Bridging the Data Preparation Gap: Healthcare

Published: January 2016 **Report Number:** A0205

Analysts: James Haight, Research Analyst;
Hyoun Park, Chief Research Officer

The Case for Big Data in Healthcare

The healthcare industry accounts for over $3 trillion in annual spend just in the United States of America. Because of this, the Big Data revolution is one that presents unique promise for the healthcare industry in developing smarter data-driven insights to improve patient treatment.

A variety of external factors are expanding demand for data analysis. Regulatory mandates expanding the need for electronic health records (EHR) and the digitization of insurance claims and physician notes are creating a deluge of new data that is overwhelming physicians, nurses, and other healthcare personnel. Furthermore, the proliferation of medical and remote patient monitoring devices is creating new data sources with unprecedented specificity into patient health. Coupled with these trends are technical advances, such as the plummeting costs of genome sequencing, and new clinical trial techniques that are expanding treatment possibilities and efficacy.

AT A GLANCE

Industry

Healthcare

Core Data Challenge

Conducting medical-relevant analytics by integrated internal data sources with third-party patient data, such as patient monitoring devices and non-standardized health records.

Opportunities for Data Preparation

Reduce the technical barrier for extracting insights from unstructured and inconsistent data sources in order to lower costs and improve health outcomes.

Despite these opportunities to scale and evolve patient treatment, the healthcare industry encounters significant headwinds in leveraging data analytics. Heavy regulatory environments and compliance standards such as HIPAA (Health Insurance Portability and Accountability Act of 1996) present challenges for organizations to share and manipulate data. Data analytics fluency is now a necessity to meet compliance and outcomes expectations. This complexity, in tandem with shifting incentive structures and changing standards for practitioners, researchers, and related service providers, has created an environment where data analytics holds tremendous promise to uncover inefficiencies and personalize treatment.

The Healthcare Opportunity for Big Data

Healthcare practitioners now have access to detailed patient information about individuals, potentially even at the genetic level. Real-time feeds from monitoring devices tracking biometrics (such as heart rate), as well as the ability to understand the composition of an individual's DNA, allow treatment to be personalized for each patient. Healthcare is quickly moving from a "one size fits most" approach to one of personalized treatment based off of the distinct attributes of the individual.

Healthcare organizations can improve their ability to support patient needs by uniting disparate data sources to create unified visibility regarding how a patient's unique medical history and biological indicators fit within the

larger context of advancements, treatment possibilities, and research. This personalization means that a litany of treatments used simply to avoid potential medical payment liability can be replaced by the single most effective treatment. This shift drives improved outcomes, reduces complications, and lowers costs.

Healthcare organizations can also reduce regulatory and compliance burdens by shifting incentives and payment models to an outcome-based approach where payments are released on the basis of patient results rather than on the services provided to the patient. Using data to maximize treatment efficacy not only removes deadweight from the system from redundant or non-optimal treatments, but also presents the opportunity to save significantly on costs and to stay ahead of compliance mandates. For example, this approach can allow healthcare providers to avoid penalties for outcomes such as readmissions to a hospital within 30 days of discharge.

This level of data analytics requires a truly broad-based approach to incorporating data from a variety of sources. Hospitals, for example, must bring together operational data with an assortment of health information systems. External data sources from research of scholarly outcomes and population studies must be assessed and analyzed. Even social media, demographic, and census data come into play for those working on population-wide studies or disease prevention. Ultimately, healthcare service providers must combine practice-based data with external third-party data. Figure 1 outlines some key sources of data in the healthcare industry:

Figure 1: Key Healthcare Sources

Structured Data Sources	Unstructured Data Sources
• CRM systems • Electronic health records • ERP systems • Medical device data feeds • Proprietary health information systems • Remote patient monitoring device feeds	• Clinical trial outcomes • Insurance exchange data and payment history • Personal health records • Physician notes • Prescription history • Regulatory changes and drug approvals • Scholarly research findings • Social media feeds

Source: Blue Hill Research, January 2016

Challenges of Transforming Healthcare Data into Insight

The healthcare industry faces unique challenges in attempts to leverage data, starting with regulatory and compliance requirements. Personally-identifiable information (generally) and health records (specifically) face stringent privacy requirements. Health records cannot be altered, and all data storage tools must be

HIPAA-compliant. Restrictions in sharing and disseminating health records present a challenge for collaboration between parties, such as research initiatives between multiple institutions or geographies.

