Bigger Data for Better Healthcare

Silvia Piai   Massimiliano Claps
September 2013

IDC HEALTH INSIGHTS OPINION

Big Data holds great potential to change the whole healthcare value chain, from drug discovery, to personalization of care for patients, to industrialization of healthcare provider processes for improved clinical outcomes and increased efficiency, to safer public health management, to more effective and inclusive reimbursement of care. The four Vs of Big Data (volume, variety, velocity, and value) define the key characteristics of Big Data as much in healthcare as in other industries, if not more so. The healthcare industry is moving from reporting facts to discovery of insights, toward becoming data-driven healthcare organizations. The aim is to turn data and information into actionable insights to prevent inefficiencies and adapt workflows for improved healthcare outcomes across the end-to-end patient journey.

Looking at the current and future applications of Big Data in healthcare it is interesting to see how they further enhance and accelerate the convergence between the activities of clinicians, administrators, policy makers, payers, and researchers by saving costs, creating greater efficiencies based on outcome comparison, reducing risks, and improving personalized care. Patients can benefit from this convergence too because, besides being the ultimate "producers" of healthcare data, they can make more informed decisions about their health, playing a much more proactive role in their care paths.

The potential for Big Data, however, is still generally untapped. Technology developments are progressing rapidly, but in practice only 3% of potentially useful data is tagged and even less is analyzed. And it is not "just" a matter of semantics and data interoperability, it is more holistically a matter of understanding what set of methodologies, skills, regulatory, and organizational changes are necessary to leverage the benefits of Big Data. Analyzing the various technological evolutions and the organizational and process challenges posed by Big Data, IDC Health Insights identifies a number of specific actions that healthcare organizations that want to explore and leverage Big Data need to consider. Leveraging the Big Data opportunity will require an end-to-end strategy where IT is the technical enabler but where new process and organization aspects are led by key executives that will also set the overall business objectives. Such a comprehensive strategy needs to be developed through a step-by-step approach to be integrated into the healthcare organization information strategy as soon as possible.

SITUATION OVERVIEW

Defining Big Data

IDC defines Big Data technologies as a new generation of technologies and architectures, designed to economically extract <u>value</u> from very large <u>volumes</u> of a wide <u>variety</u> of data produced every day, by enabling high <u>velocity</u> capture, discovery, and/or analysis.

- In terms of **volume** it is a real data deluge: in 2011 we created 1.8 zettabytes of data. In layman's terms that's equivalent to 200 billion high-definition movies that are at least 120 minutes long. It would take one person 47 million years to watch all those movies. And all of this data that we create doubles every two years.

- **Variety** is driving a paradigm change, from traditionally structured data archived in relational databases running transactional business systems to unstructured content in the form of video, audio, images, text notes, sensor feeds, and social media feeds; for example more than 11,600 #qldfloods tweets were sent on January 12, 2011, at the height of the Brisbane flooding, while YouTube users upload 48 hours of new video every minute of the day. 15% of information today is structured information, while unstructured information, such as email, video, blogs, call center conversations, and social media, makes up about 85% and presents challenges in deriving meaning with conventional business intelligence tools.

- On the one hand **velocity** is affecting the production and capture of data: for example, when using 100% quality and 5 frames per second technology, a video-surveillance system should deal with the capture of 5.7MB generated over an hour. On the other hand, velocity is also impacting discovery and analysis processes: for instance, decoding the human genome took 10 years the first time it was done, in 2003, but can now be done in one week. From a business outcome perspective, data velocity can translate into faster execution of tasks because end users can act and decide more quickly.

- It is, however, the **value** component that could generate a step-improvement in customer service, operational efficiency, employee satisfaction, and risk reduction for commercial and public sector organizations across multiple industries. Big Data enables organizations to analyze business problems in the context of a more complete view of processes and their interactions, analyzing a greater number of scenarios quicker and more cheaply. In the U.S., for instance, through the analysis of the FDA Adverse Events Reporting Systems (FAERS) large database, Big Data helps pharmacovigilance researchers to get new insights into drug-drug interactions that have not been fully understood or foreseen during clinical trials.

#IDCWP25V

Healthcare Is Where It All Started

In 1854 there was a sudden and serious outbreak of cholera in London's Soho. Doctor John Snow mapped the cases, with the map essentially representing each death as a bar and clearly showing that cases were clustered around the pump in Broad (now Broadwick) Street. After further investigation Snow discovered that a 59-year-old woman collected water every day from the Broad Street pump; the water was taken on Thursday, August 31, and she died of cholera on the Saturday. At a local brewery and a local workhouse, workers were surrounded by cases but appeared unaffected because they had their own water supply. Eventually Snow discovered that the pump in Broad Street was polluted by sewage from a nearby cesspit where a baby's nappy, contaminated with cholera, had been dumped. Snow's approach had not only changed the understanding of cholera transmission, but had also changed data analysis and visualization.

Fast-forward to the 21st century and Big Data holds great potential to change the whole healthcare value chain, from drug discovery, to personalization of care for patients, to industrialization of healthcare provider processes for improved clinical outcomes and increased efficiency, to safer public health management, to more effective and inclusive reimbursement of care. The four Vs of Big Data (volume, variety, velocity, and value) define the key characteristics of Big Data as much in healthcare as in other industries, if not more so:

● The **volume** of worldwide healthcare data in 2012 was 500 petabytes, equal to 10 billion four-drawer filing cabinets. That is estimated to grow in 2020 to 25,000 petabytes, equal to 500 billion four-drawer filing cabinets — a 50-fold increase from 2012 to 2020. The use of multimedia such as medical images and high-resolution video has significantly impacted the growth of healthcare data volume. With the progress in medical photography, the resolution of medical images is increasing, with the image files ranging from MB to GB. The size of high-resolution medical images, such as 64/128-slice CT (Computed Tomography) scans, 3.0T MRI (Magnetic Resonance Imaging), and PET (Positron Emission Tomography), often exceed 100MB. According to the OECD (Organization for Economic Co-operation and Development), for example, in 2007, in France alone there were 7.6 million CT examinations, while in 2011 there were 10 million. Healthcare providers are also increasingly using high-resolution videos that generate 25 times the data volume of even the highest resolution still images. In addition to that, the volume issue is further exacerbated by the long period of data retention mandated by regulations on patient records. In Europe, it is not uncommon for retention requirements to last decades, and some providers' organizations retain patient information indefinitely for research purposes.

