Coupling of Big Data with Healthcare-What is Big Data & why it Requires

Arun Kumar Singh

Asst. Professor, College of Computing and Informatics,
Saudi Electronics University, Kingdom of Saudi Arabia-KSA

Abstract - Medicine and healthcare are assuming profound changes. Full-genome sequencing and high-resolution imaging technologies are key handlers of this speedy and vital transformation. Technological revolution combined with automation and miniaturization has triggered an explosion in data production that will soon reach Exabyte proportions. The potential of "big data" for cultivating health is enormous but, at the same time, we face a wide range of encounters to overcome urgently. However, exploitation of the data made available through advances in genomic medicine, imaging, and a wide range of mobile health applications or connected devices is obstructed by numerous historical, technical, legal, and political obstacles. There is a lack of synchronisation of data formats, processing, analysis, and data transfer, which leads to inconsistencies and lost chances. Legal frameworks for data sharing are evolving. Clinicians, researchers, and citizens need upgraded methods, tools, and training to generate, analyse and query data efficiently. Adopting these blocks will contribute to creating the Single Market for health everywhere, which will improve health and healthcare for all citizens. Big data analytics has helped healthcare improve by providing personalized medicine and authoritarian analytics, clinical risk intervention and predictive analytics, waste and care variability reduction, automated external and internal reporting of patient data, standardized medical terms and patient registries and fragmented point solutions. The level of data created within healthcare systems is not trivial. With the added adoption of mHealth, eHealth and wearable technologies the volume of data will renew to increase. There is now an even greater need for such surroundings to pay greater attention to data and information quality. "Big data very often means `dirty data' and the fraction of data inaccuracies increases with data volume growth." Human examination at the big data scale is unmanageable and there is a desperate need in health service for intellectual tools for accuracy and credibility control and handling of information wasted.

I.   INTRODUCTION

Big Data refers to the huge amount of information created and which proves complicated to analyse in real time, meaning that traditional databases are not equipped to store and analyse it [1]. The three V's [2] are normally cited when mentioning Big Data and some authors [3-4] list five Vs, which are: volume, velocity, variety, veracity and value. One of the main commitments of Big Data focuses on healthcare. The aim is to use Big Data and the increasingly powerful computational systems to shift from generic to personalised medicine. The intention is to decode each patient's genetic sequences in order to personalise and know for certain which treatment would be most efficient for each case. Thanks to skill development and Big Data, one of the biggest steps forward is that the investment required for sequencing has dropped from the initial cost of millions to several thousand dollars per genome [5].

Another aim of Big Data is control of epidemics. A georeferenced factor is added to traditional databases and, together with the information obtained from official as well as informal means such as blogs, searches, social media sites, etc., an estimation of the progression of the illness [6] can be made. For instance, Google developed the Google Flu Trends [7] tool, which took into account the amount of keyword searches and comments on social media sites such as Twitter and Facebook to monitor the progression, initially of the H1N1 flu [4].

An enormous amount of information is collected in healthcare and, for this reason, computer analysis techniques are not sufficient, requiring that new algorithms and methods be found to carry out these analyses more efficiently [8-9]. Tools such as Mahout's Apache Hadoop, Google's MapReduce and the NoSQL databases have been developed more as format than a tool [10-11]. It is also necessary to access databases to store the information. Relational databases (such as SQL) often cannot be used with Big Data. One solution is to use non-relational databases (NoSQL). Examples of these include MongoDB, CouchDB o SimpleDB [12-13]. Genomics is one of the main fields in which Big Data is playing an increasingly larger role. It is constantly evolving, with many new discoveries being made. It is in this field that efforts are being directed to discovering and analysing each person's individual genome and, as we mentioned at the beginning, achieve personalised healthcare. Study of DNA and why bacteria's are resistant to antibiotics are also examined in this field [8]. Hospitals are another area where Big Data analysis of patients' tests may give better early diagnostic results. In this manner, patients' drug dosage could be lowered. When said flu virus broke out, some companies were

already using Big Data to analyse its progression. Google, for example, analysed via geolocation and the number of hits matching the search where the virus was spreading with the purpose of finding a model to predict its progression and warn the Centers for Disease Control and Prevention (CDC). Finally, although Big Data use is on the rise in every field of healthcare, it is important to mention diabetes [9–11]. The disease is affecting an increasingly larger number of persons across the world. It is the cause of a considerable number of deaths in the non-communicable disease category. The onset of diabetes can appear at any time, even at birth, and requires lifelong treatment [12]. For this reason, Big Data is being used to explore for a cure as there is a large amount of data and evidence available [13].

cutting-edge Big data in healthcare: In what way, big and better data help healthcare distribution and patient consequences?

