



## Big Data Analytics in Healthcare- Its Benefits, Phases and Challenges

Jasleen Kaur Bains

Department of Computer Science and Applications, Panjab University,  
Chandigarh, India

---

**Abstract:** *With the advent of technology and growing amount of data (Big Data), need is felt towards implementing effective analytics techniques (Big Data Analytics) to analyse this big volume of data for unknown and useful facts, patterns, associations and trends which can give birth to new line of treatment of diseases and provide high quality healthcare at lower cost to all. This paper gives a wide insight and know how about the various Big Data analytics (BDA) initiatives taken to improve healthcare worldwide. It also explains the various phases involved in BDA process and depicts its benefits and challenges with focus on healthcare industry.*

**Keywords:** *Big Data, BDA, Healthcare, BDA Process, BDA benefits*

---

### I. INTRODUCTION

Big Data is a buzzword which is reigning the innovation market from quiet sometime. An enormous amount of data often referred to as Big Data is getting generated everyday by diverse segments of industries like business, finance, manufacturing, healthcare, education, research and development etc. The traditional DBMS's and RDBMS's in market are incapable of storing such vast amount of data. We cannot take full advantage of the hidden knowledge and information from this data as the traditional data mining algorithms do not work effectively on this enormous data. So there is need of developing and using effective, innovative tools and technologies offered by Big Data.

*"If you want to find out how Big Data is helping to make the world a better place, there's no better example than the uses being found for it in healthcare."* [9]

The healthcare industry has generated enormous amount of data till date which is measured in petabyte /exabyte scale. According to [3], with such fast growing rate of growth of data, U.S. healthcare alone will soon reach the zettabyte ( $10^{21}$  gigabytes) scale. The goal of healthcare industry is to analyse this big volume of data for unknown and useful facts, patterns, associations and trends with help of machine learning algorithms, which can give birth to new line of treatment of diseases. The aim is to provide high quality healthcare at lower cost to all. This will benefit the community and nation as a whole.

The 5 V's of Big Data relevant to Healthcare are:

**Volume:** As mentioned earlier, healthcare industry generates prodigious data at staggering rate. The report from EMC and the research firm IDC anticipates an overall increase in health data of 48 percent annually. According to the report the volume of healthcare data in 2013 was 153 exabytes and it may increase to 2,314 exabytes by 2020.[22]

**Variety:** In the past, the emphasis had been on generating clinical data for patients with similar symptoms, storing and analysing it to derive the most effective course of treatment for the admitted patient. Now the healthcare industry is focusing on complete healthcare, by providing an effective treatment through analysis of a patient's data from various other sources too. This refers to variety.

**Variety:** The varied health care data generally falls into one of the three categories i.e. structured, semi structured and unstructured. Generally the following data is gathered: clinical data from Clinical Decision Support systems (CDSS) (physician's notes, genomic data, behavioural data, data in Electronic Health Records (EHR), Electronic Medical Records (EMR)), machine generated sensor data, data from wearable devices, Medical Image data (from CT scan, MRI, X Ray's etc), medical claim related data, hospital's administrative data, national health register data, medicine and surgical instruments expiry date identification based on RFID data[6], social media data like twitter feeds, Facebook status, web pages, blogs, articles.[3]

**Velocity:** It refers to the speed at which new data is generated and moves around. The sensor devices and wearables collect real time physiological data of patients at a rapid pace or velocity. This new data being generated every second poses a big challenge for data analysts. Social media data also adds to velocity as the users views, posts, feeds scale up in seconds to enormous amount in case of epidemics/national disasters.

**Veracity:** It refers to trustworthiness of data. Analysing such a voluminous, variable and fast paced data is a hurricane task. There is no scope for error especially in critical healthcare solutions where patient's life is at stake. The primary aim is to ensure the data is reliable. It is quiet difficult to ensure the reliability of unstructured data for e.g. usage of different terms/abbreviations for disease/symptoms, misinterpreted prescriptions due to bad handwriting, false/fake comments on social portals, improper readings from faulty machines/sensors, incomplete/inaccurate data filled by patients etc.