● The **variety** of healthcare data formats and sources is widely known. A vast array of structured data is produced every day at the point of care and in the back office by clinical and administrative

systems. This flow of structured data is joined by an increasing amount of unstructured data such as images, videos, text notes from clinical and nursing staff, and information from wearable sensors and monitoring equipment — creating a wider mix of data formats and types. And, in the near future, patient records are expected to integrate new streams of data from fitness devices, social media, genetics and genomics, etc. Combining all of these types of information — for example, a large number of patient records with published medical research and genomic data to find the best treatment for a particular patient — is a complex technical challenge. Thanks to Big Data technologies, healthcare organizations are now starting to combine this varied information to extract insights for improved healthcare outcomes on both an individual and a population level.

- **Velocity.** The shift from very accurate analysis of relatively small samples of data and clinical and business performance indicators, to high-velocity, on-the-flight, real-time analysis of a "universe" of data is already changing drug discovery and clinical care, for instance the use of real-time streaming data is already used in ICUs (intensive care units) to prevent life threatening episodes. Researchers at the Hospital for Sick Children (Sick Kids) in Toronto (Canada) have introduced a health analytics platform called Artemis that leverages high-speed physiological data (more than 1,000 recordings per second) with other electronic health record data enabling retrospective data mining on a dataset of nearly two years of 30-second spot readings obtained from 1,151 patients. The algorithms are used to predict when a baby is at risk of infection by detecting subtle changes in physiological measures. Physicians are alerted before a child shows signs of illness: early intervention improved treatment efficacy along with consequential shorter hospital stays and reduced costs. However, today healthcare providers mostly rely on analytics and datawarehouse solutions that analyze information in "batch" and can barely scale to cope with real-time analytical outputs of high-frequency data. Clinical and administrative end users, therefore, do not make decisions based on the latest information. Use cases such as clinical decision support have a real return only if end users have a complete and up-to-date view of the patient with the latest information. **Velocity** is also offering new opportunities to improve outcomes by integrating the various parts of the health value chain; for example through a much tighter integration between laboratories and wards in highly specialized university hospitals driving toward a financially and operationally sustainable bench-to-bed paradigm.

- Ultimately it is the **value** that will make **Big Data** a real force of change in healthcare. With the advent of changing financing and care delivery, analytics has taken on new importance. Healthcare organizations are looking at:

- Operational efficiencies, to reduce costs, waste, and fraud through more efficient methods for data integration, management, analysis, and service delivery.

- Business process enhancements, to find new ways of delivering care while efficiently allocating services to enable sustainable management of the population health.

The healthcare industry is therefore moving from reporting facts to discovery of insights, toward becoming data-driven healthcare organizations. Turning data and information into actionable insights to prevent inefficiencies and adapt workflows for improved healthcare outcomes across the end-to-end patient journey will be paramount. It will enable better collaboration among clinical professionals, empowering public health authorities to manage mass individualized prevention campaigns, and payers and providers to implement new business models, such as pay-for-performance, that could make healthcare services more financially sustainable, thus accessible to a wider share of the population.

Big Data Technology Landscape

Figure 1 represents the Big Data technology stack (infrastructure, data organization and management, analytics and discovery, and decision support and interface automation), which is composed of four macro-layers that IDC uses to describe the Big Data technology market. Within each of these, there are more granular layers of components and capabilities that form Big Data architecture in real life.

Decision Support and Automation Interface

Data analysis is about closed loop decision-making models that at a minimum include steps such as track, analyze, decide, and act. Supporting decision making and ensuring that action is being taken based on decisions that were influenced by the analysis requires collaboration, scenario evaluation, risk management, and decision capture and retention. IDC defines two decision support and automation software categories: transactional decision management and project-based decision management software.

- Transactional decision management is highly automated, application embedded, real-time, streaming, automated, rules-based, and prescriptive functionality, such as those used in fraud detection, clinical decision support systems, and remote patient monitoring.

- Project-based decision management is characterized by standalone, ad hoc, exploratory, discovery, and predictive functionality. Examples include applications for patient segmentation, pharmaceutical R&D, and health population management systems.

Analytics and Discovery

This layer is broadly segmented into:

- Software that supports offline, ad hoc, discovery, and deep analytics.

- Software that supports dynamic, real-time analysis and automated, rules-based transactional decision making.

Big Data analytics and discovery tools can also be segmented by the type of data being analyzed. Examples include data, text, audio, video mining software, and network analysis software (i.e., evaluation of links among entities). Another dimension that identifies specific analysis and discovery software is the complexity of analysis. There are tools for highly complex descriptive and predictive analysis and tools that simply help with basic aggregation and access to information. Both are needed but require different features and functionality. Tools such as Hadoop, MapReduce, and Dryad (along with methodologies including graph analysis) are proving adept at expediting searches through the large, irregular data sets that characterize Big Data initiatives.

Data Organization and Management

The data organization and management layer of the Big Data technology stack refers to software that processes and prepares all types of structured and unstructured data for analysis. This layer extracts, cleanses, normalizes, tags, and integrates data. These technologies include relational databases, NoSQL databases, key value stores, text analytics, ontologies, categorizers, schema extractors, search indexes, parallel file systems, complex event processing engines, graph databases, and other technologies that organize and manage data at rest and in flight.

There are discrete technologies that focus on a particular data type or source, a particular data schema (or lack thereof), and use case or workload (it is important to mention that not all Big Data uses cases are analytics focused; for example, operational workloads include the use of Hadoop for archiving).

