Big Data Technology – valuable tool for healthcare research

Daniela Marcu[1], Mirela Danubianu[1] and Gheorghe Pentiuc[1]
[1]Stefan cel Mare" University of Suceava,
Department of Electrical Engineering and Computer Science, Suceava

Summary
Big Data, the actual concept concerning data analytics demonstrates its utility for many areas of life.
This paper presents two directions to applied the Big Data technologies and algorithms related to healthcare domain: in genomics and prevention of epidemics. The analysis of such information could dramatically decrease the cost of creating different treatments for the same disease, depending on the patient's medical history.
Key words:
Big Data, healthcare.

1. Introduction

In the last decade, there has been a massive increase in the amount of data generated collected processed and used in previous domains. This is thanks to new technologies that allowed the expansion of social networks, GPS devices, sensors to monitor plants' evolution and mobile applications that inform people of prevalent contagious diseases in certain areas.
Each of us contributes to generating Big Data by: sending emails, social media interactions, making bank transactions, recording sports performance with dedicated applications, and so on.
Data may come from a variety of sources, such as:

- Meteorologists use huge sets of data in order to predict catastrophes and extreme conditions in certain places.
- Biological research scientists collect large sets of data to better understand the development through different stages of a disease and how to improve the corresponding cures or treatments, based on the patient's medical history.
- Online shopping is a very popular source of information regarding the behaviour of a large set of users [1].

Consequently, the types of these data are very diversified. Exploiting large volumes of data with heterogeneous types is a topical issue.
Researchers are currently developing algorithms for Big Data with the aim of discovering new data models (data minning) or analyzing data (analytics).
Such algorithms have been successfully used in sales. Discovering patterns in data, correlating different sets of data, analyzing consumer actions have made it possible to predict certain behaviors. They have led to the establishment of appropriate marketing strategies for making profits.
The benefits of possessing knowledge so valuable quickly motivated companies to invest in the field of Big Data. Unsurprisingly, ever since 2012, when the concept of a Data Scientist was invented, the need for people with skills in this domain has been constantly increasing. Now it is one of the most popular career paths.
Every year, LinkedIn publishes top 25 most needed skills for a successful career. Using several billion data points from the half a billion users, in 2018, the top 2 most popular skills sets were: Cloud computing and Statistical analysis and data mining [2].
Areas such as education and health are in the pioneering phase of creating big data analysis algorithms and technologies. In this paper we propose to present two examples of their use in medicine:

- the use of genomics in cancer treatment;
- prevent the epidemics.

2. Big Data

2.1 Concept definition and characteristics

The term Big Data is used for a collection of data that is too large for traditional models to extract information. This could either be because they cannot cope with the size or it will take too long to obtain results.
The most important characteristics of Big Data are knows as the 5 Vs: volume, variety, velocity, veracity and value.

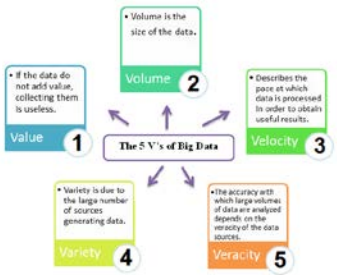

Fig. 1 The 5 V's of Big Data.

Manuscript received April 5, 2019
Manuscript revised April 20, 2019

- Volume is the size of the data. The volume of the data encourages engineers to find new storage solutions and more appropriate analysis algorithms. This is quickly becoming a necessity, as processing such huge amounts of data in a reasonable period of time is getting increasingly difficult. In the medical field, the volume of data is steadily increasing by gathering data from: clinical trials, patient data, administrative data, and, more recently, data from human genome segmentation. [3].
- Variety is due to the large number of sources generating data. For example, in medicine there is data everywhere: medical imaging, history of patients, test results, various sensors, and so on. The majority of this information is unstructured which led to new technology such as: parallel processing, cloud computing and non-relational databases. Google came up with MapReduce, a concurrent model that can extract certain information from Big Data [4].
- Velocity is a characteristic of Big Data which describes the pace at which data is processed in order to obtain useful results. One tool that provides real time data processing is Google Analytics. It can be used in order to obtain details about the users of a website and their behaviour throughout their web journey. Online marketing knowledge and information about the users such as: age, hobbies, location, language and personal devices used to navigate, can be used to provide the user with a tailored, more attractive and efficient online experience. Also, Google search can provide real-time information about the public's interest for a specific content on the site. Below you can find a screenshot showing the analysis of the users' journey through our site https://www.modinfo.ro.

