Big Data in Healthcare Hype and Hope

Authors: **Bonnie Feldman**
Ellen M. Martin
Tobi Skotnes

Date: **October 2012**

Dr. Bonnie 360°
Business Development for Digital Health

About the Authors

Bonnie Feldman, D.D.S., M.B.A.

As principal of DrBonnie360 (formerly Feldman Stakeholder Relations), Bonnie brings a 360-degree view of private and public healthcare to her consulting work, which includes market research and business development in newly emerging markets.

She has earned a broad and deep understanding of the players and the playing field in Health 2.0/3.0, life science capital markets, and healthcare practice. Prior to this original research on the emerging Big Data landscape, she created a first of its kind industry overview *"Mobile, Social and Fun: Games for Health,"* published by MobiHealthNews, which was well received as a presentation at the Games for Health Conference.

On the analytic side, after working as a health services researcher at the Rand Corporation, she worked on Wall Street as a buy-side and sell-side equity research analyst. She has provided investor relations services both inside companies and on a professional consulting basis.

In clinical practice, as an entrepreneur, she built and owned two dental practices, managing finances, business development, staffing, operations and patient care as well as performing claims review consulting for Prudential.

She holds a BA in Economics, a Doctor of Dental Surgery, and an MBA in Finance from the University of California, Los Angeles.

Ellen M. Martin, M.B.A.

For more than 20 years, Ellen has provided professional communications services to life sciences and high-tech organizations including biopharmaceutical, medical device, healthcare, bioinformatics, genomics tools and IT firms. She is affiliated with Kureczka|Martin Associates, DrBonnie360, and is an owner and an employee of Haddon Hill Group Inc.

Based on her background and ongoing interests in multiple science fields, she concisely communicates complex ideas to a variety of audiences. Recently published writing assignments include articles on Big Data in Healthcare, Legacy Systems in Biomanufacturing, and Mobile Health Games.

She led corporate communications for five years at XOMA, and was an early adopter of the Internet for investor relations. Earlier, she led communications for DNA Plant Technology, and for the Bay Area Bioscience Center (forerunner to BayBio), while consulting to bio- and high-tech clients, solo and as an associate with J. Kureczka Associates.

Ellen holds an MBA in marketing and finance from Northwestern University's Graduate School of Management (now Kellogg), a BS in physical anthropology from the University of Illinois (Champaign-Urbana), and pursued graduate studies in and taught archeology, genetics, osteology and paleopathology.

Tobi Skotnes

Tobi Skotnes is a senior undergraduate at University of California, Los Angeles, with a math major and a Spanish minor. She is interested in pursuing a career in public health, using analytic and quantitative methods to address public health problems around the world. She worked as a volunteer in a medical clinic in Nicaragua for a short period in 2010, and spent 2011 living in Granada, Spain. Summer 2012 she worked under Professor Sally Blower, Head of the Disease Modeling Group at the David Geffen School of Medicine, at the Semel Institute for Neuroscience and Human Behavior at UCLA, on modeling the spread of HIV in Lesotho, Africa.

Table of Contents

About the Authors 2

Introduction 5

What is Big Data? 7

Who Cares and Why? 13

The Companies: An Evolving Ecosystem 17

I. Supporting Research - Genomics and Beyond 17

II. Transforming Data to Information (and Information to Data) 22

III. Supporting Self-Care 26

IV. Supporting Providers, Improving Patient Care 29

V. Increasing Awareness 32

VI. Pooling Data to Build a Better Ecosystem 35

Issues and Challenges of Big Data 38

Three Trends for the Use of Big Data within an Emerging Ecosystem 42

The Future of Big Data in Healthcare 44

Bibliography 47

Acknowledgements 53

Introduction

Digitized information is ubiquitous, a digital flood creating puddles and lakes, creeks and torrents, of data: numbers, words, music, images, video. Most recently, the rapid growth in the use of mobile devices—smart phones, laptops, tablets, personal sensors—is generating a data deluge; most of the world's data has been created within the last two years.[1] For the more than two billion people[2] who use the Internet for email, Facebook (70 petabytes and 2700 multiprocessor nodes itself)[3], LinkedIn, Twitter, commenting, blogging, or downloading information and entertainment, digital data flows in a deepening river through our everyday lives, feeding an ocean of global information and noise.

Extremely large data volumes at high velocities (known as Extreme or Big Data), were originally the realm of supercomputers, nuclear physics, military simulations and space travel. Late in the 20th century, bigger and faster data proliferated in airline and bank operations, particularly with the growth of credit cards. Starting in 1990, The Human Genome Project was the moon launch of Big Data in healthcare, a data-intensive research effort that pushed the limits of available data processing technology. Increasingly powerful hardware and software, improvements in IT data management and integration, new analytics tools, and accumulating experience using Big Data in finance, research, entertainment and consumer marketing, are building a foundation for the increasing use of Big Data and analytics in healthcare.

The potential of Big Data allows us to hope to slow the ever-increasing costs of care, help providers practice more effective medicine, empower patients and caregivers, support fitness and preventive self-care, and to dream[4] about more personalized medicine. Yet, as with the Internet, social media, and cloud computing, early enthusiasts are creating hyperbolic expectations about how and how quickly Big Data will transform healthcare.

A number of issues challenge the adoption and success of healthcare Big Data, including privacy and security, who owns the data, and the regulatory labyrinth. Furthermore, real advances depend on better ways to exploit the disconnected puddles and lakes of existing data (e.g., health records, clinical trial data, actuarial information) as well as better ways to generate, capture, analyze and make use of the streams of new kinds of data (genomics, sensor readings, population and disease tracking) that are about to flood healthcare.

This report will introduce readers to Big Data and explore how it is becoming a growing force in the changing healthcare landscape. Using the power of the Internet, we researched the coming of Big Data to healthcare, and then interviewed, in person, by phone and via email, more than 30 companies in the emerging healthcare Big Data ecosystem.

1 http://www-01.ibm.com/software/data/bigdata/
2 http://www.internetworldstats.com/stats.htm
3 http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-healthcare-analytics.html
4 http://radar.oreilly.com/2012/08/data-health-care.html

New Streams of Data

Over the next 3 years

+1 billion
smart phones will enter service

3 billion
IP-enabled devices by 2015

By 2016

4.9 million
patients will use remote health monitoring devices

3 million
patients will use a remote monitoring device via smartphone hub

142 million
healthcare and medical app downloads

The Healthcare Data Explosion

2012
500 petabytes

Worldwide healthcare data is expected to grow to **50 times** the current total

2020
25,000 petabytes

What is Big Data?

"Big Data" is a hot topic.

A recent New York Times article[5] discusses the evolution of the term "Big Data." Another[6] shows the power of Big Data in consumer marketing, enabling Target to identify women who were likely to be pregnant in an effort to secure them as long-term customers. A third[7] identifies Big Data as the next wave of technology change, as revolutionary as personal computers in the 1980s, the Internet in the 1990s and smart phones today.

A sure sign of topicality is a Colbert[8] satirical view of Big Data and, a Dilbert[9] comic strip. Moreover, there was intense media coverage[10] of IBM's Watson's successful debut on Jeopardy (demonstrating powerful new natural language capabilities in a computer).

"Big Data" is a catch phrase with multiple definitions:

- **Wikipedia**: "...data sets so large and complex that [they are] awkward to work with using on-hand database management tools. Difficulties include capture, storage, search, sharing, analysis, and visualization."[11]
- **O'Reilly Radar**: "... data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures. To gain value from this data, you must choose an alternative way to process it."[12]
- **ZDNet**: "In simplest terms, the phrase refers to the tools, processes and procedures allowing an organization to create, manipulate, and manage very large data sets and storage facilities."[13]

Experts interviewed for this paper brought other perspectives:

- **Stephen Gold**, VP of Marketing for IBM's Watson: "Every day, we create 2.5 quintillion bytes of data — 90% of the data in the world today has been created in the last two years alone. Big Data is the fuel. It is like oil. If you leave it in the ground, it doesn't have a lot of value. But when we find ways to ingest, curate, and analyze the data in new and different ways, such as in Watson, Big Data becomes very interesting."

[5] http://www.nytimes.com/2012/08/12/business/how-big-data-became-so-big-unboxed.html?_r=1
[6] http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all
[7] http://www.nytimes.com/2012/09/09/technology/data-driven-discovery-is-techs-new-wave-unboxed.html?_r=1
[8] http://e-inteam.com/BI/stephen-colbert-explains-big-data-and-predictive-analytics/
[9] http://dilbert.com/strips/comic/2012-07-29/
[10] http://blogs.ngm.com/blog_central/2011/02/jeopardy-20-ill-take-computers-who-form-questions-for-answers-for-100.html
[11] http://en.wikipedia.org/wiki/Big_data
[12] O'Reilly Radar (http://s.tt/1kHFU)
[13] http://www.zdnet.com/blog/virtualization/what-is-big-data/1708

- **Don Jones**, Vice President of Global Strategy & Market Development at Qualcomm Life suggests that "because we are bringing together sources of data that have never been brought together before, even if the amount of data isn't particularly large, it is Big Data, because you never had it all in one place."
- **Martin Leach**, the Chief Information Officer at The Broad Institute of MIT and Harvard, suggested that "Big is a relative term; now Big Data is about accessibility of data and how to bring it together to create value."

There are four main "dimensions" to Big Data, commonly referred to as the Four Vs (or three, or five, depending on the source):

1	**Volume**	=	quantity, from terabytes to zettabytes
2	**Variety**	=	structured, semi-structured and unstructured
3	**Velocity**	=	from any-time batch processing to real-time streaming
4	**Veracity**	=	quality, relevance, predictive value, meaningfulness

How does each of these dimensions apply to healthcare data?

Volume: New healthcare data streams swell exponential growth

The volume of global data overall is increasing exponentially, from 130 exabytes (an exabyte is 10^{18} bytes of data) in 2005 to 7,910 exabytes in 2015.[14] By 2020, there will be 35 zettabytes (10^{21} bytes) of digital data—a stack of DVD's that would reach halfway from the Earth to Mars.[15]

However, only 20% of the world's data is structured (suitable for computer processing), with unstructured data (e.g., handwritten notes, untagged text, audio and video files) growing at 15 times the rate of structured data.[16] In the next 3 years, more than 1 billion smartphones will enter service, 400 million new tablets will connect to the Internet and there will be 1 billion active personal computers in the world.[17]

In healthcare, growth comes both from digitizing existing data and from generating new forms of data. The already daunting volume of existing healthcare data includes personal medical records, radiology images, clinical trial data, FDA submissions, human genetics and population

[14] http://www.hsph.harvard.edu/news/features/files/big_data.pdf
[15] http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-healthcare-analytics.html
[16] https://www.ibm.com/developerworks/wikis/display/db2oncampus/FREE+ebook+-+Understanding+Big+Data
[17] http://pro.gigaom.com/2012/03/a-near-term-outlook-for-big-data

data, genomic sequences, etc. Newer forms of big byte data, such as 3D imaging, genomics and biometric sensor readings, are also fueling this exponential growth.

The volume of worldwide healthcare data in 2012 is 500 petabytes (10^{15} bytes) = 10 billion four-drawer file cabinets. That is estimated to grow in 2020 to 25,000 petabytes = 500 billion four-drawer file cabinets—a fiftyfold increase from 2012 to 2020.[18]

Advances in data management, particularly virtualization and cloud computing, are facilitating the development of platforms for more effective capture, storage and manipulation of large volumes of data. Storing information "in the cloud" for access by desktop PCs and mobile devices allows small devices and single locations to become windows into a universe of information.

Many companies (not all of them particularly focused on healthcare) are working to further advance data management platforms and frameworks. This includes traditional IT vendors like IBM, Cisco Systems Inc., and Oracle Corporation; platform companies like Google Inc. and Amazon.com, Inc., open source groups like The Apache Software Foundation (Hadoop), The Linux Foundation, Mozilla Foundation and Corporation, plus a myriad of smaller organizations and individual developers.

In the universe of companies interviewed for this paper: DNAnexus, Appistry, NextBio and Genome Health Solutions are building products and services that rely on and enable their customers to manage extreme data volumes.

Variety: healthcare data sources and complexity

The enormous variety of data—structured, unstructured and semi-structured—is a dimension that makes healthcare data both interesting and challenging. Historically, the point of care generated mostly unstructured data: office medical records, handwritten nurse and doctor notes, hospital admission and discharge records, paper prescriptions, radiograph films, MRI, CT and other images.

Structured data is data that can be easily stored, queried, recalled, analyzed and manipulated by machine (although humans may not so easily read or interpret them). Historically in healthcare, structured and semi-structured data include electronic accounting and billings, actuarial data, (some) clinical data, (some) laboratory instrument readings and data generated by the ongoing conversion of paper records to electronic health and medical records.

Already, new data streams, structured and unstructured, are cascading into the healthcare river from fitness devices, genetics and genomics, social media, research and other sources. Relatively little of this data can presently be captured, stored and organized so that they can be manipulated by computers and analyzed for useful information. Healthcare applications

[18] http://visual.ly/future-healthcare

particularly need more efficient ways to combine and convert varieties of data, including automating conversion from structured to unstructured data.

