

# Big Data, Bigger Outcomes

[Save to myBoK](#)

*By Lorraine Fernandes, RHIA; Michele O'Connor, MPA, RHIA, FAHIMA; and Victoria Weaver, RHIA*

---

Healthcare is embracing the big data movement, hoping to revolutionize HIM by distilling vast collections of data for specific analysis

---

One only needs to open a recent conference brochure, read an electronic newsletter, or preview marketing materials to appreciate that "Big Data" is getting a lot of buzz in healthcare-as well as many other sectors of the global economy. Big Data tries to make sense out of information overload, and provides new insights from the growing volumes and sources of data with the goal of answering business, operational, and clinical questions in near-real time. As technology grows, the various types of data available for research grow with it. Big Data solutions aim to harness large and complex collections of digital data and extract focused knowledge and insights from it. In healthcare, experts say Big Data empowers caregivers, scientists, and management to make better decisions that have the potential to save lives, improve efficiencies, and decrease costs. Big Data also has the potential to revolutionize the way health information management (HIM) professionals collect, store, and transmit data.

"Today's episode-oriented discrete data does not allow us to be as prescriptive as we need to be in delivering better healthcare and empowering consumers," says Lisa Khorey, vice president of enterprise systems and data management, information technology at the University of Pittsburgh Medical Center. "Medicine can get closer to the action when it is prescriptive, predictive, and precise. Big Data allows organizations to focus on wellness and standardize care processes."

## Big Data Basics

Big Data can be defined by reviewing its basic characteristics, sometimes referred to as the 3 Vs: volume, velocity, and variety.

- **Volume** refers to the rapid rate at which data is growing. In 2020 it is estimated there will be 44 times more data than in 2009-35 zettabytes compared to 800,000 petabytes. Big Data techniques and software work to manage large data blocks and make sense of the information.
- **Velocity** represents the increasing frequency with which data is delivered. Data such as social media, monitoring and sensing devices, and embedded chips- now in every imaginable device from refrigerators and airplanes to bodily implants-all add to the growing mounds of available data.
- **Variety** signifies the many forms in which data exists. In healthcare this includes unstructured data in text format, scanned documents, streams of data from monitoring devices, email or text messages, and audio and video from images and procedures that add to the wide variety of existing structured healthcare data.

The intrigue of Big Data technologies in many industries, including healthcare, is its promise to transform how an industry operates. Scott Schumacher, PhD, IBM chief scientist and distinguished engineer, says these technologies can allow physicians to have predictive analytics that can lead to both long-term and immediate care decisions.

"Technologies aimed at the first V, volume, support the analysis of the large quantity of data required for meaningful statistics and finer grained personalization," Schumacher says. "The second V, velocity, delivers the transformational promise of Big Data through predictive analytics tied to real-time measurements.

"The third V, variety, leverages natural language processing, semantic normalization using standard ontologies, and image and video extraction to bring more and varied evidence into analytic systems."

# How Big Data Helps Healthcare

Big Data has tremendous potential to add value in all healthcare settings. Big Data solutions can help organizations personalize care, engage patients, reduce variability and costs, and improve quality. Once Big Data is managed and integrated, organizations can apply analytics to better understand the clinical and operational states of their business based on historical and current trends, and predict what might occur in the future with a trusted level of reliability.

Personalization, whether based on genomic data, standard test data, or a combination of the two, requires the integration and analysis of much larger volumes of data than is used today, Khorey says.

"Big Data provides a rich context to shape many areas of healthcare, especially genomics where massive amounts of data are required and costs are rapidly decreasing," she says.

While these technologies center on vast collections of data, they can also be used for select and specific analysis. For example, Big Data can be used to define patient populations at a level of granularity previously unobtainable, according to Dr. Richard Tayrien, DO, FACOL, chief health information officer for the Hospital Corporation of America. By referencing a patient to a cohort of several million similar patients, aligned by hundreds of clinical features and modeled through numerous therapeutic pathways, Big Data tools can be used to find outcomes that are predicted with a high degree of sensitivity and specificity, Tayrien says.

"Big Data solutions can result in personalized medicine that makes a dramatic difference by redirecting the care of a patient toward the most favorable outcome before predictably sustaining an adverse clinical event," he says.

Big Data solutions will benefit healthcare providers, payers, research, and government organizations. The following is an overview of what Big Data delivers for each of these sections of the healthcare industry.

## Providers Get Patient-Specific Best Practices

Healthcare providers have massive amounts of unstructured data in the form of images, scanned documents, and encounter or progress notes. Big Data solutions enable providers to analyze unstructured data in its native state, integrate it with structured data, and address priorities based on their findings. Priorities may include care pattern identification that aids in process modifications; predictive identification of risk factors to avoid never or sentinel events and untoward outcomes; and comparisons of images, procedures, and surgeries to improve education, research, and care.