With the world's population is accumulative and everyone living longer, healthcare budgets continue to upsurge. But simulations of treatment delivery are quickly changing, and many of the decisions behind those changes are being determined by data. This is helping to decrease costs. Big Data in healthcare is being used to expect epidemics, cure disease, improve quality of life and evade preventable deaths. The effort now is to understand as much about a patient as possible, as early in their life as thinkable – with any luck picking up warning signs of serious illness at an early enough stage so that healing is more simple and less expensive. For the patient, it will mean healthier, faster and inexpensive diagnosis and treatment, and better info about health care suppliers.

Analysing dissimilar and highly active data will benefit different fields like epidemiological research or early detection and avoidance of diseases. By moving from a reporting approach (what has happened?) to a predictive approach (what will happen?), big data is creating a new awareness era in the world of medical care." The use of big data to healthcare offers new ways to quicken research. Huge amounts of data on applicants will allow researchers to gather the best topics. Recently, data-sharing engagements between the pharmaceutical hulks has led to breakthroughs such as the finding that desipramine, commonly used as an anti-depressant, has potential uses in curative types of lung cancer.

At the Science Business Network meeting, they explore the right balance for healthcare, whereas also concentrating other related questions:

- As healthcare costs worldwide grow ever larger, how quickly can big data in healthcare be applied to realise huge cost savings through a more precise identification of unnecessary procedures or duplication of tests?
- The medical industry gathers an enormous amount of data by different doctors' surgeries, hospitals, clinics and administrative departments. How can acceptable data chomping and data analytics applications be used to efficiently search, process and analyse massive datasets?
- How can medical and data professionals cooperate to enable them to peer into the future and identify problems before they happen?
- Can open access to data rapidity up healthcare transport? What are the barriers to open access?
- Can big data become a planned advantage for healthcare companies?

II.   BIG DATA

The Big data term has been in routine since the 1990s. Big data usually incorporates data sets with sizes beyond the ability of commonly used software tools to take, curate, manage and process data within an acceptable passed time. Big data "size" is a continually moving target, as of 2012 ranging from a few dozen terabytes to many petabytes of data. Big data needs a set of techniques and technologies with new forms of incorporation to reveal insights from datasets that are diverse, complex and of a massive scale.

"Big data" has a wide range of definitions in health research [14-15] and to generate a single meaning for all uses ("one size fits all" approach) may be too abstract to be useful. However, a practical definition of what big data means for health research or at least a harmony of what this term means was proposed during the workshop in Luxembourg. "Big data in health" encompasses high volume, high diversity biological, clinical, environmental and lifestyle information collected from single entities to large partners, in relation to their health and wellness status, at one or several time points. Big data can only be allocated with by adopting a strong governance model and best exercises of new technologies, e.g., in large-scale data invention compliant with community based quality standards, coupled with interoperable data storage, data integration and advanced analytics solutions [16]. Another goal of the workshop was to develop an EU action plan for research funders concerning the integration of big data into policy expansion, biomedical research, and clinical practice in health and wellness management. Big data comes from a variety of sources, such as clinical tests, electronic health records (EHRs), patient registries and databases, multidimensional data from genomic, epigenomic, transcriptomic, proteomic, metabolomics, and microbiomic measurements, and medical imaging. More recently, data are being incorporated from social media, socioeconomic or interactive indicators, occupational information, mobile applications, or environmental checking [17]. Big data comes in a wide range of setups.

Data streams have to be measured and understood in a timely manner to benefit patients affected by diseases and to help citizens stay in good health [14-18].

III. TECHNOLOGICAL APPLICATIONS OF BIG DATA IN HEALTH-CARE

Big data has improved the demand of information management specialists so much so that Software AG, Oracle Corporation, IBM, Microsoft, SAP, EMC, HP and Dell have disbursed more than $15 billion on software organisations specializing in data management and analytics. The world's operative capacity to interchange information through telecommunication networks was 281 petabytes in 1986, 471 petabytes in 1993, 2.2 exabytes in 2000, 65 exabytes in 2007 and forecasts put the amount of internet traffic at 667 exabytes annually by 2014. Agreeing to one guess, one third of the globally stored information is in the form of alphanumeric text and still image data, [19] which is the format most useful for most big data applications. This also shows the possible of yet unused data (i.e. in the form of video and audio content). While many merchants offer off-the-shelf solutions for Big Data, experts endorse the development of in-house solutions custom-tailored to solve the company's problematic at hand if the company has enough technical capabilities [20].