The structured data in electronic medical records (EMRs) and electronic health records (EHRs) include familiar input record fields such as patient name, date of birth, address, physician's name, hospital name and address, treatment reimbursement codes, and other information easily coded into and handled by automated databases. The need to field-code data at the point of care for electronic handling is a major barrier to acceptance of EMRs by physicians and nurses, who lose the natural language ease of entry and understanding that handwritten notes provide. On the other hand, nearly all providers agree that an easy way to reduce prescription errors is to use digital entries rather than handwritten scripts.

IBM is an obvious instance of a big company tackling the problem of using varied data sets. Watson, with its unique natural language capabilities, is the primary example. Also in this paper's universe, Health Fidelity is using natural language processing to convert unstructured into structured data. Other companies dealing with data variety include Explorys, Practice Fusion, athenahealth Inc., Humedica, and One Health.

The potential of Big Data in healthcare lies in combining traditional data with new forms of data, both individually and on a population level. We are already seeing data sets from a multitude of sources support faster and more reliable research and discovery. If, for example, pharmaceutical developers, can integrate population clinical data sets with genomics data, they may move closer to getting more and better drugs approved in the first place, and more importantly, to getting the right drug to the right patient at the right time.

Velocity: healthcare data at rest and in motion

The constant flow of new data accumulating at unprecedented rates presents new challenges. Just as the volume and variety of data that is collected and stored has changed, so too has the velocity at which it is generated and the speed needed to retrieve, analyze, compare and make decisions using the output. The migration from checks to credit cards is a familiar example of the move from slow, batch-processed data handling to real-time data processing.

Most healthcare data has traditionally been quite static—paper files, X-ray films, scrips. But in some medical situations, real-time data (trauma monitoring for blood pressure, operating room monitors for anesthesia, bedside heart monitors, etc.) become a matter of life or death. In between are the medium-velocity data of multiple daily diabetic glucose measurements (or more continuous control by insulin pumps), blood pressure readings, and EKGs.

Future applications of real-time data in the ICU, such as detecting infections as early as possible, identifying them swiftly and applying the right treatments (not just broad-spectrum antibiotics), could reduce patient morbidity and mortality or even stop hospital outbreaks. Real-time streaming data can already monitor neonates in the ICU, to predict life-threatening

infections sooner.[19] Being able to perform real-time analytics against such high-volume data in motion could revolutionize healthcare.

Medical device companies, like Baxter International, Boston Scientific Corporation, Hospira, Inc., Medtronic Inc., and Zoll Medical Corporation have been at the forefront of capturing and displaying real-time data in ambulances, operating rooms, hospitals, and increasingly, at home. Smaller companies such as Abiomed Inc., Alere Inc., and ResMed Inc. are particularly focused on home monitoring devices. Diagnostics is an arena where many new devices, techniques and algorithms are being developed, even if few have reached the market.

In this paper's company universe, Ginger.io, Zeo, Predixion Software, Asthmapolis, SickWeather and Sproxil are all developing ways to cope with and get more meaning from faster data.

Veracity - Data of varying quality, relevance and meaning

Traditional data management assumes that warehoused data is certain, clean, and precise. However, as anyone who has suffered bank, credit or insurance errors knows, data is sometimes uncertain, imprecise or just plain wrong.

Data quality issues are a particular concern in healthcare for two reasons: 1. It matters—life or death decisions depend on having the information right. 2. The quality of healthcare data, especially unstructured data, is highly variable and all too often incorrect. Unreadable handwritten prescriptions are perhaps the most infamous example.

Veracity in healthcare data faces many of the same issues as in financial data, especially on the payer side: Is this the correct patient, hospital, payer, reimbursement code, dollar amount? Other veracity issues are unique to healthcare: Are diagnoses, treatments, prescriptions, procedures, outcomes correctly captured?

Improving coordination of care, avoiding errors and reducing costs depend on high-quality data, as do advances in drug safety and efficacy, diagnostic accuracy and more precise targeting of disease processes by treatments. However, high Variety and Velocity hinder the ability to cleanse data before analyzing it and making decisions, raising issues of data "trust."

Underlying statistical issues bedevil large data sets, not just "garbage in, garbage out". The emergence of safety issues in marketed drugs that successfully completed large clinical trials shows that rare events may not be visible in any but the largest data sets. A deeper concern is the dawning realization that treatments targeting the average patient will, by definition, **mis**treat a significant portion of the population. The predictive value of analytic tools will not be realized if the data sets being analyzed are low quality or represent irrelevant measures.

While many organizations are concerned with data quality in healthcare, they are focused on traditional IT issues, e.g., data management, warehousing, compliance, audit, fraud prevention,

[19] http://www.youtube.com/watch?v=ZiqY7p1v950

error reporting and regulatory compliance. There appears to be less focus on the underlying veracity (truthfulness, relevance, predictive value) of the data, an issue that will be of increasing importance in realizing the vision of healthcare Big Data analysis and personalized medicine. For example, Watson's ability to provide accurate provider decision support relies on the completeness and accuracy of the data it can access.

Companies working on data quality issues include IBM and Google, as well as many open source and standards organizations. Companies interviewed for this paper with a focus on data quality and veracity include NextBio, Appistry, Explorys Inc., Humedica and Predixion Software.

Who Cares and Why?

Healthcare costs are unsustainably increasing worldwide. Today, the portion of US GDP spent on healthcare (~17%) is about the same as that spent on entertainment (although the high cost of entertainment is not yet a contentious public policy issue)[20]. But most estimates of future costs show healthcare growing disproportionately in all national economies. Why should this be? Several forces are widely acknowledged:

- **Demographics**: aging populations, more chronic conditions, increasingly intense and expensive end-of-life care
- **Technology**: advanced high-tech medicine is more expensive to deliver than the lower-tech breakthroughs of the 20th century
- **Quality of care**: uncoordinated care, inefficient workflows, medical errors, readmissions, hospital-acquired infections, nurse and physician shortages
- **Structural issues**: institutional inefficiencies, fraud, waste, market distortions (third-party payers, reimbursement regimes), regulatory overhead, defensive medicine

The growth rate of US healthcare expenditures, increasing annually by nearly 5% in real terms over the last decade, is unsustainable and a major contributor to the high national debt levels projected over the next two decades.[21]

"Valuing Health Care: Improving Productivity and Quality", a Kauffman Report, points to a combination of insufficient information, poor incentives for cost control, and inefficiencies in healthcare R&D all leading to waste.[22] By one estimate, of the $2.5 trillion spent on healthcare in the US in 2010, $700 billion was not necessary.[23]

Based on the widely accepted position that the continued acceleration of costs is unsustainable, the Report to the President Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward, December 2010,[24] suggests that capturing, storing and analyzing medical information electronically could facilitate better quality of care, improved population-based knowledge and the development of new tools for medicine.[25]

Overall, most stakeholders agree on the need to reduce costs and improve the efficiency and effectiveness of the US healthcare system. Nevertheless, reducing costs is not the only goal.

[20] http://hbr.org/2011/09/how-to-solve-the-cost-crisis-in-health-care/ar/1
[21] http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation
[22] http://www.kauffman.org/uploadedfiles/valuing_health_care.pdf
[23] http://www.kauffman.org/uploadedfiles/valuing_health_care.pdf
[24] http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf
[25] http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf

The vision is to enable patient-centered care, where researchers understand the causes of disease and developers design effective treatments that can be tuned to each individual's biology. This includes rapid, precise and accurate diagnostics as well as the right match between the diagnosis, the patient's biology and the treatment. Oncology in particular, where the cancer and the patient are genetically unique and the therapeutic window is narrow, needs better ways to diagnose, characterize and match the treatment to the patient and disease. Farther out, the dream is to provide coherent and accurate decision support to caretakers, doctors, nurses, and patients themselves, for disease treatment and management or, even better, for effective prevention.

Healthcare Stakeholders envision different goals

Different healthcare stakeholders have different goals and hopes for Big Data and analytics:

- ***Patients*** want their everyday use of technology to flow seamlessly into their medical care. Some want to comparison shop for medical treatment as they do for consumer products. Everyone wants customer-friendly service, one-stop shopping, and better coordination of care between themselves, caregivers and various providers, with an ultimate goal of error-free, compassionate and effective care.
- ***Providers*** want real-time access to patient, clinical and other relevant data to support improved decision-making and facilitate effective, efficient and error-free care. They want technology to be a transparent tool, not an encumbrance.
- ***Researchers*** want new tools to improve the quality and quantity of workflow – e.g., predictive modeling, statistical tools and algorithms that improve the design and outcome of experiments and provide a better understanding of how to develop treatments that meet unmet needs while successfully navigating the regulatory approval and marketing process.
- ***Pharma companies*** want to better understand the causes of diseases, find more targeted drug candidates, and design more successful clinical trials to avoid late failures and market safer and more effective pharmaceuticals. Once in the market, they want accurate formulary and reimbursement information to customize their marketing efforts, as well as less costly post-marketing surveillance.
- ***Medical device*** companies, many of which have been collecting data for some time from hospital and home devices for safety monitoring and adverse event prediction, are beginning to wonder what to do with this data, and how to integrate it with old and new forms of personal data.
- ***Payers*** are moving from fee-for-service to pay-for-performance, and want to use Big Data to help stratify population risk, guide them to more sustainable business models such as expanding into wellness management, data analytics and API platforms.
- ***Governments*** are trying to reduce costs, enforce regulations and maximize the social value of data. Examples include incentives outlined in the HITECH Act (designed to

accelerate the adoption of EHR systems among providers), and the creation of prizes and incentives such as Datapalooza. The Blue Button initiative, available to veterans, uniformed service members, and Medicare beneficiaries, has enabled almost a million people to download their own health information.

- ***Software Developers*** see opportunities to serve an enormous and growing market that is a decade behind other industries (banking, telecommunications, consumer operations and marketing, travel reservations, insurance) that IT has already transformed.

Everyone agrees that reducing waste and inefficiency are worthwhile goals. And while there are still conflicting incentives, all stakeholders would pay at least sincere lip service to the goal of rapid, error-free, compassionate care.

Data pools and problem solving

The health data universe falls into six big, disconnected pools of data owned by different stakeholders with competing demands and aspirations, all of which could benefit from better integration of these data.

1. **Providers**: clinical/medical data (EHRs)
2. **Payers and Providers**: claims and cost data
3. **Researchers**: academic, independent, government
4. **Developers**: pharma and medical device R&D, including clinical trials
5. **Consumers and Marketers**: patient behavior and sentiment data
6. **Government**: population and public health data

McKinsey[26] estimates that Big Data can enable more than $300 billion savings per year in US healthcare, with two-thirds of that through reductions of around 8% to national healthcare expenditures. Clinical operations and R&D are two of the largest areas for potential savings, with $165 billion and $108 billion in waste respectively. McKinsey believes Big Data could help reduce waste and inefficiency in the following three areas:

1. Clinical Operations

- Comparative effectiveness research to help determine more clinically relevant and cost-effective ways to diagnose and treat patients. However, Medicaid and Medicare cannot yet apply comparative effectiveness, leaving a big gap in its impact. Moreover, there are caveats about comparative effectiveness as currently envisioned (e.g., in the ACA), which will probably improve care for patients with common diseases and be woefully inadequate for those with rare ones.

[26] http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation

- Clinical decision support systems to enhance the efficiency and quality of operations; i.e., providing real-time information to emergency technicians, nurses and doctors to improve triage, diagnosis, treatment choice, prevent iatrogenic infections and readmissions, prescription and other medical errors.
- Other areas include increasing transparency about medical data, remote patient monitoring, and predictive analytics to identify individuals who would benefit from proactive care.

2. R&D

- Predictive modeling could help produce a leaner, faster, more targeted, and lower attrition R&D pipeline in drugs and devices.
- Statistical tools and algorithms could improve clinical trial design and patient recruitment to better match treatments to individual patients, thus reducing trial failures and speeding new treatments to market.
- Analyzing clinical trials and patient records to identify follow-on indications and discover adverse effects before products reach the market.
- Ultimately, personalized medicine, emerging from the analysis of large datasets to help match the right medicine to the right patient at the right time.

3. Public Health

- Analyzing disease patterns and tracking disease outbreaks and transmission to improve public health surveillance and speed response.
- Faster development of more accurately targeted vaccines, e.g., choosing the annual influenza strains.
- Turning torrents of data into actionable information that can be used to identify needs, provide services, and predict and prevent crises, especially for the benefit of lower income populations.[27]

The goals of organizations such as Global Viral Forecasting, UN Global Pulse, Ushahidi, and Google Flu Trends, will be discussed in another report.

[27] http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf

The Companies: An Evolving Ecosystem

An increasing number and variety of organizations, from independent not-for-profit initiatives, to entrepreneurial enterprises and even large payers/providers, are beginning to harness Big Data to address multiple different healthcare challenges. In this section, we will summarize information gleaned from more than 30 interviews with companies and organizations that can be placed in six groups, reflecting six ways Big Data could help healthcare:

I. Support Research - Genomics and Beyond
II. Transform Data to Information
III. Support Self-Care
IV. Support Providers - Improve Patient Care
V. Increase Awareness
VI. Pool Data to Build a Better Ecosystem

I. Supporting Research - Genomics and Beyond

Genomics has been the cutting edge of the Big Data revolution in the life sciences, one that holds considerable (if yet-to-be-delivered) promise for enabling personalized medicine. While all of these companies are genomics-focused, each of them is taking a different approach to the data, hoping to accelerate translational research and ultimately, transform treatment development and medical practice.