Kristen Wilson-Jones, vice president of data and online services for Sutter Health, describes Big Data as a means for provider organizations to apply "mass personalization" principles to healthcare in ways similar to those used in consumer product design and manufacturing.

"Big Data will allow traditional claims and procedure data to be integrated with data created outside of healthcare to break down artificial barriers between healthcare settings," Wilson-Jones says. "For example, data from grocery store purchases, social media, and personal preferences can be integrated to better understand what impacts individual and population health."

These new insights can improve health at many levels, Wilson-Jones feels. With Big Data, best practices are more readily identified, variability decreases, and costs and quality are enhanced by providers, delivering a truly personalized patient experience.

## Payers Leverage Data Pool

Payers have massive amounts of claims data they would like to harness to provide insights that improve wellness, patient compliance, fraud detection, and enable early warning to negative patient trends. Whether they are private payers or the government, payers increasingly use incentive programs to reward better outcomes while controlling costs. Many also want to utilize social media as a wellness and patient intervention tool that drives lifestyle changes, improves care, and reduces costs. Big Data solutions enable payers to integrate high volumes of different varieties and sources of data to enable these diverse initiatives.

## Research Enabled with Unprecedented Reach

Research that requires the integration of large amounts of data has historically been underserved due to computational limitations. With Big Data solutions, researchers can contextually integrate and correlate large amounts of information automatically to gain faster insights.

For example, the State University of New York (SUNY) at Buffalo has deployed a Big Data solution to better understand the complex causes of multiple sclerosis. The system combines and analyzes variables such as diet, exercise, living, and working conditions, as well as clinical and genetic data. This approach used to take days of computing time, but now takes minutes due to the advanced computing power of today's systems.

"Big Data allows us to take our research to a new level," says Dr. Murali Ramanathan, PhD, lead researcher at SUNY Buffalo. "We can now rapidly analyze larger data sets including thousands of genetic variations, many environmental factors, and the interaction between them to gain valuable new insights that weren't possible before."

## Benefit of Using Vast Government Data Stores

Government organizations may be the biggest beneficiary of Big Data solutions. Organizations already have vast stores of data sitting in data warehouse silos. With Big Data solutions these data silos can be quickly integrated to provide valuable insights such as detection of fraud and abuse patterns, identification of best practices for safer and more efficient care delivery, and better epidemiology surveillance.

"Proceeding with the implementation of Big Data healthcare solutions requires organizations to make a cultural commitment to use data to improve quality and reduce waste," Wilson-Jones comments. "Information must be recognized as the strategic enterprise asset it is, and must be mastered and governed to break down the large number of silos and barriers in today's healthcare systems."

## Information Management Tasks Supported by Big Data



Source: IBM Corporation, 2012

# Ensuring Success for Big Data Solutions

Big Data solutions can provide significant benefits, but to ensure their successful implementation healthcare organizations need to take the following four steps:

## 1. Establish data governance, define data objectives

Before organizations implement Big Data solutions, stakeholders should convene an executive council made up of senior leadership to develop an information governance model that clearly defines Big Data objectives and expected outcomes, as well as drives Big Data initiatives.

"Data must be managed and treated as a strategic enterprise asset, and data governance or active management of the data should be vital, especially in light of Big Data," Wilson-Jones says.

An effective Big Data governance program should include the basic tenets of people, process, and policies. Specific people that should be included are data stewards, who can assist with the interpretation and use of data, and a data governance council that provides representation for key stakeholders across the organization. Special consideration must be given to the new automated processes, inferences, metrics, and monitoring tools provided by Big Data solutions. Policies and procedures will also be required that govern the use of data, define the required actions and quality control processes, and optimize, secure, and leverage information as an enterprise asset by aligning the objectives of multiple functions.

## 2. Identify data and information requirements

Once an organization has established an information governance model, its next step is to identify where all of the required data resides, what information should be gleaned from data, and how data will be leveraged to help prevent adverse situations, improve care, and keep patients healthy. Most structured healthcare data, estimated to make up 20 percent of all data in a healthcare facility, resides in automated systems such as the hospital information system, the radiology information system, laboratory systems, etc. The remaining 80 percent of healthcare data consists of rich unstructured data that historically has only been leveraged using labor-intensive processes or, more commonly, has not been leveraged at all.

Big Data solutions provide healthcare organizations with the ability to access and analyze unstructured data to assist them in making more informed decisions and reducing errors and missed opportunities. However, unstructured data introduces new challenges for data stewards, specifically verifying that new information is extracted correctly (i.e., proper handling of negations such as "... tests indicate lack of evidence of ...") and that individual patient records are accurate. To properly identify and remediate errors, organizations will need to develop and deploy new data mining tools.

Organizations need to understand what data they will use today, and any potential data that they may want to access in the future. This can include data from mobile or remote devices, implanted devices, text messages and e-mails between patients and providers, and data from third parties or health information exchanges. Organizations will also need to establish a data acquisition roadmap based on business and analysis priorities.