A. Government-
The use and acceptance of big data within governmental developments is beneficial and allows competences in terms of cost, productivity, and innovation, [21] but does not come without its mistakes. Data analysis often needs multiple parts of government (central and local) to work in association and create new and innovative progressions to deliver the desired result. Below are the thought leading examples within the governmental big data space.

United States of America
In 2012, the Obama administration declared the Big Data Research and Development Initiative, to discover how big data could be used to address significant problems faced by the government. [22] The initiative is collected of 84 different big data program extent across six departments. [23] Big data analysis played a large role in Barack Obama 's successful 2012 re-election campaign. [24] The United States Federal Government owns six of the ten most powerful supercomputers in the world. [25] The Utah Data Center has been built by the United States National Security Agency.

India
Big data analysis was in part trustworthy for the BJP to win the Indian General Election 2014.The Indian government consumes several techniques to ascertain how the Indian electorate is replying to government action, as well as ideas for policy amplification.

United Kingdom
Data on prescription drugs: by concerning origin, location and the time of each prescription, a research unit was able to demonstrate the significant delay between the release of any given drug, and a UK wide adaptation of the National Institute for Health and Care Excellence guidelines. This recommends that new or most up-to-date drugs take some time to filter through to the general patient. [26] Joining up data: a local authority mixed data about services, such as road gritting rotas, with services for people at risk, such as 'meals on wheels'. The connection of data allowed the local authority to evade any weather-related delay.

B. International development-
Research on the operative usage of information and communication technologies for development (also known as ICT4D) proposes that big data technology can make significant influences but also present distinctive challenges to International development. [27] Developments in big data analysis offer cost-effective chances to improve decision-making in critical development areas such as health care, employment, economic productivity, crime, security, and natural disaster and resource management.

C. Manufacturing-
Based on TCS 2013 Global Trend Study, enhancements in supply planning and product quality offer the greatest benefit of big data for manufacturing. Big data delivers an infrastructure for transparency in manufacturing industry, which is the skill to unravel doubts such as inconsistent component performance and availability. Predictive manufacturing as an applicable tactic toward near-zero downtime and transparency requires massive amount of data and advanced forecast tools for a systematic process of data into useful information. [28] A theoretical framework of predictive manufacturing begins with data attainment where different type of sensory data is available to acquire such as acoustics, vibration, pressure, current, voltage and controller data.

D. Healthcare-

Big data analytics has helped healthcare expand by providing personalized medicine and prescriptive analytics, clinical risk intervention and predictive analytics, waste and care inconsistency reduction, automated external and internal journalism of patient data, standardized medical terms and patient registries and split point solutions. The level of data generated within healthcare systems is not inconsequential. With the added adoption of mHealth, eHealth and wearable technologies the volume of data will remain to rise. There is now an even greater need for such surroundings to pay greater care to data and information quality. [29] "Big data very often means `dirty data' and the fraction of data imprecisions increases with data volume growth." Human assessment at the big data scale is incredible and there is a desperate need in health service for intellectual tools for accuracy and believability control and handling of information missed.

E. Education-

A McKinsey Global Institute study uncovered a shortage of 1.5 million highly trained data professionals and managers and a number of universities including University of Tennessee and UC Berkeley, have formed master's programs to meet this requirement. Private bootcamps have also advanced programs to meet that demand, including free programs like The Data Incubator or paid programs like General Assembly.

F. Media-

To know how the media applies Big Data, it is first compulsory to provide some context into the apparatus used for media process. It has been proposed by Nick Couldry and Joseph Turow that specialists in Media and Advertising approach big data as many actionable points of information about millions of individuals. The industry looks to be moving away from the old-fashioned approach of using specific media situations such as newspapers, magazines, or television shows and instead taps into consumers with technologies that reach aimed people at prime times in optimal locations. The final aim is to serve, or convey, a message or content that is (statistically speaking) in line with the customer's mind-set.

G. Information Technology-

Specifically, since 2015, Big Data has come to eminence within Business Operations as a tool to help employees work more proficiently and streamline the gathering and circulation of Information Technology (IT). The use of Big Data to attack IT and data collection issues within an enterprise is called IT Operations Analytics (ITOA). By relating Big Data principles into the concepts of machine intelligence and deep computing, IT departments can predict possible issues and move to provide solutions before the problems even happen. [30] In this time, ITOA businesses were also commencement to play a major role in systems management by contribution platforms that brought individual data storage tower together and generated insights from the whole of the system rather than from remote pockets of data.