Genome Health Solutions applies its expertise and network of physicians and technology providers to integrate personal genomics and streamline care delivery to make possible a new standard of care for improving patient outcomes in cancer and other diseases.

- Currently GHS serves as an integrator to design and implement genomics oncology workflows to get patients to the right doctors at the right time with the right diagnosis and treatment options.
- Pilot program for self-pay patients and compassionate use cases to design, prototype and develop more automated and scalable processes to demonstrate comparative effectiveness, cost avoidance and better outcomes in large patient populations.
- Creates continuing medical education resources and decision-support systems for healthcare professionals as well as educational materials for “empowered patients.”
- GHS provides expert personalized genomic medicine consulting and customized, cost effective solutions to healthcare delivery organizations including “build, buy or outsource” alternatives, and the design and management comparative effectiveness research programs.

Genome Health Solutions' founder, Dr. Mark Boguski of Harvard Medical School, has a passionate desire to bring the power of human genome knowledge and technologies to improve patient care. He proclaims, "In the past genomics has over-promised and under-delivered with respect to influence on medical practice and improving human health. But we're now in the 'third wave' of genomic medicine which I firmly believe will lead to better health outcomes through precision diagnosis."

GNS Healthcare builds mathematical cause-and-effect models to determine drivers of outcomes.

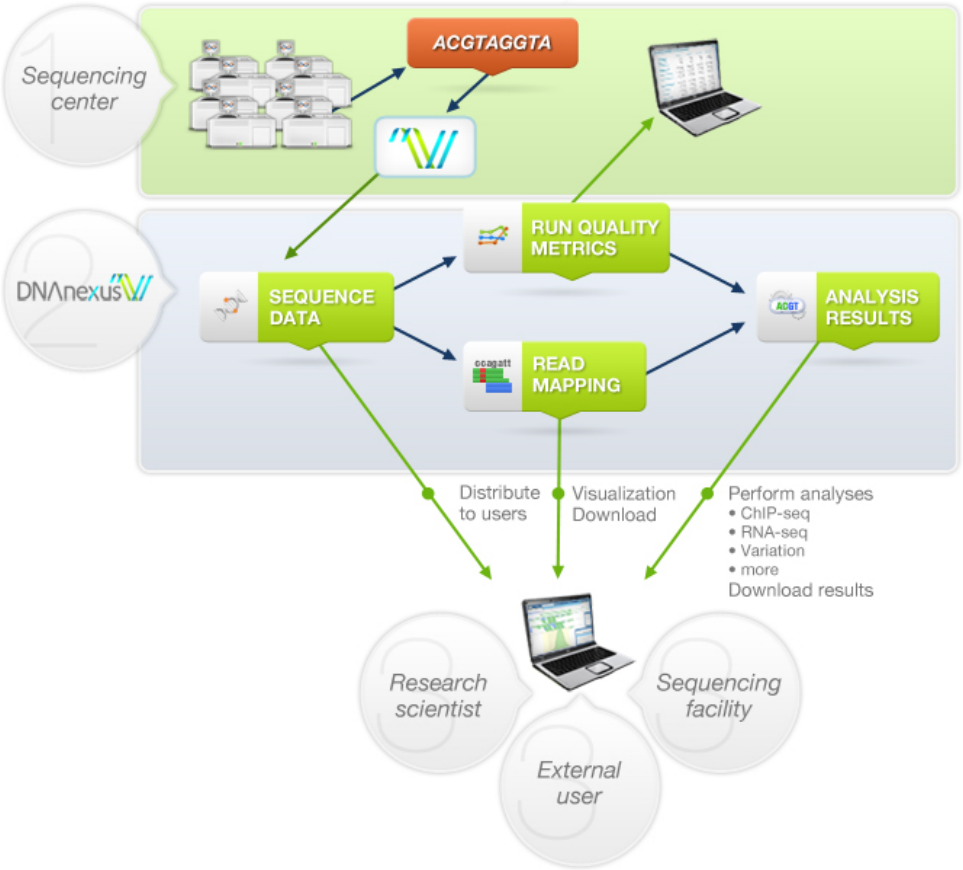
- Uses REFS™ (Reverse Engineering Forward Simulation), a proprietary, scalable supercomputer-backed framework to infer causal network models directly from observational data and create visual interactive simulations.
- Allows researchers to dynamically explore outcomes for different interventions, which speeds time to insight.
- Directed towards pharmaceutical product development, payers and providers.
- Current work with Aetna Innovation Labs will apply REFS™ technology platform to rapidly predict and match at-risk profiles with possible intervention combinations for metabolic syndrome.

Carol McCall, the Chief Strategy Officer of GNS, enthusiastically describes their approach as "What we do is Models-in-a-Jar with embedded analytics. You put in the data and get a high resolution interactive visualization that represents the underlying complex structures (causation) that are implied by the data."

DNAnexus offers a cloud-based, community-inspired, collaborative and scalable data technology platform that provides next-generation sequencing (NGS) data management, analysis, and visualization.

- Enables customers to store, manage, analyze, and visualize next-generation DNA sequencing (NGS) data through a web-based cloud service model.
- Building a community-inspired cloud infrastructure from the ground up, to create a collaborative and scalable data platform.
- Encryption, firewalls, and other safeguards support enterprise security and compliance with HIPAA, CLIA, and other regulations.
- Customers include research scientists and clinical research partners in pharma and biotech, as well as physicians using genomics in individual cases for diagnosis and treatment guidance.

Marc Olesen, President and COO, emphasizes their open approach. He told us, "We plan to stay focused on sequence data as we strive to drive this community-powered ecosystem so that developers can develop an app, and then researchers around the world can take advantage of having the information (tools) in one place."

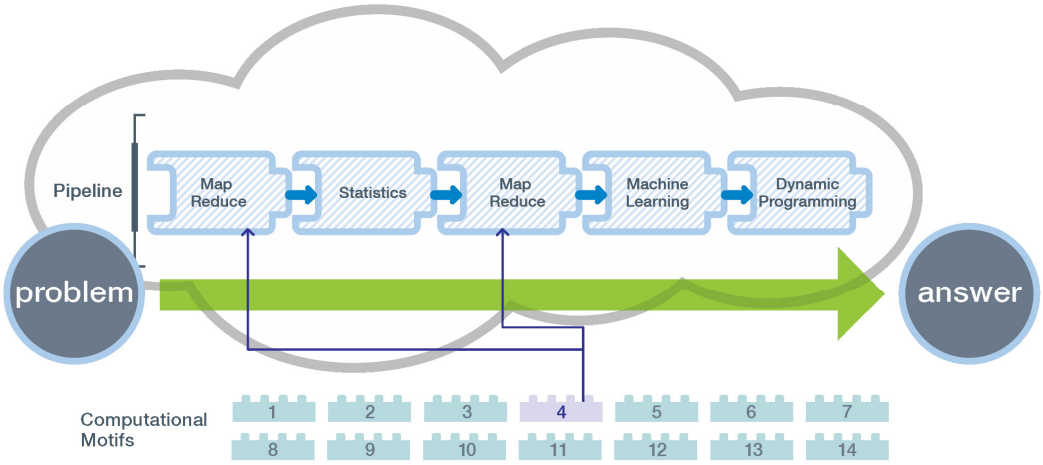


Appistry Inc. adapts learning from FedEx and the Department of Defense to streamline the storage, management, analysis and interpretation of Big Data in genomics, including custom development.

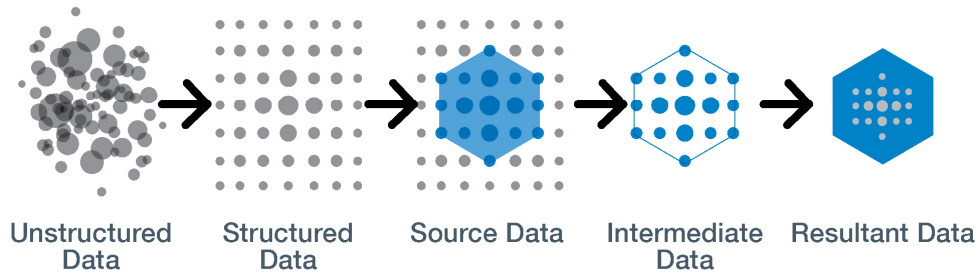
- Brings computation and storage together in a private cloud to improve the speed and performance of genetic analysis.
- Produces easy-to-read reports that can be used by doctor's, patients, pathologists and hospitals.
- Partners include Brown University, Stanford and CU Boulder.

"Genetic information by itself is useless, unless we can put it into context for the patient, commented Sultan Meghi, the Vice President of Product Strategy of Appistry, "With government funding in genomics decreasing, it is up to the private sector to use the available genetic information and make the information useful, [so] we are reinventing how the science is deployed—making it entirely automated and scaled—partnering with academic and research institutions so that they can focus on the science."

THE AYRRIS™ APPROACH



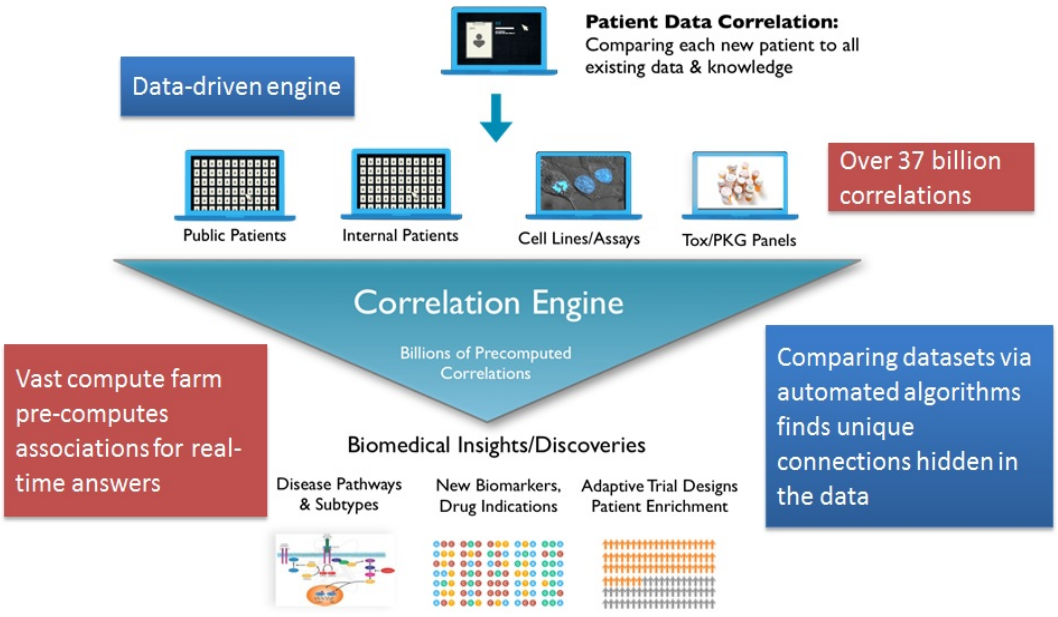
UNSTRUCTURED DATA TO RESULTS



NextBio combines large public with private datasets to enable new -omics discoveries.

- Assembles vast amounts of curated and annotated clinical and molecular data enabling clients to make unique discoveries that would not be possible with their own private datasets alone.
- Uses Big Data technology to make correlations between the billions of data points from the public domain with private genomic and clinical data sets.
- Delivered as Software as a Service (SaaS).
- A rich set of APIs enable clients to integrate NextBio within their workflows.
- Current clients include Pharmaceutical R&D and academic medical centers.
- Initial focus on oncology, now expanding into metabolic and autoimmune diseases.

"The future of medicine and medical research lies in genomics and the use of other "-omics" data, said Dr. Alpana Verma-Alag, the Head of Clinical Development of NextBio. "Researchers can bring their own private data, which gets correlated with the semantically enriched public datasets to identify new biomarkers, drug targets and mechanisms of disease. We are currently expanding [from oncology] into metabolic and autoimmune diseases."



II. Transforming Data to Information (and Information to Data)

Given the growing flood of healthcare data, and the late-adopting nature of the field, a big unmet need is to better manage this data. A key aspect is transforming data to usable information**.** Making unstructured data structured for machine management is an important stepping-stone to enabling data-driven healthcare. Perhaps counter-intuitively, in some cases turning unstructured information (medical charts and provider notes) into data is a needed first step. Explanatory analytics uses a collection of tools based on data mining, cluster analysis, statistics, data visualizations, artificial intelligence machines, text analytics, and Natural Language Processing (NLP) to mine data for patterns and meaning.