There are also software platforms or packages that combine features and functionality to address multiple types of Big Data including unified access and analysis software, and software for processing both structured and unstructured data. In each case, the information is processed and valuable attributes are extracted and then stored. It is interesting to see that new development in this area allows information crossover: text extractions may be stored in databases or datawarehouses, while cleansed data may be stored in a search index. Furthermore, new, usually text analytics-based, technologies now normalize across both the data and the content so that at access, analysis, and discovery time relationships among disparate sources can be uncovered.
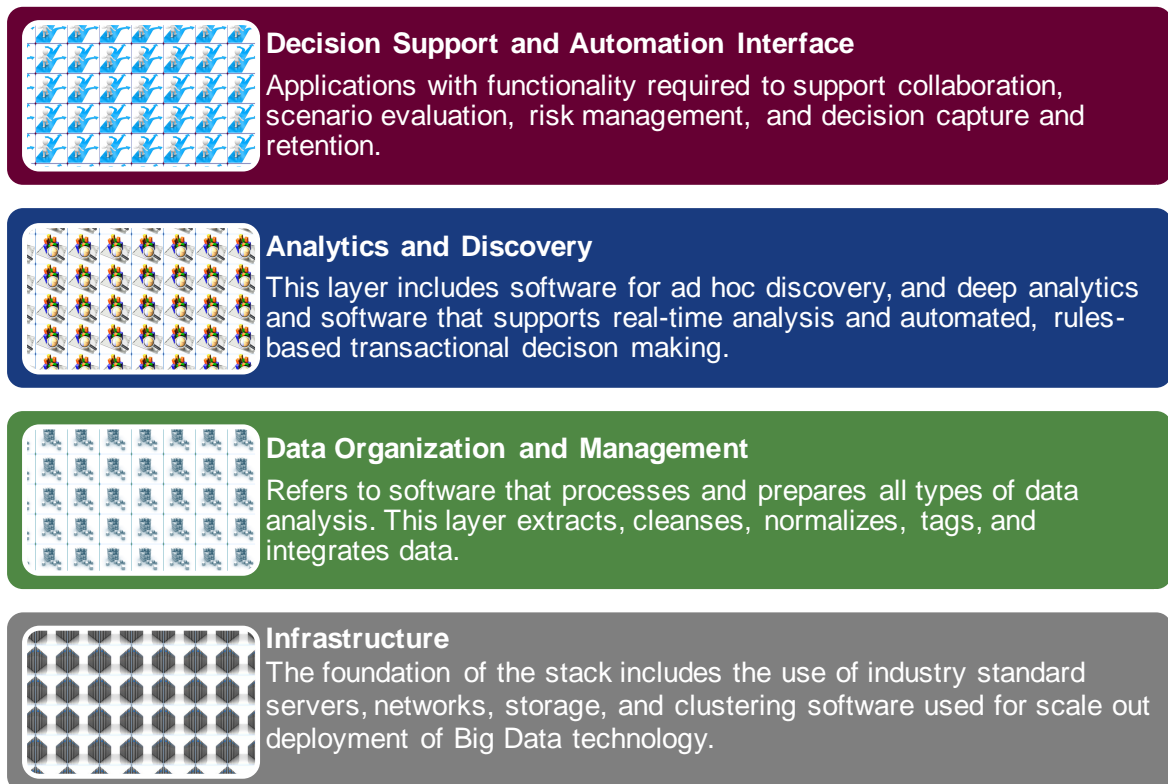
Infrastructure

At the foundation of the Big Data stack is the infrastructure. The use of industry standard servers, networks, storage, hypervisors, and clustering software has been a major driver of the value component of Big Data. Moore's Law has driven significant capabilities into standard, scalable, affordable, and highly available hardware, such as x86 servers and 10Gb Ethernet, that may have required specialized infrastructure or supercomputers in the past.

Looking more specifically at storage, an evolving trend toward the use of object-enabled, content-addressable, and/or scale-out file systems is becoming the norm for Big Data infrastructures. From a compute perspective, most data-intensive jobs can be partitioned and can be run efficiently on standard clusters with memories that are physically and logically distributed. A smaller number of problems (in healthcare generally encountered by research and public health institutions) that are less uniform and more communications dependent need logically shared global memory spaces and capable interconnects that are addressable by the high-performance computing (HPC) market. With increased capabilities and affordability of multigigabyte (and in cases multiterabyte) memory footprints, there is a movement toward in-memory databases.

FIGURE 1

Big Data Technology Stack

Decision Support and Automation Interface
Applications with functionality required to support collaboration, scenario evaluation, risk management, and decision capture and retention.

Analytics and Discovery
This layer includes software for ad hoc discovery, and deep analytics and software that supports real-time analysis and automated, rules-based transactional decison making.

Data Organization and Management
Refers to software that processes and prepares all types of data analysis. This layer extracts, cleanses, normalizes, tags, and integrates data.

Infrastructure
The foundation of the stack includes the use of industry standard servers, networks, storage, and clustering software used for scale out deployment of Big Data technology.

Source: IDC, 2013

FUTURE OUTLOOK

Big Data Driven Healthcare Scenarios

Looking at the current and coming applications of Big Data in healthcare it is interesting to see how they further enhance and accelerate the convergence between the activities of clinicians, administrators, and policy makers, payers and researchers along the lines of cost saving, greater efficiency based on outcomes comparison, reduction of risks and personalized care. Patients can benefit from this convergence, because besides being the ultimate "producers" of healthcare data, they can make more informed decisions about their health, playing a much more proactive role in their care paths. The following paragraphs describe benefits and opportunities driven by Big Data for the various healthcare stakeholders.

Research

● Life-science research optimization. Big Data is supporting healthcare research organizations in optimizing operations and strategic planning. Pharmaceutical companies are using Big Data to predict disease trends to model future demand and costs and make more appropriate strategic investment decisions. Applications in life science research have shown how Big Data can improve clinical trial design and results analysis, speeding up research times and result accuracy. Big Data and analytics are at the core of many [projects](#) within the [Innovative Medicine Initiative](#) (IMI), a joint undertaking between the European Union and the pharmaceutical industry association EFPIA (European Federation of Pharmaceutical Industries and Associations). With a €2 billion budget, IMI supports research projects in predictive safety and efficacy, knowledge management, and education and training. The research consortia participating in IMI projects consist of large biopharmaceutical companies as well as small and medium-sized enterprises, along with partners such as patients' organizations, universities and other research organizations, hospitals, and regulatory agencies. IMI projects are especially aimed at the development of a new "toolbox" (including toxicology tests, biomarkers, clinical trials protocols, etc.) for drug developers to reduce the risk of failure of new medicines in the drug development process (pre-clinical and clinical phases) available to all researchers (industry and academic) and to translate research results into methods and technologies for industry and regulatory practice.

