BIG DATA IN HEALTH

Gaurav Goel
09/29/2015

AGENDA

- BIG DATA IN GENERAL
- DATA CAPTURE THROUGH SENSORS
- What can BIG DATA DO IN HEALTHCARE ?
- BIG DATA SOURCES
- GOVT&INDUSTRY ENDORSMENTS IN BIG DATA
- BIG DATA OPPORTUNITIES AND CHALLENGES
- BIG DATA ALGORITHMS IN REAL WORLD
- HARDWARE/TECHNOLOGY FOR BIG DATA
- BIG DATA STORING AND PROCESSING TECHNOLOGY

BIG DATA ? AND WHY DO WE NEED IT ?

Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, and information privacy.

Often highcosts, lost opportunity, poor decision making, governance and customer service are the pitfalls of poor data capture and management.

But datacapture and management is not so simple and it involves several overlapping fields to get a data insights and make real difference.

Paradigm Shift

Standard medical practice is moving from relatively ad-hoc and subjective decision making to evidence-based healthcare.

BIG DATA ? AND WHY DO WE NEED IT ?

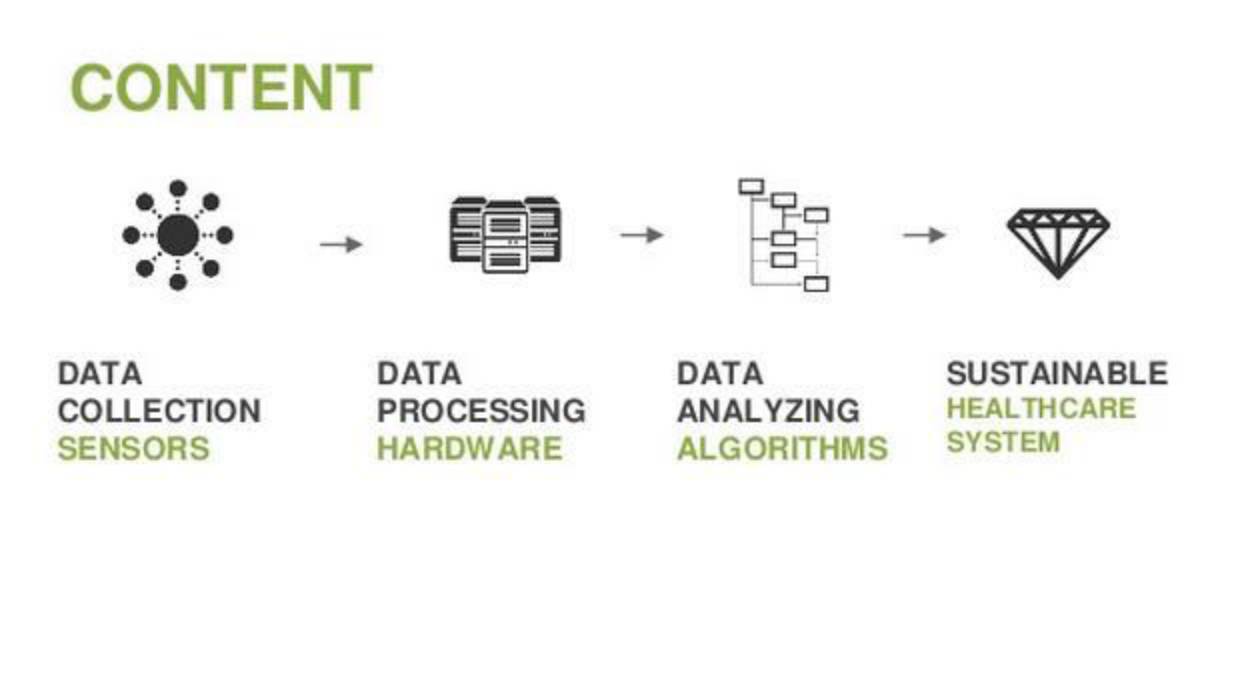
Economic Advantage

It is widely believed that the use of information technology can reduce the cost of healthcare while improving its quality by making care more preventive and personalized and basing it on more extensive (home-based) continuous monitoring. McKinsey estimates [McK2011] a savings of **300 billion dollars** every year in the US alone.

Indirect Social Benefits

It will improve social mood and less physical violence, more innovations,Better policy formulations in general