The challenges of working with healthcare data extend beyond regulatory concerns to the challenges of basic data standardization. Even "standard" data sources such as EHRs are extremely irregular due to the vast variety of formats in the industry. As a result, healthcare organizations must pursue extensive integration to optimize data for further analysis. Vital data is often noted as unstructured text from doctor's notes, medical journal findings, or prescription histories. Beyond the limitations of turning volumes of text into analyzable data, there is often no standard semantic logic between various sources. For instance, a heart attack patient may have "heart attack," "myocardial infarction," or just "MI" denoted in their EHRs, depending on the recording system or practitioner.

Remote patient monitoring devices and other devices that capture real-time patient information present fundamentally different data sources that produce *streams* of real-time and in-the-moment information. Healthcare professionals place emphasis on analyzing variations rather than on the absolute values of the data. The growing complexity of analysis also creates challenges for utilizing data, since understanding medical outcomes requires analysis of a complex web of cause and effect amidst noisy background data. Traditional multivariate regression techniques are often insufficient because they lack medical context. Instead, graph analysis demonstrating the interconnectivity of relationships and trends amongst the data is far more useful. Graph analysis permits data to be analyzed "as is" without being placed into rigid structured schemas. This mode of analysis is particularly useful when processing large amounts of text-based data.

This analysis requires next-generation data storage environments, such as Hadoop, that can handle enormous volumes of data, both structured and unstructured. Furthermore, healthcare requires advanced, predictive, and prescriptive analytics tools to perform analysis that would test traditional business intelligence offerings.

Industry Analytics Use Cases

Analytics can provide sustained value to healthcare organizations through advanced analysis techniques that help map genomic data to patient treatments, improve disease prevention, and promote better clinical research efficiency. Hospitals, in particular, have a unique opportunity to improve their operations using data analytics. Every data initiative has the potential to help hospitals address new business, support daily operations, and stay within compliance requirements. Leaders must be able to answer core questions such as:

- the rate of patients being readmitted within 30 days of discharge
- physician performance in terms of cost and outcomes,
- which patients are at high risk of future complications.

Payment and incentive models have shifted to reward practitioners for established outcomes rather than services performed. Thus, hospitals are rewarded not for *volume* of treatments but *efficacy* of treatments. Optimizing patient outcomes, maintaining high visibility into operations, and emphasizing preventative approaches are top priorities. Leading hospitals utilize data analytics to address these objectives by:

- Bringing together a patient's medical records and genomic information to personalize treatment by identifying the most impactful methods for more targeted care, improved outcomes, and lower costs
- Using remote patient monitoring devices to provide real-time alerts as leading indicators to potential problems, and to analyze historical data feeds to understanding future trends and recommended actions
- Analyzing operational data to understand emergency room patient flow, and equipment usage patterns to predict bottlenecks and better allocate staff and resources for more efficient treatment
- Analyzing EHR (Electronic Health Records) data in real time during patient treatment to identify patients at risk of readmission, and recommend further actions for high-risk patients
- Monitoring research findings and drug approvals through analysis of third-party data feeds and combining insights with patient records to identify new treatment opportunities
- Analyzing outcomes across patient demographics such as age, gender, and payment types with prescriber information such as location, education, or specialty to uncover bias and outliers

Data Preparation Opportunities for Healthcare

The myriad data sources and formats necessary to fully use Big Data challenge healthcare organizations. Legacy technologies are unable to handle the complexity of the analysis required, nor able to handle the unstructured data types associated with critical healthcare data. Organizations face a gap in the skills and manpower required to unify all of the necessary data for analysis at scale. (For more detail, read Blue Hill analyst James Haight's blog: "What U.S. Agriculture and the Big Data Skills Gap Have in Common.")

In this context, a new breed of data preparation solutions presents a compelling value proposition for overcoming the challenges in data scale and complexity. These new high-level graphical products are known as data preparation solutions, and utilize machine learning and automated recommendation algorithms to augment IT teams, data analysts, data scientists, and business analysts. Ultimately, these solutions allow data analysts to build and manage data products and transformation scripts more effectively and on demand.