**Value:** It refers to the quality of data. Normally data from EMR's and EHR's are recognized as high value data. But it is hard to ascertain the value of data from social media. The effective analysis of high value data can lead to better quality ,effective healthcare solutions and innovations

## **II. WORLDWIDE BDA INITIATIVES IN HEALTHCARE INDUSTRY**

There have been upsurge in the number of big data initiatives around the globe. Some of the initiatives are as follows:

New Zealand's Ministry of Health in collaboration with experts from the New Zealand Society for the Study of Diabetes (NZSSD) have used SAS data analysis capabilities, for establishing a Virtual Diabetes Register (VDR) that combines and filters various sources of health information to more accurately determine how many people are diagnosed with the diabetic condition, as well as predicting who is likely to develop diabetes in the future. [12]

McKinley Children's Center based in San Dimas, California, is a premier provider of child welfare services for Los Angeles County and Orange County. The organization serves more than 700 children annually through residential care, foster care and adoptions, special education, and mental health services. The center has launched an innovative big data and analytics initiative to help staff identify the variables that impact each child's success and identify the right combination of programs to improve outcomes. [13]

The Data Science Institute of Columbia University, New York in collaboration with the New York City Department of Health and Mental Hygiene (NYC DOHMH), is working on a project that focuses on the detection of disease outbreaks in New York City restaurants. The goal of the project is to identify and analyze the unprecedented volumes of user-contributed opinions and comments on social media sites such as Twitter, Face book and Yelp, which host massive amounts of content by users about their real-life experiences and opinions about restaurants. It will help to extract reliable indicators of otherwise-unreported disease outbreaks associated with the restaurants. The NYC DOHMH analyzes these indicators, as they are produced, to decide when additional action is required to be taken. This project is developing non-traditional information extraction technology --over redundant, noisy, and often ungrammatical text-- for a public health task of high importance to society at large.[14]

WellPoint, Inc. is an Indianapolis-based health benefits company with approximately 37 million health plan members and processes more than 550 million claims per year. They wanted to reduce the waste of resources (money) by improving the utilization management (UM) process, which governs the preapproval of healthcare insurance coverage for many medical procedures. Its goals were to accelerate processing of physicians' treatment requests, save members' time and improve efficiencies in the approval process, while continuing to base UM decisions on medical evidence and clinical practice guidelines(for ensuring consistency in process).It was a very big challenge considering the volume of data that is analyzed in making UM decisions.

WellPoint teamed up with IBM on a new approach to UM: using the cognitive system IBM Watson to provide approval suggestions to nursing staff based on clinical and patient data. WellPoint trained Watson with 25,000 historical cases. The system uses hypothesis generation and evidence-based learning to generate confidence-scored recommendations that help nurses make decisions about UM. The new system provides responses to all requests in seconds, as opposed to 72 hours for urgent pre-authorization and three to five days for elective procedures with the previous UM process. Encouraged with success of the system, today 15,835 healthcare provider offices use it. [15]

Seattle Children's Hospital and Regional Medical Center, is a 250-bed children's hospital in the Laurelhurst neighbourhood of Seattle, Washington. It treats 350,000 patients annually. It is using big data analytics as part of its Clinical Standard Work (CSW) program, which defines patient populations and recommends an ideal protocol for each population, allowing ensuring that every patient at the hospital receives the same standard of care. The CSW program gets the enormous data from enterprise data warehouse (EDW) which currently integrates data from 10 sources across the hospital, including electronic medical records (EMRs) and billing systems. With the help of CSW program, the doctors and nurses gets complete information based on thousands of data points of each patient, they get answers to complex queries about potential treatments and procedures, and identify pathways of care for patients with particular needs, regardless of provider. Clinicians can also evaluate treatment protocols for determining the resources which need to be allocated by hospital. [16]

Seton Healthcare Family, based in Texas, Austin has over 12,000 employees; over 2600 credentialed physicians; 1.35M outpatient visits; over 68,000 admissions to its hospitals; over 265,000 emergency room visits per year; 38 clinical locations; 10 acute care hospitals. It has used the IBM Content and Predictive Analytics for Healthcare solutions. The system gives an integrated view of relevant clinical and operational information to drive more informed decision making. By teaming unstructured content (History and Physical, Discharge Summaries, Echocardiogram Reports, and Consult Notes) with predictive analytics, Seton is able to identify patients likely for re-admission and introduce early interventions to reduce cost, mortality rates, and improved patient quality of life.[17]