BIG DATA WORKFLOW



CONTENT

DATA COLLECTION
SENSORS

→

DATA PROCESSING
HARDWARE

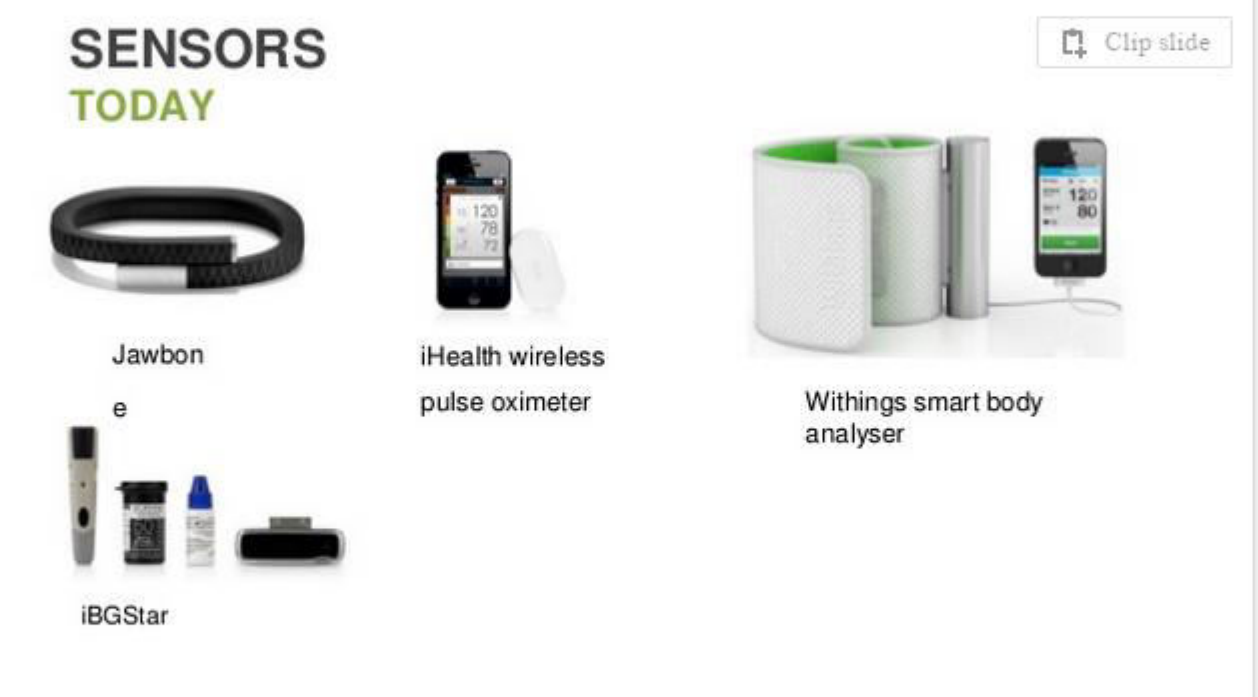
→

DATA ANALYZING
ALGORITHMS

→

SUSTAINABLE
HEALTHCARE SYSTEM

HEALTH DATA CAPTURE THROUGH WEARABLE DEVICES



SENSORS TODAY

Jawbon

e

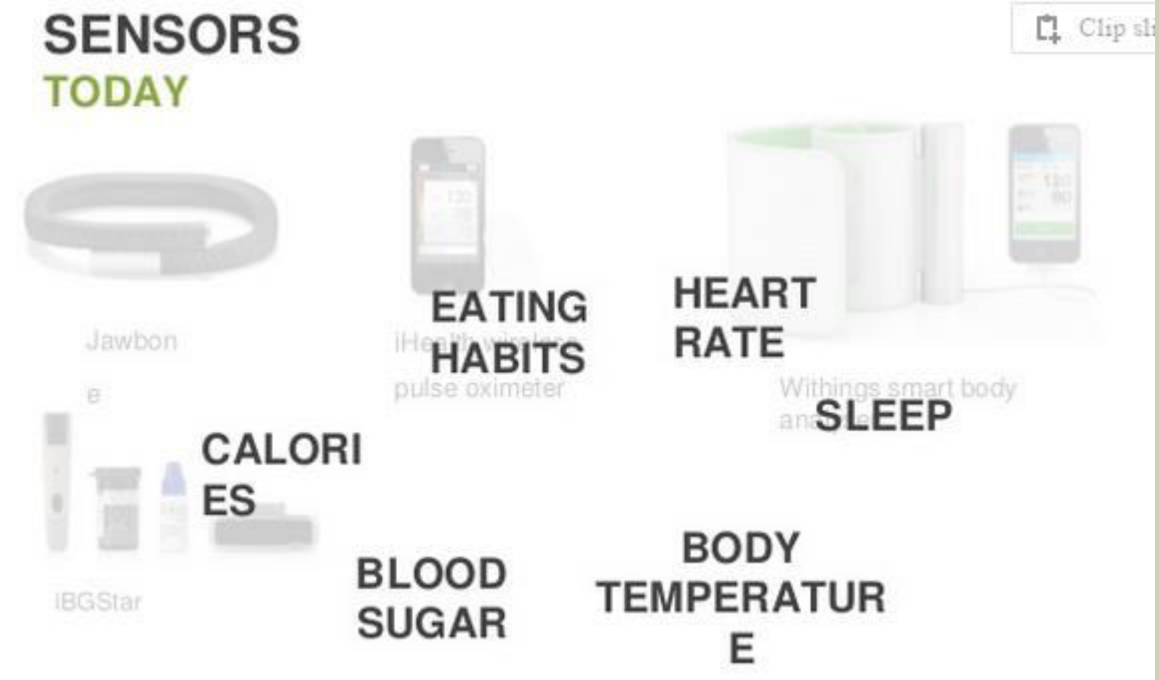
iHealth wireless

pulse oximeter

Withings smart body analyser

iBGStar

CURRENT DATA CAPTURING SENSORS



SENSORS
TODAY

Clip sli

Jawbon
e

iHealth wireless
pulse oximeter

EATING
HABITS

HEART
RATE

Withings smart body
analyzer

SLEEP

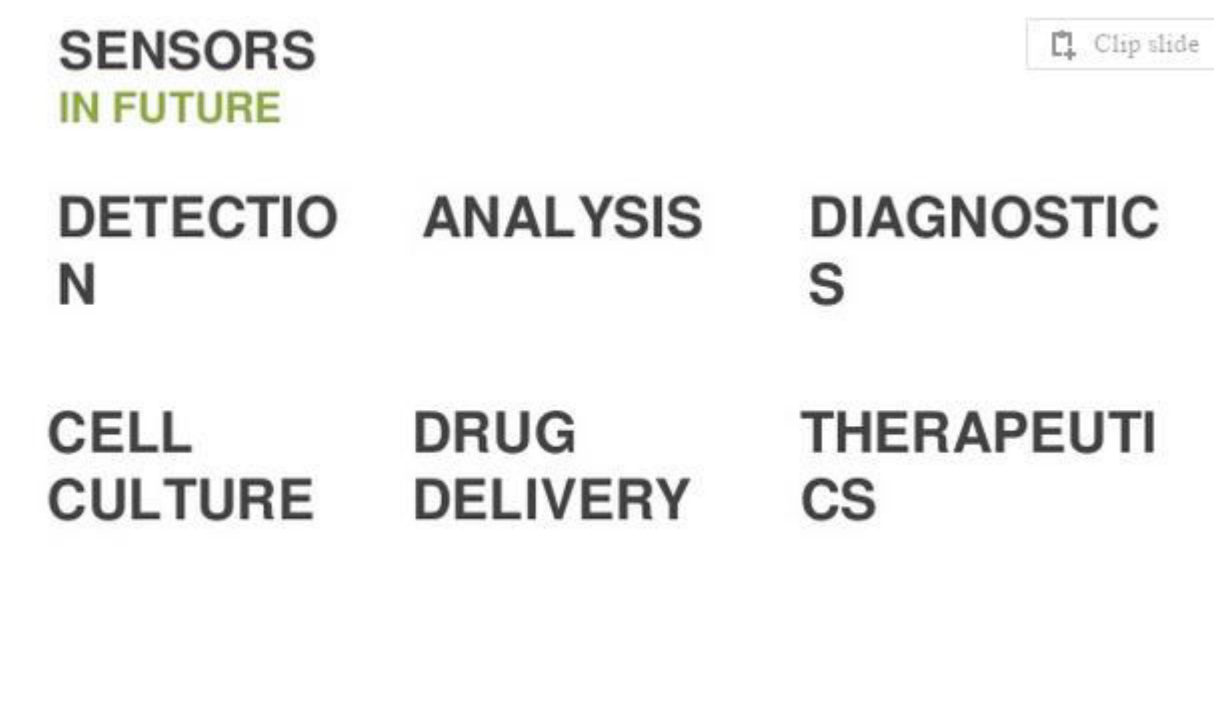
CALORI
ES

iBGStar

BLOOD
SUGAR

BODY
TEMPERATUR
E

FUTURE DATA CAPTURING SENSORS



SENSORS
IN FUTURE

DETECTION

ANALYSIS

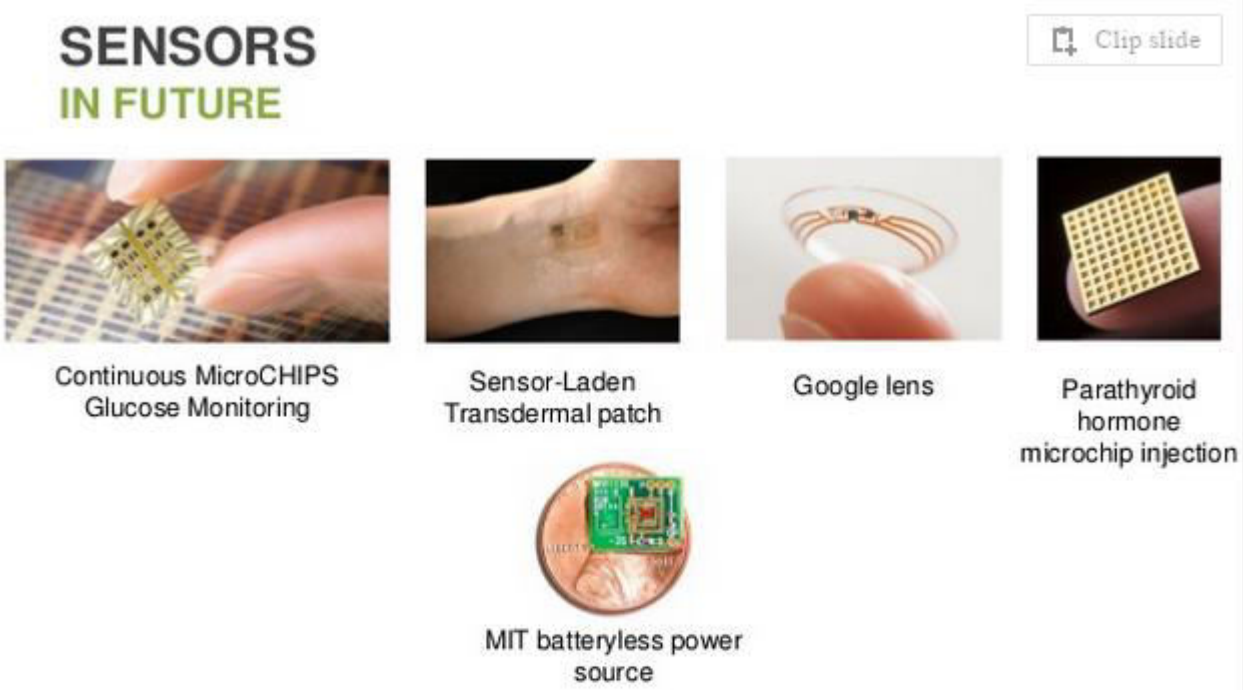
DIAGNOSTICS

CELL CULTURE

DRUG DELIVERY

THERAPEUTICS

FUTURE DATA CAPTURING SENSORS



SENSORS
IN FUTURE

Continuous MicroCHIPS Glucose Monitoring

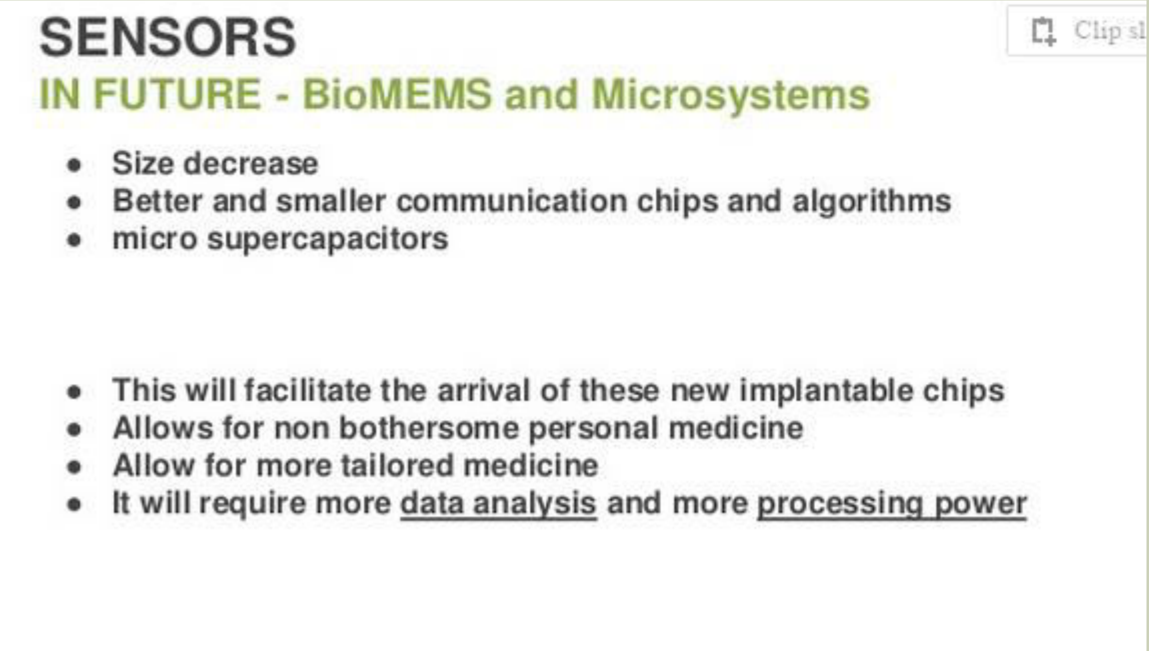
Sensor-Laden Transdermal patch

Google lens

Parathyroid hormone microchip injection

MIT batteryless power source

FUTURE DATA CAPTURING SENSORS



SENSORS

IN FUTURE - BioMEMS and Microsystems

- Size decrease
- Better and smaller communication chips and algorithms
- micro supercapacitors

- This will facilitate the arrival of these new implantable chips
- Allows for non bothersome personal medicine
- Allow for more tailored medicine
- It will require more data analysis and more processing power

WHAT CAN BIG DATA DO IN HEALTHCARE ?

1. **Preventing Organ Failure**
2. **Cost-Effective Treatment Of Chronic Disease**
3. **Reducing Drug Reactions**