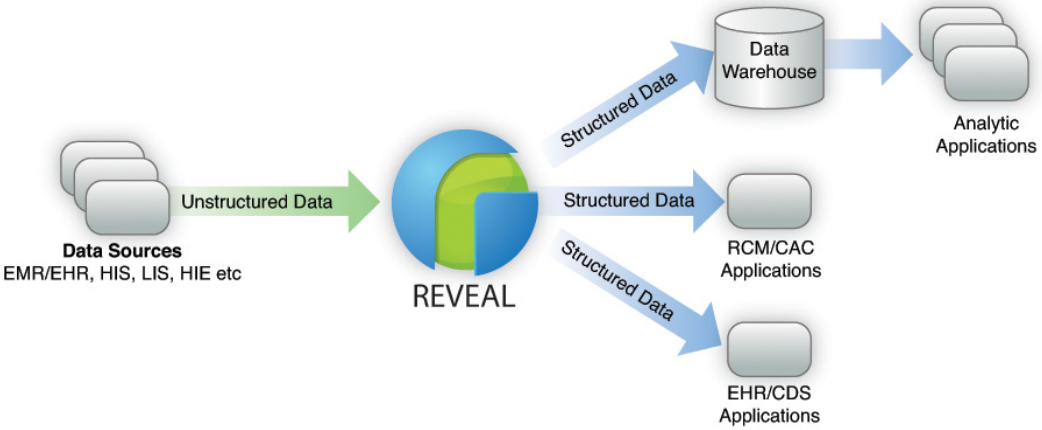
Predixion Software uses cloud-based predictive analytic software to explain patterns in hospital datasets to reduce readmissions and prevent hospital-acquired conditions.

- Pulls data from a variety of sources, using data mining, machine learning and mathematical algorithms to power predictions.
- Uses a predictive analytics algorithm to risk score patients upon admission and throughout their hospital stay, to identify those at risk of readmission before they leave the hospital, with 86% accuracy.
- Current project is applying analytics to prevent MRSA infections and deaths in the hospital setting.
- Working to use predictive analytics as a tool for prevention of chronic disease – e.g., diabetes.

"Predictive analytics allow you to aggregate this data to see what patterns are realistically making a difference in the decisions you make" says Jamie MacLennan, Cofounder and CEO. He also argues that "predictive analytics does not do our technology justice, because our technology is not just predictive, but is also explanatory."

Health Fidelity is using NLP to turn unstructured data (e.g., narrative medical records) into structured data suitable for computer management, to address needs in revenue cycle management, compliance, and analytics.

- Health Fidelity's NLP technology converts complex and specialized medical narratives and breaks out critical content to make it available in real time. It runs multiple data streams in many formats—note types, domains, linguistic forms, jargon, grammatical relationships and contexts—through NLP for comprehensive data extraction.
- This complex and unique process was initially funded by the National Institute of Health and National Science Foundation. Because of its roots in academia, there is already a broad range of peer-reviewed literature supporting the technology.
- Clients include healthcare IT vendors that serve medical practices, provider networks, and large healthcare organizations.
- Early use cases focus on revenue cycle management, (including ICD-10 conversion), compliance, (including meaningful use and accountable care), and analytics focused on cost reduction and quality improvement.

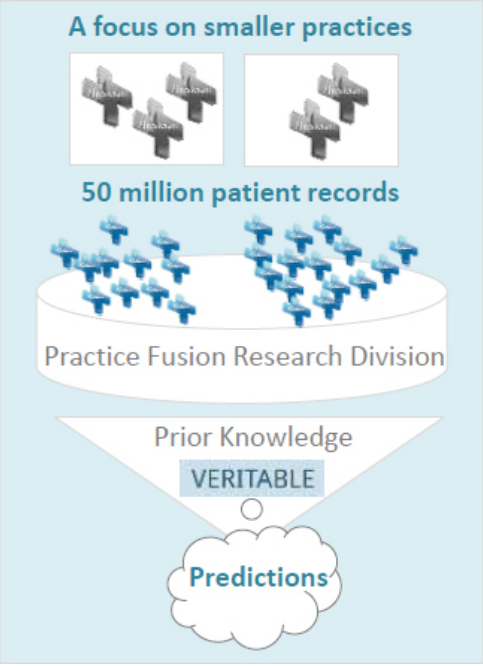


Dan Riskin, the CEO of Health Fidelity, sees NLP as an important foundation to data-driven healthcare. "We're trying to get at the 80% of unstructured healthcare data and make it usable for a broad array of applications in revenue cycle management, compliance, and analytics." He enthusiastically continues, "What is now entirely manual in healthcare should be automated and turned into review steps. NLP offers the opportunity to implement this workflow while massively expanding the amounts of usable and useful data available to improve care."

Practice Fusion is a free, cloud-based EMR platform for medical practices that also aggregates population data across multiple sites to improve clinical research and public health analysis.

- Offers a free plug-and-play EMR platform for smaller practices including e-prescribing, labs, Meaningful Use, charting and scheduling.

- Analyzes de-identified, aggregated data from the EMR system (and other public data sets) to monitor health on a population level, including outbreak detection, and provide research-based insight (never raw data) to partners.
- Works in 4 main areas, using data from multiple sites to produce a bigger data set than Kaiser on a national level:
  1. Health Population Surveillance and Education e.g. flu, asthma
  2. Post-Market Surveillance of drugs
  3. Public Health Research
  4. Plan of Care and Best Practice Development
- Customers include smaller practices and providers
- Collaborators include Prior Knowledge and Stanford Center for Biomedical Informatics Research
- Current focus on cancer, moving into heart disease



Lauren Fifield, Senior Health Policy Advisor expressed Practice Fusion's approach and hope: "Traditionally, EMRs have been centered in one hospital and aren't accessible by other hospitals. What makes this interesting from an analytical standpoint, is having real-time health information in a shared database."

"There have been good cancer surveillance systems that take data from hospitals and surveys. But there is no equivalent system for heart disease, and that is where we hope to come in."

athenahealth, Inc. is a cloud-based EMR and analytics company that focuses on standardizing health record systems across providers.

- The EMR system updates itself over time as technology improves, making improvements, such as new mobile platforms, available to customers.
- The platform provides deep analytics and hot spotting for physicians, giving them better insight into managing specific types of patient populations.
- Other services include practice management, patient communications and care coordination.
- Customers include 35,000 physicians in 48 states, who pay monthly, giving them flexibility.

Jeremy Delinsky, Senior Vice President and Chief Technology Officer told us "Healthcare, from a technology perspective, is at least a decade behind the rest of the world. But now, everyone is getting EMR systems for a reason, and hopefully a huge part of that reason is to increase quality of care. While we sell to doctors, we know patients are a huge part of the future landscape."

Camden Coalition of Healthcare Providers is a group of social workers, community organizers and nurses brought together by Dr. Jeffrey Brenner, the Executive Director and founder, to use predictive analytics to target the sickest members of the community in order to decrease hospital costs and readmissions.

- Analogous to "hot spotting" for crime, but aimed to improve care of the neediest patients, decrease readmissions, and reduce hospital costs.
- The Coalition goes to individuals and to hot-spot locations to address environmental and socioeconomic factors that affect health and hospital usage.
- One of the success cases is a patient named Derek W., who reduced his hospital trips from 35 to 2 in 6 months.
- The patients are the "customers".

Dr. Brenner, Executive Director and Founder of the Coalition, quoted in the New Yorker[28] "We are not going to cure poverty. The question is, how can you take the current situation the patients are in and improve it enough to make them a little bit healthier and lower their unnecessary hospital use and make them have a more productive interaction with the healthcare system."

[28] http://www.newyorker.com/reporting/2011/01/24/110124fa_fact_gawande

III. Supporting Self-Care

Another group of companies is using Big Data in new ways to help us help ourselves. Combining the convenience of mobile phones with the power of Big Data, the "worried well" can gather medical information and track sleep, while the chronically ill can better manage diabetes, heart disease and asthma, and everyone can better understand behavior patterns and motivations to make changes that can prevent, forestall or mitigate disease.

Humetrix's iBlueButton® is a mobile health information exchange app system that gives consumers and their providers an easy and secure way to access and exchange medical records.

- Delivers mobile health information exchange to Veterans Administration, Aetna and others.
- Provides device-to-device communication and data exchange at the point of care between patients and providers.
- Pools data from multiple devices in real time to inform providers of ongoing care from different providers.
- Security includes password protection with two-way secure encryption, HIPAA-compliant.

Dr. Bettina Experton, the CEO of Humetrix, describes the iBlueButton apps. "With iBlueButton in their hands, veterans and millions covered by Medicare can now be protected from prescription errors or redundant tests and procedures."

Ginger.io is a cloud-based platform for collecting real-time passive and active behavioral data from patients' cell phones (movement, communication, mobile usage, patient input, etc.) to help doctors, nurses, family members, and patients manage their health, beginning with chronic diseases. With patient consent, collected data and analytics are made available to providers and researchers through a HIPAA-compliant dashboard.

- Novel approach combines clinical insight and medical intuition and uses machine learning and predictive modeling to recognize subtle changes in normal behavior and lifestyle patterns.
- Initial focus on mental health, chronic conditions.
- Particular emphasis on patient privacy includes not only HIPAA compliance, but multiple layers of security. Based on the belief that users own their own data, they ask user permission before sharing (opt-in).

"The biggest advantage of our system is that we are not asking you to change anything about your behavior. With passively collected data, our system can determine how a patient is doing on a daily or weekly basis for the times in between appointments," explained Anmol Madan, PhD, CEO and Co-Founder of Ginger.io.

100Plus uses public and private data to motivate consumers to take small healthy steps to change daily habits via a mobile application.

- Pioneer in bringing the power of predictive analytics directly to the consumer.
- Target market is the worried well, primarily between 20 and 40.
- SXSW Booze Quiz uses CDC data on drinking behavior to help partiers avoid bingeing.

"We saw the convergence of Big Data sets, clinical data sets, government data and sensor data and believe that this combination of information could give consumers powerful information" said Chris Hogg, the Founder and CEO of 100 Plus.

ZEO, Inc. is analyzing over a million nights of data to help consumers improve their sleep.

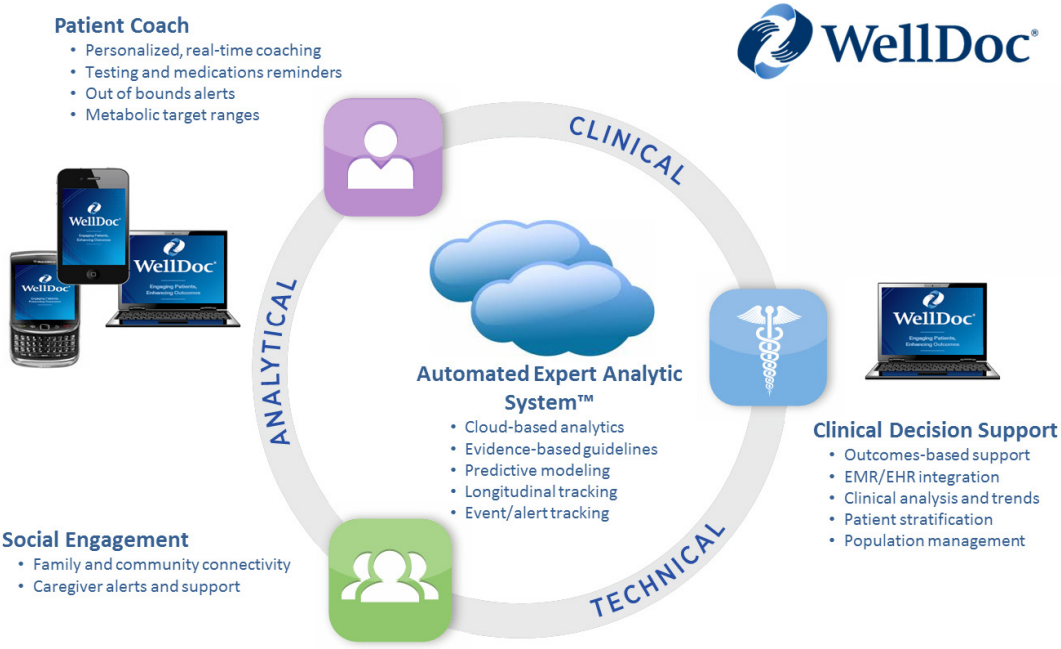
- The personal sleep coach device tracks the quality of users' sleep and gives personalized advice on how to improve sleep.
- As data philanthropists, they have shared sleep data with academic institutions to further the collective understanding of sleep.

"Right now, the biggest challenge is that we still only have sleep data," states Ben Rubin, the Co-Founder and Chief Technology Officer at Zeo. "We need someone to start a company that combines this with data on blood pressure, weight, heart rate, and other measures, aggregates it, and republishes it."

Welldoc® is using automated, real-time coaching that integrates behavioral and clinical messaging, as one element of its FDA Class II medical device platform, to help patients manage chronic diseases, such as diabetes.

- WellDoc captures a wide variety of patient-reported structured and unstructured data on clinical and behavioral aspects of the patient's health using the patient's mobile phone or web-based application. Hundreds of data points on individual patients become millions of data points as the application is scaled to large populations of patients.
- Among the features of the WellDoc coaching is real-time feedback (RTFB) in response to data entered. As an example, if a patient enters a low blood glucose value, a message, based on evidence-based guidelines, will provide treatment instructions and then prompt the individual 15 minutes later to re-test to determine if the treatment worked.
- Applying this example to the domain of Big Data; as more and more of these real-time events and response interactions populate the dataset, the data can be used to model more effective messaging and treatment.

Malinda Peeples, the Vice President of Clinical Advocacy at WellDoc states, "We are just beginning to apply Big Data principles to our growing dataset." Bharath Sudharsan, Senior Analytics Associate, adds, "WellDoc is actively taking efforts to understand Big Data – its challenges, opportunities and relevant technologies, as well as its applicability for mHealth."

