they depend on making business rule decisions before you know what you want to do with it. It would be expensive to warehouse goods with the thought in mind that you would store everything you could ever want in the future. So you're paying for all the storage space and the overhead that comes along with it. But you're not using it.

“Late-Binding™ allows you to aggregate data quickly and develop business rules on the fly so users can develop hypotheses, use the data to prove them right or wrong, and continue the discovery process until they are able to make scientific, evidence-based decisions.”

Traditionally other industries look ahead at what business questions they'll want to answer. They know exactly what information they'll need. Their data warehouses, then, store everything they need in the way that they need it.

Healthcare is not like those industries where business rules and definitions are fixed for long periods of time. The volatility of healthcare data means a rule set today may not be a best practice tomorrow. The industry is filled with instances of EDW projects that never deliver results or even come close to completion because the rules and definitions keep changing.

A better approach is to use a [Late-Binding™ Data Warehouse](#). With this schema, data is brought into the EDW from the source applications as-is, and placed into a source data mart. When you need to turn it into information, it is then transformed into exactly what the analysis requires. If there is a change to the business rules or definitions, such as what constitutes an at-risk patient, that change can be applied within the application data mart rather than having to transform and reload all the data from the source.

That is how Late-Binding™ supports the discovery process so important to healthcare. When frontline business users enter into a clinical analysis of the data, you want them to start free of any pre-conceived data models.

Late-Binding™ allows you to aggregate data quickly and develop business rules on the fly so users can develop hypotheses, use the data to prove them right or wrong, and continue the discovery process until they are able to make scientific, evidence-based decisions.

Resources

- Healthcare Analytics Adoption Model <http://www.healthcatalyst.com/healthcare-analytics-adoption-model/>
 - Late-Binding™ Data Warehouse Platform <http://www.healthcatalyst.com/late-binding-data-warehouse-platform>
-



About the Author

Dan has been developing and implementing the core products and services of Health Catalyst since February of 2011. He started as a data architect, moved into a technical director role and is now a Vice President of Client and Technical Operations. Prior to joining Health Catalyst, Dan owned and operated a management consultancy for five years that assisted ambulatory practices in the implementation of electronic health records and data-driven management methodologies. In this venture he served as data architect, business-intelligence developer, and strategic advisor to physicians and practice owners in the strategic management and growth of their practices. Dan holds Master's degrees in Business Administration and Health-Sector Management from Arizona State University and a Bachelor of Arts degree in Economics from Brigham Young University.

ABOUT HEALTH CATALYST®

Health Catalyst is a mission-driven data warehousing, analytics, and outcomes improvement 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 50 million patients for organizations ranging from the largest US health system to forward-thinking physician practices.

For more information, visit www.healthcatalyst.com, and follow us on [Twitter](#), [LinkedIn](#), and [Facebook](#).