Leveraging Big Data and Analytics in Healthcare and Life Sciences: Enabling Personalized Medicine for High-Quality Care, Better Outcomes
With the cost of mapping an individual human genome poised to break the $1,000 barrier – bringing personalized medicine closer to reality – the healthcare and life sciences industries are now grappling with managing the explosive growth of data. Applying downstream analytics in a volatile data environment, overseeing data storage and movement, and transforming the data to improve patient outcomes are just some of the challenges stakeholders face. Market drivers, however, are mitigating those issues. In the private sector, more sophisticated, innovative and robust information technology is being developed to aggregate, manage, analyze and share big data. In March 2012, the Obama Administration launched a $200 million “Big Data Research and Development Initiative,” which aims to transform the use of big data for scientific discovery and biomedical research, among other areas.

Now stakeholders need to come together to take advantage of market conditions and realize the value of the incredible amount of important data in a more efficient and expeditious manner. “It is critical to collaborate with researchers and the technology ecosystem to develop innovative solutions to seemingly intractable problems emerging in healthcare and life sciences today,” said Ketan Paranjape, director of life sciences and healthcare for Intel. In the spirit of collaboration and as part of the Intel Innovation Summit, the global technology leader in computing convened four thought leaders to share how their respective industries are meeting challenges and utilizing big data.

"It is critical to collaborate with researchers and the technology ecosystem to develop innovative solutions to seemingly intractable problems emerging in healthcare and life sciences today.”

Ketan Paranjape
Director of Life Sciences and Healthcare
Intel

Developing Technologies to Meet Big Data Challenges

The challenge of big data for John Halamka, MD, CIO of Beth Israel Deaconess Medical Center (BIDMC) and CIO and Dean of Technology at Harvard Medical School, is capturing data in structured form, normalizing it for analysis and exchanging it at the right place and time – all the while keeping it private and secure to avoid federal penalties for data breaches. With the ability to store and analyze ever increasing quantities of data in near real time, Graham Hughes, MD, CMO for SAS Institute’s Center for Health Analytics and Insights, is interested in the big clinical and operational insights that can be gained as the traditional siloes of provider, payer and life sciences data are brought together for analysis.

Martin Leach, CIO of the Broad Institute of MIT and Howard, which was formed out of the human genome project, wants to be able to integrate and ask questions of the data and then collaborate and move information around in the big data space. “The size of the data isn’t the challenge; it’s more the variety, accessibility and how you actually join that data together to get some value out of it,” Leach said. As senior director at Pfizer Global R&D for High Performance Computing, Mike Miller and his group work with a full spectrum of data. While protecting the data is paramount, access control of data is just as important, he said. Managing access at scale and when data is being integrated are pain points.

While these issues are today’s challenges, innovations being developed by industry leaders – such as the Broad Institute, which creates tools and methods to advance research in genomics; Pfizer, which builds innovative technologies to develop next-generation therapeutic agents; SAS, which assembles data analytics into scalable products; and others – will take down these barriers, enabling stakeholders to move forward in advancing personalized medicine.

The Impact of Big Data in Life Sciences and Pharma

With growing collaboration in the pharmaceutical industry between pharmaceutical peers, pharmaceutical firms and medical institutions, and pharmaceutical firms and academia, stakeholders need to manage data access over time, Miller said. For some of the larger or longer-term collaborations, Pharma uses peer-to-peer networks and dark fiber to move data. For early-stage or shorter-duration collaborations, Pfizer is looking at distributed computing – moving the compute to the data as opposed to moving the data to be able to compute – which eliminates privacy, security and access control risks. Cloud computing is another option that relieves the burden of hosting for participants but allows them to apply emerging analytic techniques. “Organizations may not be ready to share fully, but in a collaborative shared environment it gives them an opportunity for sharing while giving them enough control over the situation,” he said.

Existing analytical techniques can be applied to the vast amount of existing patient and consumer data, which currently remains unanalyzed, to help gain a deeper understanding of outcomes using “real-world data.” Those resulting insights can be applied at the point of care in determining the most appropriate treatment plan for each individual, according to Graham. Ideally, physicians would be able to engage in shared decision making with their patients, to determine the most appropriate treatment option based on a combination of individual and population data. Just as Amazon leverages consumer data to target offers to individuals – “people like you tend to buy books like these” – care providers should be able to leverage the power of analytics to analyze an individual’s medical data signature and compare that to insights gained from analysis of outcomes in large populations.