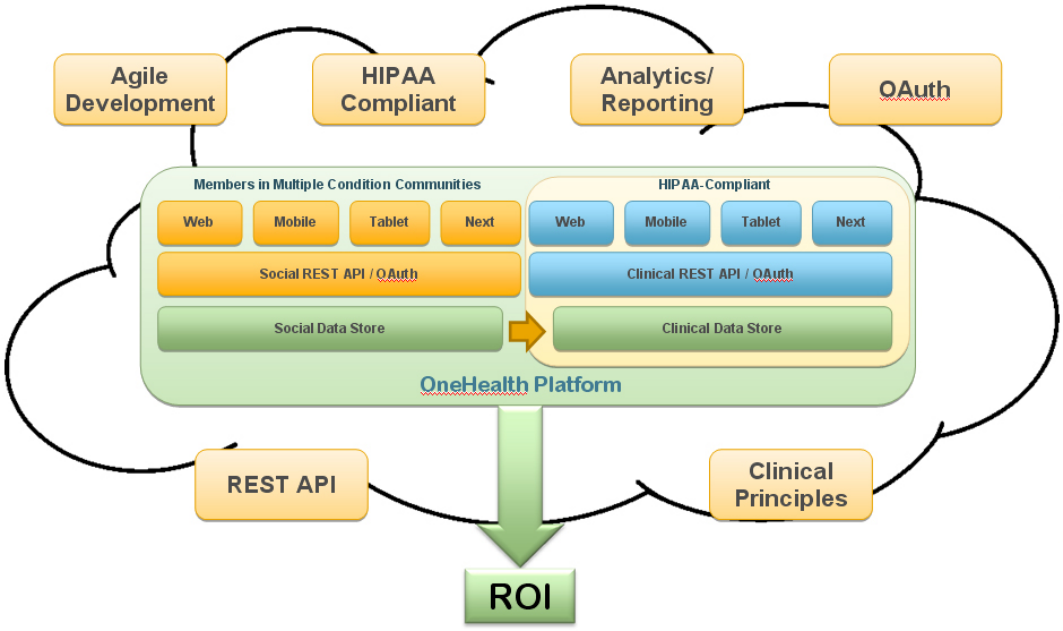


IV. Supporting Providers, Improving Patient Care

Care providers face increasing pressure: less time and money to do more with a growing torrent of information while remaining compassionate and not making errors. Provider support is one of the most exciting areas where Big Data may help. However, provider resistance to change and user-unfriendly interfaces remain major challenges in this arena. Various companies are taking different approaches to building provider support systems that are easy to use, save money and improve outcomes while giving providers more time to be compassionate.

OneHealth Solutions, Inc, is combining social and clinical data streams with flexible APIs to create the first real-time behavioral health records (snapshots).

This is the first real-time behavioral health platform built with the patient at the center; with a front-end designed to drive engagement and a flexible back end to enable comprehensive sharing of existing data and new data streams while preserving privacy.



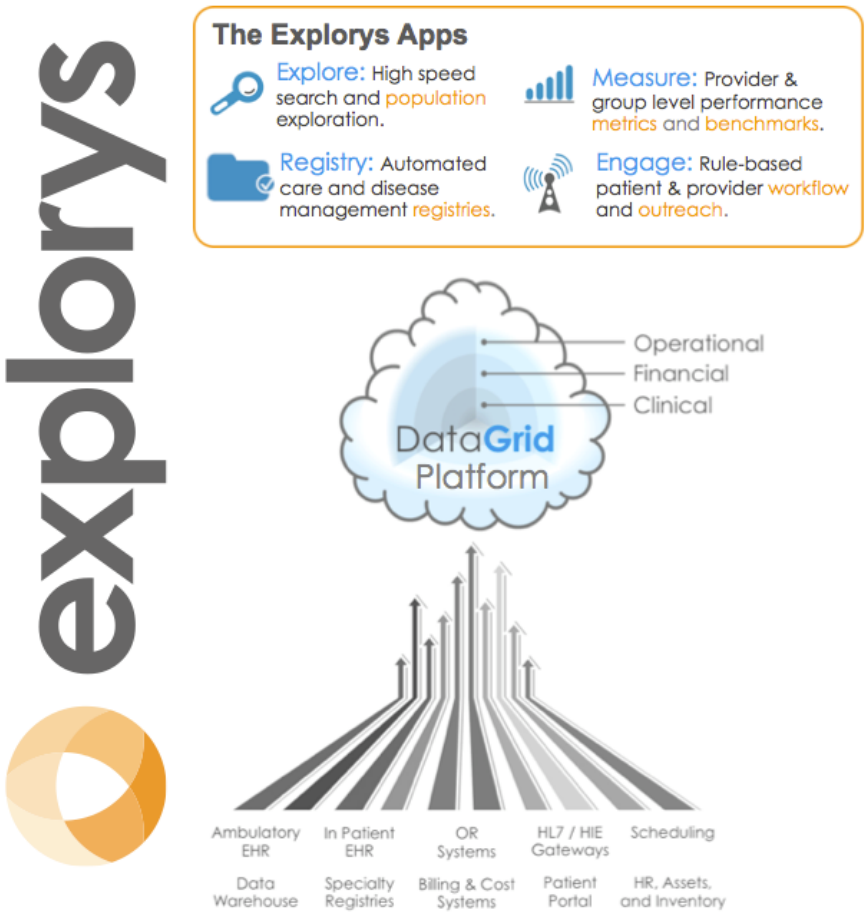
- Building on previous experience as OneRecovery supporting addiction-related disorders (smoking, alcohol, drugs), the platform has grown to encompass chronic disease management (obesity, diabetes, depression, respiratory conditions), wellness (fitness, healthy eating, stress management), and caregiver/family support (codependency, autism family support, cancer support).
- Social side of the API: the front end looks like Facebook, with cleverly integrated evidence-based clinical tools; on the back end, the API can collect user-generated

content to help customize and personalize the user experience and close the behavioral feedback loop.

- Clinical side of the API: HIPAA-compliant and has a flexible back end which allows information to be published/integrated selectively for various types of partners.
- Current customers include health plans, employers, providers (hospitals) and third party vendors.

"We have always put the patient in the center" explains Drew Paxton, the Chief Marketing Officer at OneHealth. "By closing the feedback loop on behavioral health, we are helping the medical community get better information about the patient." "True Big Data will happen when we can correlate physical symptoms with emotional states to help users better manage their day to day health decisions," adds John Shade, the CEO.

Explorys, Inc. has built a cloud-based computing platform that aggregates large amounts of data from many disparate sources including financial, operational and clinical data from multiple partners.

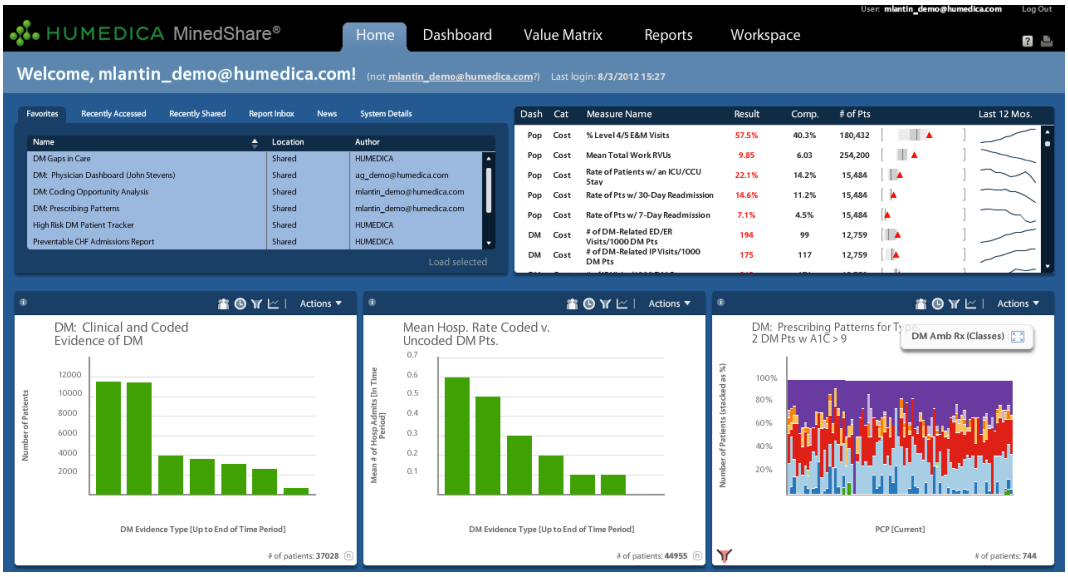


- Health Data Gateway pulls health data from client sites and sends it via a secure VPN to a Data Grid platform designed from the ground up to accommodate volume, velocity and variance challenges.
- Data is standardized and normalized to find patterns in disease, treatment and outcome to help clients manage populations.
- Customers include 12 healthcare systems and academic medical centers - with over **44** billion data elements, spanning more than **15** million patients, **114** hospitals, and thousands of providers and ambulatory venues.
- Clinical trial use case "opening new doors in clinical research, we now can search across large patient pools to facilitate faster, cheaper, better clinical trial recruitment."

"Using the same underlying technology as Facebook, Google, Amazon and Yahoo, we are processing healthcare data more quickly and easily than has been done in the past," said Jason Gilder, PhD, Director of Analytics and Informatics at Explorys. "We are working on the bleeding edge of Big Data technology. We are not only using the latest versions of these open source products, but we are contributing back to the open source community."

Humedica, Inc. is a clinical informatics company that provides SaaS business intelligence by connecting clinical and patient information across varied settings and time periods to generate longitudinal and comprehensive views of patient care.

- Provides more accurate and detailed predictive models: normalizes and cleans data to produce more accurate and precise inputs over longer timeframes (e.g., in congestive heart failure).



- Collective data sharing and benchmarking: de-identifies and shares data as part of AMGA community.
- Collaborative quality improvement and best practice sharing through Anceta, the medical informatics subsidiary of AMGA.
- Clients include: providers, hospitals and health systems, and life science companies.

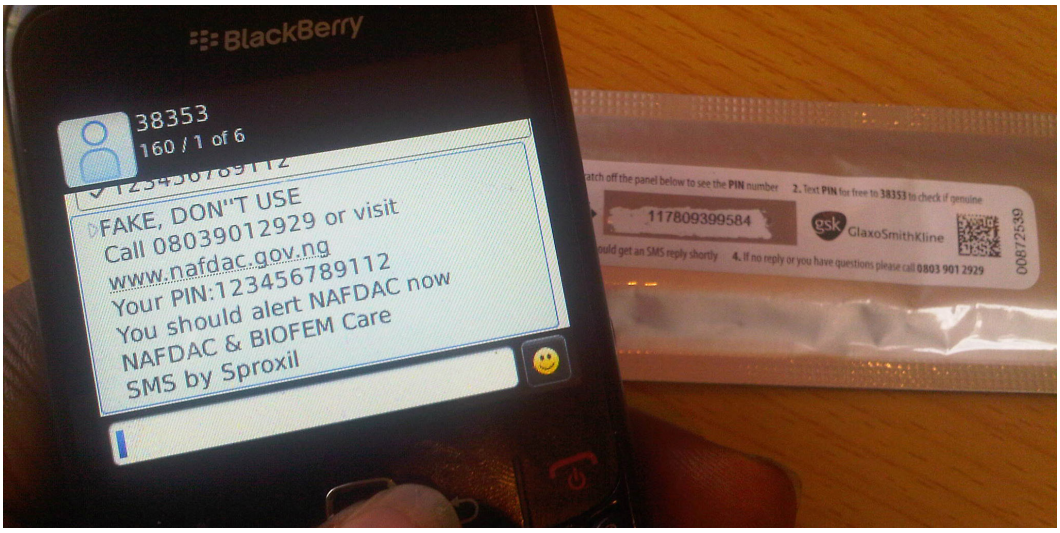
"Our data asset is close to 25 million contracted lives across the country in more than 30 states. These volumes make our predictions better. It is almost as though you go from covering one eye and seeing the world to opening both eyes to have 20/20 vision", said Allen Kamer, the Vice President of Corporate Development and Marketing at Humedica.

V. Increasing Awareness

Big data is an obvious toolkit to increase knowledge to begin solving a variety of data-based problems: identifying counterfeit drugs, tracking environmental issues that trigger asthma, predicting disease outbreaks, helping developing countries make better policy decisions, and prioritizing global public health issues.

Sproxil uses Big Data to identify counterfeit drugs, to protect patient health and enable pharmaceutical companies to track drug distribution and prevent theft.

- PIN codes added to each individual drug product package identify whether the drugs are real or counterfeit.
- Working alongside drug manufacturers, a service allows individuals to call or text in serial numbers to determine drug status.