● Personalized medicine. In personalized medicine, the application of **Big Data** solutions to genome datasets analysis, and to a wider number of "data-rich" patient records (including structured data such as biometric information from medical devices and unstructured data such as clinicians' notes), is used to examine the relationships between genetic variation, predisposition to specific diseases, and specific drug responses. This type of research will enable early detection and diagnosis before a patient develops

disease symptoms. Therapies will also be more effective because patients with the same diagnosis can be segmented according to their genetic variation, to adjust drug dosages to minimize side effects and maximize response. But future scenarios for personalized medicine are not "limited" to the human genome. DNA sequencing technologies are also applied to a new field of research, called metagenomics, allowing comprehensive examination of microbial communities. Within the body of a healthy adult, microbial cells are estimated to outnumber human cells 10 to one, and researchers involved in the NIH Human Microbiome Project are focusing on the study of the human microbial communities, trying to deliver a comprehensive characterization of the human microbiome and analyzing its role in human health and disease. Research on the human microbiome will add a further aspect to personalized medicine, understanding the relationships between microbes and between the microbiome and clinical parameters that underpin the basis for individual variations that may ultimately be critical for understanding microbiome-based disorders and developing novel prophylactic strategies such as the application of prebiotics and probiotics to improve human health.

Healthcare Providers

- Evidence-based care. Being able to quickly analyze a wider set of patient information, healthcare providers will be able to accurately apply the latest findings of medical research, thus providing personalized and evidence-based care services, not only being able to react to current patient-specific conditions but also to efficiently prevent complications and new disease developments. Clinical decision support systems already help healthcare providers compare patient information against research literature and medical guidelines, alerting them to potential errors such as adverse drug reactions and enhancing the efficiency and quality of care; for example, chest pain can result in approximately 100 different diagnoses; decision support systems that can narrow down the alternatives can still leave the final decision to the physician, but will greatly speed up the process. The reliability and the comprehensiveness of support offered to healthcare providers will be further improved as these solutions mature and include other capabilities, such as image analysis. Charité Hospital in Berlin (Germany), in partnership with SAP and Hasso Plattner Institute, launched a new stratified medicine research initiative called HANA Oncolyzer to support decisions on treatment of patients suffering from cancer. HANA Oncolyzer combines structured and unstructured data from patient records including data on the biological and genetic mutations of cancer, with all available research findings and approved care guidelines, to adjust therapies to individual patients and cancer forms.

- Improved chronic disease management. Personalized medicine and evidence-based practices will also better integrate with chronic disease management programs. Thanks to Big Data platforms, healthcare providers will be able to better leverage information

coming from remote patient monitoring (RPM) systems to monitor patient adherence to prescriptions and to improve future treatment options and reduce complications. Leveraging information from RPM systems, healthcare providers will be able to reduce inpatient stays, limit emergency department visits, and improve the effectiveness of home care and outpatient appointments. Applying advanced analytics, such as segmentation and predictive modeling, to patient profiles, healthcare providers can detect anomalies and identify patients who are at high risk of developing a specific disease or complications and would benefit from a preventive care program or a disease management program.

Efficiency and Cost Savings

- Risk reduction. The increasing use of clinical decision support and evidence-based medicine is also expected to reduce the incidence of medical errors and adverse events, easing the costs related to medical insurance and mitigating the risk of lawsuits.

- Process optimization. Analyzing datasets from the "patient journey" as well as other back-office flows, healthcare provider managers will be able to identify inconsistencies, bottlenecks, and misuse of resources. Mapping processes, healthcare providers will have greater visibility of areas where operations need to be streamlined. Leaner process will reduce costs, liberate untapped capabilities, and deliver a better patient service. For example, Nice University Hospital (France) uses an RFID-based system for the management of its approximate 57,000 biological samples in the hospital's biobank. Before, the hospital relied on a paper-based traceability process that was time-consuming and error-prone, and could result in lost samples and compromised security. Thanks to the analysis and the integration of the workflows of pathology and the biobank center, the new system provides more than a 50% time saving, increased traceability, and timely delivery of biospecimen samples.

- Spend management. Healthcare authorities across Europe are moving toward a more transparent and coordinated approach to procurement of goods and services for national healthcare services. For instance, the U.K. National Health Service recently published its [procurement strategy](). The strategy calls for improving procurement capability and performance through the sharing of best practices and the adoption of standards for procurement. The Department of Health will create a new Centre of Procurement Development (CPD) to support the procurement development program and ensure best practice is embedded throughout the NHS. A key enabler for this strategy will be improving data quality and transparency. Only aggregating and comparing information on procurement from the various providers the NHS expects to determine what really defines a best practice and what drives performance, by understanding the variation of the cost per unit of a certain product or service category, and what is driving these variations. For example, the report highlights that there are differences in the use of temporary staff, with Trusts ranging from

apparently zero use of temporary staff to those where temps are around 10% of the total workforce. Big Data and analytics tools can provide new insights and different perspectives, for example on the impact on productivity of certain thresholds of temporary staff.

Healthcare Payers

- Optimal treatment pathways. Healthcare authorities and healthcare funds across Europe are starting to realize the benefits of Big Data. Many studies have shown that wide variations exist in healthcare practices, outcomes, and costs across different providers, geographies, and patients. Agencies such as the National Institute for Health and Care Excellence (NICE) in the United Kingdom, IQWIG (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen — the Institute for Quality and Efficiency in Health Care) in Germany, and AIFA (Agenzia Italiana del Farmaco — the Italian Medicines Agency) in Italy are working on risk sharing programs that base reimbursements for providers and how pharmaceutical companies are paid on effectiveness and patient outcomes. These initiatives will not only increase the transparency of the cost of different services and drugs, but will also help to determine the so-called "optimal treatment pathways" — by analyzing comprehensive patient outcome data it will be possible to compare the effectiveness of various interventions and thus reduce the effects of over-treatment and under-treatment. This will also empower patients to interact with clinicians and make the best decisions. Population segmentation can also help public and private payers identify categories of patients that are likely to develop chronic diseases that currently account for the vast majority of health expenditure. In Germany, since 2006 the government, aiming to contain health insurance costs, has introduced some elements of competition in the reimbursement mechanisms for health funds. AOK, a statutory health fund covering 24 million people, developed and implemented the Oscare platform based on SAP HANA, analyzing patterns of data from ERP, CRM prescriptions, and other insurance data, identifying, among other patterns, those patients that are likely to develop chronic diseases. Thanks to Oscare, AOK is now able to offer customized prevention plans and more accurate disease management programs to its customers, reducing the cost of treatment.

