Harnessing Big Data to transform clinical trials

From protocol to patient recruiting, data analytics can yield valuable insights

Imagine a multinational clinical trial in which patients in each country were selected based on genomic factors, in-depth medical histories and physician notes, and sponsors could select the best sites and Principal Investigators (PI).

Such a trial would be brought to you by Big Data, through which complex and disparate information—a person's genome, electronic hospital and medical records with doctor comments, even prescription data—is linked together to determine eligibility for a clinical trial.

This knowledge would be the result of a sponsor electronically scouring parts of the world with significant numbers of potential patients who match the protocol requirements, and where sites and PIs have been evaluated based on past performance and capabilities.

Big Data is not just volumes of data; it is a shorthand label for a combination of technology, measurement and decision making.

Although the concept of Big Data has been around for years, now more than ever before there are seemingly never-ending volumes of company-generated and third-party data available. Those data include content from social media, public web sites and data streamed 24/7 from mobile devices, sensors and technology-enabled devices. Current software and emerging sophisticated tools are used to organize, manage and analyze this data.

In addition, Big Data also can refer to the use of specialized tools and techniques to better understand and extract meaningful insights from massive amounts of information.

Said Jack Collins, director of the Advanced Biomedical Computing Center at the National Cancer Institute, Big Data is all about asking the right questions. "It's about being able

Big Data facts

At any given time, more than five billion people are calling, texting, tweeting and browsing on mobile phones worldwide.

100 terabytes of data are uploaded daily to Facebook.

Facebook stores, accesses and analyzes 30+ petabytes of user-generated data.

YouTube users upload 48 hours of new video every minute of the day.

571 new websites are created every minute of the day.

Walmart handles more than one million customer transactions every hour, which are imported into databases estimated to contain more than 2.5 petabytes of data.

Decoding the human genome originally took 10 years to process; now it can be achieved in one week.

The volume of business data worldwide, across all companies, doubles every 1.2 years.

By 2020, the number of both business-to-business and business-to-consumer transactions on the internet will reach 450 billion per day.

Estimated size of the Big Data market
Total spending in $U.S. billions on database projects of 5 petabytes or more worldwide

Year	Value
2010	$3.2
2013	$8.7
2015	$16.9
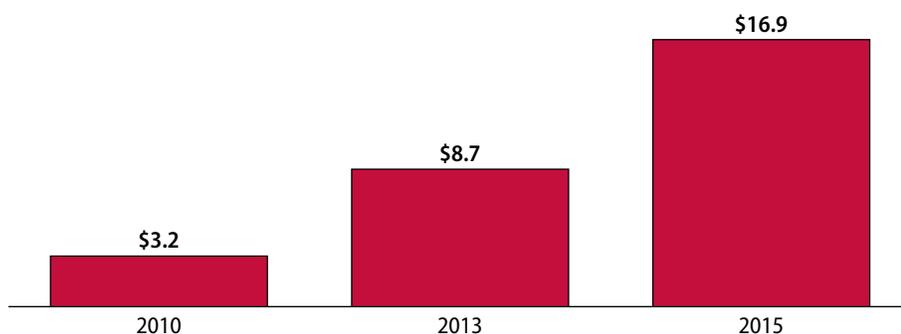
Source: Deloitte

Source: SmartPlanet and IDC