H. Science-

The Large Hadron Collider (LHC) experiments signify about 150 million sensors distributing data 40 million times per second. There are approximately 600 million collisions per second. After filtering and refraining from recording more than 99.995% of these streams, there are 100 collisions of interest per second. As a result, only working with less than 0.001% of the sensor stream data, the data flow from all four LHC experiments represents 25 petabytes annual rate before replication (as of 2012). This becomes nearly 200 petabytes after replication. The Square Kilometre Array is a radio telescope made of thousands of antennas. It is predictable to be operational by 2024. Jointly, these antennas are expected to gather 14 exabytes and store one petabyte per day. It is considered one of the most ambitious scientific projects ever undertaken. [31] Science and research: When the Sloan Digital Sky Survey (SDSS) began to gather astronomical data in 2000, it amassed more in its first few weeks than all data collected in the history of astronomy previously. The DNA sequencers have divided the sequencing cost by 10,000 in the last ten years, which is 100 times lower than the decrease in cost forecast by Moore's Law. [32] The NASA Center for Climate Simulation (NCCS) stores 32 petabytes of climate observations and simulations on the Discover supercomputing cluster.

I. Sports-

Big data can be expended to improve training and thoughtful competitors, using sport sensors. It is also likely to predict winners in a match using big data analytics. Future performance of players could be predicted as well. Thus, players' value and salary is resolute by data collected through the season. The movie MoneyBall exhibits how big data could be used to scout players and also classify underrated players. In Formula One races, race cars with hundreds of sensors produce terabytes of data. These sensors collect data points from tire pressure to fuel burn effectiveness. Then, this data is conveyed to team headquarters in United Kingdom through fiber

optic cables that could carry data at the speed of light. [33] Based on the data, engineers and data analysts resolve whether adjustments should be made in order to win a race.

IV.   IMPORTANCE OF PATIENT REGISTRIES

Patient registries have for decades assisted as a key tool for evaluating clinical outcome and clinical and health technology implementation. Rare disease registries pool data to accomplish an adequate sample size for epidemiological and/ or clinical research. The European Organization for Research and Treatment of Cancer (EORTC) [34] began a prospective registry for patients with melanoma in June 2015 [35]. The European Network of Cancer Registries (ENCR) [36], founded within the framework of the Europe Against Cancer Programme of the EC, helps collaboration between cancer registries, defines data collection standards, provides training for cancer registry people, and regularly disseminates information on incidence and mortality from cancer in the

European Union and other European countries.

Patient registries deliver important potential for research and public health developments in the EU, owing to the large volume of patients in each registry and the variability of quality medical information related to each patient. Patient registries are progressively important to supervise patients' treatments and for safety evaluation and the identification of trends in translational medicine (e.g., registry-based clinical trials, personalized medicine). Patient registries permit informed policy decisions at the local, regional, national, and, in some cases, the international level. As a result, hundreds of registries have been set up that range from national to international rare disease initiatives, coupling clinical and genetic data and biobanks. However, for various reasons, counting data protection and the crumbling of monitoring frameworks, the combination of these disparate information sources to guide health research and decision-making in the clinic has so far lagged behind the use of large-scale, big data assemblies in other sectors. Other disciplines, such as electronic and mechanical engineering, and whole industries, such as construction, airplanes, weather forecasting, or robotics, have exhibited computational modeling and simulation as an important component that is based on data sharing and their experience could help overcome the blocks experienced in health research [37].

Registries in INDIA

While the developed world has realized the value of registries in epidemiology and research, India still lags in the number of registries and the information stored in them. Very few established registries exist such as the National Cancer Registry, The Indian Transplant Registry etc. The data contained is restricted to clinically relevant data with no insights into the economic implications of treatment. In UK and US, the healthcare is funded by the respective government. As a result, optimization of healthcare budgets becomes imperative for the government. Hence, Health Technology Assessments (HTA) and recording of data to serve the needs of HTA becomes a necessity in the western world. In India, on the other hand, as healthcare is not funded by the government, it sees little incentive in assessing healthcare technologies and building systems to generate the relevant data. As a result, registries, databases, medical records and other sources of RWD have remained undeveloped in our country. Hence, in India we are losing out on the tremendous wealth of RWD, which could have been generated, had our population been exposed to appropriate clinical registries. The best source of RWD in the Indian context would be Electronic Health Records (EHRs). However, EHRs in India are fraught with many challenges. [38] Though the government has shown interest in EHRs, apart from a handful of premier institutes using them, they are yet to be accepted by the Indian Medical fraternity. Awareness, training and development of tools to capture RWD are the need of the hour for establishing a HTA system in India. Unfortunately, while the developed nations are recognizing the value of HTAs and RWD, Indian government has shown no will to progress on this front.