4. **Avoiding Readmissions**
5. **To Prevent Fraudulent Behaviour of overbilling patients**

By aggregating and analyzing health data from disparate sources, such as clinical, financial and administrative data, the outcome of treatments in relation to the resource utilization can be monitored.

DATA SOURCES IN HEALTH CARE

- Clinical data is a staple resource for most health and medical research. Clinical data is either collected during the course of ongoing patient care or as part of a formal clinical trial program. Clinical data falls into six major types

- **Electronic health records**
- **Administrative data**
- **Claims data**
- **Disease registries**
- **Health surveys**
- **Clinical trials data**

ELECTRONIC HEALTH RECORDS

- An **electronic health record** (**EHR**) is a digital version of a patient's paper chart. EHRs are real-time, patient-centered **records** that make information available instantly and securely to authorized users.

What EHR contains ?

- Contain a patient’s medical history, diagnoses, medications, treatment plans, immunization dates, allergies, radiology images, and laboratory and test results
- Allow access to evidence-based tools that providers can use to make decisions about a patient’s care
- Automate and streamline provider workflow

BENEFITS OF EMRS

- **Electronic medical records (EMRs)** are a digital version of the paper charts in the clinician's office. An EMR contains the medical and treatment history of the patients in one practice. EMRs have advantages over paper records. For example, EMRs allow clinicians to:
- Track data over time
- Easily identify which patients are due for preventive screenings or checkups
- Check how their patients are doing on certain parameters—such as blood pressure readings or vaccinations
- Monitor and improve overall quality of care within the practice
- But the information in EMRs doesn't travel easily *out* of the practice. In fact, the patient's record might even have to be printed out and delivered by mail to specialists and other members of the care team. In that regard, EMRs are not much better than a paper record.

BENEFITS OF EHRS

- With fully functional EHRs, all members of the team have ready access to the latest information allowing for more coordinated, patient-centered care. With EHRs:
- The information gathered by the primary care provider tells the emergency department clinician about the patient’s life threatening allergy, so that care can be adjusted appropriately, even if the patient is unconscious.
- A patient can log on to his own record and see the trend of the lab results over the last year, which can help motivate him to take his medications and keep up with the lifestyle changes that have improved the numbers.
- The lab results run last week are already in the record to tell the specialist what she needs to know without running duplicate tests.
- The clinician’s notes from the patient’s hospital stay can help inform the discharge instructions and follow-up care and enable the patient to move from one care setting to another more smoothly.

ADMINISTRATIVE DATA

These are primarily hospital discharge data reported to a government agency like AHRQ.

