A sponsored whitepaper.

Big Data Will Have a Big Impact on Life Sciences

Darryl Williams,
Head of Global
MDM and Platform
Solutions, Health
Care, for LexisNexis

Big Data—the combination of familiar, structured data points like facts and transactions plus unstructured data such as transitions through different health care settings and social media posts—is already transforming retail, entertainment and other industries. The transformational power of Big Data that has already changed the way in which manufacturers, sellers, marketers and buyers of retail products operate is beginning to move into the Life Sciences industry. The immediate potential is to eliminate $300 billion of annual costs in the U.S. pharmaceuticals segment alone. The longer-term potential is to dramatically rework, rearrange and reinvent the ways Life Sciences manufacturers, marketers, providers, regulators, payers, consumers and other stakeholders interact to improve the lives and health of individuals.

Darryl Williams, Head of Global MDM and Platform Solutions, Health Care, for LexisNexis Risk Solutions discussed the impact of Big Data in the Life Sciences with Fierce Markets on December 8, 2015. This white paper is an expansion of the discussion exploring how and why Big Data will have a big impact on Life Sciences in 2016 and beyond.

What is Big Data?

Big Data has become a business buzzword. And like many buzzwords, the meaning can change depending on the situation. In terms of data management, Big Data is the combination of mushrooming volumes of data from a growing variety of sources that is being collected at ever faster velocity. Big Data is a collection of new types of data, collected from novel sources, visualized in new ways that can be manipulated to gain insights and guide decision-making in real-time.

In business terms, Big Data is simply a different paradigm, a different way of conceptualizing data: what types of data are collected, how data are collected, where data are stored, who can manipulate data, when data can be transformed into intelligence and why data are becoming an essential tool to meet challenges in research, development, marketing, sales, compliance and overall profitability. Big Data means more data than ever, from more sources than ever, collected faster than ever, and manipulated in new ways to give decision makers more granular and more immediate information to guide business decisions.

Big Data is a set of circumstances and concepts that models the ways in which decision makers in all sectors of the Life Sciences universe will interact with information in the future.

Volume, Variety, Velocity

Big Data can be defined by three Vs: volume, variety and velocity.

Volume: In 2015, forbes.com noted that more data has been created in the past two years than in the entire previous history of the human race. The accumulated universe of digital data in September 2015 was 4.4 zettabytes; by 2020, that universe is projected to grow to 44 zettabytes, or 44 trillion gigabytes.

Variety: The second V is variety. Every new technology has allowed data to be collected from new sources. Words and numbers are the most familiar forms of data in the business world and can be captured in either structured or unstructured ways.

Digital tools such as Word, Excel and Access capture and store data based on attributes, often analyzing them in rows, cells and columns with assigned values and categories. Words and numbers also exist in unstructured formats such as written documents and publications. It is easy enough to

read documents, but not so easy to automate the analysis and comparison of the data contained in the documents.

Numbers add another layer of complication. Numbers usually represent measurements, typically cost, volume, quantity, speed or some other discrete value. Key performance indicators and other metrics are usually computations or formulas which take the input of raw measurements to generate a meaningful result for assessing rankings, ratings, performance and the like.

Data capture has grown far beyond familiar instruments such as computers and cell phones. The Internet of Things (IoT) creates a whole new channel for even more diverse sources of data to be collected, stored and retrieved to generate and discover meaning from new intelligence which heretofore did not exist in an effectively accessible and interactive format. Smart watches and personal sensors such as FitBits can collect data as long as they have power. And a growing number of other daily items—cars, traffic signals, surveillance cameras, automated teller machines, elevators, roadways and more—have embedded sensors collecting data.

The IoT is just one example of the expanding universe of data collection devices and data sources. Even sound has emerged as a major type of data. Music, podcasts, webinars, audio feeds and more can collect, store and disseminate both raw data and data that have been processed into information.

The growing popularity of cell phone cameras and other Internet-enabled cameras is adding to the store of visual data. Facebook, Snapchat and other applications that enable and encourage the sharing of pictures add another type of data to the growing stream. So do videos, from recorded entertainment to live events and video feeds.

Velocity: This growing variety of information is being collected at increasing velocity, the third V. More sensors, faster Internet connections and more networks of people and devices all increase the velocity of data creating and capture. It is estimated that by 2020, about 1.7 megabytes of new data will be created for every human being on the planet every second.

Velocity is simply the speed with which data can be collected and stored. Information was once stored and shared orally; then via paintings on cave walls; then by clay tablets, scrolls, and books; and now in digital form as strings of 0s and 1s. The methods used to collect, store and share

information have changed dramatically over the millennia, but one thing has not changed: every new shift in technology—from the introduction of cave paintings, the creation of written language, the development of printing, and the adoption of digital technology—has increased the sheer volume of information.

Two more Vs: Veracity and Value

Veracity: From an enterprise information management perspective, there are two other Vs which are equally worthy of note in describing Big Data. One is the veracity or trustworthiness of the data being collected. If knowledge workers do not trust the accuracy, completeness and consistency of the information they are accessing and analyzing, they will place low confidence in the conclusions and insights drawn from it, regardless of the polish and attractiveness of the visualization tools used to display it.

Value: The final V is value. Since information is a huge corporate asset, it requires diligent maintenance to keep it in working order. That means business rules, policies, processes and procedures must be established, practiced and enforced internally (with managers and consumers of the information) and externally (with suppliers and partners) to maintain the trust of users in the enterprise information assets. With proper guidance, knowledge workers will continue to drink from the information well of trusted, up-to-date, enriched market-relevant information.

Big Value

For LexisNexis, the five Vs have created a data stream of about 3.2 million bits per hour that is being managed and used by customers. That equates to 77 million transactions daily, 2.3 billion transactions annually and 28 billion different types of data being collected, stored, organized, structured, shared and manipulated every year which helps give data a voice and enables better health care decisions in real-time. As health care sector product manufacturers, service providers, insurance companies and benefit managers continue to more closely align service delivery with patient outcomes and quality of life, the need to quickly distill value from massive volumes and varieties of data to achieve organizational success and viability becomes more urgent and crucial.