Doctors at UCLA with help of IBM Watson Foundations have recently started using data streaming technology in order to make more informed decisions about brain functions and abnormalities. Using IBM Watson Foundations physicians are able to gather data from sensors to analyze brain functions in real time(data is analyzed by the computer system as it is fed into that system).As a result of the use of this technology, patient care can be substantially improved, and doctors have more time to serve more patients.[18]

The U.K.'s National Health Service is using cloud analytics software to pluck numerical and text data on health-care facilities from spreadsheets and databases and presents it in plain English on its website, NHS Choices. This endeavour is helpful to its citizens, as they can make better choice about their care based on information of about 50,000 health care facilities. NHS Choices is using **Narrative Science Inc. software, called Quill**, a cloud application running in Amazon

Web Services. The software uses natural language generation techniques, to examine through structured data and automatically present it in story form.[19]

US State of North Carolina processes about 88 million claims totalling about \$12 billion annually from 66,000 providers who treat the state's two million Medicaid patients., The State's Department of health and human services in collaboration with IBM used big data analytics to help identify suspicious billing patterns by healthcare providers. Using three years' worth of North Carolina Medicaid claims data, IBM data mining software, which featured special algorithms and modelling capabilities, was applied to detect common fraud and abuse schemes. Almost 90% reduction in fraud was achieved.[20]

i2b2 (Informatics for Integrating Biology and the Bedside) is an NIH-funded National Center for Biomedical Computing based at Partners HealthCare System. The i2b2 Center is developing a scalable informatics framework that will enable clinical researchers to use existing clinical data for discovery research and, when combined with IRB-approved genomic data, facilitate the design of targeted therapies for individual patients with diseases having genetic origins.[21]

Human Connectome Project led by Washington University, University of Minnesota, and Oxford University (the WU-Minn HCP consortium) is working to map the human brain by making a comprehensive connectivity diagram. It will yield invaluable information about brain connectivity, its relationship to behavior, and the contributions of genetic and environmental factors to individual differences in brain circuitry and behavior. This will help to figure reasons why certain people have certain brain disorders, help the physicians to easily diagnose and in certain cases prevention of mental or physical ailments. Over a 3-year span (2012-2015), the Human Connectome Project (HCP) is scanning 1,200 healthy adult subjects. The data thus generated is made publicly accessible quarter wise.[7]

In January 2015, President Obama had officially announced a new biomedical research project "Precision Medicine" that will use the power of big data to help with the development of specialized drugs to treat diseases like cancer and diabetes. The program will collect genetic data on one million Americans so scientists could develop drugs and treatments tailored to the characteristics of individual patients.[22]

### III. PHASES IN THE BDA PROCESS

We can map steps taken up while performing BDA Process to the data mining knowledge discovery steps as follows:

- 1. Data acquisition and storage:** As already mentioned the data is fed to the system through many external sources like clinical data from Clinical Decision Support systems (CDSS), EMR, EHR, machine generated sensor data, data from wearable devices, national health register data, drug related data from Pharmaceutical companies, social media data like twitter feeds, Facebook status, web pages, blogs, articles and many more.[3] This data is either stored in databases or data warehouse. With advent of cloud computing, it is convenient to store such voluminous data on the cloud rather than on physical disks. This is more cost effective and manageable way to store data.
- 2. Data cleaning:** The data which has been acquired should be complete and should be in a structured format, for performing effective analysis. Generally it is seen in that healthcare data from flaws like, many patients don't share their data completely like data about their dietary habits, weight and lifestyle. In such cases the empty fields need to be handled appropriately. Another example can be for e.g.: for field like Gender of person, there can be at most one of two values i.e. male or female. In case any other value or no value is present then such entries need to be marked and handled accordingly. The data from sensors, prescriptions, medical image data and social media data need to be expressed in a structured form suitable for analysis.[2]
- 3. Data integration:** The BDA process uses data accumulated across various platforms. This data can vary in metadata (the number of fields, type, and format). The entire data has to be aggregated correctly and consistently into a dataset which can be effectively used for data analysis purpose. This is a very challenging task, considering the big volume and variety of big data.
- 4. Data querying, analysis and interpretation:** Once the data is cleaned and integrated, the next step is to query the data. A query can be simple query like for eg: What is mortality rate in a particular region ? or complex query as how many patients with diabetes are likely to develop heart related problems in next 5 years? Depending on the complexity of the query, the data analyst has to choose appropriate platform and analysis tools.