- Healthcare Cost & Utilization Project (H-CUP)
- HCUPnet is a free, on-line query system based on data from the Healthcare Cost and Utilization Project (HCUP). It provides access to health statistics and information on hospital inpatient and emergency department utilization.
- Nationwide Inpatient Sample
  Kids Inpatient Database
  State Inpatient Databases
  State Ambulatory Surgery Databases
  State Emergency Department Databases

CLAIMS DATA
DATA SOURCES CONTD.

- Claims data describe the billable interactions (insurance claims) between insured patients and the healthcare delivery system. Claims data falls into four general categories: inpatient, outpatient, pharmacy, and enrollment. The sources of claims data can be obtained from the government (e.g., Medicare) and/or commercial health firms (e.g., United HealthCare).

- Basic Stand Alone (BSA) Medicare Claims Public Use Files (PUFs) This is the Basic Stand Alone (BSA) Public Use Files (PUF) for Medicare claims. This is a claim-level file in which each record is a claim incurred by a 5% sample of Medicare beneficiaries. Claims include inpatient/outpatient care, prescription drugs, DME, SNF, hospice, etc. There are some demographic and claim-related variables provided in every PUF.

- Medicaid Statistical Information System :MSIS is the basic source of state-submitted eligibility and claims data on the Medicaid population, their characteristics, utilization, and payments and is available by clicking on the link on the left-side column.

DISEASE REGISTRIES DATA SOURCES CONTD.

- Disease registries are clinical information systems that track a narrow range of key data for certain chronic conditions. The most common conditions captured include cancer, diabetes, heart disease, and asthma. Registries often provide critical information for managing patient conditions.

- Global Alzheimer's Association Interactive Network (GAAIN)

The Global Alzheimer’s Association Interactive Network (GAAIN) is a collaborative project that will provide researchers around the globe with access to a vast repository of Alzheimer’s disease research data and the sophisticated analytical tools and computational power needed to work with that data.

- National Cardiovascular Data Registry (NCDR)

The NCDR® is the American College of Cardiology’s worldwide suite of data registries helping hospitals and private practices measure and improve the quality of cardiovascular care they provide. The NCDR encompasses six hospital-based registries and one outpatient registry. There are currently more than 2,400 hospitals and nearly 1,000 outpatient providers participating in NCDR registries.

HEALTH SURVEYS
DATA SOURCES CONTD.

- In order to provide an accurate evaluation of the population health, national surveys of the most common chronic conditions are generally conducted to provide prevalence estimates. National surveys are one of the few types of data collected specifically for research purposes, thus making it more widely accessible.
- Medicare Current Beneficiary Survey The Medicare Current Beneficiary Survey (MCBS) is a continuous, multipurpose survey of a nationally representative sample of the Medicare population. The central goals of MCBS are to determine expenditures and sources of payment for all services used by Medicare beneficiaries.
- National Health & Nutrition Examination Survey (NHANES) The National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations.
- National Long Term Care Survey SThe National Long Term Care Survey is funded through a Cooperative Agreement between the National Institute on Aging (NIA) and Duke University. It is a longitudinal survey designed to study changes in the health and functional status of older Americans (aged 65+). It also tracks health expenditures, Medicare service use, and the availability of personal, family, and community resources for caregiving.

CLINICAL TRIAL DATABASES DATA SOURCES CONTD.

- ClinicalTrials.gov Registry and results database. Information on publicly and privately supported clinical studies from around the world.
- Current Controlled Trials Registry of randomized controlled trials. Can search by ISRCTN.
- European Union Clinical Trials Database Protocol and results information on interventional clinical trials conducted in the EU from May 1, 2004 onwards.
  o Good source of pediatric drug development trials.
- IFPMA Clinical Trials Portal (Pharmaceutical Manufacturers & Associations)o Search portal for ongoing trials and trial results from ClinicalTrials.gov, Current Controlled Trials, Japan Pharmaceutical Information Center, and member company corporate websites.
  Best single place to search for pharmaceutical company sponsored clinical trials.
- WHO International Clinical Trials Registry Platform Access to data provided by clinical trial registries around the world that meet WHO criteria for content and quality.

BIG DATA IN ACTION

Government Initiatives

Medicare Penalties: Medicare penalizes hospitals that have high rates of readmissions among patients with Heart failure, Heart attack, Pneumonia.

Hospitalizations account for more than 30% of the 2 trillion annual cost of healthcare in the United States. Around 20% of all hospital admissions occur within 30 days of a previous discharge. – not only expensive but are also potentially harmful, and most importantly, they are often preventable.

BRAIN Initiative: Find new ways to treat, cure, and even prevent brain disorders, such as Alzheimer's disease, epilepsy, and traumatic brain injury. A new bold $100 million research initiative designed to revolutionize our understanding of the human brain.

Industry Initiatives

Heritage Health Prize: Develop algorithms to predict the number of days a patient will spend in a hospital in the next year. http://www.heritagehealthprize.com

GE Head Health Challenge: Methods for Diagnosis and Prognosis of Mild Traumatic Brain Injuries. Develop Algorithms and Analytical Tools, and Biomarkers and other technologies. A total of $60M in awards.