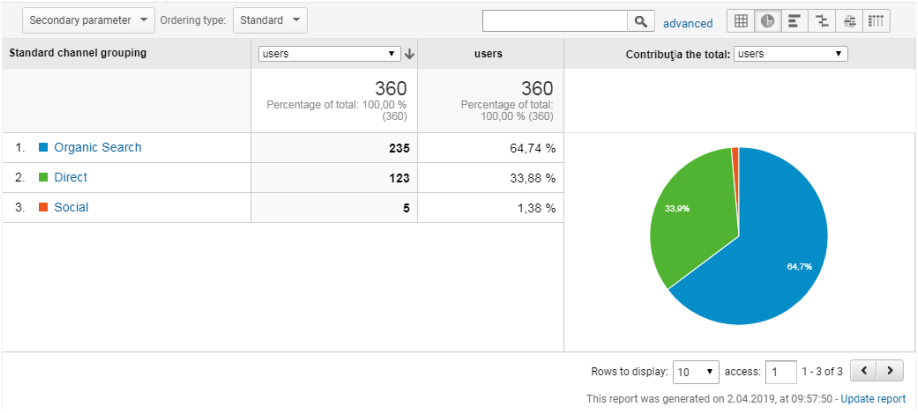

Fig. 2 The analysis of the users' journey.

- Veracity refers to the confidence of data presents. The accuracy with which large volumes of data are analyzed depends on the veracity of the data sources. In medicine, veracity is closely linked to data variability. Presenting a symptom to the physicians with different specializations may lead to different interpretations of the diagnosis, but also to recommendations on the patient's health investigations [4] [5].
- Value. Only accessing Big Data is not enough to get the right predictions that lead to measurable improvements. If the data do not add value, collecting them is useless. Consequently, data value is one of the most important features to be taken into account before investing in the IT architecture needed to exploit large data volumes [5].

There are currently some controversies and confusion about the number V's of Big Data. Starting with 2001, when the concept of 3V's (Volume, Variety, Velocity) was introduced, have been added other characteristics to the Big Data paradigm. The most widely accepted are 5V's, presented above [6].

2.2 Big Data systems and technology

Special problems need special solutions. So, due to its complexity, Big Data requires special architectures. Open-source frameworks have been developed to cope with the large amount of data, obtained from clusters of computers. One of the most popular ones is Hadoop, created by Apache. This was adopted by some of the biggest companies on the market: Twitter, Facebook, Adobe, LinkedIn, Google, etc. The core of this ecosystem is

formed of two components: a framework for processing complex calculations on large sets of data and a distributed file system [7].
In 2002, Google published an article about MapReduce. It is a concurrent model that is able to run on huge sets of data [8]. The way it works is split in two:

- Breaking the problem into smaller instances of the same problem and passing those to operational nodes that can solve them
- Collecting the results and combine them in order to obtain something meaningful.

Running this parallel offers a great leap in performance over the most traditional, sequential algorithms. MapReduce includes a JobTracker that delegates the nodes with tasks from nearby clusters. In order to reduce the traffic on the main part of the network, JobTracker will assign some of the tasks to other available nodes [9].
Hadoop was launched in December 2011 and is written in Java. This system can be used by other applications such as Apache Mahout and Hive Data Warehouse [10].

3. The Use of Big Data Technologies in Healthcare

Healthcare is a domain that generates huge volumes of data. Exploitation of Big Data generates knowledge, analysis and models that can bring significant benefits to scientific research and innovation in healthcare. The benefits of health-related Big Data currently have been demonstrated in some areas like: prevent disease, identify modifiable risk factors for disease, and design interventions for health behavior change [11].
Big Data's potential has already been demonstrated in: analyzing data models in preventive medicine, personalizing treatments, minning data from medical literature, detecting vulnerabilities within patient populations during disease outbreaks or disasters [12], reducing number of hospital re-admissions, preventing adverse drug effects, genomics, preventing epidemic outbreaks by analyzing anomalies in Internet searches.