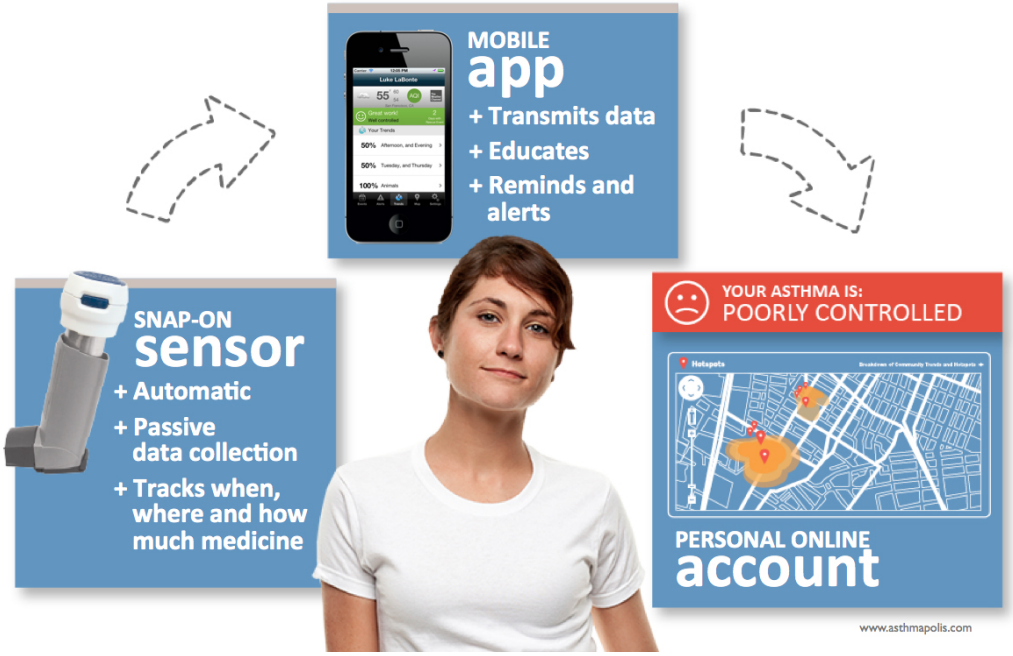
- An NLP algorithm that is currently being adapted to understand multiple languages.
- Customers include pharmaceutical companies, non-profits and ultimately the consumer.

"We feel we are using technology to empower both patients and companies to have smoother and more efficient transactions which has tremendous benefit to both the patient and pharmaceutical manufacturers. We are partnering with several types of organizations, including government agencies, that are serving the healthcare needs of their people."

Asthmapolis collects data from patients and provides them with feedback which helps them better manage their asthma.

- A mobile sensor tracking device attaches to asthma inhalers to monitor the time and location of events, plus triggers and symptoms through an iOS/Android app.
- Helps patients control their asthma, helps providers and payers identify at-risk patients
- Aggregates real-time data for epidemiological research and public health use.

Mark Gehring, the president of Asthmapolis told us, "Conservative estimates are that 10 billion of the 50 billion dollars spent on treating asthma is unnecessary, and the cost would diminish if we can bring asthma under control. One of our goals is to create a community index of the level of risk of asthma using the population and our data."



Sickweather LLC scans social media (Facebook, Twitter) to track outbreaks of disease, offering forecasts to users, similar to weather forecasting.

- Offers real time sickness forecasting services to keep individuals aware of outbreaks in their area.
- Allows individuals to be members and input information about what is happening in their area.
- Supported by advertising links to low-cost medications.
- Currently works in English-speaking countries with future plans to expand globally.

Founder Graham Dodge proclaims the premise of Sickweather "Social media is an untapped source of data for tracking health, diseases and illnesses. I see the service evolving like weather reporting—where initially people thought they did not need weather forecasting and now it is a global multibillion dollar market."

The Institute for Health Metrics and Evaluation (IHME) is gathering a large number of data sets globally for data analysis and health measurement that can guide policy decisions to improve population health.

- This independent global research center at the University of Washington, funded by the Bill and Melinda Gates Foundation and the State of Washington, has a mission to answer three questions in global health:
  1. What are the world's major health problems?
  2. How well is society addressing these problems?
  3. How do we best dedicate resources to maximize health improvement?
- Analyzes data from disparate sources including censuses, surveys, vital statistics, disease registries, hospital records and others to create evidence for policy and decision making.
- Collaborates with and provides information for inter-governmental agencies, governments, and other public and global health organizations.
- Currently finishing one of their major research projects, The Global Burden of Disease.

Peter Speyer, Director of Data Development at IHME, explains that **"**we are pushing the envelope on what is possible on collating and analyzing population health data and providing policy and decision makers with better evidence. As an example, the Global Burden of Disease project synthesizes all available information on morbidity and mortality to estimate burden from more than 240 causes and 60 risk factors in 187 countries in the world."

VI. Pooling Data to Build a Better Ecosystem

An interesting application of Big Data is to bring disparate data sets together for the first time in ways that could enable new kinds of analyses and facilitate the answers to big questions, some of which haven't been asked yet.

Qualcomm Life is enabling a global wireless health **connectivity** platform (2net™) and open ecosystem that brings healthcare data—new and existing biometric data sources—together in ways that have never been done before.

- Uses a secure cloud-based platform with open APIs to create a developer's environment allowing data to be pooled and accessed by creative applications, software and analytics programs.
- Pooling creates an environment for device companies, software and analytics companies, where the mixing and matching of data is more "valuable" than the individual pieces of data.
- Customers are device companies that get more varied connectivity options, software companies and service providers that get access to the data from devices and analytics companies who get new data sources.
- Companies own their own data, but they license an de-identified version into the data pool where they have access to the data. In the near future, consumer data will be included.
- Early use cases include a blood pressure cuff company who might want to also understand how consumers and patients are using a bathroom scale to manage a specific disease like diabetes.

"We are bringing new data sets into the playing field, making them easier to manipulate, so all participants could be potential customers," explained Don Jones, Vice President of Global Strategy & Market Development at Qualcomm Life.

"Data pooling works because it is a win-win, where my data is more valuable when I combine it with other peoples' data. This is a true network effect model. Our frame of reference is consumer electronics, where the phone in your hand is more valuable when it is connected with more things."

Factual wants to democratize access to healthcare data

- Setting out to develop a definitive data platform for healthcare in an effort to maximize data accuracy, transparency, and accessibility.
- Based on their previous work with global places, restaurants, hotels and products, and healthcare providers.
- Using machine learning and other techniques to produce clean, accessible data at lower cost, so that 10% of the largest users support an ecosystem where 90% of the smallest companies can get the data for free.
- Started with provider locations, now adding doctor data.

According to Eva Ho, the Vice President of Marketing & Operations at Factual, "We believe that open data is the notion that data is more accessible, less encumbered and a lot more affordable, but not necessarily free. If you look at the world in terms of entities on a spreadsheet, there are three categories: places, products and people. So we started with places and products—millions of data points on schools, parks, hospitals, gas stations and everything under the sun."

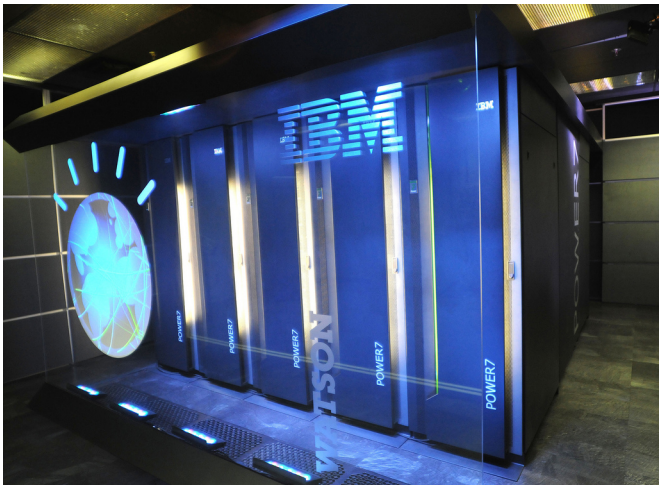
Watson – a peek into the future –

physicians' diagnosis and treatment assistant supercharged with Big Data and analytics

Watson, a compilation of 21 supercomputer subsystems, is the first of a new class of industry-specific analytical platforms and decision support systems that use deep content analysis, evidence-based reasoning and natural language processing to support faster and more precise diagnostics and clinical decision making.

- Watson is the most advanced natural language processor on the planet, as demonstrated by its performance on Jeopardy.
- With 16 terabytes of memory, twice the amount in the Library of Congress, Watson can store huge amounts of data, ranging from patient health records to the latest publications about cutting-edge treatments.
- A doctor typically spends about 10 hours a week reading the latest advances in medical journals, but Watson can read 200 million pages of text in three seconds and remember every word.
- Watson takes in data from patient history, family history, symptoms and test findings and produces a list of disease suggestions ranked by confidence, to assist the physician in diagnosis and treatment.
- WellPoint is working on a Watson project with Cedars-Sinai Medical Center in Los Angeles to build decision-making tools for oncologists treating breast, colon and lung cancer.

"In cancer there are over 75 new clinical trials starting every day with over 15 million catalogued articles and patient longitudinal records," said Stephen Gold, a Vice President of Worldwide Marketing for IBM. "Watson is currently ingesting de-identified longitudinal information, starting with lung and breast cancer. Over time, it will learn from its mistakes and it never forgets. It is going to discover patterns that will shed light on how medicine is practiced."



"We have historically spent our time thinking about structured data, but all of a sudden we find ourselves with the majority of information being unstructured data. Watson is uniquely capable of making use of unstructured data, a capability that is particularly valuable in healthcare."

"Watson represents part of a cognitive system that can leverage and incorporate a massive amount of data—it can not only ingest it, but curate it. Curation is key to annotating the information so that it can be used."

Issues and Challenges of Big Data

It's easier to write about the promises of Big Data than to realize them. Several issues and challenges have been widely recognized as major barriers to the successful implementation of Big Data in healthcare.

Data and Information Privacy

Privacy issues have become increasingly urgent recently, as Internet transactions and communications, cloud storage, social media and mobile devices expose more and more personal data to potential misuse. While online and social media users have been rather inconsistent about the privacy implications of their own behavior, increasing publicity and intense discussion makes it clear that most people are becoming concerned about protecting and controlling their personal data, especially health and medical data.

Healthcare is a special case for a number of reasons, including:

- The tradition (and legal status) of doctor-patient confidentiality (and the related tradition of providers controlling or blocking access to patients' own medical records)
- Concerns by individuals about disclosure of personal health information to third parties:
    - payers and other insurers (an issue that came to a boil during the AIDS crisis),
    - caregivers, next-of-kin, spouses/partners acting on behalf of patients
    - outsiders such as the media, criminals, etc.
- The conflicting desire of third parties (insurers, employers, etc.) to access data about potential employees or insurees for actuarial, pricing, employment, and other decisions
- Government regulations intended to address health data privacy, particularly HIPAA (Health Information Portability and Accountability Act) and the Affordable Care Act, as well as ARRA stimulus moneys to pay to digitize health records (HITECH Act).

Most of us consider health data privacy—control of one's personal health information and who sees it—a fundamental right. It is a visible and politically potent issue as the HIPAA Act showed. But the HIPAA regulations[29] while well-intentioned, frequently make access more difficult to patients and caregivers, and add layers of regulatory compliance that complicate and slow the process of communicating data from providers to providers and from providers to patients.

The Markle Foundation's Connecting for Health is a public-private collaboration organized to advance healthcare through information technology, and tackle the challenges of creating a

[29] http://privacyruleandresearch.nih.gov/healthservicesprivacy.asp

networked health information environment that enables secure and private information sharing. Their Health in a Networked Life survey on privacy in 2011 found that 80% of the public and doctors agree that privacy safeguards for health information are important.[30]

Similarly, every company interviewed was concerned with privacy. Each declares, at the very least, adherence to HIPAA requirements, and many claim more.

"The privacy horse has already left the stable," proclaimed Eileen Bartholomew, the Vice President of Prize Development of X-Prize, "right now you walk around with a gps tracking device all day long."

"We have invested heavily in security and privacy and we are very serious about these issues, said Allen Kamer of Humedica. "We adhere to what we call "HIPAA plus," following all 18 HIPAA requirements and also de-identifying the physician's name." In addition, they have established a set of data principles which discloses to their customers what they do with the data.

Of note, Ginger.io has a particularly progressive view of data privacy, which includes the philosophy that patients own their own data, they can opt-in, choose when and how to share their data and can discontinue data sharing at any time.

Health data privacy remains a difficult and contentious issue that is bigger than a data management concern, with political, regulatory, business practice and even cultural dimensions. Big Data raises particular concerns around data security (see below), de-identification, use of cloud, and pooling of data sets.

Data Security

In addition to privacy concerns all the companies interviewed were concerned about data security, unintentional exposure or loss of data to unauthorized parties. Use of the Internet, cloud computing and pooling of data all raise the data security stakes.

"Healthcare data contains the intimate details of a person's life and we must respect and protect it with the highest security possible," declares Jason Gilder of Explorys. "One of our biggest barriers to adoption is trust, so we have security audits and make sure we exceed all of the current compliance and legislation."

Jeremy Dilinsky of athenahealth has similar concerns. "In any sales with a large customer, you spend a lot of time discussing their concerns about security and privacy." Apparently, there is still resistance to moving healthcare data to the cloud. "Some CIOs find the idea of putting their protected health information in the cloud is scary."

[30] http://www.markle.org/publications/1443-public-and-doctors-agree-importance-specific-privacy-protections-health-it

Who owns the data?

Another approach to these questions is to ask who owns and who should own the data?

