The last decade has witnessed a veritable explosion in the amount of data made available to healthcare providers and federal agencies. In 2012, experts estimated that more than 500 petabytes (1 petabyte = 1 thousand terabytes) of health data was created – and that figure is projected to grow to 25,000 by 2020.¹ The proliferation of health data from electronic health records (EHR), research trials, physician notes, medical devices, insurance claims, and social media feeds is driving a paradigm shift toward more innovative models of health service delivery, promising to reduce costs and improve the quality of care for all Americans.

**Big Opportunities for Big Health Data**

The key challenge facing federal health agencies is deploying the right set of tools to allow them to sift through immense targets of structured and unstructured data. Real-time analytics enabled by in-memory data platforms could provide government data scientists with the capabilities they need to separate the signal from the noise and derive actionable insights in areas ranging from public health to personalized care.

**Public Health Research**

Budgetary investment in the current fiscal climate proves the value that federal leaders are placing on big health data. In July 2013, the National Institutes of Health (NIH) announced a commitment of $96 million over four years to establish Big Data to Knowledge (BD2K) Centers of Excellence aimed at supporting data science to advance biomedical research. BD2K is a top NIH priority according to Director Francis S. Collins, who stated that its “goal is to help researchers translate data into knowledge that will advance discoveries and improve health, while reducing costs and redundancy.”²

A number of federal agencies are already putting big data analytics to use to improve public health outcomes. For instance, after a series of high-profile pharmaceutical recalls, the Food and Drug Administration (FDA) deployed Sentinel, a system that identifies latent safety risks in post-marketed medications by cross-referencing 125 million Americans’ EHRs with pharmaceutical trials and insurance claims databases.³ Between November 2012 and July 2013, Sentinel helped the FDA re-issue safety guidelines that are estimated to have saved hundreds of lives.⁴ In another part of the Department of Health and Human Services, officials from the Centers for Disease Control and Prevention (CDC) are pioneering ways to leverage...
data from medical records, social media, and over-the-counter antiviral medication sales to track the spread of influenza and allocate vaccine supplies accordingly.5

But few initiatives can match the scale of the Congressionally-launched Patient-Centered Outcomes Research Institute (PCORI), which aims to aggregate 30 million Americans’ complete medical records by September 2015. By tapping into data from millions of potential samples, PCORI researchers hope to compare the effectiveness of different treatments for conditions like cystic fibrosis, multiple sclerosis, and certain cancers.6

**Genomics and Personalized Medicine**

Significant technological advances have reduced the cost of sequencing the human genome by a factor of a thousand, creating opportunities for both public and private research institutions to uncover the root causes of chronic conditions like diabetes and Alzheimer’s disease.7 However, unlocking the human genome’s power to improve healthcare quality and accelerate innovations in personalized medicine will require powerful analytic tools capable of recognizing molecular patterns in thousands of samples, each containing more than 3 billion base pairs of genetic code. Applying new technologies to genome sequencing could exponentially reduce diagnostic timelines as well as identify targeted protocols for treatment.

One field with perhaps the greatest potential for innovation is the mapping of genetic signatures of aggressive cancer strains. In January 2014, the National Cancer Institute (NCI) announced the launch of a public cloud to house its 2.5-petabyte centralized database, the Cancer Genome Atlas, for further collaborative research. This data has already proven incredibly useful; NCI data was at the center of recent findings that cancer treatment may be more effective when targeting the genetic cause of mutation rather than its tissue type.8

Medical genetics can also lay the foundation for more personalized and preventive – and therefore less expensive – care. In an article in the *New England Journal of Medicine*, leaders from NIH and FDA described a not-so-distant future in which patients’ entire genomes could be integrated into their EHR, allowing doctors to more accurately prescribe medication and replace expensive DNA testing with a simple electronic query.9

**Evidence-Based Medicine**

Garnering better insight from harnessing data can also help federal health agencies adopt the practice of evidence-based medicine (EBM). EBM involves supplementing physicians’ medical expertise with real-time access to clinical data from hundreds of thousands of similar cases to pinpoint the most accurate diagnosis and effective course of treatment for each patient.10 For example, the Department of Veterans Affairs (VA), the largest primary healthcare provider in the federal government, is working to develop interoperable data models for the Veterans Health
Agency and Defense Health Agency to provide doctors with a complete picture of veterans’ medical histories spanning the length of their service careers. Data shared between the VA and DoD could be a valuable asset in cataloguing and monitoring chronic conditions like traumatic brain injury and post-traumatic stress disorder, allowing physicians to better employ EBM to improve long-term outcomes.

In much the same way, the Centers for Medicare and Medicaid Services (CMS) is collaborating with other public and private institutions to explore the application of advanced analytics in comparative-effectiveness research. By pooling its claims database alongside surgical records or pharmaceutical trials, CMS can compare alternative procedures or medications longitudinally and evaluate relative risks and benefits before issuing best practice recommendations.

Translating Health Data into Health Knowledge with In-Memory Data Platforms

The cases noted above illustrate the opportunities available for federal agencies that adopt a forward thinking, data-centric mindset and invest in analytics. However, big data is only as powerful as the platform it runs on. For true real-time analytics capabilities, federal agencies will need the speed and computing power supplied by in-memory database (IMDB) technologies. Unlike traditional database management systems, IMDBs such as SAP HANA can store massive data sets in the form of RAM and eliminate the bottlenecks that occur with disk processing. Coupled with an architecture that combines predictive text analytics, spatial processing, and data virtualization, this translates to the ability to process information at speeds never before possible. Petabytes of health data produced by EHR, genome sequencing, clinical trials, and many other sources can be analyzed in a fraction of the time taken with traditional analytic models.

The big data revolution in healthcare is here and here to stay. For federal agencies, the move to a next-generation real-time data platform could mean the difference between being overwhelmed with health data and being part of the next major medical breakthrough.
Sources


Images: National Institutes of Health and Wikipedia Creative Commons user Calleamanecer