## 3. Normalize, integrate, and organize Big Data solutions

After all data sources have been identified, a plan needs to be developed for how data will be normalized, integrated with, and organized into the Big Data solution. The plan should address technology requirements as well as business objectives, and must ensure that data are accurate and complete. Big Data solutions present even greater challenges than traditional data and analytic solutions as the volume of data is multiplied many times. The quality of many data sources accessed may have never been evaluated before.

For individually-focused analytics, most Big Data solutions require a complete view of patient and provider data. The ability to recognize relationships between patients and providers, households, payers, and organizations may also be helpful but difficult to achieve given the number of data sources. Any Big Data solution should support the systems, data, and information needs that organizations have today, but also must be configurable and flexible enough to adapt and meet future requirements.

## 4. Protect security and privacy of Big Data

Data privacy and security must also be a key component of any Big Data solution. All systems, data flows, and information lifecycles must be accounted for and the privacy of personally identifiable information protected. Organizations need to consider what types of information they expect to generate, and whether it will be individually identified or population-based. Data that are used for population-based clinical research to detect diseases or disease patterns usually masks or removes the identities of individuals before the database is populated with the clinical information. But due diligence should be taken to check if the information is de-identified before using the data.

Since healthcare Big Data solutions may use data from many different sources and be predictive and inferential in nature, there may be uncertainty within an organization about how to apply privacy and security mandates like the HIPAA requirements, the Fair Credit Reporting Act, and the Federal Trade Commission's Fair Information Practice Principles (FIPPs). The best way to address privacy concerns or requirements is for Big Data solutions to support FIPPs. FIPPs are industry-agnostic, basic information privacy principles that can guide the thorny discussions that may be required when analytic projects cross industries, data sources, and data types.

"FIPPs are a roadmap for good data stewardship and the foundation for regulations or policies understood and practiced around the globe," says Deven McGraw, JD, director of the Health Privacy Project, Center for Democracy and Technology, and member of the Office of the National Coordinator for Health IT's Health IT Policy Committee. "Since many organizations will deploy healthcare Big Data solutions that use data from outside their walls, they must be able to assure consumers that they have put the appropriate privacy practices in place and that only authorized personnel can access data."

## Big Data's HIM Opportunities

The move to Big Data solutions provides HIM professionals with significant opportunities for advancement. Those professionals who have an understanding of Big Data and know how to apply HIM principles and data management skills to Big Data implementations will have the most growth opportunities.

Big Data offers HIM professionals the chance to play a strategic role in crafting the next level of healthcare information management, and act as key stakeholders in advancing the strategic use of Big Data across the healthcare ecosystem.

As the industry transforms, it becomes essential for HIM professionals to move beyond the principles of record maintenance and documentation and develop an understanding for data transport, mapping processes, and other Big Data characteristics. Continuing education can help to expand individual knowledge and expertise in health informatics, data management, clinical vocabularies, and data standards—all important aspects of Big Data solution planning. For example, being well-versed in key concepts such as the Systematized Nomenclature of Medicine (SNOMED) classification system and the Logical Observation Identifiers Names and Codes (LOINC) can empower an HIM professional to champion the use of data across systems and facilitate interoperability.

From a broad perspective, HIM professionals should ensure that industry leadership understands the value that HIM brings to Big Data. Not only is it important for HIM professionals to get involved in Big Data planning, but they must come prepared to work with the organizational team and address data and information on a whole new level.

## References

Office of Science and Technology Policy, Executive Office of the President of the United States of America. "Obama administration unveils 'Big Data' initiative: announces \$200 million in new R&D investments." March 29, 2012. [http://www.whitehouse.gov/sites/default/files/microsites/ostp/big\\_data\\_press\\_release\\_final\\_2.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf).

US Department of Health and Human Services National Institutes of Health. "1000 Genomes Project data available on Amazon Cloud." March 29, 2012. <http://www.nih.gov/news/health/mar2012/nhgri-29.htm>.

Federal Trade Commission. "Fair Information Practice Principles." <http://www.ftc.gov/reports/privacy3/fairinfo.shtm>.

Lorraine Fernandes ([lfernand@us.ibm.com](mailto:lfernand@us.ibm.com)) is global healthcare industry ambassador and Michele O'Connor ([moconno@us.ibm.com](mailto:moconno@us.ibm.com)) is global MDM sales at IBM. Victoria Weaver ([victoria.weaver@hcahealthcare.com](mailto:victoria.weaver@hcahealthcare.com)) is assistant vice president, clinical data management at HCA.

**Article citation:**

Fernandes, Lorraine M.; O'Connor, Michele; Weaver, Victoria. "Big Data, Bigger Outcomes"  
*Journal of AHIMA* 83, no.10 (October 2012): 38-43.

---

Driving the Power of Knowledge

Copyright 2022 by The American Health Information Management Association. All Rights Reserved.