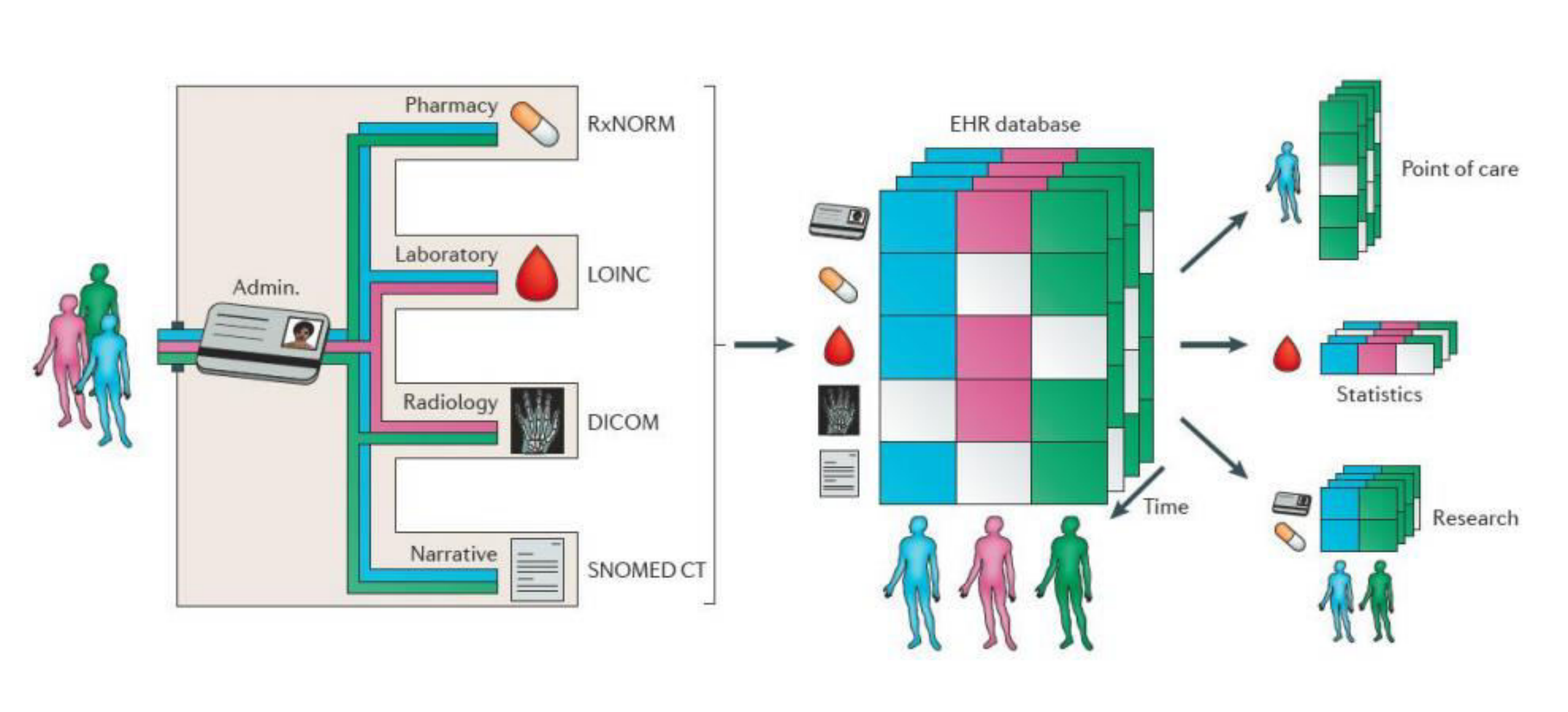
GOVT BIG DATA ENDORSEMENTS

- The US President unveiled a new bold $100 million research initiative designed to revolutionize our understanding of the human brain. BRAIN (Brain Research through Advancing Innovative Neurotechnologies) Initiative.
- Find new ways to treat, cure, and even prevent brain disorders, such as Alzheimer's disease, epilepsy, and traumatic brain injury.

- *"Every dollar we invested to map the human genome returned $140 to our economy... Today, our scientists are mapping the human brain to unlock the answers to Alzheimer's."*

-- President Barack Obama, 2013 State of the Union.

PATIENT DATA TRAIL



BIG DATA OPPORTUNITIES AND CHALLENGES

- Analysis of data sets can find new correlations, to "spot business trends, prevent diseases, combat crime and so on." Scientists, business executives, practitioners of media and advertising and governments alike regularly meet difficulties with large data sets in areas including Internet search, finance and Business informatics. Scientists encounter limitations in e-Science work, including meteorology,genomics,connectomics, complex physics simulations,and biological and environmental research.
- Data sets grow in size in part because they are increasingly being gathered by cheap and numerous information-sensing mobile devices, aerial (remote sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers, and wireless sensor networks.Work with big data is necessarily uncommon; most analysis is of "PC size" data, on a desktop PC or notebook that can handle the available data set.

BIG DATA OPPORTUNITIES AND CHALLENGES

- Digital **curation**

It is the selection, preservation, maintenance, collection and archiving of digital assets. Digital **curation** establishes, maintains and adds value to repositories of digital data for present and future use.

BIG DATA OPPORTUNITIES AND CHALLENGES

- Semantic Annotation
- Data heterogeneity (reports, lab reports, images, sensor data, etc.).
- The International Data Corporation (IDC) market research institute estimates that in the upcoming years 90% of health data will be provided in unstructured formats [6]. Semantic annotation is described as a possible solution for processing heterogeneous unstructured data seamlessly.
- The reading and interpretation of such data is accomplished manually by individual clinicians. Without semantic annotations, it is not possible to process the content of unstructured data automatically.

BIG DATA CHALLENGES AND OPPORTUNITIES

Data Sharing

- As of today, a lot of health data is stored in data silos. A seamless exchange and aggregation of the data often relies on individualized solutions due to the lack of standards and flexible interfaces as well as the heterogeneous nature of the data. On average, for instance, in Germany less than 23%, in the UK less than 46% and in the US less than 36% of the healthcare providers use healthcare information exchange technology. (International Classification of Diseases), which is broadly accepted, is used in country-specific adaptations only.
- Although it is feasible from a technical point of view to exchange health data by e.g. using HL76 CDA (Health Level 7 – Clinical Document Architecture), as of today health data is hardly shared across organizations due to non-technical reasons. Standardized data models for clinical data as well as coding schemes for labeling content need to be agreed upon.