A.   *WHY REGISTRIES ARE IMPORTANT FOR Health Technology Assessments (HTA)-Registries can be called as 'storehouses of RWD'-*

Compared to all other sources of RWD, registries form a tool which can collate information from multiple sources. They can also be linked to other sources of RWD such as Electronic health records (EHRs) and administrative databases. Registries can be linked across nations, thus giving a perspective on changing trends across nations. This is of particular importance in cases of diseases common to a certain race or rare diseases. The volume of data stored in registries is immense. Registries track data spanning decades, something which is beyond the scope of traditional RCTs. Patients are followed up in time after enrolling. Thus, they give insights into changing trends of a particular disease, therapy and demography over time. Registries not only track clinical and drug related efficacy, but can also be developed so as to track economic data such as resource utilization and QOL. Thus, they can be used for getting a holistic view from the data collected. The number of patients covered by registries is huge, unlike RCTs. The large number of patients makes statistical analyses and tests of significance more valid. Owing to the numbers involved there is also greater likelihood of detecting the

relatively rare events which may be missed during RCTs. Though registries don't have a control on the treatment given by a clinician, they can reveal the degree to which clinicians are managing a particular disease in accordance with the principles of evidence-based medicine. This information is of great use to healthcare administrators and planners, to get an idea as to how a drug or therapy is accepted in the Real World. Registries provide a detailed view of the morbidity, mortality and resource utilization associated with a particular disease entity. This data is of prime importance in coming to decisions on coverage of a drug or treatment. The collation of information is also quick and efficient owing to better methods of data management. Most modern databases are equipped with sophisticated data processing software and technologies.

B. The potential benefits of big data for healthcare-

Big data in health can be used to advance the efficiency and effectiveness of estimate and stoppage strategies or of medical interventions, health services, and health policies. Access to well-curated and high quality health-related data will probably have a number of welfares in a diverse range of circumstances. In clinical practice, these data will expand results for individual patients through personalization of predictions, earlier diagnosis, better treatments, and improved decision sustenance for clinicians in cyclic procedures. Cyclic procedures are usually composed of the definition of policy/decision options, the selection of the best alternate, and the subsequent operation and validation of this option. Incorporating feedback from a constant evaluation on the process completes this cycle. These enhancements should ultimately lead to lowered costs for the healthcare system. Similarly, the addition of fragmented information systems into the clinical life cycle will permit the detection of medically pertinent associations, early signals, or changed disease trajectories and should, therefore, enable better patient management policies and better-quality and safety of care. This in turn should lead to upgraded treatment responses for biologically or clinically defined patient subgroups, which will also evade needless rejection of strong drugs and devices. As a result, patient groups will benefit and the unmanageable trend of escalating costs in hospital and community care management as well as diagnostic and drug expansion costs by the biopharma industries will stop or slow down.

Big data also has many possible benefits for translational research into health and well-being. Integrated data sets should expand models of common disease to better know the evolution of rare diseases [39]. They may also assist the detection of population-level effects, such as the off-target and opposing effects of drugs. Biomarkers constitute a key building block of precision medicine, yet the development and clinical validation of new biomarkers is a lengthy process and comparatively few such markers have yet touched routine clinical practice. However, a sizeable number of biomarkers are now widely used in routine clinical diagnostics, which include—but are not limited to—targeted cancer therapy [36–39]. Multidimensional signatures that take into account a wealth of prior information, both from the patient's earlier life history and state-of-the-art information from the literature and relevant databases, will positively deliver a much higher prognostic power than the single biomarkers used today. There is also potential for research on the impact of healthcare involvements and monitoring trends in infectious diseases to inform public health policies [40].

C. The challenges ahead for the effective use of big data in healthcare-

To refine the references for an EU action plan, we acknowledged the main challenges that exist for the use of big data in healthcare research and in the clinic. The challenges have been conveyed elsewhere and include clinical, technical, legal, and cultural obstacles. These challenges differ depending on whether the data are preclinical based on cellular or animal models or from patients in clinical settings, on the intended type of analysis and interpretation, on cross-cultural aspects of privacy, and on ethical and legal thoughts. The incorporation of structured and unstructured data, using natural language processing and other sophisticated machine learning tools, is being tested and it is hoped this will lead to a new level of integration of prior information with up to date clinical information. Data achievement is undergoing quick change. Wearable devices, integrated sensors, and continuous monitoring capabilities are available for all scales of measurements [41].