Although logically, most people would assume that they own their own healthcare data, this may not always be the case. Stories of patients or their caretakers struggling to get timely access to crucial data from providers or payers abound, as in Regina Holiday's struggle to get access to her dying husband's medical record, and others, posted on the US government HHS website.[31]

These concerns have led to patient advocacy groups such as the e-patient movement, where patients help each other to become active participants in their own care alongside doctors. Started by the late Tom Ferguson, some of current evangelizers include Hugo Compos, Regina Holiday and e-patient Dave. There is even a song entitled "give me my damn data."

To further give voice to the active and engaged patient, the Society of Participatory Medicine is a cooperative model of healthcare that encourages and expects active involvement by all connected parties (patients, caregivers, healthcare professionals, etc.) as integral to the full continuum of care. The mission of its peer-reviewed, open access journal, the Journal of Participatory Medicine, is to advance the understanding and practice of participatory medicine among healthcare professionals and patients.

Siloed data and the need for integration

Another reason that patients and providers have trouble accessing all of their data is that there are six major unintegrated data pools controlled by different stakeholders:

1. Providers: clinical/medical data (EHRs)
2. Payers and Providers: claims and cost data
3. Researchers: academic,
4. Developers: pharma and medical device R&D, including clinical trials
5. Consumers and Marketers: patient behavior and sentiment data
6. Government: population and public health data

"Building the architecture for integration is challenging," said Jason Gilder of Explorys.

In a PWC Health Research Institute (HRI) survey of more than 600 providers, health insurers and pharmaceutical/life science professionals on their clinical informatics perspectives, 73% of respondents said integrating data from multiple sources is their top clinical goal in the next 2 years. Yet only 17% of provider respondents and 16% of health insurance companies were confident that their HIE vendor could meet their analytic and integration needs over the next 2 years.

[31] http://www.healthit.gov/patients-families/health-it-stories

Providers noted both cultural and organizational barriers to effectively analyzing clinical data.

The biggest cultural barrier was accepting the clinical system as a work flow productivity tool. In addition, organizational barriers included data being kept in silos and preventing information overload.

Call for sharing data/collaboration to help liberate the data

Despite the privacy, security and integration challenges, every company interviewed wanted to gain more access to data. Each was interested in what others were doing with their data and every one wanted opportunities to collaborate with one another.

Several of the companies interviewed see the open science movement as a solution to some of the challenges. Concepts similar to "open science" include open data, open source and data sharing.

"The open-source community is here to give everyone a shot to build something great. Two years ago we were two guys in garage. We have used many tools and given back to the community." said Jason Gilder of Explorys."

Eva Ho, of Factual, states: "We believe open data is the notion that data is more acceptable and accessible - less encumbered and a lot more affordable."

Other examples of data sharing include "data philanthropy" where Zeo gives its data to research scientists and crowd-sourced competitions employed by Kaggle. The WellPoint and "Watson" partnership as well as a series of other informatics partnerships such as AstraZeneca and IMS Health are additional examples of this data-sharing trend.[32]

[32] http://www.pwc.com/us/en/health-industries/publications/needles-in-a-haystack.jhtml

Three Trends for the Use of Big Data within an Emerging Ecosystem

In the course of more than 30 interviews with Big Data healthcare enterprises, we found a growing variety of creative ways to channel and exploit the flood of Big Data in healthcare.

Three trends emerged:

1. Working with limited data sets
2. Combining a greater variety of data
3. Pooling data for bigger and better data sets to facilitate more powerful analysis and insights

We think that these trends reflect how the ecosystem is emerging over time.

Some early experimenters are using a few data streams, or isolated digital puddles to tackle discrete healthcare problems:

- **Health Fidelity**'s NLP technology enables healthcare IT vendors to help medical practices, providers and other healthcare organizations convert their 80% of unstructured data to structured data.
- **DNAnexus** provides genomic data management, analysis, and visualization using its collaborative and scalable data technology platform
- **Predixion Software** analyzes a hospital's data to find patterns predictive of readmission and iatrogenic infections.

Others are using multiple data streams:

- **NextBio** assembles vast amounts of curated and annotated molecular data enabling clients with private data to make discoveries not possible with their own datasets.
- **Explorys** aggregates large amounts of data from a many disparate sources—financial, operational and clinical data—in a cloud-based open source computing platform.
- **OneHealth** combines social and clinical data streams with two flexible APIs to create the first real-time behavioral health snapshots to help people manage addictions and chronic diseases.
- **Practice Fusion** aggregates population data across multiple medical practices from their free cloud-based EMR platform to improve clinical research and public health analysis.

Still others are looking to build open ecosystems with opportunities for large amounts of data pooling in new ways:

- **Qualcomm Life** brings healthcare data—new data sources and existing data—together through a global wireless connectivity platform.
- **Factual** is leveraging large-scale data aggregation and community exchange to drive and accelerate innovation via an open data platform for application developers.

The trend toward using more data types in more complex ways reflects how the ecosystem is developing over time. The new companies coming into the ecosystem are mostly using limited data streams, while looking for collaborations to help them combine a greater variety of data to facilitate more powerful analysis and insight.

The Future of Big Data in Healthcare

In the course of more than thirty interviews, we found an emerging ecosystem of companies interested in using Big Data to improve healthcare in six ways:

1. Support Research: Genomics and Beyond
2. Transform Data to Information
3. Support Self-Care
4. Support Care Providers
5. Increase Awareness
6. Pool Data to Expand the Ecosystem

All of the companies were eager to learn more about each other, as they look for new ways to gain access to additional data.

Among this sample of companies three trends in data usage emerged:

1	Working with limited data sets
2	Combining a greater variety of data
3	Pooling data for better results

Most of these companies want to move from working with limited data sets to combining a greater variety of data to facilitate more powerful analysis and insights. These early trends may well reflect how the ecosystem will evolve over time.

Two of the interviewed experts, Don Jones of Qualcomm Life and Eileen Bartholomew of X-Prize, think that consumer demand is going to drive these Big Data trends.

“We are in a period of exponential growth driven by consumer demand,” said Eileen.
Don Jones explained, “Right now most of the companies are focused on business-to-business applications, but in 5 years, I expect to see a business-to-consumer model where software is made for the individual.”

How such consumer demand will manifest—what products and services, how big and how fast, —remains to be seen. For comparison, the iPhone 4 attracted 600,000 pre-orders in the first 24 hours, setting a benchmark for the most consumer demand for a technology product.

The dream of consumer-driven, patient-centered healthcare came up repeatedly throughout the interviews. Yet, given the issues and challenges of privacy, security, data ownership, regulation and integration, the question remains as to how much and how quickly Big Data can help those dreams become reality.

Several key healthcare improvement goals clearly emerged from the research and interviews:

- Increasing provider and payer efficiencies, reducing errors and costs
- Enabling comparative effectiveness research for current treatments and to inform R&D
- Moving toward patient-centered, outcome-oriented medicine
- Empowering consumers - "Health 2.0," participatory healthcare
- Making personalized medicine possible for everyone

Most agreed about the need to improve the efficiency and effectiveness of the US healthcare system and to halt the growth of healthcare costs worldwide. However, reducing costs was not the only goal.

The vision is a world of patient-centered care, where researchers understand the causes of disease and developers design treatments tuned to each individual's biology. This vision includes rapid, precise and accurate diagnostics as well as the right match between diagnosis, patient genomics and treatment modality. Beyond that, the vision is of seamless decision support for caretakers, doctors, nurses, and patients themselves, not only for disease treatment and management but, even better, for effective prevention and delay of disease.

Big Data is only one tool among many in healthcare IT, which in turn, is only one of many specialized disciplines critical to healthcare. Nevertheless, it's clear that a tsunami of healthcare Big Data is already upon us. Beyond the current hype, riding that wave adeptly will be a critical contribution to safer and more effective healthcare for every one of us.

The Cost of Healthcare - Trending up at an Unsustainable Rate

$2.9 trillion
overall spending in 2009

17.6% GDP
in 2009

25.0% of GDP
in 2025 (projected)

$1.875 trillion
(Annual cost in 2009) spent on Chronic Disease

96/100,000 patients
die from preventable conditions annually in the US

+1/5 patients
are readmitted after 30 days

Potential Savings from Using Big Data

$300 billion
in annual savings to US Healthcare spending according to a McKinsey estimate

Bibliography

A MIX OF EXISTING DATA AND DATA STREAMS CREATES POSSIBILITIES FOR IMPROVEMENT

Anderson, Janna Quitney, and Lee Rainie. "Big Data: Experts Say New Forms of Information Analysis Will Help People Be More Nimble and Adaptive, but Worry over Humans' Capacity to Understand and Use These New Tools Well." Pew Research. Pew Research Center's Internet & American Life Project, 20 July 2012. Web. 22 July 2012. http://www.greenplum.com/sites/default/files/PIP_Future_of_Internet_2012_Big_Data.pdf

Bloem, Jaap, Menno Van Doorn, Sander Duivestein, Thomas Van Manen, and Erik Van Ommeren. "Creating Clarity with Big Data." *Sogeti VINT*. Sogeti, Aug. 2012. Web. 5 Sept. 2012. http://blog.vint.sogeti.com/wp-content/uploads/2012/07/VINT-Sogeti-on-Big-Data-1-of-4-Creating-Clarity.pdf

Bollier, David. "The Promise and Peril of Big Data." Communications and Society Program. The Aspen Institute, 2010. Web. 24 Aug. 2012. http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf

Brown, Brad. "The McKinsey Quarterly." Are You Ready for the Era of 'big Data'? N.p., Oct. 2011. Web. 20 June 2012. https://www.mckinseyquarterly.com/Are_you_ready_for_the_era_of_big_data_2864

Data, Data Everywhere. The Economist. The Economist Newspaper, 25 Feb. 2010. Web. 26 July 2012. http://www.economist.com/node/15557443

Dumbill, Ed. "What Is Big Data?" What Is Big Data? - O'Reilly Radar. O'Reilly, 11 Jan. 2012. Web. 01 Aug. 2012. http://radar.oreilly.com/2012/01/what-is-big-data.html

Dumbill, Ed. *Planning for Big Data*. N.p.: O'Reilly Media, n.d. *O'Reilly Strata*. O'Reilly Media, Mar. 2012. Web. 5 Sept. 2012. http://shop.oreilly.com/product/0636920025559.do?cmp=il-radar-ebooks-planning-for-big-data-radar-announcement

Gilbert, George. "A Guide to Big Data Workload - Management Challenge." GigaOM Pro. GigaOM, May 2012. Web. 01 Aug. 2012. http://pro.gigaom.com/2012/06/a-guide-to-big-data-workload-management-challenges

A Near-term Outlook for Big Data. GigaOM. GigaOM Pro, Mar. 2012. Web. 01 Aug. 2012. http://pro.gigaom.com/2012/03/a-near-term-outlook-for-big-data/

Jablonski, Joey, and Aurelian Dumitru. "Taking Control of Big Data". Dell, 2011. Web. 26 July 2012. http://i.dell.com/sites/doccontent/business/solutions/power/en/Documents/ps4q11-20110460-hadoop.pdf

Lesser, Adam, Derrick Harris, Jody Ranck, David Card, and Michael Wolf. "Connected World: The Consumer Technology Revolution." GigaOM Pro. GigaOM, 10 Nov. 2011. Web. 05 Sept. 2012. http://pro.gigaom.com/2011/11/connected-world-the-consumer-technology-revolution/

Manyika, James. "Big Data: The next Frontier for Innovation, Competition, and Productivity. "McKinsey & Company. N.p., May 2011. Web. 20 June 2012. http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation

Stanford Medicine 2nd ser. 29 (2012): 1-42. *Stanford Medicine*. Stanford University, Summer 2012. Web. 5 Sept. 2012. http://stanmed.stanford.edu/2012summer/documents/medmag_2012summer.pdf

Zikopoulos, Paul Z., Chris Eaton, Dirk DeRoos, and Thomas Deutsch. "Understanding Big Data." IBM. IBM, 2012. Web. 24 Aug. 2012. http://public.dhe.ibm.com/common/ssi/ecm/en/iml14296usen/IML14296USEN.PDF

BIG DATA FACILITATING POSSIBILITIES TO INCREASE QUALITY OF CARE

Cortada, James W., Dan Gordon, and Bill Lenihan. "The Value of Analytics in Healthcare." IBM Global Business Services. IBM Institute for Business Value, Jan. 2012. Web. 24 Aug. 2012. http://www.ibm.com/smarterplanet/global/files/se__sv_se__healthcare__the_value_of_analytics_in_healthcare.pdf

Fraser, Heather, YangJin Kwon, and Margaret Neuer. "The Future of Connected Health Devices." IBM. IBM Global Business Services, Mar. 2011. Web. 29 July 2012. http://public.dhe.ibm.com/common/ssi/ecm/en/gbe03398usen/GBE03398USEN.PDF

Golden, Jim. "Hacking Healthcare." Forbes. Forbes Magazine, 19 June 2012. Web. 26 July 2012. http://www.forbes.com/sites/jimgolden/2012/06/19/hacking-healthcare/

Helping Accelerate Treatment for Pediatric Cancer with Dell Cloud Technology. Dell. Dell, n.d. Web. 26 July 2012.
http://content.dell.com/us/en/gen/d/corp-comm/pediatric-cancer

Howard, Alex. "The Risks and Rewards of a Health Data Commons - O'Reilly Radar."The Risks and Rewards of a Health Data Commons - O'Reilly Radar. O'Reilly, 9 Aug. 2012. Web. 24 Aug. 2012.
http://radar.oreilly.com/2012/08/health-data-commons.html

O'Reilly, Tim, Julie Steele, Mike Loukides, and Colin Hill. "Solving the Wanamaker Problem for Healthcare." *O'Reilly Radar*. O'Reilly Radar, 14 Aug. 2012. Web. 05 Sept. 2012.
http://radar.oreilly.com/2012/08/data-health-care.html

McNickle, Michelle. "5 Basics of Big Data | Healthcare IT News." 5 Basics of Big Data | Healthcare IT News. In Partnership with HIMSS, 13 June 2012. Web. 26 July 2012.
http://www.healthcareitnews.com/news/5-basics-big-data .