- Antifraud. Healthcare authorities will be able to identify fraud, reimbursement system anomalies, and regulation breaches by implementing automated systems for compliance assurance and financial control. For example, statistics shows that the number of caesarean sections (a procedure that has a high reimbursement rate) in certain Italian regions is higher than the national average, suggesting potential fraud or malpractice. If in the future, thanks to Big Data, this information is processed in real time, it would be easier for healthcare authorities to implement more timely and effective controls and deliver significant savings while ensuring more adequate care to the population. To improve transparency in

public services, European governments have launched open-data initiatives that include data from national health systems. For instance, in the U.K., two startups, Mastodon C and Open Health Care UK, and doctor Ben Goldacre, using data published on data.gov.uk, looked at the regional patterns in the prescription of statins by family doctors, comparing data on patented and generic prescribed drugs. According to NHS prescribing guidelines, doctors can prescribe patented drugs only if they have good clinical reasons. The analysis showed a good deal of local variation in prescriptions that is difficult to explain solely on clinical grounds, providing further evidence for the necessity for better guidelines and control mechanism over prescriptions.

Public Health

● Prevent and react. Being able to process data from the national health systems and from other public organizations such as social services and being able to detect disease patterns and health trends, public health authorities will be able to analyze new facets of healthcare demand and discover new correlations and dynamics and consequently be able to plan care services and resources with greater accuracy and timeliness. Public health authorities will be able to more rapidly assess the efficacy of the initiatives and policies implemented and, possibly, to adjust actions and funding without waiting for the traditional statistics that require a longer time. In the U.K., Leeds Teaching Hospitals have been testing a new syndromic surveillance proof of concept that creates a standardized approach to healthcare data analysis to support emergency preparedness and service planning. Leeds generates up to half a million records each year in its Accident and Emergency Department system and approximately 1 million unstructured case files each month. Using the system, the hospital was able to look at patterns in the data to identify potential outbreaks of infectious disease and effectiveness of vaccination programs, as well as trends that are typically more difficult to analyze, such as alcohol-related visits to the emergency room and injuries from accidents in the home.

Patients

● Patient empowerment. Chronic diseases are the leading cause of mortality and morbidity in Europe, placing a heavy burden on the future socioeconomic development of the region. For example, WHO research shows how cardiovascular diseases were the cause of 52% of all deaths in the European region — with a disease burden of 34.42 million DALYs (Disability-Adjusted Life Years: one DALY equals one year of healthy life lost). Effective treatment of these long-term conditions requires greater integration among providers involved in the care process and greater involvement of the patient in his/her care by ensuring adherence to prescriptions and adoption of healthy behavior. Patients have a more active role in their care when clear information on the effectiveness of treatment and prevention empower them to know that their actions can make a difference. A few years ago, the

Scottish NHS launched the online platform "My Diabetes, My Way" aimed at supporting self-management for diabetic patients. According to the Scottish ehealth strategy, the platform is evolving toward greater integration of patients' records on the SCI-DC (Scottish Care Information Diabetes Collaboration) clinical information systems and patient behavior information. Information will be increasingly presented in a personalized way, allowing comparison with national statistics and showing patients' level of adherence to care guidelines.

● Customer experience management. Patients' opinions on healthcare services are also a valuable source of information on care services quality. "Ensuring that people have a positive experience of care" is one of the five outcome "domains" that are to be used to hold the English NHS Commissioning Board to account for the outcomes it delivers through commissioning health services in 2013. Although for the first few years the Department of Health has understandably chosen to rely on the existing national surveys to provide data for formal assessment, there are plans to integrate information from initiatives such as Care Connect (a multichannel patient feedback service developed by NHS England, which recently went live in 18 trusts in London) and the NHS friends and family test (since April 2013, patients discharged from hospital were asked whether or not they would recommend their hospital to family and friends, with answers ranging on a five-point scale from "very unlikely" to "very likely") and social media sentiment analysis.

What Should Healthcare Stakeholders Do to Be Ready for Big Data?

The potential for Big Data is generally untapped and not just in healthcare. The 2012 IDC-EMC "New Digital Universe Study" reveals that 23% of the information in the digital universe (or 643 exabytes) would be useful for Big Data if it were tagged and analyzed; however, technology is far from where it needs to be, and in practice only 3% of the potentially useful data is tagged and even less is analyzed. And it is not just a matter of semantics and data interoperability, it is more holistically a matter of understanding what set of methodologies, skills, regulatory, and organizational changes are necessary to leverage the benefits of Big Data.

IDC research confirms that the ability to grasp the magnitude of the required change is still limited among healthcare providers; in fact, of 158 healthcare IT and non-IT executives interviewed in Western Europe in September 2012, 43% thought they could deal with Big Data by simply increasing their storage capacity; more worryingly 18% thought Big Data will have limited to no effect on their organizations and only 30% understand that they will need to reassess their information management processes. That is a strong indication of a technically-centric and volume-biased approach to Big Data.

Technology is the Enabler, Technology is the Barrier

Technological advances have made Big Data and analytics accessible to a wider number of organizations. In a sort of virtuous circle technology can be seen not only as an enabler but also a driver for advanced analytics.

The market for Big Data solutions is expanding rapidly, providing a healthy, competitive environment that will spur product and service innovation and pinpoint financially sustainable business models and vendors. For example, IDC forecasts the cost of storage per GB will drop from $4 in 2010 to less than $0.50 in 2020, spurred by a maturing cloud computing market. Dr. David Sibbald, of Aridhia, already estimates that the cost of storing the raw sequence of a human genome is approximately £100. On top of the falling cost of storing data, the cost of analyzing data is being slashed thanks to scientific advances, so for example sequencing the human genome cost $28.7 million in January 2004 and $7,666 in January 2012 (www.genome.gov). Quantum computing, albeit at the very early stages of product development, promises to bring those costs down even further.

Moreover, thanks to the rapid spread of IT solutions, such as electronic health records (EHRs), hospital information systems, departmental systems, and PACS, the "liquidity" of healthcare data has increased, driving the adoption of clinical and business intelligence applications. IDC surveys on Western European healthcare providers show high current adoption rates and sustained investment plans in these areas, facilitating the collection, archiving, and analysis of patient information.

When looking at technology, however, we need to consider other attributes, such as the appropriateness, applicability, integration, support for standards, and performance of technology and IT architecture for the relevant workloads. The fragmented approach to design and implementation of IT systems, largely adopted by healthcare, can limit the effectiveness of Big Data initiatives.

High adoption rates of EHRs and other clinical information systems should not make us jump to the conclusion that data is completely accessible and that current systems and infrastructures are Big Data ready.

In most healthcare organizations, the reality is that most clinical data is still siloed in disparate systems that can't communicate with one another, or that can exchange only limited subsets of information. Hospitals and other healthcare providers have been slow to develop the datawarehouses they need to aggregate and normalize the data from their multiple clinical and administrative systems.

