Volume 14 - Issue 3, 2014 - Matrix

Big Data

Author

Kelly Ann Callahan

Editor, Healthmanagement

Big Data: two short words with unlimited complexity, indicating that a set of information is too voluminous or varied to be processed using traditional means. Technically-speaking, it is comprised of the multi-formatted records of what everybody does these days, from every log-in to every keystroke and every save, until the sign-out process and even beyond.

The creation and saving of more and more data over the years has been facilitated by cheap storage, despite the absence of any master plan behind its accumulation. As such, Big Data is not necessarily a new idea, but the scope and scale of its potential power has ballooned into the clouds, quite literally. NSA and competing agencies have long understood its dormant power. Indeed, "Big Brother" is dwarfed by Big Data.

Big Data In Healthcare

The now-buzzworthy phrase Big Data pervades industries ranging from banking to business, from finance to forensics, from marketing to medicine. Indeed, its applications in healthcare are as unlimited as the data being generated every second by academia, clinics, insurers, patients and vendors.

As vast amounts of healthcarerelated data are digitised, organised and analysed, what may once have been the technical terrain of engineers has attracted the attention of industry leaders eager to capitalise on system improvements, cost savings and profit hikes. According to the McKinsey Global Institute, the US could save $300 billion -- that amounts to approximately $1,000 per person -- by the efficient integration and analysis of data from sources such as clinical trials and insurance transactions.

Historical Cases

The accumulation of data that are too abundant or too abstract to manipulate by hand is not a new phenomenon, but sophisticated tools did not always exist for its manipulation. Napoleon took data to the battlefields by using mathematical models to make strategic decisions. It did not always work in his favour; there is a famous map of his army's losses in 1812 and 1813 that presents a graphic analysis of information that would be otherwise difficult to envision.

Big Data is not new to healthcare, either. A cholera outbreak in London in 1854 was traced to a public well after an inquisitive doctor mapped the addresses of the 600 people who died. The fatalities clustered around a single point, a public water well, and when the handle of the pump was removed, cholera cases diminished. The story illustrates the value of looking at data not just as numbers in isolation and out of context, but finding interconnections and patterns that relate facts and behaviour.

Using Big Data to Forecast the Future

Data-driven visualisation of trends such as the ones described above can facilitate the determination of a cause for a particular effect, but it also has predictive power when structured effectively. The ability to forecast the behaviour a population of consumers has obvious value to marketers. The potential to predict the dimensions of disease or the success of therapies can save lives. Whereas the mapping of the London cholera epidemic solved a medical mystery at the level of a single neighbourhood, Big Data now supports the timely tracking of epidemics across the globe.

One notable example is Google Flu Trends. Over the past five years, Google engineers have monitored the correlation of search terms that might be used by people feeling under the weather, such as "flu symptoms" and "local pharmacies", to show locations of outbreaks. Initially, Google's number- rich but theory-free analysis was as accurate as the tracking of influenza cases by the Centers for Disease Control, but much faster.

Figure 2.
The map originally accompanied the second edition of Snow's Of the Mode of Communication of Cholera (London: John Churchill, 1854).

Snow's decisive, iconic map, showing how cholera deaths clustered around the Broad Street water pump. Using a commercial map of the Soho District, Snow stacked his black line symbols that represented individual deaths inward from the street address. This visual innovation combined an accurate location with a measure of intensity.

Big Data Challenges

But in the winter of 2012, as news outlets began reporting on the dangers of the flu, Google Flu Trends showed cases of the flu where they did not exist, according to the later CDC report. The failure was due to healthy people using the same search terms as sick people owing to the media hype about the topic, and Google not knowing the difference.

Google Flu Trends is not the only example of Big Data failures. Prior to September 11, 2001, intelligence existed in computerised databases which could have raised a flag about the imminent threat of hijacking. However, without having some direction about what to look for, it can be difficult to separate the signal from the noise.

Industry leaders and data scientists continue to grapple with how to reign in the vast volumes of information and to structure it for efficient analysis and value extraction. In healthcare, the size of data sets is not the only problem; the variety of file types (audio, video, scans, electronic health records, lab notes, clinical trial results) and related issues of compatibility are real challenges.

Like physical obesity, information overload interferes with normal function by slowing some processes, such as metabolism and movement, and preventing others altogether.