What Healthcare Can Learn from Big Data in Life Sciences

The opportunities and challenges that big data provides health and life sciences organizations are not unique and there are lessons learned in other industries that can be leveraged as best practices. In the retail world, rapid processing of large quantities of consumer data, such as socio-demographics and point-of-sale buying patterns, can be used to provide tailored offers to consumers as well as to reposition and re-price their products for maximum impact. Global financial organizations have begun to take a more holistic view of risk and are looking to model dozens of scenarios in near
“With growing collaboration in the pharmaceutical industry between pharmaceutical peers, pharmaceutical firms and medical institutions, and pharmaceutical firms and academia, stakeholders need to manage data access over time.”

Leach emphasized that data sciences must be put in the context of data lifecycle management. “Data sciences identify the relevant data sources,” he said. In collaborative research, which involves multiple partners and big data centers, networking technology can help stakeholders organize and securely transport data. Integrating and linking the data, which includes its meta-data and annotation, are also important. “Someone has to curate and annotate the data and build the dictionaries and semantic data representation,” he said. “As the volume gets bigger and as connections become more esoteric as we try to cross these silos, it might be the rise of semantic data representation.” Once data is integrated, according to Leach, users can ask much smarter questions to conduct analyses.

“Global financial organizations have begun to take a more holistic view of risk and are looking to model dozens of scenarios in near real time.”

Stakeholders must learn how to manage the storage, archiving, retrieval and reuse of data. Once a project is completed, organizations should be able to go back and know how the data was generated, how many times it has been processed and pre-processed, and what the data provenance was. With different subsets of data having different use agreements, organizations also need to be able to reconcile how data is combined and reused, which requires digging into the meta-data. “In virtually every company you look at today, trying to track down and figure out how can I reuse these samples as part of a bio-banking strategy, data aggregation and analysis is detective work,” said Leach. Next-generation technology may solve that pain point.

Merging Genomic and Clinical Data for Personalized Medicine

As healthcare integrates genomic and clinical data, the industry will be able to advance personalized medicine. Already, clinicians want to know what treatments were dispensed and outcomes occurred for patients who have the same condition as well as the same genomes, body characteristics and care preferences as their patients so they can determine what course to prescribe to get the best outcomes. “Personalized medicine is decision support based on evidence that’s driven by protocol and takes into account patient preferences and experience,” said Halamka.
As big data helps the industry deliver personalized medicine, it can also help healthcare organizations comply with healthcare reform mandates today. Provisions within the Affordable Care Act encourage the adoption of value-based care in an Accountable Care Organization (ACO) model. “Accountable care requires us to coordinate care and share data for different purposes and different care managers, patients and families,” Halamka said. With much of healthcare data in unstructured form, analytics need to be able to support who is doing what to whom and to what effect, he said. “We have a lot of gray areas around the data that needs to be cleaned up with more sophisticated natural language processing and semantic understanding of the techniques,” Halamka said.

Sometimes data analytics require centralization, although centralizing data increases the scope of potential data breaches and resulting HIPAA penalties. “It’s very hard to analyze a population unless you centralize some aspects of their data,” he said. BIDMC’s hospitals, clinics, labs and pharmacies are mandated to send certain data elements and meta-data to central repositories at the end of patient encounters so both retrospective and prospective analyses can be performed. Distributed and federated data mining techniques are becoming more popular and for many questions they are adequate. However, such a process would take a year to get through RIB (Revenues Increasing the Budget) process, meta-data consent and re-use issues. “We need to build a front door to all our data for two million patients that can parse any question using a series of ontologies,” he said. Because this new technique, which is being used across the country, involves the use of numerators and denominators rather than patient-identified data, RIB approval would not be required. “Big data is not going to be just centralized data; it’s going to be distributed data with more protections but yet better flow,” Halamka said.

Envisioning Big Data in the Near Future

Hughes hopes that the industry will focus on big value and big insights – the ability to take any question to the ever-increasing volumes of data without having to spend hours preprocessing and preparing those data, and to be able to embed the results of the analytics directly into clinical workflow and into the hands of patients. Leach is looking for a Google-like search capability that allows natural-language searches, straight-text searches, image-based searches, and structural and other types of searches. In the near future, he wants to be able to find the data he needs from various sources and seamlessly bring them together in an environment that allows him to expose the data to a whole ecosystem of applications through a common data exchange. What's learned in the clinics needs to be moved forward to the early drug discovery stage in order to customize and tailor medicines for populations of patients, according to Miller. Federating and integrating disparate bits of data will help speed the process and development of personalized medicine. For Halamka, the eventual product of big data will be event-driven medicine: when the data combines to signal a potentially emergent situation, clinicians are alerted and the data drives an actionable event that keeps the patient well.

As collaborations grow across disciplines and industries, new issues will arise in the use of big data. Collaborations on the technology side, however, will develop innovative solutions that will address current and emerging challenges. Innovation will realize these visions for all stakeholders, and ultimately help transform big data into actionable information to improve individual and population health, deliver efficacious therapies and bend the cost of care.