Addressing the issue of healthcare data interoperability is a priority for the European Union and national ehealth strategies. The EPSOS project establishes a framework for interoperability for summary records and prescriptions that has been adopted as a reference across Europe. For instance, the standards required for the French EHR, the

DMP (dossier medical informatisé), follows the scheme developed by the EPSOS project. Aiming at greater consistency on infrastructures, NHS Scotland has established a technical Design Authority and has published a set of architectural principles along with a standards development framework aimed at ensuring value-based convergence around common IT systems and approaches to the delivery of care.

Another aspect that needs to be taken into consideration is the departmental approach used when implementing analytic applications in healthcare, especially in healthcare provider organizations. Generally analytic applications are focused on tasks, data domains, and business opportunities and risks that are associated with specific departmental requirements, not enterprise ones.

Healthcare providers generally rely on their clinical information systems for reporting and analysis. But with changes in organizations and the recent healthcare reforms across Europe, most providers have found themselves handicapped by the lack of capabilities in the analytics tools embedded in their clinical applications, increasing the amount of disruption providers already face.

Not all business intelligence/datawarehouse (BI/DW) technology stacks are designed for advanced analytics. In most organizations, users have designed and optimized their BI/DW technology stacks for reporting, performance management, and OLAP. Such stacks are also capable of satisfying most departmental requirements for reporting and OLAP. This optimization is invaluable for "big picture" reports and analyses that span enterprisewide processes (especially financial ones). But in many organizations, the users did not design and deploy the BI/DW technology stack to satisfy departmental requirements for advanced analytics and Big Data. For example, a significant portion of clinical documentation is by nature unstructured, and this has thwarted efforts by providers to optimize the highly structured datawarehouses common to other industries. Clinical data is also extensive, inconsistent, and multidimensional; different types of analysis, such as population-based or episode-based analysis, rely on different dimensions for data models but often are called upon in the same analytics environment, creating issues for warehouse designers working with rigid schemas.

Some departments deploy their own platforms for Big Data and analytics. They do this when the department has a strong business need for analytics with Big Data, plus the budget and sponsorship to back it up. In summary, in many organizations, Big Data analytics is a departmental affair, implemented by the department's team on a departmentally owned platform. Such a narrow approach can limit the benefits that a Big Data investment might bring to the whole organization.

Beyond Technology: Challenges for Big Data Ready European Healthcare

The advantages brought by the Big Data technology stack (infrastructure, data organization and management, analytics and discovery, and decision support and interface automation) are potentially enormous. Today, healthcare has limited experience with the use of these technologies. All but the very largest healthcare organizations are approaching Big Data through pilot or proof-of-concept projects. To date, few healthcare organizations have adopted Big Data technology across the enterprise. Achieving the benefits that can be brought by deploying Big Data demands a bold and comprehensive vision focused not only on the required IT and dataset investments, but also on the development of analytical capabilities, of a culture of transparency and comprehensive information assurance, and on an extensive rethinking of the incentives that regulate the way healthcare is provided and funded, so that information sharing and evidence-based medicine are rewarded.

These challenges can be framed around four areas, which if addressed can help healthcare stakeholders to take advantage of the Big Data revolution: collaborative information governance, data veracity and contextual meaningfulness, information assurance and organizational change.

Establish Information Governance That Favors Collaboration

The potential of Big Data in healthcare lies in combining traditional data with new forms of data, both at the individual patient level or for the whole population of a community or region or country. But the structure and the organization of European healthcare systems is generally highly decentralized, generating a fragmented environment with clinical and administrative data residing in siloed systems and in multiple formats. Data fragmentation is the result, not only of technical interoperability issues, but also of different interests and business incentives and incompatible regulations across the health value chain. Big Data projects depend on data access and transparency: a type of access that often goes beyond the domain of the single department or the single organization. Benefits driven by a Big Data project for a single organization often clash with the interests of another one. For example, initiatives on greater transparency of costs and clinical outcomes lead to more informed decisions for patients and healthcare authorities, but might find resistance from professionals within (or outside) the organization that benefit from opacity. Even within the same organization different people will have different priorities:

- CIOs want to reduce the cost of technology needed to support business requirements, while being compliant with regulations.

- CISOs want to protect critical information at all costs.

- CMOs want to improve quality of care, while maintaining autonomy of decision making and power.

- CFOs and payers want to optimize resource allocation and reduce fraud and abuse.

Healthcare executives that want to capture value from Big Data and analytics initiatives need to pay attention to incentive alignment across the involved ecosystem.

Focus on Data Veracity and Contextual Meaningfulness

The more care becomes patient centric, the more the volume, speed, granularity, and variety of formats of clinical and administrative data become complex. Therefore, while there is much more rich potential for useful information, the risk of errors increases. Patient information needs to be reliable, complete, and relevant to the context in which it will be used, so that clinicians, administrators, and patients can trust it. In the Big Data paradigm, defined by high variety and velocity, data quality and trustfulness might become an issue because of their high variability. When physicians consult a patient record, they ask themselves key questions such as, is this the right patient? Has the information from previous providers (maybe relying on unstructured formats such as handwritten clinician notes) been correctly captured? This might lead to decisions, such as rerunning tests, that have already been done, nullifying the value that Big Data tools added to the care process.

It is not just a question of garbage in and garbage out. In healthcare the correct perspective is as important as the formal precision of data. In many circumstances, to understand a clinical phenomenon the snapshot of patient conditions needs to be integrated with the bird's eye view of his or her clinical history. The time dimension should not be neglected, because in a population increasingly affected by chronic diseases, what affects us today and what will affect us tomorrow have changed slowly over our lifetime. Long-data theorists emphasize how "Big Data puts *slices* of knowledge in context, but to really understand the big picture we need to place a phenomenon in its longer historical context" (see *Stop Hyping Big Data and Start Paying Attention to "Long Data"* — Samuel Arbesman, Wired, 2013). Clinical research has already been taking advantage of long data analysis. For example, in 1948, 5,209 adult residents of Framingham (Massachusetts, U.S.) enrolled in what would become the longest-running epidemiological study ever. The participants would submit to medical checkups every year, giving researchers a large-scale, long-term picture of human health. The study is now on its third generation; much of what is now considered conventional wisdom about heart disease and stroke — such as the links to cigarette smoking, high cholesterol, and high blood pressure — were actually groundbreaking insights from the Framingham study (http://www.framinghamheartstudy.org/).