A large no. of open source and proprietary platforms and tools are available in market. Some of them are Hadoop, MapReduce, Storm, GridGrain. Big data databases like Cassandra, HBase, MongoDB, CouchDB, OrientDB, Terrastore, Hive etc. Data Mining tools like RapidMiner, Mahout, Orange, Weka, Rattle, KEEL etc. File systems like HDFS and Gluster. Programming languages like Pig/PigLatin, R, ECL. Big data search tools like Lucene, Solr etc. Data Aggregation and transfer tools like Sqoop, Flume, Chukwa. Other tools like Oozie, Zookeeper, Avro, Terracotta. Some open source platforms are also available like Lumify, IKANOW.[11]

The criteria for platform evaluation can vary for different organizations. Generally the ease of use, availability, the capability to handle voluminous data, support for visualization, high quality assurance, cost, security can be some of the variables to decide upon the platform and tool to be used.

### IV. BENEFITS OF USING BIG DATA ANALYTICS IN HEALTHCARE

The tremendous amount of varied data gives an opportunity to researchers in field Health informatics professional to use tools and techniques for unlocking the hidden answers. BDA tools and techniques when applied effectively to this volume data can be beneficial in following ways:

- 1. For individual's/patients:** Generally while deciding any line of treatment for a patient, historical data(of a set of similar patients) about the symptoms, drugs used, outcome/response of different patients is taken into account. With help of BDA, the move is towards formulating a personalised line of treatment for a patient based on his genomic data, location, weather, lifestyle, medical history, response to certain medicines, allergies, family history etc. When the genome data is fully explored some kind of relation can be established between the DNA and a particular disease. Then the specific line of treatment can be formulated for this subset of individuals. The patients will benefit in following ways:
  - Correct and effective line of treatment.
  - Better informed health related decisions.
  - Preventive steps can be taken in time.
  - Continuous health monitoring at patients place using wearable wireless devices.
  - Designing personalised line of treatment for patient.
  - Increase in life expectancy and quality.
- 2. For Hospitals:** By using effective BDA techniques on the data available the hospitals can reap following benefits:
  - Predict the patients which are likely to stay for longer time or going to be readmitted after the treatment.
  - Identifying patients that are at risk for hospitalization. Thus, health care providers could develop new healthcare plans to prevent hospitalization.
  - Various questions can be answered by analysing the data using BDA tools and techniques like will a patient respond positively to a particular treatment?, Is a surgery required to be done on patient and will he/she respond to it?, Is the patient prone to catching disease after treatment?, what is his likelihood of getting affected from the same disease in near future?
  - The hospital authorities can take better informed and managed administrative decisions. Like if no. of patients are not getting cured early and no. of readmissions are increasing because patients become ill again after treatment, then diagnose the root cause of problem, hire more competitive and experienced staff, invest in better drugs/instruments which can aid in effective treatment, increase the cleanliness of hospital, make the treatment more timely, engage more staff on floor, plan for more frequent post treatment follow ups etc
- 3. For Insurance Companies:** A large amount of expenditure is done by governments for giving medical claims to patients. By using BDA we can analyse, identify, predict and minimize the possible frauds related to medical claims. [3]
- 4. For Pharmaceutical company:** By using BDA techniques effectively, the R&D can help to produce in a shorter time ,drugs/ instruments/tools that are most effective for treating a specific disease.
- 5. For government:** The government can use demographic data, historical data of disease outbreak, weather data, data from social media over disease keywords like cholera, flu etc. They can analyse this massive data to predict epidemics, by finding correlation between the weather and likely occurrence of disease. Therefore preventive measures can be used for avoiding the same. The BDA can help in improving the public health surveillance and speed up the response to disease outbreaks.[3]

## V. CHALLENGES

The promises of big data are huge for healthcare, but there are quite a number of challenges which need to be taken up.