Data storage issues include security, accessibility, and sustainability.

There are anxieties about trusting health-related data to public clouds, there is a strong need to come up with substitutes. The decades of involvement in big data management for the particle physics community at the European Organization for Nuclear Research (CERN) that headed to the development of the World Wide Web will be valuable. However, many aspects that are specific to big data in health research essential to be taken into account, such as data heterogeneity, institutional and legal fragmentation, and strong data protection standards. There will be an enormous increase in big data invention in all areas of biomedical research, which includes studies at the preclinical levels, such as animal or cellular models and translational studies, but also clinical research that includes patients or public health research. To make the maximum use of the information produced, several technical challenges should be spoken, such as the combination of structured data, such as

genotype, phenotype, and genomics data, with semi-structured and unstructured data, e.g., medical imaging, EHRs, lifestyle, environmental, and health economics data [42].

D. Technical challenges related to the management of electronic health records (EHRs)-

Implementation of EHRs across Europe varies greatly. Estonia and the Valencia Community in Spain (Josep Redón i Màs, personal communication) have changed entirely to EHRs. Incorporation is reinforced with auxiliary systems, for instance drug–drug interaction alert systems that notify physicians and pharmacists about potential prescription clashes, clinical risk groups scheming and costs (e.g., Valencia region, Spain), and drug–gene interaction alert systems that monitor physicians to adjust the dose of a prescribed drug in aberrant drug metabolizers. The USA have taken phases towards a "patient-driven economy". In such a situation, the patient owns his/her data. This ownership wants the development of a suitable health-record infrastructure but delivers a wide range of new health service business chances with major economic potential. Additionally, patient records can be computationally opaque. Data in different languages and inefficient searches and identification are other important barricades. There are some best practices for the management of EHRs. For example, the International Rare Diseases Research Consortium (IRDiRC) [43] develops and implements standards and harmonized methodology crossways diseases and medical cases [44]. Numerous European collaborative projects, such as the European project p-medicine, have created IT infrastructures that will accelerate translational research and the development of personalized medicine [45]. ELIXIR, one of the European infrastructures for life sciences [46], has facilitated the collection, quality control, and archiving of large amounts of life science data such as translational medicine data [47].

E. Technical challenges related to data analysis and computing infrastructures-

Elementary as well as clinical researchers want new computational tools to expand data access and aid user-friendly data analysis for proficient decision making in the clinic. Clinicians want new tools that track, and provide fast feedback for individual patient care. Researchers necessity tools that can be modified for different data sets and analyses such as those used in a wide range of EU-funded projects through the Innovative Medicines Initiative (IMI)-funded eTRIKS consortium platform [48]. Retrieving tool repositories to search for the best tool to answer specific research or clinical questions will be a prerequisite. Likewise important are traceable computational environments that maintain data provenance information from patient to sample and from sample to clinically actionable results. In December 2012, the UK declared the 100,000 Genomes Project [49], which aims to sequence 100,000 genomes, from around 70,000 people, with the focus on patients with rare diseases or cancer. Data remain to increase at an exponential rate and the requirement for cross-border exchange of biomedical and healthcare data, cloud-storage, and cloud computing is inevitable [50].

F. Data quality, acquisition, curation and visualization-

A major encounter for preclinical and clinical research is to get and attain access to sufficient high quality, informative data. Due to a lack of harmonized procedures, in most cases health data cannot be directly used for secondary purposes, such as quality of care, pharmacovigilance, safety and efficacy of treatments, health technology valuation, and public health policy. Data curation is often ignored but vitally important to permit high-quality, informative data. Research funders need to make definite that adequate attention is paid to data quality at the experimental and study design stages, for example, by ensuring data management plans and appropriately reviewed data sharing procedures are in place for all funded research. "Seeing is believing" This expression is related not only for high-resolution microscopy and imaging technologies but also for the presentation and visualization of health-related data.

V. CONCLUSION

The digital transformation is proceeding; a number of industries have already altered their activities or have now become defective. The driving forces are miniaturization, automation, and now increasingly the convergence of artificial intelligence, deep learning, and robotics. Healthcare will not leakage these developments. In fact, big data as a powerful force will play an even more significant role than in most industries. In Europe, working across borders is the only way to master the challenges of this scientific, technological, and industrial transformation. The single most important factor is the workforce/staff. Countries that are ahead in Information and Communication Technology (ICT) capability and have an thoughtful of cultural differences, an ability and willingness to work collected have the best chance to succeed.