BIG DATA CHALLENGES AND OPPORTUNITIES

- Data Privacy and Security

Big Data security and privacy challenges, which include secure computations in distributed programming frameworks, secure data storage and transaction logs, real-time security and compliance monitoring, scalable and composable privacy preserving data managing and analysis approaches and granular access control and audits.

BIG DATA ALGORITHMS IN REAL WORLD

- Genetic algorithm, Support vector machines, Decision tree, Neural network and Cluster Analysis, to disclose the hidden patterns inside the large data set.
- **Genetic algorithm (GA)** is a search heuristic that mimics the process of natural selection. This heuristic (also sometimes called a metaheuristic) is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover.
- By wiki defn. **genome** is the genetic material of an organism
  Gene is hereditary unit transferred from parent to offspring.
-

- Genetic Algorithms are explored in medical applications to characterize patterns and results.

For example, optimizing image analysis such as, assessing classes of cells in blood cell microscope images or for facilitating magnetic resonance tomography (MRT) treatment planning and 3D visualization of image data. Genetic algorithms developed for mammography were adapted for mining patient's having abdominal aortic aneurysms by analyzing abdominal computed tomography (CT) scan reports for common patterns and features of successful and unsuccessful surgeries. Genetic algorithms can be used for optimizing pharmaceutical products. Recently, it was shown that Genetic Algorithms were able to identify additional anti-bacterial peptides with a high activity during a study.

Finally, it was shown that Genetic Algorithms enhance the precision of artificial neural networks (ANNs) such as for hip-bone fracture prediction or for optimizing efficient search strategies of ANNs to predict and discriminate pneumonia within a training group.

GENETIC ALGORITHMS

- Genetic Algorithms for preparation of Antibiotics
- They refer to the class of problems where the solution set runs in exponential time. where n is the number of inputs, in our case the sample space.Order o growth is O(2expn). Let there be 20 rules to create an antidote. Now due to constantly using this antidote, there harmful bacteria have become resistant to basic formula. So, some 20 additional rules are generated to counter such bacteria. Now each individual 20 rules must be matched with additional 20 rules, there by resulting total combination of magnitude 2020 i.e. 1048576000000000000000000. Now let us try to solve this with sun ultra sparc processor having simultaneous 20 processor running at any instant of time. Let all 2020 be run independently (which is not practically feasible as different patterns are dependent on each other). It will take about 15 days to solve this kind of combination.

NEURAL NETWORKS

- Some problems are too complex to be solved through traditional approach. Neural networks are well suited to tackle problems that people are good at solving, like prediction and pattern recognition. Neural networks have been applied within the medical domain for clinical diagnosis image analysis and interpretation, signal analysis and interpretation, and drug development .Their applications can be categorised under four headings: clinical diagnosis, image interpretation, signal interpretation and drug development.

PHYLOGENETICS/ APGAR ALGORITHM

- NEIGHBOR-JOINING: PHYLOGENETICS

“neighbor-joining” algorithm, which, when paired with genetic sequencing, allows biologists to better understand the evolutionary relationships among species or populations––to trace the phylogenetic relationships within major branches of the tree of life. Phylogenetic trees are used in drug development to, for example, identify closely related, naturally occurring chemical compounds suspected to have medicinal value.

- Scoring systems, like Apgar for evaluating a newborn’s condition at birth or APACHE for determining the severity of patients in intensive care, help physicians monitor and predict a patient’s prognosis based on a multitude of factors, from heart rate and oxygen levels to neurological reflexes. It’s a system for seeing a patient holistically, a prognosis or wellness meter.

RSA/ FOURIER TRANSFORM

- RSA: THE ENCRYPTION ALGORITHM

If it weren’t for RSA.It was one of the first practicable encryption algorithms––and encryption is key to the secure sharing of electronic health records.

- FOURIER TRANSFORM: ENHANCING OUR SENSES

It’s a mathematical technique for breaking complex signals into basic components. It allows technicians, for example, to see voltage fluctuations in a wire connecting a microphone to a loud speaker. Because it reduces a signal to a short list of numbers, it’s also used to squeeze audio and image files into portable packages (MP3’s and JPEG’s). Without it, medical imaging wouldn’t exist. Magnetic resonance and ultrasound machines couldn’t turn raw data into pictures that enable doctors to see inside our bodies to diagnose and treat bleeds and broken bones, tears, tumors and more.

HARDWARE/TECHNOLOGY FOR BIG DATA

- Enterprise-grade servers that are well suited for modern big data analytics workloads have:
- Higher compute intensity (high ratio of operations to I/O)
- Increased parallel processing capabilities
- Advanced virtualization capabilities
- Modular systems design
- Elastic scaling capacity
- Enhancements for security and compliance and hardware-assisted encryption
- Increased memory and processor utilization