Structuring data is essential to eventually being able to extract meaning; some estimates indicate that only 10 percent of all data being captured are structured. Beyond architecture and interoperability is the challenge of how to ensure data quality. Real threats include not only tampering and fraud, but verifying the integrity of information collected haphazardly or without a purpose. The confidentiality of sensitive information and system security are major concerns for Big Data in healthcare, where breaches of either type can be costly in terms of customer confidence and financial fortitude.

Data feeds knowledge. Knowledge is power. It provides a basis for action, and as such, professionals skilled in the use of scientific methods to extract or create meaning from raw data are in great demand. Last year, the New York Times declared data science a hot field that promises to transform not only individual companies but entire industries, including healthcare.

Up Next:

HealthManagement considers Big Data a game changer, and we will look into all its different aspects in due course. The following article will delve deeper into the Five V's of Big Data: volume, velocity, variety, veracity and value. Each presents its own opportunities and obstacles for healthcare.

Big Data In Healthcare: Obstacles and Opportunities

Healthcare organisations are facing an unprecedented accumulation of data from academic research, clinical trials, electronic health records (EHR) and the proliferation of mobile health and remote monitoring devices. "Big Data" is not only transforming individual companies which capitalise on its inherent yet untapped value; without a doubt, it is changing the industry. To structure the discussion of Big Data as it relates to healthcare, it is helpful to look at five factors that contribute obstacles and opportunities: volume, velocity, variety, veracity and value.

The *size* of data sets is attributable to its multiple sources, necessitating the elasticity of storage sites. *Speed* becomes a factor as the size of data sets increases and as they are interconnected and shared across authorised user groups. This is especially relevant in healthcare where rapid access to solutions can save lives. However, quantity does not guarantee quality; data become meaningful and useful only when they are structured, verified and interpreted. One of the biggest challenges to Big Data is how to manage the variety of file types that contribute perspectives to medical research and patient profiles for optimum interoperability. Another is how to verify and protect the integrity of so much information.

There are two ways to think of Big Data as valuable: financially, it has the potential to reduce healthcare costs for patients and organisations when they are empowered with structured information. From a human perspective, there is no price to put on improved outcomes from novel medical solutions generated by the access to and sharing of information. The "big" in Big Data takes on a new meaning when it refers not only to size but to the power to save lives and resources.

Volume: Abundant But Unstructured

Big Data will continue to grow in size and importance for the healthcare industry. An ageing population is increasing the demand for services, and a shortage of care providers creates an opportunity for technology to move in and help to manage the situation for organisations. So-called closed-loop systems connect home- and hospital-based devices for biometric data exchanges. Meanwhile, academic and clinical research continually contribute data about diagnostic and therapeutic methods, all of which should be as accessible as possible to improve outcomes and prevent duplicated research efforts.

To this end, some healthcare companies are hiring data scientists to manage their internal datasets, and to link them with external sources. The associated costs only seem extravagant until they are compared with the value left on the table when the stores of study results, clinical records and claims data go unanalysed and therefore remain useless.

User-Generated Data

Big Data is not confined to institutions. People are taking charge of their health through fitness apps and the home-based tracking of vital statistics, causing an explosive generation of useful data. It is not only chronically ill or elderly patients who are becoming active participants in their own care; younger generations are now empowered by technology to prevent prevalent threats to their good health.

Sensors and Sensitivity

Already, people with chronic conditions such as diabetes can monitor their blood sugar from home. Implanted pacemakers serve cardiac patients and the physicians who treat them. Through remote monitoring, physicians are able to follow patients without office visits, all thanks to the efficient uploading of data from home-based devices. And it is not only prescribed devices that capture data: smartphone applications are changing the way healthcare is administered thanks to sensors built in to wearable technology such as the fitness-related FitBit and Jawbone. Google Glass assists surgeons by allowing them to augment operative procedures while the patient is still in the OR, and the company is now developing contact lenses with sensors to monitor diabetic blood sugar levels.

Cloud Elasticity

The infrastructure required for healthcare organisations to house and process data go well beyond the capacity of file rooms and warehouses. That is why no discussion of Big Data can avoid another buzzword: cloud computing. Essentially, the phrase refers to the storage and access of data and IT platforms not on individual hard drives but in "the cloud", a web-based space accessible via the Internet. Cloud-based data and platforms synchronised with remote sets and systems serve authorised users in ways never before imaginable. Of course, elasticity is essential as data accumulates and connections increase. According to a 2013 survey by global management consulting firm McKinsey & Company, the market for cloud computing is on a path toward generating $100 billion per year.