In this era of "Big Data", researcher's world over is realizing the potential of Real World Data stored in clinical registries. With incentives from the US and UK governments, Real World Data could revolutionize Health Technology Assessments (HTAs), giving a better and holistic view of outcomes of new technologies. With advances, there is need to ensure that the potential behind clinical registries is unlocked and better registries and better data capture tools are developed.

REFERENCES

[1]   Martínez Sesmero, J.M., BBig Data^; Aplicación y utilidad para el sistema sanitario. Farm. Hosp. 39(2):69–70, 2015.
[2]   Shin, D., Sahama, T., and Gajanayake, R., Secured e-health data retrieval in DaaS and Big Data. Presented at: IEEE 15th International e-Health Networking, Applications & Services (Healthcom) 255–259, 2013.
[3]   Chang, V., A model to compare cloud and non-cloud storage of big data. Futur. Gener. Comput. Syst. 57:56–76, 2016.
[4]   Huang, T., Lan, L., Fang, X., An, P., Min, J., and Wang, F., Promises and challenges of big data computing in health sciences. Big Data Res. 2:2–11, 2015.
[5]   Costa, F., Big data in biomedicine. Drug Discov. Today. 19(4):433– 440, 2014.
[6]   Parra Calderón, C.L., Big data in health in Spain: Now is the time for a national strategy. Gac. Sanit. 30(1):63–65, 2016.
[7]   TingWong, H., Yin, Q., Qi Guo, Y.,Murray, K., Hau Zhou, D., and Slade, D., Big data as a new approach in emergency medicine research. J. Acute Dis. 4(3):178–179, 2015.
[8]   O'Driscoll, A., Daugelaite, J., and Sleator, R., 'Big data', Hadoop and cloud computing in genomics. J. Biomed. Inform. 46:774–781, 2013.
[9]   Merelli, I., Pérez-Sánchez, H., Gesing, S., and D'Agostino, D., Managing. Analysing, and integrating big data in medical bioinformatics: Open problems and future perspectives. Biomed. Res. Int. 2014:1–13, 2014.
[10]   Blanke, T., Big data collecting. Digit. Asset Ecosyst.:87–117, 2014.
[11]   Cunhaa, J., Silvaa, C., and Antunesa, M., Health twitter big bata management with Hadoop framework. Procedia Comput. Sci. 64: 425–431, 2015.
[12]   Ahmad, A., Paul, A., and Rathore, M., An efficient divide-andconquer approach for big data analytics in machine-to-machine communication. Neurocomputing. 174:439–453, 2016.
[13]   Chen, M., Mao, S., and Liu, Y., Big data: A survey. Mobile Netw. Appl. 19:171–209, 2014.
[14]   Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst. 2014;2:3.
[15]   Baro E, Degoul S, Beuscart R, Chazard E. Toward a literature-driven definition of big data in healthcare. BioMed Res Int. 2015; 2015:639021.
[16]   Meldolesi E, van Soest J, Damiani A, Dekker A, Alitto AR, Campitelli M, et al. Standardized data collection to build prediction models in oncology: a prototype for rectal cancer. Future Oncol Lond Engl. 2016; 12:119–36.
[17]   Fernández-Luque L, Bau T. Health and social media: perfect storm of information. Healthcare Inform Res. 2015; 21:67–73.
[18]   Hood L, Price ND. Demystifying disease, democratizing health care. Sci Transl Med. 2014;6:225ed5.
[19]   "An Error Occurred Setting Your User Cookie".
[20]   Rajpurohit, Anmol (11 July 2014). "Interview: Amy Gershkoff, Director of Customer Analytics & Insights, eBay on How to Design Custom In-House BI Tools". KDnuggets. Retrieved 2014-07-14. Dr. Amy Gershkoff:
[21]   "The Government and big data: Use, problems and potential". Computerworld. 21 March 2012. Retrieved 12 September 2016.
[22]   Kalil, Tom. "Big Data is a Big Deal". White House. Retrieved 26 September 2012.
[23]   Executive Office of the President (March 2012). "Big Data Across the Federal Government" (PDF). White House. Retrieved 26 September 2012.
[24]   Lampitt, Andrew. "The real story of how big data analytics helped Obama win". Infoworld. Retrieved 31 May 2014.
[25]   Hoover, J. Nicholas. "Government's 10 Most Powerful Supercomputers". Information Week. UBM. Retrieved 26 September 2012. Bamford, James (15 March 2012). "The NSA Is Building the Country's Biggest Spy Center (Watch What You Say)". Wired Magazine. Retrieved 2013-03-18.