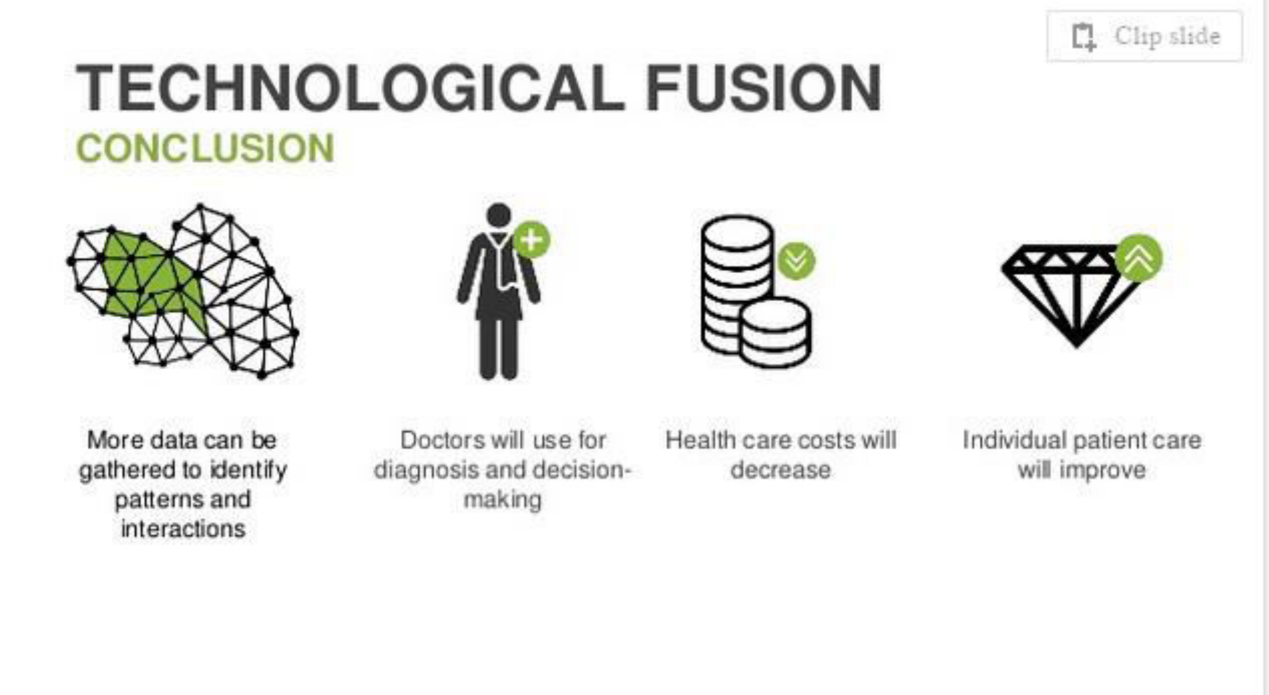
Important Tier 1 vendors such as **Cisco, Dell, HP, and IBM.**

BIG DATA STORING AND PROCESSING TECHNOLOGY

- Apache **Hadoop** is an open-source software framework written in Java for distributed storage and distributed processing of very large data sets on computer clusters built from commodity hardware.
-

  HDFS is like the bucket of the Hadoop system: You dump in your data and it sits there all nice and cozy until you want to do something with it, whether that's running an analysis on it within Hadoop or capturing and exporting a set of data to another tool and performing the analysis there.
- Hadoop is more of a data warehousing system - so it needs a system like MapReduce to actually process the data and it also doesnot use SQL like queries.
- On a Hadoop cluster, the data within HDFS and the MapReduce system are housed on every machine in the cluster. This has two benefits: it adds redundancy to the system in case one machine in the cluster goes down, and it brings the data processing software into the same machines where data is stored, which speeds information retrieval.

CONCLUSION



TECHNOLOGICAL FUSION

CONCLUSION

More data can be gathered to identify patterns and interactions

Doctors will use for diagnosis and decision-making

Health care costs will decrease

Individual patient care will improve

REFERENCES

- http://www.slideshare.net/Funk98/big-data-and-health-care-46960131
- http://www.infoworld.com/article/2610477/big-data/big-data-demands-more-than-commodity-hardware.html
- http://www.ijser.org/researchpaper%5CGenetic-Algorithm-and-its-application-to-Big-Data-Analysis.pdf
- http://readwrite.com/2013/05/23/hadoop-what-it-is-and-how-it-works
- http://www.informationweek.com/big-data/hardware-architectures/10-hadoop-hardware-leaders/d/d-id/1234772
- http://www.infoworld.com/article/2610477/big-data/big-data-demands-more-than-commodity-hardware.html
- http://arxiv.org/ftp/arxiv/papers/1307/1307.5437.pdf
- http://guides.lib.uw.edu/hsl/data/findclin
- https://www.healthcatalyst.com/big-data-in-healthcare-made-simple

REFERENCES

- http://blog.readydock.net/bid/395388/Big-Data-And-4-Ways-It-Is-Reducing-Healthcare-Costs
- https://datafloq.com/read/three-innovative-ways-big-data-will-improve-health/165
- http://www.mckinsey.com/insights/health_systems_and_services/the_big-data_revolution_in_us_health_care
- http://healthsciences.utah.edu/innovation/tenalgorithms/index.php
- http://www.ijera.com/papers/Vol3_issue6/EZ36937941.pdf
- https://hbr.org/2014/12/why-health-care-may-finally-be-ready-for-big-data
- http://www.healthcareitnews.com/news/report-patients-social-media-data-may-be-next-thing-big-data
- http://ihealthtran.com/big-data-in-healthcare
- http://www.businessinsider.com/big-data-and-healthcare-2015-7
- http://www.forbes.com/sites/gilpress/2013/12/12/16-1-billion-big-data-market-2014-predictions-from-idc-and-iia/
- http://www.prnewswire.com/news-releases/the-future-of-healthcare-is-today-how-prepared-are-organizations-to-leverage-big-data-300157860.html
- http://www.intel.com/content/www/us/en/healthcare-it/big-data-in-healthcare.html
- https://www.coursera.org/course/bigdataanalytics