Velocity: Patients and Impatience

As the volume of data increases, so does the threat of system crashes from too much influx and the slow transfers of outgoing data. The speed at which users can access health data impacts user satisfaction, to be certain, but can also be a matter of life or death. The challenge for healthcare organisations in the process of capitalising on Big Data is how to allow authorised customers to quickly access or transfer the right information, in a user-friendly manner, without sacrificing safety and security.

The Power of Now

Healthcare IT workers are familiar with the frustrations of pleasing demanding doctors and industry professionals, many of whom were not trained to be tech-savvy. Despite a real lack of training, clinicians and managers increasingly rely on web-based communications, devices and platforms. To take advantage of Big Data without slowing service delivery, apps must be as user-friendly as possible. For example, the surgical app DocSpera provides members of a coordinated care team with a platform for the secure exchange of patient information. Team members can work together from remote locations; a nurse might use the platform to upload a cardiac patient's diagnostic test results upon hospital admission, which are then shared with a consulting cardiologist, who confers with a surgeon across town, while an operating room is being reserved in the appropriate facility for the necessary medical procedure.

Big Data in the pharmaceutical industry is improving the speed of care solutions. A day can feel like a lifetime for someone waiting for an effective treatment for a life-threatening illness. However, according to a recent study, 55 percent of prescribed medications do not work for patients, and the number is closer to 70 percent for cancer therapies. Considering what is now known from epigenetic research, it should not be surprising that there is not a blanket cure for common illnesses. Individual genetic expressions, turned on or off by environmental factors, will always vary from patient to patient. When healthcare becomes personalised through the combination of a patient's genetic record and known triggers in the environment, physicians will be able to use the combined data to prescribe highly individualised, and likely more effective, treatments and prevention plans.

Learning From Other Industries

Care administration is becoming ever-more portable with mobile-health apps that deliver results rapidly, but traffic can disrupt the flow of information. Fortunately, healthcare is learning something from the way the financial industry handles its Big Data. Soon, mobile health services may adopt streaming technologies like the ones which empower economic trading. The hope is that eventually, more people will benefit from a new kind of healthcare experience. Providers will be able to administer advice or care solutions immediately upon the receipt of relevant information, through the real-time analysis of patient data.

Until then, WebMD no longer has a monopoly on medical advice in the virtual world for patients seeking expedited answers to their health questions. Internet searches for symptoms and treatments have replaced phone calls to busy doctors, and clinicians themselves use social networking services such as Facebook and the healthcare-specific Doximity and SharePractice to instantly share and store experiences at the point-of-care. The quick and easy connections of communities of people with similar health complaints may not replace prescribed treatment by qualified professionals, but the immediate sense of compassion and understanding found in such forums are therapeutic in their own right. According to Mediabistro, more than 40 percent of consumers admit that social media discussions influence their personal health decisions.

Variety: From Complexity To Simplicity

The configuration of different types of data into structured platforms that convey a spectrum of information to customers is a real challenge in healthcare IT. Raw data are typically disorganised, informal or otherwise not edited for end users. Incoming files take the form of texts, audio recordings, videos, laboratory notes, email exchanges, and multi-formatted research results. The complexity increases as the types of collected data multiply. Physicians who want to understand the complete profile of a patient will one day consider institutional data as well as information from remote sources such as as wearable technology and mobile health devices.

Moreover, the integration and interoperability of systems must be assured. Global collaboration demands a more uniform formatting of data that complies with the needs of users as well as regulatory bodies. Currently, the Institute of Electrical and Electronics Engineers (IEEE) is crafting a "Guide For Cloud Portability and Interoperability Profiles" to standardise definitions, formats and interfaces of independent and incompatible cloud providers. The organisation seeks to guide the healthcare industry in mediating cloud-to-cloud data exchange by standardising things like units of measurement. As healthcare players become connected around the world, metrics may vary by region and must be kept constant if they are to have cumulative value. For example, similar but separate studies which show blood pressure results for a drug trial might include readings taken by technicians using different procedures or gauges.

The ability to visualise vast amounts of information can clarify what is otherwise hidden in a haze of numbers. Visualisation also allows things to be seen in context, which can be much more meaningful than knowing that something costs a certain amount of money per year, but not how that figure relates to an entity's overall expenditures.

Veracity: Controlling the Quality of Big Data

Inconsistency in the formats and metrics of data gathered from disparate sources is not the only challenge for meaningful analysis. If Big Data will serve healthcare as it has other industries, it will be necessary to recruit and to train data scientists to structure it in ways that permit its verification; only then do interpretations have meaning and value. At the same time, controls must be put in place to restrict access in order to protect data integrity. Some people claim that the sheer volume of data permits computer algorithms to take over the former role, but machines still lack human intuition required to ensure the latter.