"Needles in a Haystack: Seeking Knowledge with Clinical Informatics." *PwC*. PwC Health Research Institute, Feb. 2012. Web. 5 Sept. 2012.
http://www.pwc.com/us/en/health-industries/publications/needles-in-a-haystack.jhtml

Quoyeser, Jared. "Intel® Premier IT Professionals: Healthcare: Health IT Trends: HP Identifies Big Data and Patient Participation." Intel® Premier IT Professionals: Healthcare: Health IT Trends: HP Identifies Big Data and Patient Participation. Intel, 4 June 2012. Web. 27 July 2012.
http://premierit.intel.com/community/ipip/healthcare/blog/2012/06/04/health-it-trends-hp-identifies-big-data-and-patient-participation

Quoyeser, Jared. "Intel® Premier IT Professionals: Healthcare: Health IT's Role in Security, Big Data and Mobility." Intel® Premier IT Professionals: Healthcare: Health IT's Role in Security, Big Data and Mobility. Intel, 29 May 2012. Web. 27 July 2012.
http://premierit.intel.com/community/ipip/healthcare/blog/2012/05/29/health-its-role-in-security-big-data-and-mobility

Valuing Healthcare: Improving Productivity and Quality. Kauffman. Kauffman Task Force on Cost-Effective Healthcare Innovation, Apr. 2012. Web. 29 July 2012.
http://www.kauffman.org/uploadedfiles/valuing_health_care.pdf

Redefining Value and Success in Healthcare. IBM. IBM Healthcare, Jan. 2012. Web. 26 July 2012.
http://public.dhe.ibm.com/common/ssi/ecm/en/hlw03013usen/HLW03013USEN.PDF

THE POWER OF PERSONAL DATA

Adams, Jim E., Edgar L. Mounib, and Amnon Shabo. "IT-enabled Personalized Healthcare: Improving the Science of Health Promotion and Care Delivery." IBM. IBM Global Business Services, 2011. Web. 29 July 2012.
http://public.dhe.ibm.com/common/ssi/ecm/en/gbe03299usen/GBE03299USEN.PDF

Personal Data: The Emergence of a New Asset Class. World Economic Forum. World Economic Forum In Collaboration with Bain & Company, Inc, Jan. 2011. Web. 26 July 2012.
http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf

Rethinking Personal Data: Strengthening Trust. World Economic Forum. World Economic Forum in Collaboration with The Boston Consulting Group, May 2012. Web. 26 July 2012.
http://www3.weforum.org/docs/WEF_IT_RethinkingPersonalData_Report_2012.pdf

Touching Lives through Mobile Health: Assessment of the Global Market Opportunity. PWC. PWC, Feb. 2012. Web. 29 July 2012.
http://www.pwc.com/in/en/assets/pdfs/telecom/gsma-pwc_mhealth_report.pdf

A GLIMPSE INTO THE TECHNICAL PROWESS OF BIG DATA

IBM Forms Watson Healthcare Advisory Board. IBM News Room. N.p., 01 Mar. 2012. Web. 20 June 2012.
http://www-03.ibm.com/press/us/en/pressrelease/36989.wss

Pepitone, Julianne. "IBM's Watson Computer Gets a Healthcare Job with WellPoint."CNNMoney. Cable News Network, 12 Sept. 2011. Web. 20 June 2012.
http://money.cnn.com/2011/09/12/technology/ibm_watson_health_care/index.htm

Watson for a Smarter Planet." IBM -. IBM, n.d. Web. 20 June 2012.
http://www-03.ibm.com/innovation/us/watson/watson-for-a-smarter-planet/index.html

WellPoint and IBM Announce Agreement to Put Watson to Work in Healthcare. IBM News Room. IBM, 12 Sept. 2011. Web. 20 June 2012.
http://www-03.ibm.com/press/us/en/pressrelease/35402.wss

Yuan, Michael J. "Watson and Healthcare." Watson and Healthcare. IBM, 12 Apr. 2012. Web. 24 Aug. 2012.
http://www.ibm.com/developerworks/industry/library/ind-watson/

THE IMPORTANCE OF THE CLOUD

Quoyeser, Jared. "Intel® Premier IT Professionals: Healthcare: The Intersection of Cloud and Big Data in Health IT." Intel® Premier IT Professionals: Healthcare: The Intersection of Cloud and Big Data in Health IT. Intel, 19 June 2012. Web. 27 July 2012. http://premierit.intel.com/community/ipip/healthcare/blog/2012/06/19/the-intersection-of-cloud-and-big-data-in-health-it

Doyle, Kerry. "Virtualization 101." *Global Knowledge*. Global Knowledge, 2012. Web. 5 Sept. 2012. http://www.globalknowledge.com/training/whitepaperdetail.asp?pageid=502&wpid=942&country=United+States

BIG DATA COMES TO PUBLIC HEALTH

The Promise of Big Data. Harvard School of Public Health, n.d. Web. 24 Aug. 2012. http://www.hsph.harvard.edu/news/features/files/big_data.pdf

Howard, Alex. "Data for the Public Good." *Strata*. O'Reilly Strata, Feb. 2012. Web. 05 Sept. 2012. http://strata.oreilly.com/2012/02/data-public-good.html

PUBLIC HEALTH: A GLIMPSE INTO THE GLOBAL POSSIBILITIES

Big Data, Big Impact: New Possibilities for International Development. World Economic Forum. World Economic Forum, 2012. Web. 26 July 2012. http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf

Big Data for Development: Challenges & Opportunities. UN Global Pulse. UN Global Pulse, May 2012. Web. 6 Aug. 2012. http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf

Novillo-Ortiz, David, and Alejandro R. Jadad. "The Global People-centred EHealth Innovation Forum." *Ministerio De Sanidad Politica Social E Igualdad*. Affinity, BMJ Group, Feb. 2011. Web. 5 Sept. 2012. http://group.bmj.com/group/affinity-and-society-publishing/Satellite_18_Global_People_Centred_eHealth.pdf

Touching Lives through Mobile Health: Assessment of the Global Market Opportunity. PWC. PWC, Feb. 2012. Web. 29 July 2012. http://www.pwc.com/in/en/assets/pdfs/telecom/gsma-pwc_mhealth_report.pdf

Well-being and Global Success. World Economic Forum, 2012. Web. 26 July 2012. http://www3.weforum.org/docs/WEF_HE_GAC_WellbeingGlobalSuccess_Report_2012.pdf

Dillow, Clay. "How The Largest Health Surveillance System Ever Created Is Preventing An Olympic-Size Pandemic." Popular Science. Popular Science, 3 Aug. 2012. Web. 24 Aug. 2012. http://www.popsci.com/science/article/2012-07/how-it-works-worlds-largest-health-surveillance-system

Mokdad, Ali. "Can London Learn from New York Public Health Policy?" The Guardian. Guardian News and Media, 10 Aug. 2012. Web. 24 Aug. 2012. http://www.guardian.co.uk/healthcare-network/2012/aug/10/london-new-york-chronic-disease

Petersen, Chris. "EXPERT ADVICE Big Data and the London Olympics Cybersecurity Challenge." Technology News: Cybersecurity: Big Data and the London Olympics Cybersecurity Challenge. TechNews World, 27 July 2012. Web. 24 Aug. 2012. http://www.technewsworld.com/story/75754.html

THE GOVERNMENT'S CONTRIBUTION

Holdren, John P., Eric Lander, and Harold Varmus. "Report to the President Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward." President's Council of Advisors on Science and Technology. Executive Office of the President, Dec. 2010. Web. 5 Sept. 2012. http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf

"Digital Government: Building a 21st Century Platform to Better Serve the American People." *Digital Government*. Executive Office of the President of the United States, 23 May 2012. Web. 05 Sept. 2012. http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html

STRADDLING THE PRIVACY FENCE

Cavoukian, Ann, and Jeff Jonas. "Privacy by Design in the Age of Big Data." Privacy by Design. N.p., 8 June 2012. Web. 5 Sept. 2012. http://privacybydesign.ca/content/uploads/2012/06/pbd-big_data.pdf

Lauren Boyles, Jan, Aaron Smith, and Mary Madden. "Privacy and Data Management on Mobile Devices." Pew Internet. Pew Research Internet, 5 Sept. 2012. Web. 5 Sept. 2012. http://pewinternet.org/Reports/2012/Mobile-Privacy.aspx

Acknowledgements

The following contributed to this report:

Name	Title	Company
Chris Hogg	Co-Founder & CEO	100 Plus
Romer Rosales, PhD	Chief Data Scientist	100 Plus
Gordon Norman, MD, MBA	EVP & Chief Innovation Officer	Alere
Sultan Meghji	Vice President of Product Strategy	Appistry
Mark Gehring	Founder & President	Asthmapolis
Jeremy Delinsky	Senior Vice President & Chief Technology Officer	athenahealth
Dave Chase	Founder & CEO	Avado
Chris Long	EVP & COO	BeyondCore
Arijit Sengupta	Founder & CEO	BeyondCore
Jack Challis, PhD	Founder	CliniCast
Andrew Litt, MD	Chief Medical Officer	Dell
Amanda Engler	Global Communications	Dell
Hayley Jordan	Marketing Thought Leadership	Dell
Mark Olesen	President & Chief Operating Officer	DNAnexus
Angela Anderson	Marketing Specialist	DNAnexus
Jason Gilder, PhD	Director of Data Curation	Explorys
Laura Panzica	Marketing Manager	Explorys
Eva Ho	Vice President of Marketing & Operations	Factual
Aaron Crow	Software Engineer	Factual

Name	Title	Company
Dave Shackleton	VP of Product Strategy	Frost Venture Partners
Mark Boguski, MD, PhD	Founder	Genome Health Solutions
Anmol Madan, PhD	Co-Founder & CEO	Ginger.io
Aaron Perlut	Managing Partner	Go Elastic
Carol McCall, FSA, MAAA	Chief Strategy Officer	GNS Healthcare
Philip Schaadt	President & CEO	Haddon Hill Group
Dan Riskin, MD, MBA	Co-Founder & CEO	Health Fidelity
Charlie Bess, P.E.	Fellow, HP Labs	Hewlett-Packard
Allen Kamer	Vice President, Corporate Development & Marketing	Humedica
Bettina Experton, MD, MPH	President & CEO	Humetrix
Nina Hein	Project Manager	Humetrix
Stephen Gold	Vice President of World Wide Marketing	IBM Watson
Kristen Dattoli	Global External Communications	Business Analytics & Watson Solutions
Peter Speyer	Director Data Development	IHME
Michael Kim	Founder & CEO	Kairos Labs
Guy Cavet	VP Life Sciences	Kaggle
Yan Chow, MD, MBA	Director, Innovation	Kaiser Permanente Information Technology
Katie Vukas	Senior Account Executive	KNB Communications
Joe Hardy, PhD	Vice President of R &D	Lumos Labs
Tom Lawry	Director World Wide Health at Microsoft	Microsoft
Bill Crounse, MD	Senior Director of World Wide Health, World Wide Public Sector	Microsoft
Trish Hickey	Simplicity Consulting	Microsoft

Name	Title	Company
Alpana Verma-Alag, MD	Head of Clinical Development	NextBio
John Shade	CEO	One Health
Drew Paxton	Chief Marketing Officer	One Health
Lauren Fifield	Senior Health Policy Advisor	Practice Fusion
Jake Marcus	Data Scientist	Practice Fusion
Stephen Cheung	Web Analytics	Practice Fusion
Mark McNally	Head of Business Development	Predixion Software
Jamie MacLennan, PhD	Co-Founder & CTO	Predixion Software
Jennifer Dodos	Director of Marketing	Predixion Software
Don Jones	Global Strategy and Market Development	Qualcomm Life
Jamie Eisinger	Marketing Communications	Qualcomm Life
The Rock Health Team		Rock Health
Joshua Rosenthal, PhD	Co-Founder & CSO	Rowdmap
Graham Dodge	Co-Founder & CEO	Sickweather
Jeff Loucks	PR Manager	Sproxil
Alden Zecha	CFO & Strategist	Sproxil
Kris Singh	President	The Service Research and Innovation Institute
Martin Leach, PhD	Chief Information Officer	The Broad Institute
Myla Wagner	President	Myla Wagner Public Relations
Malinda Peeples	Vice President of Clinical Advocacy	WellDoc
Bharath Sudharsan	Senior Analytics Associate	WellDoc
Eileen Bartholomew	Vice President of Prize Development	X-Prize
Ben Rubin	Co-Founder & CTO	Zeo

Dr. Bonnie 360°