[26]   "Survey on Big Data Using Data Mining" (PDF). International Journal of Engineering Development and Research. 2015. Retrieved 14 September 2016.
[27]   "White Paper: Big Data for Development: Opportunities & Challenges (2012) – United Nations Global Pulse". Retrieve 13 April 2016.
[28]   Lee, Jay; Wu, F.; Zhao, W.; Ghaffari, M.; Liao, L (January 2013). "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications". Mechanical Systems and Signal Processing. 42 (1).
[29]   O'Donoghue, John; Herbert, John (1 October 2012). "Data Management Within mHealth Environments: Patient Sensors, Mobile Devices, and Databases". 4 (1): 5:1–5:20. doi:10.1145/2378016.2378021. Retrieved 16 June 2016 – via ACM Digital Library.
[30]   Solnik, Ray. "The Time Has Come: Analytics Delivers for IT Operations". Data Center Journal. Retrieved June 21, 2016.
[31]   "Australia's bid for the Square Kilometre Array – an insider's perspective". The Conversation. 1 February 2012 Retrieved 27 September 2016.
[32]   Delort P., OECD ICCP Technology Foresight Forum, 2012.
[33]   Dave Ryan. "Sports: Where Big Data Finally Makes Sense". www.huffingtonpost.com. Retrieved 12 December 2015.
[34]   EORTC: European Organisation for Research and Treatment of Cancer. http://www.eortc.org. Accessed 6 May 2016.
[35]   EORTC opens prospective registry for patients with Melanoma. http://www. eortc.org/news/eortc-opens-prospective-registry-for-patients-withmelanoma. Accessed 8 May 2016.
[36]   ENCR: European Network of Cancer Registries. http://www.encr.eu. Accessed 6 May 2016.
[37]   Delfino A, Faure Ragani A, Telpis V, Tilley J, McKinsey & Company. Mature quality systems: what pharma can learn from other industries. Pharm Manuf. 26 Feb 2015; http://www.pharmamanufacturing.com/articles/2015/mature-qualitysystems- what-pharma-can-learn-from-other-industries/. Accessed 20 May 2016.
[38]   Dang A, Vallish BN.Electronic Health Record System in India: Implications for Health Technology Assessment (HTA), Status and Challenges. The Pharm Rev. Nov-Dec 2014. Available from: http://kppub.com/articles/nov2014/electronic_ Health_Record_System_in_India.html (Accessed on 7th Aug' 2915.
[39]   Espay AJ, Bonato P, Nahab FB, Maetzler W, Dean JM, Klucken J, et al. Technology in Parkinson's disease: challenges and opportunities. Mov Disord Off J Mov Disord Soc. 2016. doi:10.1002/mds.26642.
[40]   Hay SI, George DB, Moyes CL, Brownstein JS. Big data opportunities for global infectious disease surveillance. PLoS Med. 2013;10, e1001413.
[41]   Yaman H, Yavuz E, Er A, Vural R, Albayrak Y, Yardimci A, et al. The use of mobile smart devices and medical apps in the family practice setting. J Eval Clin Pract. 2016;22:290–6.
[42]   Di Meglio A. Big data management–from CERN/LHC to personalized medicine. Ajaccio, France: MEDAMI; 2016. doi:10.5281/zenodo.50739.
[43]   IRDiRC: International Rare Diseases Research Consortium. http://www.irdirc. org. Accessed 8 May 2016.
[44]   RARE-Bestpractices. http://www.rarebestpractices.eu/home. Accessed 8 May 2016.
[45]   p-medicine - from data sharing and integration via VPH models to personalized medicine. http://www.p-medicine.eu. Accessed 8 May 2016.

[46] ELIXIR: A distributed infrastructure for life-science information. https://www. elixir-europe.org. Accessed 6 May 2016.
[47] Ison J, Rapacki K, Ménager H, Kalaš M, Rydza E, Chmura P, et al. Tools and data services registry: a community effort to document bioinformatics resources. Nucleic Acids Res. 2016;44:D38–47.
[48] eTRIKS: European Translational Research Information and Knowledge Management Services. https://www.etriks.org. Accessed 6 May 2016.
[49] Genomics England 100,000 Genomes Project. http://www.genomicsengland. co.uk. Accessed 6 May 2016.
[50] Rosenthal A, Mork P, Li MH, Stanford J, Koester D, Reynolds P. Cloud computing: a new business paradigm for biomedical information sharing. J Biomed Inform. 2010;43:342–53.