Technology such as clinical decision support systems are constantly evolving in order to use the largest and the longest patient datasets, but in order to gain immediate benefits from the currently available

solutions it is important to train clinicians and other healthcare staff on the sources and the rationale behind the technology. This will establish the right expectations among medical staff and information systems actively supporting the care process.

Evolve Information Assurance to Avoid Data Protection Bans

"First, do no harm" is one of the principal precepts of medical ethics: in a highly integrated care system, protecting patient safety and privacy is paramount for patient welfare. Maintaining the high confidentiality, integrity, and authenticity of patient information is paramount to preventing the disclosure of information to unauthorized individuals or systems and avoiding the risk of patient discrimination based on their clinical profiles. In a highly collaborative environment such as Big Data initiatives, the typical siloed-based organizations of healthcare make security and privacy compliance more error-prone and resource-intensive than in other sectors.

In Western Europe, protection of sensitive data is the subject of a series of national and local laws which have been incrementally developed over a long time to align with the guiding principles of the European Data Protection Directive (Directive 95/46/EC). In addition to the basic framework offered by the data protection directive, there are many more specific regulations contained in laws relating to hospitals and cancer registries, in civil, criminal, and social codes and in codes of medical conduct and ethics. These regulations have been designed to protect the security and privacy of patient data in specific settings. As Big Data requires data sharing across departmental boundaries, these regulations can collide and prevent the necessary collaboration. For instance, for several months now, conflicting regulations on patient data have prevented NHS England, clinical commissioning groups (CCGs), and commissioning support units (CSUs) from handling patient identifiable data from the "secondary uses service" for purposes other than the direct commissioning of patient care. These purposes include core mission activities for CCGs and CSUs such as the analysis of health outcomes for patients with a particular condition or investigations into population health and monitoring provider performance.

With Big Data becoming more relevant to the healthcare sector, policy makers will have to reassess these regulations and modify them to ensure that access to data is available in a safe and secure way that also enables healthcare outcomes to be optimized. Big Data and analytics can be part of the solution as they are also key enablers of more sophisticated cyber security strategies to protect sensitive data, for instance to enable anonymous analysis of large data sets, or to predict and prevent the source of malicious attacks and malpractice.

Recruiting the Right Talent and Developing the Right Culture

Recruiting and retaining the right talent required to leverage the value of Big Data opportunities can be a critical challenge. Coping with the information tsunami will require a renewed focus on the competencies of data scientists, data architects, and database administrators, and the adaptation of data architectures to new data models, visualization techniques, and semantics emerging from the middle out, especially from the needs of clinicians and nurses rather than from a top-down definition of an ontology. Therefore on one side healthcare providers will need to hire more data scientists and IT specialists. On the other side, however, hiring data scientists will only be part of the solution. To demonstrate value, when launching a Big Data project it is necessary to hire managers who can understand business and operations requirements, translate these needs into analytical terms, and then communicate and contextualize Big Data findings in the healthcare business processes. As this type of complex mix of talents is not abundant in the market, salaries for these types of professionals can be a barrier, especially for organizations in the public sector due to current salary caps and career progress freeze policies caused by budget cuts. Another element strictly connected with talent and the ability to gain maximum benefit from Big Data is about creating the right culture within medical staff. Medical leaders who understand the value of data and the willingness of employees to use data-driven insights are essential. Comparative research on clinical processes can require physicians to change their routine in care delivery according to guidelines derived from Big Data. But studies in many clinical areas show that physicians often resist "change." Therefore, when embarking on Big Data, healthcare executives need to address not only staffing and organizational issues but also focus on training for physicians and nurses to develop soft skills, such as recognizing instances where they apply confirmatory biases.

Addressing the Issues Through a Shared Services Model

When considering a Big Data project, the investment required to address people, process, and technology issues might be challenging for healthcare organizations, especially European public healthcare providers that are subject to the severe scrutiny and limits posed by spending review targets. This is why only the largest healthcare providers, generally those with strongest connections with medical research, are able to invest in Big Data.

Healthcare authorities, however, are starting to realize that the cost of not adopting Big Data and analytics can be daunting over the long term. All healthcare reforms across Europe are focused on a more appropriate distribution of care services and resources, tailoring services around patients' needs, integrating the activity of care of the various healthcare and social services providers. Their aim is to prevent dynamics that will affect access to services and quality of care, while keeping costs under control.

To achieve these goals it is necessary to have a comprehensive, accurate, and forward looking view of demand and supply of healthcare services and resources. Big Data is an essential enabler here. To make investments in Big Data tools and methodologies more affordable and improve collaboration, Western European healthcare providers have increased adoption of shared services. Shared services are used in this context as a broad term that includes a variety of business models, from full-blown consolidation of assets and people from separate entities to create competencies centers, to more lightweight collaborative governance agreements, for example between hospitals that can offer expertise in clinical decision support and primary care centers that have data and competencies around public health. Shared services are sometimes funded by local health authorities or by national government programs that might not mandate them but still offer strong incentives for collaboration. The benefits of shared services include:

- Spreading the cost and risk of investing in and maintaining new Big Data solutions.

- Attracting and nurturing the skills of people with advanced competencies that meet rising end-user demand.

- Facilitating access to competencies and insights for smaller organizations that might have less advanced requests, but most likely cannot afford even the most basic solutions.

- Clarifying lines of accountability for data quality, security, and compliance. And sometimes overcoming the challenges of exchanging data between legally separate entities — in France, for example, the regional health agencies (ARS), in cooperation with the Ministry of Health, the various health funds, and the HAS (Haute Autorité de Santé — the Higher Health Authority), are monitoring the financial and clinical performance of healthcare providers in their territory. This information provides the basis for the reorganization of services and their integration along care pathways and for the implementation of pay-for-performance initiatives included in the recent healthcare reform legislation.

Find Value at the Nexus: Big Data Converging With the Other Three Pillars of Third-Generation Platforms

IDC expects the worldwide Big Data technology and services market to significantly expand and evolve, growing from $6 billion in 2011 to $23.8 billion in 2016. This represents a CAGR of 31.7%, or about seven times that of the overall ICT market. Healthcare represents about 6% of this. New offerings are introduced to the market on a regular basis, sometimes defining new functional segments and new deployment options, addressing specific use cases with increased performance in terms of speed and scalability of analysis. In the long term, this means analytics functionality will be made available to the

largest possible number of clinical and back-office staff as well as managers.