- 1. Unstructured data and provenance of data :**As has been discussed before, the BDA process uses data from varied sources. Most of the data is unstructured data like medical prescriptions, blogs, tweets, status updates, and comments. We need to generate right metadata for this data and transform it into a structured format. The image and video data should be structured for semantic content and search. Provenance of data along with its metadata should be carried through data analysis process so that it is easier to track the processing steps in case of error.[3] Some intelligent processing techniques should be devised to handle the data input from sensors and wearables in memory. This will help to filter/derive the meaningful data, which can then be stored on permanent storage. Therefore it will save space.
- 2. Missing or incomplete data:** It is seen through practices that the patient tends to hide some of the personal facts or choices about his/her lifestyle while filling up forms or oral interview by physicians. When this data is stored in digital format then many fields remain empty. Sometimes it happens that some of the fields carry wrong values. When analysis is done on the entire data such empty or wrong fields may or may not get processed. In both the cases they produce wrong results. If we are leaving some records as they are empty then our analysis is not on cumulative data. If we consider wrong value fields then the again the analysis is incorrect and unreliable. Such issues need to be addressed.
- 3. Quality of data:** We need processes to ensure that data from sources is a valid data or not and is of good quality. Determining the validation and quality of data in case of social media data is another big challenge.
- 4. Technical challenges:**
  - Data aggregation from different database systems is also a challenge in BDA. It can be made easier by devising certain standard database design practices meant for a specific domain like healthcare, financial sector etc.[3]We need to come up with a lot more technological standards and protocols for different data systems to integrate seamlessly.

- The traditional algorithms for data mining processes or analysis have to be scaled up to handle the big volume of data. Another aspect is parallelisation of algorithms, the processors speed has come to a point beyond which it's hard to increase. Therefore the trend is moving towards multi-core processors. In such a scenario we need statistical algorithms which can be parallelised, else their computing performance will decrease when they handle complex big volume data.[9] Apart from this scaling complex query processing techniques to terabytes, while enabling interactive response times is another big problem.[3]
- An analysis is useful only when a non technical person is able to understand and interpret it. Considering the volume and variety of data used in BDA, it is very hard to depict it visually in a more understandable and easier way. Going one step further than this, a user should be able to perform the repeated analysis with different set of assumptions, data sets and parameters. It will help the user to better understand the analysis process and verify whether the system works in a desired way or not.
- With market flooded with so many open source and proprietary platforms and tools for BDA, we need careful evaluation to use the best platform and tool for our problem.

5. **Data Security:** As more and more data is digitized, the role of securing data is gaining importance. Large no. of people are not willing to share their personal data with a fear of security breach. In case they are assured of full security then this problem can be tackled. We need strict Government policies and norms which should be adhered to regarding what data can be shared and what not. In addition to this, strong technological hardware and software level security measures should be implemented to discourage the hackers and malicious people.
6. **Lack of Experts:** There is a dearth of qualified and experienced data scientists in the world. We need to create an expertise in the field of data science, so that in future we have data scientists who can help to turn the promises of big data into reality.

## VI. CONCLUSION

BDA is a process which has many steps like data acquisition, cleaning etc and each step has a unique challenge. We need to address these challenges and formulate new technological standards, protocols so that at least technologically we become competent enough to manage and analyse such volume of complex data. As has been seen in existing studies, the BDA has shown remarkable outcomes in many healthcare organizations. In the future, with even more advancements in the BDA processes we expect that healthcare cost will come down drastically, life expectancy will increase, and we will see much healthier population as compared to now with people taking more accountability and charge of their health using technological advancements. The future of healthcare is promising.