3.1 Big Data, genomics and cancer treatment

The progress made in calculus, nanotechnology and biotechnology is crucial for the evolution of medical technology, and also makes it possible to go from the principles of 3P to the ones of 6P: personalized, precise, participatory, preventive, predictable and patient-focused.
The effectiveness of therapeutic and preventive approaches to a particular disease may be influenced of the lifestyle, the environmental factors and the patient's genetic composition.
In 2016, the US National Institutes of Health launched the Precision Medicine Initiative (PMI) project for which $ 130 million was allocated to use and analyze clinical and genomic data from over 1 million American patients. The National Cancer Institute received $ 70 million in the same project to develop cancer genomics [13]. One of the National Cancer Insitute (NCI) initiatives is the Genomic Data Commons (GDC). This project centralizes, standardizes and sharing genomics data. Some of the largest sets of cancer genomics data are generated from two programs: TCGA (The Cancer Genome Atlas) and TARGET (Therapeutic Applicable Research to Generate Effective Treatments) [14].
Sequencing the entire human genome is very expensive. Also this generated at least two issues: data size, since sequencing of a single whole genome generates more than 100 GB in BAM format (the binary version of sequence alignment / map) that is processed in about 5 hours [15] and data privacy policy.
Sequencing parts of or even the whole genome of the patient would make it possible to identify the orthographical differences when compared to the genomes of healthy patients.
In the last several years, there have been a few breakthroughs in assembling the human genome, the last one performed with the nanopore sequencer MinION which can detect around 1 million base pairs by analysing huge DNA fragments [16]. It is currently estimated that by 2025 over 100 million human genomes will be possible sequenced [17].
Big Data applications for genomics provide statistics that associate the expression changes of the studied genes with anticipated functional effects.
In the absence of proper infrastructure, the exploitation of genomic data may take days or even months. An example of an open-source framework for implementing Big Data technologies is the Hadoop system. With the MapReduce concurrent model, data is fragmented and distributed across multiple nodes. These will be processed in parallel, which will lead to higher speed. SeqHBase is an example of Big Data toolset designed to analyze sequensing data created with the aim of discovering inherited homozygous, and compound heterozygous mutations [15][18].
The exponential increase in the amount of genomic data available for processing (only the researchers from Broad Institute of MIT and Harvard generate about 20 TB of data sequences per day [14]) led to the development of initiatives aimed at developing methods and software tools to promote Big Data science. One of these is the Big Data Genomics project. It is supported by the National Institutes of Health (NIH) BD2K and the NIH Cancer Cloud Pilot. The project main goal is to build a scalable API for processing genomics data (ADAM).

ADAM is a genomic analysis platform, written in the Scala. It works on a wide range of Apache Spark and Apache Hadoop versions. Adding support for Spark SQL makes it possible to use APIs for Python and R. ADAM can be used to run queries of data stored in the most common genomic formats: BAM/SAM/CRAM, BED, VCF, NarrowPeak, FASTA and so on. In March 2018, using ADAM, Cannoli and Avocado, the researchers managed to genotype more than 19 TB of gVCFs in about 6 hours using 1024 cores [19].

With the aim of identifying molecular anomalies within a tumor and personalizing treatment for the ill, making it more efficient and eliminating as many side-effects as possible, the National Cancer Institute of France is sequencing the genome of 70000 new patients from all around the world every year [20]. Not all patients suffering from cancer react to treatment in the same way. This usually depends on the characteristics of the tumor but also of the patient's genomics. Currently, for the first 3 stages of cancer, the selected patients go through various experiments based on their own genome's profile in order to evaluate the outcome of the treatment and predict the way the disease is going to evolve. However, this is not a medical routine yet [21].

At the moment, using the analysis of Big Data to study the curing of cancer through genomics is not very common. There are a few ongoing projects in Europe: one in Germany (The German Cancer Research Center), one in France (study on the predisposition to mammary and ovarian cancer as part of the France Genomic Medicine 2025 plan) and two more in England and Switzerland [22].

In Vienna, a few months ago, the representatives of each country from the European Commission, signed a document in order to merge their efforts into connecting all the genomics databases and also share the resources with patients from EU. This move was supported by the Austrian minister of health, Beate Hatinger-Klein, underlying the importance of this collaboration in order to improve the medical services [23]. The main purposes of the coalition are to develop personalized cancer treatment, analyze the predisposition to a disease and also to prevent them. The health insurance companies would benefit from this as well, as the huge funds for cancer treatment might not be needed anymore. On top of this, the European commissioner for public health, Vytenis Andriukaitis, mentioned that “We are far from benefitting from a digitalised health ecosystem. Open discussions to create opportunities and eliminate the necessary barriers will be required [23]."

The need for genomic data sharing within healthcare communities for the benefit of human health has led to the emergence of the non-profit international alliance Global Alliance for Genomics and Health (GA4GH). GA4GH has developed a strategic plan by 2022. One of the main objectives is the approval of the Standard GA4GH Data Use Ontology (DUO), developed in the W3C Ontology Language to allow users to semantically tag genomic datasets [24].