However, the "Big Data revolution" cannot be sufficiently understood as a standalone phenomenon within IT, but as a core ingredient of a larger transformation of the industry triggered by 3rd Platform technologies: mobile, social business, cloud services, and Big Data and analytics. The convergence of 3rd Platform technologies and the greater acceptance of interoperability standards will make data acquisition and preparation processes, which today account for the biggest share of Big Data technology and staffing budgets, simpler. This evolution will allow healthcare CIOs to focus on data quality and governance, policy makers to create greater harmonization in data privacy and security regulation, and data scientists to concentrate on hypothesis generation, algorithm (and data) selection, and model development, testing, and execution. At the nexus of the four pillars of 3rd Platform technologies, Big Data will benefit from:

- Mobile computing can greatly enhance the capture of accurate and timely information at the point of care, which is critical to enhance the ability to cope with the velocity and veracity of Big Data requirements. At the other end of the service delivery value chain, mobile solutions will deliver insights at the point of care to provide more timely advice to patients and reduce mistakes. Because of the wide variety of devices brought into enterprise environments by healthcare professionals and patients to enhance convenience and usability, healthcare IT executives should carefully consider the computing and bandwidth constraints, and decide how much data and intelligence in the Big Data solution should actually reside on a downloadable app, or how much should be made accessible via a web-enabled application that resides on a remote cloud platform.

- Social business will make an impact first of all because most deployments of Big Data will increasingly incorporate multistructured data from multiple sources. Social media analytics will play an increasing role as a percentage of the population will opt to share behavioral, demographic, purchasing, financial, and other personal data relevant for care and prevention in exchange for services enabled by Big Data solutions. Consumer social media will also increasingly offer some valuable insights to patients, for example to prevent risky behavior for the chronically ill, allowing healthcare providers to focus their Big Data investments on more complex and sensitive business requirements. A slower but equally profound impact of social media will be felt in the long term in the business practices and organizational behavior that are key to Big Data — traditionally process-centric workflows such as patient discharge will be enriched by collaboration-centric practices that can speed up execution by uncovering new way of sharing and accessing data that are now common in consumer social media platforms.

- Cloud computing can speed up deployment of Big Data solutions to respond to fast evolving business insight requirements, while

lowering maintenance costs. Cloud can also help match the ever-increasing demand for higher volumes and fast-occurring peaks of data to be acquired, prepared for analysis, transferred, and archived, while avoiding the sunk cost of investing in spare capacity, especially for fast evolving technologies at the earlier stages of the product life cycle, such as column-store databases and in-memory analytics. Given the increased strategic value of data as an asset, healthcare organizations will need to carefully consider a technology's country of origin and security measures adopted by cloud providers to limit the risk of data loss or contamination. They should also understand how shared service models can help build hybrid cloud services, where the most sensitive pieces of data are at rest, or analyzed within the boundaries of a private cloud.

ESSENTIAL GUIDANCE

Analyzing the various technological evolutions and the organizational and process challenges created by Big Data, we have identified a number of specific actions that healthcare organizations that want to explore and leverage the Big Data need to consider. However, it is clear that leveraging the Big Data opportunity will require an end-to-end strategy where IT is the technical enabler but where new process and organization aspects are led by key executives that will also set the overall business objectives. Such a comprehensive strategy needs to be developed through a step-by-step approach that IDC Health Insights has summarized as follows:

- **Next Monday.** Recognize the value of untapped data assets in supporting fact-based decisions of management, healthcare staff, patients, and care givers. Recognize the implications of operating without critical information and build use cases to address the challenges. Assess the current status of the business and Big Data and analytics. Identify opportunities to use existing data, technology, and analytics in new ways. Explore opportunities to use new low-cost public cloud and open source options as they emerge. While investing in clinical information systems opt for standards based solutions, and architect and implement for future reusability in a Big Data environment. Identify the relevant technology and analytics skills among existing staff, peers, and vendors. Experiment with proof-of-concept and prototype projects.

- **Next month.** Use early quantifiable wins to demonstrate potential and justify budget allocations. Evaluate existing technology and its shortcomings. Assess skill gaps and plan to hire and/or externally source professional services. Whether technically solid or not, many IT-initiated projects demonstrate a lack of connection to business goals. Continued lack of sufficient IT and business collaboration will lead to slow adoption of solutions and difficulty in securing additional rounds of project funding. Therefore identify business sponsors and champions that will support and promote Big Data projects. Expand projects and begin to define architectural standards. Merge Big Data into security and governance policies. Formulate a Big Data strategy that includes

evaluation of decision makers' requirements, decision processes, existing and new technology, and availability and quality of data. Look for and incorporate new data sources. Start budgeting for workload-specific technology.

● **Next year.** Insights from Big Data projects will begin to affect business processes by uncovering existing inefficiencies and identifying new ways of interacting with patients, employees, suppliers, partners, and regulatory agencies. In many cases, however, Big Data and analytics potential will be slowed due to organizations' inability to change business processes fast enough. Optimized business processes will have a greater chance of emerging in new Big Data–enabled use cases and operating models. Therefore ensure that both performance management and experimentation and discovery use cases are supported with appropriate technology, staff, data, processes, and funding. Engage in business process reengineering in response to new insights from Big Data solutions. Assess progress and adjust internal investment priorities to match evolving requirements. Ensure balanced resource allocation across all dimensions of the solutions by taking an enterprisewide approach to Big Data that assigns clear responsibilities to a central organizational unit while fostering a culture of collaboration to flexibly respond to local business needs. Maintain a closed-loop learning environment based on data-driven decision making and expert judgment.

RELATED RESEARCH

● *Personalization, Integration, Industrialization: The Three Forces of Healthcare Change* (IDC Health Insights #HIOH52V, April 2013)

● *Western Europe, Hospital Sector, Business Solution Investment Trends Survey: How Hospitals Will Invest in 2013* (IDC Health Insights #HIOH02V, March 2013)

● *Western Europe Healthcare 2013 Top 10 Predictions* (IDC Health Insights #HIOH01V, January 2013)

● *Worldwide Big Data Technology and Services 2012–2016 Forecast* (IDC #238746, December 2012)

● *IDC's Worldwide Big Data Taxonomy* (IDC #231099, October 2011)

● *The Big Deal About Big Data* (IDC #226904, February 2011)

Copyright Notice