## REFERENCES

- [1] Vivek Wadhwa, "The rise of big data brings tremendous possibilities and frightening perils", April 2014. Available: <http://www.washingtonpost.com/blogs/innovations/wp/2014/04/18/therise-of-big-data-brings-tremendous-possibilities-and-frightening-perils/>
- [2] D. Agrawal et. al, "Challenges and Opportunities with Big Data", Big Data White Paper- Computing Research Association, Feb 2012. Available: <http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf>
- [3] Wullianallur Raghupathi and Viju Raghupathi, "Big data analytics in healthcare: promise and potential", *Health Information Science and Systems* 2014, 2:3, Available: <http://www.hissjournal.com/content/2/1/3>
- [4] Nambiar, R. ; Cisco Syst., Inc., San Jose, CA, USA ; Bhardwaj, R. ; Sethi, A. ; Vargheese, R., "A look at challenges and opportunities of Big Data analytics in healthcare", IEEE Conference 2013. Available: [http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6691753&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs\\_all.jsp%3Farnumber%3D6691753](http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6691753&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D6691753)
- [5] Ahmed E. Youssef, "A Framework for Secure Healthcare Systems Based on Big Data Analytics in Mobile Cloud Computing Environments", *The International Journal of Ambient Systems and Applications* 06/2014, Available: <http://airccse.org/journal/ijasa/papers/2214asa01.pdf>
- [6] J. Archenaa, E.A. Mary Anita, "A Survey of Big Data Analytics in Healthcare and Government", *Procedia Computer Science*, Elsevier, Volume 50, 2015, Pages 408-413, Big Data, Cloud and Computing Challenges, Available: <http://www.sciencedirect.com/science/article/pii/S1877050915005220>
- [7] Matthew Herland, Taghi M Khoshgoftaar and Randall Wald, "A review of data mining using bigdata in health informatics", Herland et al. *Journal of Big Data* 2014, Springer, 1:2 Available: <http://www.journalofbigdata.com/content/1/1/2>
- [8] MH Kuo, T Sahama, AW Kushniruk, EM Borycki, DK Grunwell, "Health big data analytics: current perspectives, challenges and potential solutions", *International Journal of Big Data Intelligence*, Vol. 1, Issue 1, pp.114-126.
- [9] Bernard Marr, "How Big Data Is Changing Healthcare", Available: <http://www.forbes.com/sites/bernardmarr/2015/04/21/how-big-data-is-changing-healthcare/>
- [10] "Improve healthcare win \$3,000,000", Available: <http://www.heritagehealthprize.com/c/hhp>
- [11] Cynthia Harvey, "50 Top Open Source Tools for Big Data", Available: <http://www.datamation.com/data-center/50-top-open-source-tools-for-big-data-1.html>
- [12] "Big Data provides true picture of diabetic population", Available: [http://www.sas.com/en\\_us/news/sascom/2014q1/nz-ministry-of-health.html](http://www.sas.com/en_us/news/sascom/2014q1/nz-ministry-of-health.html)

- [13] [http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=AB&infotype=PM&appname=SWGE\\_YT\\_YT\\_USEN&htmlfid=YTC03753USEN&attachment=YTC03753USEN.PDF](http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=AB&infotype=PM&appname=SWGE_YT_YT_USEN&htmlfid=YTC03753USEN&attachment=YTC03753USEN.PDF)
- [14] Health Analytics, Available: <http://datascience.columbia.edu/health-analytics>
- [15] [http://www.ibm.com/smarterplanet/us/en/ibmwatson/assets/pdfs/WellPoint\\_Case\\_Study\\_IMC14792.pdf](http://www.ibm.com/smarterplanet/us/en/ibmwatson/assets/pdfs/WellPoint_Case_Study_IMC14792.pdf)
- [16] Linda L. Briggs, "*Big data means better care at Seattle's Children Hospital*", Available: <http://tdwi.org/articles/2013/08/13/big-data-analytics-smarter-care.aspx>
- [17] "*Reducing readmissions to improve care*", Available: <http://www01.ibm.com/software/sg/industry/healthcare/pdf/setonCaseStudy.pdf>
- [18] Joe Clabby, "*Using data stream analysis in brain research at UCLA's School of Medicine*", Available: <http://www.ibmbigdatahub.com/blog/using-data-stream-analysis-brain-research-ucla-s-school-medicine>
- [19] <http://blogs.wsj.com/cio/2014/11/05/big-data-meets-ai-how-the-u-k-s-national-health-service-uses-analytics/>
- [20] Marianne Kolbasuk McGee, "*North Carolina Fights Medicaid Fraud With Analytics*", Available: <http://www.informationweek.com/administration-systems/north-carolina-fights-medicaid-fraud-with-analytics/d/d-id/1104519>
- [21] About us, Available: <https://www.i2b2.org/>
- [22] Luke Dormehl, "*President Obama's New Health Care Initiative Will Harness The Power Of Big Data*", Available: <http://www.fastcompany.com/3041775/fast-feed/president-obamas-new-healthcare-initiative-will-harness-the-power-of-big-data>