In order to identify global resources and sharing clinical and genomic data, GA4GH provided a catalog. It is allowed to apply search filters. In Table 1 there are some examples of current cancer genome research projects.

Table 1: Some examples of current cancer genome research projects

Project	Mission statement	Cohort size and description
Asian Cancer Research Group (ACRG)	ACRG aims to generate a unique genomic resource to be used by the scientists and clinicians worldwide to accelerate efforts to discover treatments for different type of cancers.	176 HCC tumours and paired normal tissues
Australian Genomics	AGHA (Australian Genomics Health Alliance) is a national network of universities and research institutes, hospitals and other organizations. AGHA is trying to make changes to the state health service system to facilitate the rapid implementation of scientific innovations in clinical practice. In the first period (2016-2018), neuromuscular disorders, mitochondrial disorders, neurodevelopmental disabilities (epileptic encephalopathies), genetic immune disorders will be studied.	1800 Cases (approximately 900 cancer, 900 rare disease)
Clinical Sequencing Exploratory Research (CSER)	Clinical Sequencing Exploratory Research (CSER) is a project initiated by the National Human Genome Research Institute (NHGRI) and the National Cancer Institute (NCI). This program bring together clinicians, informaticians, economists and others to develop and sharing the best practices in the integration of genomic sequencing into clinical care.	6000 Pediatric and adult patients with various phenotypes, healthy adults, and physicians. ~6,000 have already been enrolled through CSER. The projected enrollment (through May, 2017) is ~7,000.
deCODE Genetics	deCODE is an initiative ho leads the world in the discovery of genetic risk factors for common diseases. Some 500,000 individuals from around the globe taking part in the discovery work and proprietary statistical algorithms and informatics tools for gathering, analyzing, visualizing and storing large amounts of data.	500000 International participants
Human Genome Variation Society (HGVS)	HGVS promotes collection, documentation and free distribution of genomic variation information and associated clinical variations, together with development of necessary informatics techncologies. HGVS runs a database of locus-specific databases (LSDBs), available at: http://www.hgvs.org/locus-specific-mutation-databases.	International

Human Longevity, Inc. (HLI)	Human Longevity, Inc. (HLI) created the world's largest and most comprehensive database of whole genome, phenotype and clinical data. HLI is developing and applying large scale computing and machine learning to make novel discoveries to revolutionize the practice of medicine.	International
Treehouse Childhood Cancer Initiative	Treehouse Childhood Cancer Initiative is a data sharing initiative researching the feasibility of real-time pan-cancer analysis for difficult-to-treat pediatric cancer patients	International

What roles have the Big Data analysis algorithms play in detecting, treating and even preventing cancer? Instead of replacing the diagnostic given by the doctor, they would help with more complex cases with gene mutations. To give an overview of how popular Big Data is at the time of writing, there are approximately 200 hospitals in 28 countries from the EU that use Big Data services to interpret the genes profiles of different patients suffering from cancer.

In order to insure the prevention of such diseases, and also to improve the efficiency of the currently available treatments, it is necessary to identify the risk factors depending on the medical history of each patient, climate conditions the patient was exposed to and even diet. This is where Big Data becomes powerful.

Unfortunately, we are only scratching the surface with Big Data in genomics. The elaboration and application of analysis techniques to diagnose and generate a personalized treatment, that is more cost-effective, is relying on the accumulation of very large datasets about the human genome and how the disease evolved over time. This process would be accelerated only by unifying all the available Big Data worldwide and the lessening of the privacy rights with regards to personal data. Also, it would be good to see more training available from domain experts.

3.2 Big Data - prevent epidemics

The last decade there have been characterized by a major change in the way people interact with each other. It went from verbal to online communication, thanks to the socializing networks and personal blogs. The need for information is satisfied by the vast stream of news and online newspapers. More specific information can be found on more area focused websites with content usually written by communities and domain experts. Thanks to the massive increase in the portable devices industry, valuable health information can be found pretty much everywhere by anyone [25].

As important as transmitting information is the ability to extract information from what people usually look for online. The development of new algorithms, creation of models with scientific purposes are contributions towards much more efficient discovery of reasons behind epidemics.

Health Canada was the first step towards web analysis, when in 1990, they created a network of public health information. Aiming to collect information about a possible emergency of public health, the network uses autonomous aggregators feeding o news. The network proved to be extremely useful on many occasions: in 2002 when the south of China was bombarded with SARS which eventually travelled through airplanes to other countries infecting over 750 people, and in 2008, when Listeria monocytogenes took over 24 lives and cost the Canadian government around 250 million dollars [26].