Truth and Consequences

Data security is an issue in just about every industry, from banking to internet commerce to healthcare. Failures in the form of breaches can result in the loss of valuable data, but also a loss of customers as public or professional trust erodes; either way, it is an outcome no organisation can afford. Data encryption, malware protection and access restriction are only the front-line defences to the ever-present threats of tampering,

fraud and theft. The higher the quality of the data, the more valuable it becomes to authorised users and those who could profit from its unauthorised exposure.

Privacy Please

Hospital workers around the world are familiar with privacy policies put in place to protect patients from the accidental or deliberate sharing of their personal information. While this may safeguard individual patient files, the dangers loom much larger for Big Data. With so much confidential information in the cloud, it is not sufficient to hire security guards to patrol the perimeter of a warehouse. There is also controversy over the ownership of cloud-based data. Hiring a company to store and structure data might mean that they control the terms and services of access and usage, as Facebook does with its subscribers' uploaded photos. At the moment, there is no governing body for the administration of cloud computing services.

Data Integrity

Unfortunately, controlling the access to masses of complex and confidential data does little to ensure its integrity. Most people who are familiar with scientific and statistical methods know that representative samples are used in research since it is impractical and arguably unnecessary to survey every part of a whole entity. With Big Data, it is tempting to do away with statistics and theories given the size of data sets. After all, if every data point is being captured, sampling becomes obsolete.

That attitude ignores the bias inherent in the collection of internet-based data. A recent TED talk used the example of a smartphone app that tracked potholes around the city of Boston based on vibrations detected by drivers' phones and their GPS locations. Apps like RoidRage upload data to city planners who can then send teams to repair the holes. Theoretically, the city should receive all of the information it needs from the app in order to locate every pothole, but that thinking does not consider the status of streets in neighbourhoods where residents are less affluent and less likely to own cars and cell phones, and to know about the app.

For healthcare, the danger of a "sampling bias" is just as real. Studies based on populations in industrialised countries are not always representative of those in developing nations, where records are kept on paper, if they are collected at all, and oversight is scant. Without the technology to share data with other communities and to regularise the data collection process, it is difficult to know things as basic as birth and death rates in some third world countries. Fortunately, the ease and portability of mobile devices are changing the the world to allow the information that comprises Big Data to be more widely representative, and outcomes more widely applicable.

Value: Connections and Predictions

As with many things, value comes not from a stand-alone supply but from exchanges. The value of money is in its trading possibilities, and the value of Big Data in healthcare is in its interpretation and predictive power. As hospitals and healthcare-related organisations begin to order what is disordered, and to build connections between internal and external data sources, the value that currently lies dormant will transform the industry.

Telemedicine will connect 19 million people by 2018, according to Berg Insight's mHealth and Home Monitoring report. Incentives for remote patient monitoring extend from cost control to demographics: non-essential office visits are becoming obsolete. The market may be currently dominated by cardiac and sleep apnea device manufacturers and service providers, but the report projects that the pharmaceutical industry and so-called "health hubs" that centralise services with software solutions will soon enter the market space.

Thanks to the Big Data that goes into paired transplant programs such as kidney chains, compatible donors and desperate recipients are being connected, increasingly across state and national borders. In the United States alone, the number of people waiting for a kidney transplant exceeds 100,000, and 15 percent will not survive long enough to receive a compatible organ. Paired exchanges depend upon the building of broad databases which can identify and connect living donors with needy recipients, ultimately multiplying the number of successful transplants.

The Future of Big Data

Next month, The European Association of Hospital Managers (EAHM) will welcome best-selling author and economist Leo Nefiodow as the keynote speaker at its 25th EAHM Congress. The topic of Nefiodow's talk will be the sixth Kondratiev cycle, and how hospitals are transforming from growth-drivers rather than cost-drivers in Europe's economy. According to Nefiodow, the next long wave of economic and social development will be driven by the healthcare market.

Big Data is a key component in what will propel the healthcare industry to the forefront of economic growth. Its power to improve diagnostic accuracy, manage medical populations, optimise business outcomes, streamline care costs and facilitate revenue reimbursements represents an enormous potential. Challenges are abundant, too, beginning with the identification and training of qualified professionals who know to protect sensitive information and locate the signal within the noise.

Big Data is voluminous, fluid, and varied. As its integrity improves, so will its connective and predictive values for the healthcare industry.