Data preparation solutions today offer several major benefits to healthcare organizations:

- **Empower exploratory analytics:** Simplifying data preparation for exploratory analytics makes data useful more quickly and easily than traditional solutions. By removing legacy inefficiencies and technical requirements of analytic data prep, data preparation solutions provide a superior approach for business teams for data discovery and analytic transformation without relying on IT.
- **Increase productivity:** The most efficient data preparation systems make existing data resources more productive across the organization. These solutions present an avenue to bring volumes of unstructured text-based and inconsistent data into a cohesive format for further analysis. By reducing the time to integrate, cleanse, and prepare data, analysts can utilize more data in important analytic tasks, such as: feature construction, model construction and validation, and exploratory visualization and content augmentation.

- **Bring data analysts to raw source data on demand:** Data analysts, scientists, and business users (non-programmers) can pull data from various formats and sources directly into downstream visualization and analytics tools. As healthcare service providers demand increasing visibility into internal operations, external findings, and patient monitoring device data feeds, there is an opportunity to defray demands on IT by placing this data directly into the hands of those who will need to perform the analysis.

- **Improve organizational data usage:** New data preparation solutions accelerate the time to clean and manipulate data. Data preparation couples automated routines for anomaly/irregularity detection with visual summaries so end users can identify and fix data quality anomalies more quickly. This enables a consolidation of silos of data across health information systems and external feeds to present decision-makers with a consistent and real-time view of operations. Bringing relevant information to light in the context of the broader organization not only stands to improve operational efficiency, but to improve patient outcomes as well.

James Haight

Research Analyst

James Haight is a research analyst at Blue Hill Research, focusing on analytics and emerging enterprise technologies. His primary research includes exploring the business case development and solution assessment for data warehousing, data integration, advanced analytics, and business intelligence applications. He also hosts Blue Hill's Emerging Tech Roundup Podcast, which features interviews with industry leaders and CEOs on the forefront of a variety of emerging technologies. Prior to Blue Hill Research, James worked in Radford Consulting's Executive and Board of Director Compensation practice, specializing in the high tech and life sciences industries. Currently, he serves on the strategic advisory board of the Bentley Microfinance Group, a 501(c)(3) non-profit organization dedicated to community development through funding and consulting entrepreneurs in the Greater Boston area.

CONNECT ON SOCIAL MEDIA

@James_Haight

linkedin.com/in/jamesthaight

bluehillresearch.com/author/james-haight/

For further information or questions, please contact us:

Phone: +1 (617) 624-3400
Fax: +1 (617) 367-4210

Twitter: @BlueHillBoston
LinkedIn: linkedin.com/company/blue-hill-research
Contact Research: research@bluehillresearch.com

Blue Hill Research offers independent research and advisory services for the enterprise technology market. Our domain expertise helps **end users** procure the right technologies to optimize business outcomes, **technology vendors** design product and marketing strategy to achieve greater client value, and **private investors** to conduct due diligence and make better informed investments.

Hyoun Park

Chief Research Officer

Hyoun Park is the Chief Research Officer of Blue Hill Research, where he oversees day-to-day research operations, delivery, and methodology focused on vendor and technology selection. In addition, Park covers analytics and enterprise mobility technologies as a noted advisor, social influencer, and practitioner. Park has been named as a top 10 Big Data, analytics, and mobility influencer including quotes in USA Today, the Los Angeles Times, and a wide variety of industry media sources. Over the past 20 years, Park has been on the cutting edge of web, social, cloud, and mobile technologies in both startup and enterprise roles. Park holds a Masters of Business Administration from Boston University, and graduated with a Bachelor of Arts in Women's and Gender Studies from Amherst College.

CONNECT ON SOCIAL MEDIA

@HyounPark

linkedin.com/in/hyounpark

bluehillresearch.com/author/hyoun-park

For further information or questions, please contact us:

Phone: +1 (617) 624-3400
Fax: +1 (617) 367-4210

Twitter: @BlueHillBoston
LinkedIn: linkedin.com/company/blue-hill-research
Contact Research: research@bluehillresearch.com

Blue Hill Research offers independent research and advisory services for the enterprise technology market. Our domain expertise helps **end users** procure the right technologies to optimize business outcomes, **technology vendors** design product and marketing strategy to achieve greater client value, and **private investors** to conduct due diligence and make better informed investments.