In 2004, Gunther Eysenbach from the University of Toronto, mentions that there is a very good chance the Google searches are related to the outbreaks in Canada. The method the web searches analytics uses is far more cost-effective than the traditional ones. However, there is a small chance that some are false-positives due to panic mode being triggered. To obtain the data, the Canadian researcher created an advertising campaign on Google AdSense with a link to an educational website for patients looking for various diseases and symptoms. Based on this, he published statistics on a daily basis that proved to be correlated with the instances of flu during the following week [27].

Mid-April of 2009, CDC 24/7 - Centers for Disease Control and Prevention announced the first two cases of A/H1N1, 130 miles apart from each other. Shortly after, another 2 cases appeared in Texas and 14 in Mexico. A month later, CDC admit the pandemic of H1N1 in a large press conference with over 2500 journalists. It took almost one year for the World Health Organisation to declare the end of the 6-th wave of A/H1N1. Over 18500 people lost their lives during this year just because of the virus. This is officially, the first pandemic of our century [28].

Early detection of the virus, followed by a rapid response can dramatically reduce the impact of a pandemic: treatments become more efficient and more human lives are saved. Together with the first symptoms of the disease, people began to search related information on Google. Analysing millions of such searches from before the official announcement of the pandemic, some analysts from Google managed to build a mathematical model that can prove the relationship between the Google searches and the wide spread of the virus.

The results were published in 2008, a few months before the first 2 cases occurred [29].

During the flu season in US, 2007, the comparison between the estimations generated by the model with the results of the Department of Epidemiology, CDC to test the model.
Google searches can be processed quickly, generating results even 2 weeks before the official reports of the sick. This could accelerate the development of treatments and quarantining a virus which leads to better economy and more human lives saved.

4. Conclusion

Medicine is a science based on data. The more developed the analysis methods are, the bigger the data-set obtained by the medical system can be. In consequence, the health domain could benefit from advancements in Big Data [30].
According to the European Centre of Prevention and Control of Diseases, over 100000 patients suffer from viruses, and over 37000 die every year. A third of these cases could be prevented by strict norms of hygiene recommended by all the hospitals. On the other hand, currently available technology still lacks in performance, hence the reason why very large datasets still cannot be processed in real time.
The attempts that have been made so far, include genome studies and dates of various epidemics, for developing a component to prevent and predict the initial stages of an outbreak. This would save between 25 and 32 billion dollars only in US [31] and many human lives.
The usage of models and Big Data technology has become a priority for the medical research in Romania, and even in planning the scientific activities at the UMF, Iasi.
The genomics research center in Cluj, combines the molecular biology technology with the analytics from Big Data in order to create more personalised therapies and attempt of prevention. As this is far more cost-effective, it would allow for reallocation of remaining funds from the public health sector.
Integrating medical research upon the human genome into Big Data can yield important results in more efficient treatments and prevention of some diseases.
Big Data technology proved to be useful in the extraction of information from web searches, helping the prevention of outbreaks and reducing the mortality of epidemics. Big Data is an opportunity that companies and universities from Romania take into consideration for future developments and educational plans.

References

[1] Dr. Martin S. Martin. Big Data. Viata medicala Nr. 43 (1500), aprilie 2016. url: http://www.viata-medicala.ro
[2] LinkedIn. Resume Skills List. url: https://www.businessinsider.com/
[3] Anil Jain, The 5 Vs of Big Data, IBM Blog Post | Healthcare Data Analytics September 17, 2016 https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/
[4] Wu M. Gopukumar D. Luo J. and Y. Zhao. Big Data Application in Biomedical Research and Health Care: A Literature Review. In: Biomedical informatics insights 8 (2016), pp. 1-10. doi: 10.4137/BII.S31559
[5] Kalbandi, Ishwarappa & Anuradha, J. (2015). A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology. Procedia Computer Science. 48. 319-324. 10.1016/j.procs.2015.04.188.
[6] Patgiri, Ripon & Ahmed, Arif. (2016). Big Data: The V's of the Game Changer Paradigm. 10.1109/HPCC-SmartCity-DSS.2016.0014.
[7] Adriana Alexandru and Dora Coardos. Big Data - concepte, arhitecturi și tehnologii. In: Revista Română de Informatică și Automatică 27 (2017).
[8] Veena Priyanka, Annapoorna Manohar, Pooja Gupta and Muhammad Fahim Uddin. Big Data Application in Biomedical Research and Health Care: A Literature Review. In: Utilizing Big Data Analytics to Improve Education ().
[9] Jeffrey Hanson. An introduction to the Hadoop Distributed File System. url: https://www.ibm.com/developerworks/library/wa-introhdfs/
[10] Apache Hadoop. url: https://ro.wikipedia.org/wiki/Apache_Hadoop
[11] Barrett MA, Humblet O, Hiatt RA, Adler NE. Big data and disease prevention: From quantified self to quantified communities. Big Data 2013;1(3):168-75
[12] Swarna Latha Garapati, Dr. Sridevi Garapati. Application of Big Data Analytics: An Innovation in Health Care. International Journal of Computational Intelligence Research. ISSN 0973-1873 Volume 14, Number 1 (2018), pp. 15-27.
[13] National Institutes of Health, SUA.url: https://allofus.nih.gov/
[14] Technology Networks. url: https://www.technologynetworks.com/tn/lists/7-projects-that-are-harnessing-the-power-of-big-data-291346
[15] Karen Y. He, Dongliang Ge, Max M. He . Big Data Analytics for Genomic Medicine. International journal of molecular sciences vol. 18,2 412. 15 Feb. 2017, doi:10.3390/ijms18020412
[16] Gizmodo. url: https://gizmodo.com/scientists-sequence-human-genome-with-handheld-usb-devi-182259166
[17] Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, Iyer R, Schatz MC, Sinha S, Robinson GE. Big Data: Astronomical or Genomical? PLoS Biol. 2015 Jul; 13(7):e1002195
[18] He M., Person T.N., Hebbring S.J., Heinzen E., Ye Z., Schrodi S.J., McPherson E.W., Lin S.M., Peissig P.L., Brilliant M.H., et al. Seqhbase: A big data toolset for family based sequencing data analysis. J. Med. Genet. 2015;52:282–288. doi: 10.1136/jmedgenet-2014-102907.
[19] Big Data Genomics. url: http://bdgenomics.org/
[20] Patrice Bouexel, Catherine Commaille-Chapus, Anne Gimalac, Guillaume Jeannerod, Magali Leo, Bruno Leroy, Bernard Nordlinger, Michel Paoli, Pablo Prados, Jean-Yves Robin, Isabelle Barbier, Jeanne Bossi Malafosse. BIG

DATA AND PREVENTION FROM PREDICTION TO DEMONSTRATION. In: (2016). url: https://healthcaredatainstitute.com/wp - content / uploads / 2017/01/hdi-bigdata-prevention-2016-final.pdf.
[21] Young Jin Kim. The Future Medical Science and Colorectal Surgeons. In: (2017). doi: 10.3393/ac.2017.33.6.207. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5768474/
[22] Juan Romob, Jose L. Torrecilla. Data learning from big data. Statistics and Probability Letters. In: (2018), pp. 15-19
[23] url: https : / / reureporter . co / health / 2018 / 09 / 12 / eapm - austria -signs - up - for - genome - project - and - puts - emphasis - on - digital - health/.
[24] Global Alliance for Genomics and Healths (GA4GH).url: https://www.ga4gh.org/
[25] Nicole DiLisio, Y. Tony Yang, Michael Horneffer. Mining Social Media and Web Searches For Disease Detection. In: (2013), pp. 17-21
[26] Mahmoud A Lemon S et al. Yanzhong Huang Knobler S. THE SARS EPIDEMIC AND ITS AFTERMATH IN CHINA: A POLITICAL PERSPECTIVE. Institute of Medicine (US) Forum on Microbial Threats. url: https://www.ncbi.nlm.nih.gov/books/NBK92479/.
[27] Gunther Eysenbach. Infodemiology: Tracking Flu-Related Searches on the Web for Syndromic Surveillance. In: (2006), pp. 244{248. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839505/
[28] The 2009 H1N1 Pandemic: Summary Highlights. 2009. url: https://www.cdc.gov/h1n1flu/cdcresponse.html
[29] Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, Larry Brilliant, Jeremy Ginsberg, Matthew H. Mohebbi. Detecting influenza epidemics using search engine query data. In: 457 (). url: https://www.nature.com/articles/ nature07634
[30] European Centre for Disease Prevention and Control. url: https://ecdc.europa . eu / en / publications - data / directory - online - resources - prevention-and-control-antimicrobial-resistance-amr.
[31] Big Data: The next frontier for innovation, competition and productivity. McKinsey Global Insitute. url: https://www.mckinsey.com